IECE
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2024.876830

REVIEW ARTICLE

A Comprehensive Survey on Emerging Techniques and Technologies in Spatio-Temporal EEG Data Analysis

Pengfei Wang[1], **Huanran Zheng**[1], **Silong Dai**[1], **Yiqiao Wang**[1], **Xiaotian Gu**[1], **Yuanbin Wu**[1,*] **and Xiaoling Wang**[1,*]

[1] School of Computer Science and Technology, East China Normal University, Shanghai 200062, China

Abstract

In recent years, the field of electroencephalography (EEG) analysis has witnessed remarkable advancements, driven by the integration of machine learning and artificial intelligence. This survey aims to encapsulate the latest developments, focusing on emerging methods and technologies that are poised to transform our comprehension and interpretation of brain activity. The structure of this paper is organized according to the categorization within the machine learning community, with representation learning as the foundational concept that encompasses both discriminative and generative approaches. We delve into self-supervised learning methods that enable the robust representation of brain signals, which are fundamental for a variety of downstream applications. Within the realm of discriminative methods, we explore advanced techniques such as graph neural networks (GNN), foundation models, and approaches based on large language models (LLMs). On the generative front, we examine technologies that leverage EEG data to produce images or text, offering novel perspectives on brain activity visualization and interpretation. This survey provides an extensive overview of these cutting-edge techniques, their current applications, and the profound implications they hold for future research and clinical practice. The relevant literature and open-source materials have been compiled and are consistently updated at https://github.com/wpf535236337/LLMs4TS.

Keywords: electroencephalography (EEG), self-supervised learning (SSL), graph neural networks (GNN), foundation models, large language models (LLMs), generative models.

1 Introduction

Electroencephalography (EEG) has long been a cornerstone in the study of brain function, offering a non-invasive means to monitor electrical activity within the brain. Non-invasive are easier to implement without surgery, but they lack simultaneous consideration of temporal and spatial resolution, as well as the ability to capture deep brain information. In contrast, invasive methods like Stereoelectroencephalography (SEEG) [1] can measure these brain signals more precise with higher signal-to-noise data [2], albeit requiring surgical

Academic Editor:
Hongqi Fan

Submitted: 31 July 2024
Accepted: 10 December 2024
Published: 15 December 2024

Vol. 1, **No.** 3, 2024.
10.62762/CJIF.2024.876830

***Corresponding authors:**
✉ Yuanbin Wu
ybwu@cs.ecnu.edu.cn
✉ Xiaoling Wang
xlwang@cs.ecnu.edu.cn

Citation
Wang, P., Zheng, H., Dai, S., Wang, Y., Gu, X., Wu, Y, & Wang, X. (2024). A Comprehensive Survey on Emerging Techniques and Technologies in Spatio-Temporal EEG Data Analysis. *Chinese Journal of Information Fusion*, 1(3), 183–211.

procedures to insert recording devices. Overall, non-invasive signals are relatively safer, more portable, have greater potential for use, and are applicable to a wider population, reflecting voltage fluctuations caused by ion currents in neurons.

While our understanding of the brain deepens and computational methods advance [3, 4], the field of EEG analysis faces many challenges. The first challenge is the effective capture of representations in EEG data, particularly in the absence of labels. The second challenge involves the identification and classification of complex and subtle patterns within brain activity, requiring advanced discriminative methods that can accurately interpret the nuanced differences indicative of various brain states or conditions. Lastly, the challenge of creating meaningful visualizations or interpretations from EEG data calls for generative methods that can transform the abstract EEG signals into more tangible and comprehensible forms, such as images or text, thereby enhancing our understanding of the brain's intricate workings. Addressing these challenges collectively advances the field of EEG analysis, making it more robust, insightful, and applicable to a wider range of scientific and clinical applications.

In response to aforementioned challenges, recent developments in deep learning and artificial intelligence have paved the way for more robust and nuanced EEG analysis strategies. This paper surveys three key areas of advancement that are reshaping the field of EEG analysis:

- **Representation Learning in EEG Analysis**: Representation learning is the first fundamental step in EEG analysis, concentrate on automatically extracting useful features from EEG signals. Self-supervised learning methods are being employed to develop robust signal representations that enhance the precision and interpretability of downstream tasks. These unsupervised learning methods are naturally suited for the vast amounts of brain signal data and mimic human learning processes.
- **Discriminative EEG Analysis**: Discriminative methods focus on distinguishing between different categories or patterns in EEG signals. Advanced architectures such as Graph Neural Networks (GNNs), Foundation Models, and LLMs-based Methods are being utilized to gain deeper insights into brain activity. These architectures efficiently capture discriminative patterns, which are crucial for understanding complex neural processes.
- **Generative EEG Analysis**: Generative methods aim to generate new modalities or signal data from EEG signals. Innovative approaches such as diffusion produce images or text from EEG data are providing novel approaches to the understanding and visualization of brain activity. These generative techniques are also important applications for AI-generated content (AIGC).

This paper aims to provide a comprehensive overview of cutting-edge techniques, discuss their details, and explore the significant implications they hold for future research and clinical practice in EEG analysis. **The structure of this paper is organized according to the categorization within the machine learning community, with representation learning as the foundational concept that encompasses both discriminative and generative approaches [5].** The remainder of this paper is organized as follows: Section 2 summarizes the background and related surveys of our work. Section 3 discusses the robust representation learning strategy and its significance in EEG data analysis. Section 4 explores the emergent discriminative architecture, detailing the role of GNNs (4.1), Foundation Models (4.2), and LLMs-based Methods (4.3). Section 5 examines the innovative generative applications of EEG data. Section 6 provides an introduction of the most widely used datasets and the key metrics employed to assess the performance of various EEG analysis models. Finally, Section 7 concludes the paper and discusses potential future directions for EEG analysis.

2 Related survey

2.1 Existing Surveys on EEG Analysis

In the domain of EEG-related concepts and research, numerous review studies have provided comprehensive summaries. Hosseini et al. [4] introduced the application of machine learning in EEG signal processing, covering traditional methods such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Naive Bayes in classification scenarios. However, this review did not consider the extensive discussion of deep learning algorithms that have demonstrated superior performance. Jiang et al. [6] discussed the removal of artifacts from EEG signals, making their review more detailed in technical aspects. Nevertheless, their work did not cover deep learning algorithms and did not consider a

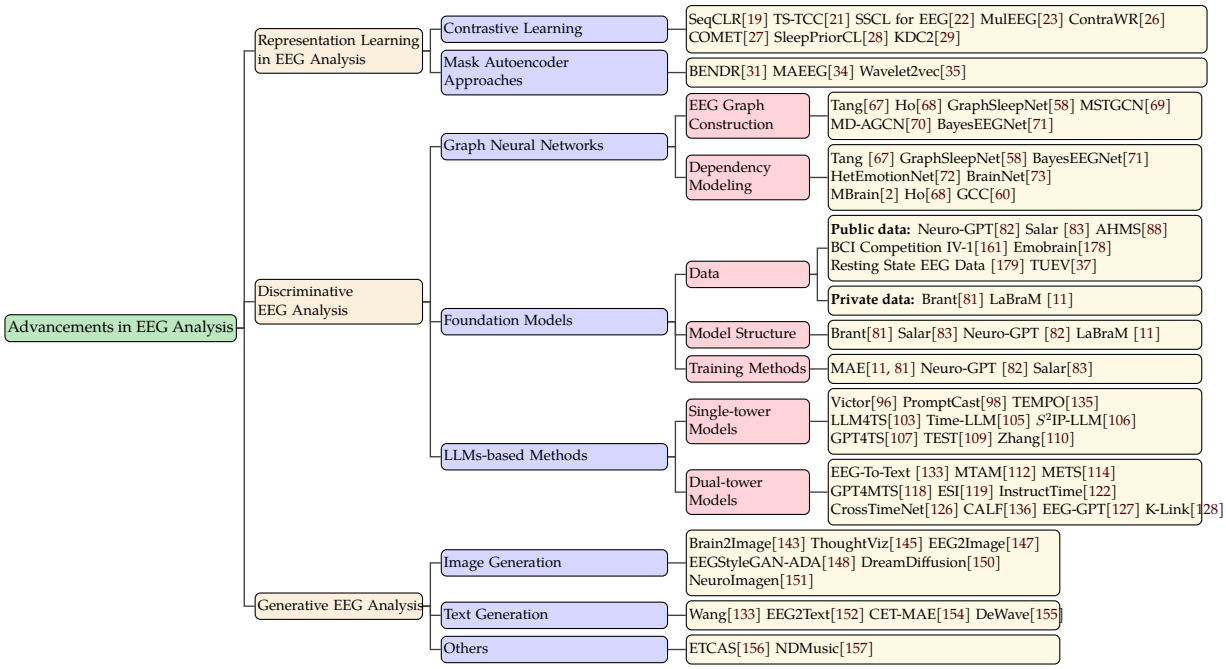


Figure 1. A comprehensive taxonomy of advancements in EEG Analysis.

broader range of EEG downstream tasks. In contrast, Zhang et al. [7] provided a more comprehensive perspective, introducing the origins and applications of Brain-Computer Interface (BCI) and discussing the integration of mainstream deep learning algorithms such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Generative Adversarial Networks (GAN) with EEG tasks. With the continuous innovation in artificial intelligence community, EEG research based on foundational models and large language models has begun to emerge. However, to the best of our knowledge, there is currently no literature that reviews EEG analysis from a more holistic frontier technology perspective, which is the gap this paper aims to fill.

2.2 Emerging Surveys on General Time-Series Analysis

In the general time series domain, a substantial body of work has summarized the application of the latest technologies in various downstream tasks. Zhang et al. [8] categorized existing self-supervised learning-based time series analysis methods into three types: generative, contrastive, and adversarial, and discussed their key intuitions and main frameworks in detail. Jin et al. [9] provided an overview of the application of graph neural networks in time series tasks such as forecasting, classification, imputation, and anomaly detection. Liang et al. [10] reviewed foundational models in time series analysis from the perspectives of model architectures, pre-training techniques, adaptation methods, and data modalities. Similarly, [11–13] systematically outlined methods and procedures for time series analysis based on large language models. Yang et al. [14] reviewed the application of diffusion models in time series and spatio-temporal data. Additionally, there are some works focusing on more specific model architectures or downstream tasks [15, 16]. We refer the reader to the corresponding publication for a more in-depth understanding.

Although numerous reviews exist within the broader time series field, few surveys concentrate exclusively on EEG data. Moreover, EEG data possesses unique characteristics, and a substantial body of related work has emerged recently. Thus necessitating a comprehensive review and synthesis, this paper seeks to offer an in-depth examination of state-of-the-art techniques, elaborate on their intricacies, and explore their profound implications for future EEG research and clinical applications.

3 Representation Learning in EEG Analysis

In recent years, deep learning has excelled in extracting hidden patterns and features of the data. Typically, feature extraction models based on deep learning rely heavily on large volumes of labeled data, a method commonly referred to as supervised learning. However, in certain practical applications, particularly in time-series data such as Electroencephalograms (EEG), acquiring extensive

labeled data is both time-consuming and costly. As an alternative, Self-Supervised Learning (SSL) has garnered increasing attention due to its label efficiency and generalization capabilities. SSL, a subset of unsupervised learning, extracts supervisory signals by solving tasks automatically generated from unlabeled data, thereby creating valuable representations for downstream tasks.

With the significant success of SSL in fields such as computer Vision(CV) [17] and Natural Language Processing(NLP) [18], its application to time-series data appears particularly promising. However, directly applying tasks designed for visual or linguistic processing to time-series data is challenging and often yields limited effectiveness. The primary reasons include:

- Time-series data possess unique attributes such as seasonality, trends, and frequency domain information, which are typically not considered in tasks designed for images or language.
- Common data augmentation techniques in computer vision, such as rotation, flipping, and cropping, can disrupt the temporal dependencies and integrity of time-series data, such as EEG signals. For instance, rotating or flipping the time points in an EEG signal could completely lose physiological significance and contextual information.
- Many time-series datasets are multidimensional, with each dimension potentially representing a different measurement channel. This contrasts with handling single images or text data, requiring synchronous analysis and processing across multiple dimensions.

To address these issues, this section summarizes two main paradigms of SSL: contrastive learning, which trains models to distinguish between similar and dissimilar pairs of data points and masked autoencoders, which aim to learn the intrinsic feature information of the data. All of the methods are summarized in Table 1.

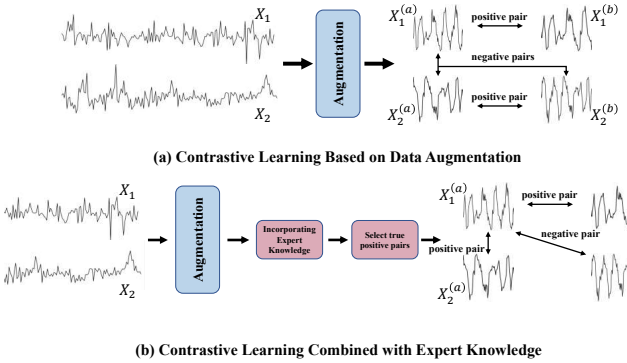
3.1 Contrastive Learning

Contrastive learning is a self-supervised learning method that acquires invariant representations of data by learning the similarities and differences between samples. This approach maps similar samples to proximate representation spaces and dissimilar samples to distant ones, thereby enabling the learning of generalized feature representations without the need for explicit label information. Formally, given a set of samples $\mathcal{X} = \{x^1, x^2, \cdots, x^N\}$, contrastive learning aims to learn a mapping function f that maximizes the similarity between positive sample pairs of the same class and minimizes the similarity between negative sample pairs of different classes. For positive sample pairs (x, x^+) and negative sample pairs (x, x^-), the objective of contrastive learning is to optimize the following loss function:

$$L\left(x, x^+, x^-\right) = -\log\left(\frac{e^{f(x,x^+)/\tau}}{e^{f(x,x^+)/\tau} + e^{f(x,x^-)/\tau}}\right) \quad (1)$$

where $f(x, x^+)$ denotes the similarity of feature representations for positive pairs, $f(x, x^-)$ for negative pairs, and τ is a temperature parameter that adjusts the scale of similarity. The intuitive interpretation of this loss function is that by maximizing the similarity of positive pairs while minimizing that of negative pairs, the model learns high-level semantic relationships between samples, resulting in more distinctive representations.

In this section, we will introduce two types of contrastive learning methods, which are contrastive learning based on data augmentation and contrastive learning combined with expert knowledge(as shown in Figure 2). All of the methods are presented in Table 1.



(a) Contrastive Learning Based on Data Augmentation

(b) Contrastive Learning Combined with Expert Knowledge

Figure 2. Two types of contrastive learning methods.

3.1.1 *Based on Data Augmentation*

Data augmentation is an indispensable component of contrastive learning. It generates different views of input samples using data augmentation techniques, and then learns representations by maximizing the similarity between views of the same sample while minimizing the similarity between views of different samples. SeqCLR [19] introduces a set of data augmentation techniques specifically for EEG and

Table 1. Summary of Self-Supervised Learning for EEG Analysis.

SSL	Method	Strategy	Backbone	Task	Datasets	Metric
CL	SeqCLR[19]	Signal transformation	CNN & GRU	Multiple tasks	THU[37], SEED[38], SleepEDF[39], ISRUC-S3[40]	Accuracy
	TS-TCC[21]	Weak & strong augmentation	Transformer	Sleep & seizure detection	HAR[41], SleepEDF[39], ESR[42], FD[43]	Accuracy, F1
	SSCL for EEG[22]	Signal transformation	CNN	Sleep stage classification	SleepEDF[39], DOD[44]	Accuracy, F1
	MulEEG[23]	Multi-view contrast	CNN	Sleep stage classification	SleepEDF[39], SHHS[45]	Accuracy, Kappa, F1
	ContraWR[26]	Non-negative contrast	CNN	Sleep stage classification	SHHS[45], SleepEDF[39], MGH[46]	Accuracy
	COMET[27]	Multi-level contrast	CNN	Disease detection	AD[47], PTB[48], TDBRAIN[49]	Accuracy, F1, AUROC, AUPRC
	SleepPriorCL[28]	Expert knowledge incorporation	CNN	Sleep stage classification	SleepEDF[39], MASS-SS3[50]	Accuracy, F1
	KDC2[29]	Cross-view contrast	CNN & GNN	Multiple tasks	SEED[38], MMI[51], CHB-MIT[52]	Accuracy
MAE	BENDR[31]	Temporal-domain mask	CNN & Transformer	Multiple tasks	MMI[51], BCIC[53], ERN[54], SSC[48]	Accuracy
	MAEEG[34]	Temporal-domain mask	Transformer	Sleep stage classification	MGH[46]	Accuracy
	Wavelet2vec[35]	Frequency-domain mask	ViT	Seizure detection	CHSZ[55], TUSZ[56]	Accuracy, BCA, F1, MAE

extends the SimCLR [20] framework to extract channel-level features from EEG data.

TS-TCC [21] generates different views of input data using both strong and weak augmentation methods. Weak augmentation employs jittering and scaling strategies, while strong augmentation uses permutation and jittering strategies, applying them to the temporal contrast module of EEG signals for temporal representation learning. This method maximizes the similarity between contexts of the same sample while minimizing the similarity between contexts of different samples. Jiang et al. [22] applies transformations such as horizontal flipping and adding Gaussian noise to EEG signals, then learns the correlation between signals by measuring the feature similarity of these transformed signal pairs. Additionally, the authors explore the impact of transformation combinations on the network's representation capability to find the optimal combination for downstream tasks. mulEEG [23] proposes a novel multi-view self-supervised method. By designing EEG augmentation strategies and introducing a diversity loss function, mulEEG effectively leverages complementary information from multiple views to learn better representations. However, these EEG data augmentation methods often lead to sampling bias [24], especially for noisy EEG data, which can significantly affect performance [25]. To address these limitations, ContraWR [26] constructs positive sample pairs using data augmentation and employs global average representations as negative samples to provide contrastive information, thereby learning robust EEG representations without labels. Additionally, ContraWR assigns greater weight to closer samples when calculating the global average.

Existing contrastive learning methods primarily focus on a single data level and fail to fully exploit the complexity of EEG signals. Therefore, COMET [27] leverages all data levels of medical time-series, including patient, trial, sample, and observation levels, to design a hierarchical contrastive representation learning framework. Its advantage lies in fully utilizing the hierarchical structure of medical time-series, enabling a more comprehensive understanding of the intrinsic relationships within the data.

3.1.2 *Combined with Expert Knowledge*

Expert knowledge contrastive learning is a relatively new representation learning framework. Generally, this modeling framework incorporates expert prior knowledge or information into deep neural networks to guide model training. In a contrastive learning framework, prior knowledge can help the model select the correct positive and negative samples during training. SleepPriorCL [28] was proposed to mitigate the sampling bias problem in data augmentation-based contrastive learning. It is well known that each sleep stage occupies a certain frequency range. The authors utilized this fact to calculate the energy of these frequency bands and used it as prior knowledge for training. Specifically, the authors calculated the rhythm energy vector $E = [E(\delta), E(\theta), E(\alpha), E(\beta)]$ for each EEG segment x, referred to as prior features, and then defined the dissimilarity $d_{i,j}$ between the anchor x_i and the sample x_j as follows:

$$d_{i,j} = \log\left(\|E_i - E_j\|_2\right) \tag{2}$$

Samples are ranked by dissimilarity, with the top K samples selected as positive samples and the rest as negative samples. Additionally, SleepPriorCL introduces a mechanism to adjust the gradient penalty strength of each sample based on its confidence as a positive or negative sample. To achieve this, each sample is assigned a customized temperature. The multi-positive contrastive loss is modified as follows:

$$\mathcal{L}(x_i) = \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(s_{i,p}/\tau_p)}{\exp(s_{i,p}/\tau_p) + \sum_{n \in N(i)} \exp(s_{i,n}/\tau_n)} \tag{3}$$

where x_i is the sleep epoch, $s_{i,j}$ is the cosine similarity between z_i and z_j, and z_i and z_j are the vectors of x_i after encoding and projection. The index i is referred to as the anchor, the index p as the positive sample, $N(i)$ is the set of all negative samples in the batch, and

the index n as the negative sample. $P(i)$ is the set of positive samples containing all true positive samples of x_i in the batch.

KDC2 [29] is based on the neural theory of EEG generation, which states that EEG signals are produced by synchronized synaptic activity that stimulates neuronal excitation, generating a negative extracellular voltage that transforms neurons into dipoles. The voltage generated by the dipoles is transmitted to the scalp via capacitive and volume conduction and is captured by electrodes as EEG signals. Therefore, the authors constructed scalp and neural views to describe the external and internal information of brain activity, respectively, and designed a knowledge-driven cross-view contrastive loss to extract neural knowledge by contrasting the same augmented samples between views. Positive sample pairs are composed of representations of the same augmented samples in different views, while negative sample pairs are composed of representations of different augmented samples in different views. By minimizing the distance between positive sample pairs and maximizing the distance between negative sample pairs, the model learns complementary features that describe the internal and external manifestations of brain activity. The designed cross-view contrastive loss can be calculated as follows:

$$\mathcal{L}_{cross} = -\frac{1}{|\mathcal{B}|} log(\frac{pair^+}{pair^+ + pair^-}) \tag{4}$$

$$pair^+ = \sum_{b\in\mathcal{B}} \sum_{i=0}^{m} exp(s(r^i_{sa,b}, r^i_{ta,b})/\tau) \tag{5}$$

$$pair^- = \sum_{b\in\mathcal{B}} \sum_{i=0}^{m} \sum_{j=i+1}^{m} exp(s(r^i_{sa,b}, r^j_{ta,b})/\tau) \tag{6}$$

where $pair^+$ and $pair^-$ represent the cross-view positive and negative pairs, respectively, $\mathcal{B}$ is the sample batch, and τ is the temperature parameter. The function $s(\cdot)$ represents the cosine similarity. The representation generated from the scalp view is denoted as r_s, and the representation generated from the inner neural topology view is denoted as r_t. r_{sa} and r_{ta} represent the corresponding augmented samples, and b indexes the samples contained in the batch.

3.2 Mask Autoencoder Approaches

Masked language modeling is a widely adopted method for pre-training in NLP. BERT [30] retains a portion of the input sequence and predicts the missing content during the training phase, which generates effective representations for various downstream tasks. MAE can be represented as:

$$x_m = \mathcal{M}(x), \quad z = E(x_m), \quad \tilde{x} = D(z), \tag{7}$$

$$\mathcal{L} = \mathcal{M}(\|x - \tilde{x}\|_2) \tag{8}$$

where $\mathcal{M}(\cdot)$ denotes the masking operation, x_m represents the masked input, $E(\cdot)$ and $D(\cdot)$ represent the encoder and decoder.

Inspired by this, BENDR [31] follows the wav2vec2.0 [32] architecture. It first encodes EEG data into temporal embeddings using 1D convolutions, then creates a mask vector to randomly mask these embeddings. A transformer-based module [33] is then used to extract temporal correlations and output the reconstructed embeddings. The contrastive loss function aims to make the reconstructed embeddings as similar as possible to the original unmasked embeddings while making them as different as possible from the remaining embeddings. It can be calculated as follows:

$$\mathcal{L} = -log\frac{exp(cossim(c_t, b_t))/\kappa}{\sum_{b_i\in B_D} \exp(cossim(c_t, b_i))/\kappa} \tag{9}$$

where c_t represents the output of the transformer module at position t, b_i represents the original vector at some offset i, B_D is a set of 20 negative samples uniformly selected from the same sequence, along with b_t, $cossim$ denotes the cosine similarity, and κ is a temperature parameter controlling the contrastive loss.

MAEEG [34] has a similar structure to BENDR but includes two additional layers to map the output of the transformer module back to the original EEG dimensions. The reconstruction loss is calculated by comparing the reconstructed EEG ($\hat{\mathrm{x}}$) with the input EEG (x) signal, using the formula $1 - \frac{\hat{\mathbf{x}}\cdot\mathbf{x}}{\|\hat{\mathbf{x}}\|\|\mathbf{x}\|}$. The key difference between BENDR and MAEEG is that MAEEG learns representations by minimizing the reconstruction loss rather than using contrastive learning.

Unlike the above two methods that mask temporal embeddings, WAVELET2VEC [35] performs masking and reconstruction tasks in different frequency bands to capture time-frequency information. Specifically, the authors apply low-pass and high-pass filtering to the raw EEG signal, recursively calculate the coefficients of each level of decomposition, and obtain wavelets in different frequency bands. They then

design an encoder consisting of six parallel ViT [36] units, each corresponding to a frequency band wavelet. Each wavelet is flattened and divided into patches, and 10% of the input patches are randomly masked. The decoder reconstructs the missing patch sequences, and self-supervised pre-training is performed by minimizing the Euclidean distance between the patch sequences of the original signal and the reconstructed patch sequences. This method forces the model to learn the time-frequency information and understand its correlations by masking the frequency patch sequences of the EEG.

3.3 Discussion

Contrastive learning and Masked Autoencoders (MAE) have demonstrated significant advantages in EEG analysis. Contrastive learning effectively extracts feature representations by exploring the similarities and differences among samples, while MAE enhances the model's understanding of data by predicting missing information. These self-supervised learning methods not only reduce dependency on large amounts of labeled data but also improve model generalization.

However, current self-supervised learning methods have some limitations. First, contrastive learning often relies on carefully designed data augmentation strategies, which may disrupt the temporal dependencies in EEG data and negatively affect the model's learning effectiveness. Additionally, the analysis and processing of multi-channel EEG data remains complex, and existing methods still face challenges in effectively handling multi-channel signals. Although MAE can capture the intrinsic characteristics of input data, its mask strategy may lead to information loss in some cases, thereby affecting reconstruction quality and downstream task performance.

Future research directions could focus on the following aspects: developing more efficient and flexible data augmentation techniques to better preserve the structural characteristics of EEG data while ensuring that augmentation does not compromise the physiological significance of the signals. Given the complexity of multi-channel time series data, researchers should continue to design self-supervised learning frameworks capable of handling multi-channel signals effectively, thereby improving model performance in practical applications. Additionally, combining contrastive learning and MAE could be explored to ensure that the reconstructed data not only resembles the original data but also forms meaningful distinctions with other samples, potentially mitigating issues related to information loss.

4 Discriminative-based EEG Analysis

For a more profound comprehension of brain activity, this survey examines advanced architectures, including: **Graph Neural Networks (GNNs)** in section 4.1: These networks capitalize on the structural information inherent in brain connectivity to offer deeper insights. **Foundation Models** in section 4.2: Models pre-trained on extensive datasets and adaptable for specific EEG analysis tasks through fine-tuning. **LLMs-based Methods** in section 4.3: Leveraging the power of large language models to improve the interpretability of EEG data.

4.1 Graph Neural Networks

EEG data is a type of multi-channel time series data, in which multiple channels (brain regions) are related to each other, with structural and functional connectivity [57]. Due to brain regions are in non-Euclidean space, graph is the most appropriate data structure to indicate brain connection [58]. In recent years, graph neural networks(GNN), represented by graph convolutional networks(GCN) [59], have developed rapidly and become a powerful tool for learning non-Euclidean data representations. They are able to capture intricate relationships inter-variable and inter-temporal, therefore emerging as one of the mainstream frameworks for modeling multivariate time series. Motivated by the success of graph representation learning, a line of studies has utilized GNNs to perform multivariate time series analysis and demonstrate promising results in many downstream tasks such as classification [60], forecasting [61], and anomaly detection [62]. The survey by Jin et al. [9] has summarized the application of GNNs in time series analysis, but it does not specifically concentrate on EEG data and only briefly outlines the application in the field of healthcare. In contrast, this paper mainly focuses on EEG data, reviews the recent advances in mainstream EEG analysis tasks with GNNs. It covers a wide range of tasks such as epilepsy detection, sleep staging, and emotion recognition, and sorts out related works from the perspective of EEG graph construction and dependency modeling(as shown in Figure 3). All of the methods are summarized in Table 2.

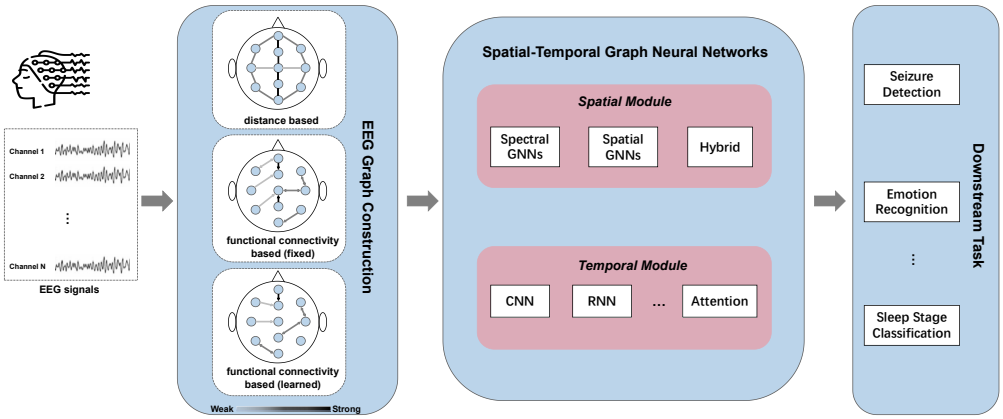


Figure 3. General pipeline for EEG analysis using graph neural networks.

Table 2. Summary of representative GNN-based Methods for EEG Analysis.

Task	Method	Graph Construction	Dependency Modeling	Training	Datasets	Metric
Sleep Stage Classification	GraphSleepNet[58]	Learned	Spectral, Attention, CNN	-	MASS-SS3[50]	Accuracy, F1, Kappa
	MSTGCN [69]	Learned	Spectral, Attention, CNN	-	ISRUC-S3[40], MASS-SS3[50]	Accuracy, F1, Kappa
Emotion Recognition	HetEmotionNet [72]	FC	Spectral, GRU	-	DEAP[74], MAHNOB-HCI[75]	Valence, Arousal
	MD-AGCN [70]	FC, Learned	Spatial	-	SEED, SEED-IV, SEED-V[38]	Accuracy
Seizure Detection	Tang et al. [67]	SC, FC	Spatial, Spectral, GRU	Generative Learning	TUSZ[37]	AUC, F1
	BrainNet [73]	Learned	Spatial	Contrastive Learning	Private data	Precision, Recall, F1, F2, AUC
	MBrain [2]	Learned	-	Contrastive Learning	Private data, TUSZ[37]	Precision, Recall, F1, F2
	EEG-CGS[68]	SC, FC	-	Contrastive Learning and Generative Learning	TUSZ[37]	AUC, Precision, F1 Sensitivity, Specificity
Sleep Stage Classification and Emotion Recognition	BayesEEGNet [71]	Learned	Spatial	-	MASS-SS3[50], SEED[38] ISRUC-S3 [40]	Accuracy, F1, Kappa

Graph Construction: "SC" and "FC" denote "structural connectivity" and "functional connectivity", respectively. "Learned" indicates that the graph structure is learned from data.

4.1.1 EEG Graph Construction

In general, each channel in the EEG signal is considered as a node in the graph. Referring to structural connectivity and functional connectivity, the methods for calculating adjacency matrix can be roughly divided into two categories. One is based on the geometry of EEG channels, the other is based on functional connectivity between brain regions. Based on the geometry between the channels, i.e., the anatomical connections between brain regions, previous studies have presented that adjacent brain regions affect each other and the strength of the impact is inversely proportional to the actual physical distance [63]. Thus, the adjacency matrix of the graph is constructed from the Euclidean distance between the electrodes, and it is worth noting that this matrix is the same for all EEG. The other is based on functional connectivity between brain regions, which captures dynamic brain connections that vary between different EEG. It is often calculated based on correlations or dependencies among signals, and the most common methods are Pearson Correlation Coefficient(PCC) [64], Mutual Information(MI) [65], and Phase Locking Value(PLV) [66].

Tang et al. [67] utilizes the above two methods to construct EEGs as graphs and only uses one type of graph as input at a time. Experimental results on the TUSZ v1.5.2 dataset show that the correlation-based graph structure can better localizes focal seizures than the distance-based graph. For a given EEG, Ho et al. [68] employs four different metrics to construct graphs, including nodes Euclidean distance, randomly connection of nodes, node features correlations, and directed transfer function. The first two are meant to capture the geometry of EEG channels and the last two are for capturing connectivity of brain regions.

Although the correlation-based graph can be used even when the physical locations of electrodes are unknown, the adjacency matrix is still fixed, which limits its performance to a certain extent. To solve this problem, a lot of research has explored adaptive graph learning strategies. For example, GraphSleepNet [58] learns the connection relationship between two nodes based on their input features. Specifically, it is implemented through a layer neural network. If the distance between the features of the two nodes is larger, the connection of the two in the adjacency matrix is smaller. And the loss function is defined to be optimized towards this direction. The superiority of adaptive (learnable) adjacency matrix is demonstrated by comparing it with fixed adjacency matrices in the experiment. MSTGCN [69] uses the adaptive graph learning

method proposed by GraphSleepNet [58], and also computes the spatial distance-based brain graph. Both views serve as the input of the model to extract features and a concatenate operation is employed to perform feature fusion on the two views. The results of the ablation experiment show that multi-view fusion is more effective than using only one single view. MD-AGCN [70] constructs temporal domain functional brain connectivity and frequency domain functional brain connectivity, respectively. Pearson's correlation coefficient is used as the connectivity index in the temporal domain. The frequency-domain adjacency matrix is divided into public part and private part. Public part is shared by all of the samples and is set to be trainable parameters, which illustrates the general functional brain connectivity patterns for emotional recognition. Private part is obtained by computing the dot product between two vertexes, and is unique to each sample. Before performing classification, functional brain connections in the two domains are combined together. By visualization of the learned graphs, the results indicate that the model can process global connectivities with the deep layers. BayesEEGNet [71] considers an electrical impulse between two nodes in the brain as a Poisson process, the countless electrical impulses generated by the brain in a period are represented as an infinite number of connection probability graphs. Then, the countless graphs are coupled into a summary graph by superposition of Poisson distributions, and the summary graph is subsequently transformed into the functional connectivity graph through two three-layer MLPs. By comparing with the adaptive learning strategy proposed by GraphSleepNet [58], the connectivity graph obtained in this paper has the best performance in downstream tasks.

4.1.2 *Dependency Modeling and Graph Representation Learning*

Once the EEG graph is constructed, it is often necessary to model the dependencies in the graph to learn the representation that is more discriminative for the downstream task. For example, Tang et al. [67] models the spatial dependency in the EEG signals by graph diffusion convolution. And to model the temporal dependency in EEGs, Gated Recurrent Units(GRUs) is employed. Also, in order to learn task-agnostic representations, a self-supervised pretraining method that predicts preprocessed signals for the next time period is proposed. For GraphSleepNet [58], a spatial-temporal convolution is designed, which consists of graph convolutions for capturing spatial features and temporal convolutions for capturing temporal context information. Moreover, the attention mechanism is applied in the spatial dimension and the temporal dimension respectively to extract valuable information. BayesEEGNet [71] also employs the spatial-based graph convolution to aggregate neighbor information directly in the spatial domain. For the emotion recognition task based on multi-modal signals, HetEmotionNet [72] first combines the temporal domain feature vector and the mutual information based adjacency matrix to form a heterogeneous spatial-temporal graph at the current moment, and then stacks the heterogeneous graphs of all time steps to form a heterogeneous graph sequence. Next, the Graph Transformer Network(GTN) is used to model the heterogeneity of multi-modal signals by automatically extracting the meta-paths from the adjacency matrix set. GCN is used to capture the correlation between multi-modal signals, and GRU is applied to extract temporal domain features from the graph sequence obtained after GCN. BrainNet [73] utilizes GCN to model two types of brain wave diffusion processes. Concretely, cross-time diffusion models the propagation of longer epileptic waves between two consecutive time segments. Meanwhile, fast signal spreading within the same time segments of each channel are captured by inner-time diffusion. The experimental results show that both diffusion processes can promote the performance of seizure detection.

There are also methods to mine patterns in a graph by designing self-supervised learning tasks. To capture the correlation patterns in space and time, MBrain [2] proposes two self-supervised tasks. Instantaneous time shift that is based on multi-channel Contrastive Predictive Coding(CPC) aims to capture the short-term correlations focusing on spatial patterns and delayed time shift is used for temporal patterns in broader time scales. In addition, replace discriminative learning is designed to preserve the unique characteristics of each channel so as to achieve accurate channel-wise seizure prediction. Ho et al. [68] leverages a random walk with restart(RWR) technique to create two positive and one negative sub-graphs for every node in every constructed EEG graph, and employs them to perform contrastive learning. Also, a generative learning module is proposed to learn the contextual information hidden in the graph through reconstructing the target node anonymized in the positive sub-graphs, using the other node features and edges of the sub-graph. To promote

spatial consistency in multiple sensors, GCC [60] proposes novel graph augmentations including node augmentations and edge augmentations, to augment sensors and their correlations respectively. Next, a graph contrasting method is designed. Node-level Contrasting is achieved by contrasting sensors in different views within each sample while Graph-level Contrasting is achieved by contrasting the samples within each training batch. Through these two contrasting procedures, robust sensor-level features and global-level features can be learned.

4.1.3 Discussion

Due to the non-Euclidean nature of EEG signals, graphs have become one of the most suitable data structures for modeling EEG data. By capturing both structural and functional connectivity within the brain, Graph Neural Networks (GNNs) can effectively model the diffusion process of brain waves across channels (or brain regions), thereby revealing different sleep patterns, emotional states, and seizure activities, among others. This highlights the importance of GNNs as a significant method in the field of EEG data analysis.

However, existing approaches still face several challenges. Most graph construction methods are heuristic and rely on prior knowledge, which in turn necessitates extensive data to experimentally validate the performance and interpretability of these methods. Moreover, considering the clinical deployment in real-world settings, the generalization performance of the model and the ethical implications of data usage must be thoroughly investigated and addressed.

4.2 Foundation Models

Foundation models (FMs) [76], often known as large-scale pretrained models, are advanced neural networks trained on extensive datasets. These models possess a vast range of general knowledge and can recognize numerous patterns. As a result, they offer a flexible and comprehensive foundation for addressing various tasks across multiple domains. ChatGPT [77] is the most famous textural foundation model that has a powerful ability to understand and generate natural language texts, and can perform a variety of natural language processing tasks, including text classification, sentiment analysis, machine translation, etc., showing extremely high flexibility and generalization capabilities. CLIP [78] and SAM [79] are representative visual foundation models, which exhibit robust general understanding and reasoning performance. Foundation models consistently demonstrate high performance in diverse domains, from natural language processing to computer vision, showcasing their versatility and the potential to revolutionize the way AI systems interact with and understand the world.

In the field of EEG data processing, researchers usually proposed specially designed methods or models for specific data or tasks. However, data annotation in the medical field is more difficult and expensive than in other fields. As a result, the size of EEG medical data sets is usually small, which greatly restricts the capabilities of the model [73, 80]. The emergence of large language models provides a new solution for the processing of biological signal data such as EEG. Recently, a lot of work has begun to draw on the ideas of large language models, using a large amount of unlabeled data and unsupervised pre-training methods to build foundation models for EEG or biological signal data [11, 81–86]. These foundation models have learned a lot of knowledge about time series signals, can well represent EEG data, have generalization capabilities that previous models did not have, and can achieve excellent performance on different downstream tasks. Below, we outline the existing work related to foundation models in the field of EEG signals, considering the three important elements: data, model structure, and training methods. While the datasets themselves are thoroughly described in Table 6, this chapter will focus on how they are used in the process of EEG foundation models established.

While the datasets are crucial and will be extensively discussed, this chapter is dedicated to the presentation of the models and training methodologies. The summary of existing foundation models is shown as Table 3.

4.2.1 Model Structure

With the rapid development of deep learning, many model structures have emerged, such as Convolutional Neural Network (CNN) [89], Recurrent Neural Network (RNN) [90], Transformers [91], Mamba [92], etc. How to design a model structure suitable for processing time series signals is the top priority in building a foundation model. A good structure can allow the foundation model to better understand and learn the information and knowledge in time series signals. Most of the existing EEG foundation models construct the main model by stacking Transformer layers or convolutional blocks. Because both structures have strong scalability and are suitable for mining

Table 3. Summary of Foundation models for EEG Analysis.

Method	Model Structure	Training	Datasets	Metric
BrainBERT [87]	Transformer blocks	Masked Autoencoder	Private data	AUC
Neuro-GPT [82]	Convolutional blocks + Transformer blocks	Future Forecast	TUH EEG corpus[37]	MSE, Accuracy
Brant [81]	Transformer blocks	Masked Autoencoder	Private data	MSE, MAE, F1, F2
BFM [83]	Convolutional blocks	Contrastive Learning	AHMS corpus[88]	AUC, MAE
LaBraMs [11]	Convolutional blocks + Transformer blocks	Masked Autoencoder	Public data + Private data	Accuracy, AUROC, F1

information in time series signals.

Brant [81] has two encoders, temporal encoder and spatial encoder. The temporal encoder contains a 12-layer Transformer encoder and the spatial encoder contains a 5-layer Transformer encoder. They are used to capture the time correlation and channel correlation in time series signals, respectively. Salar et al. [83] built the foundation model based on an EfficientNet-style 1D convolutional neural network. Neuro-GPT [82] and LaBraM [11] use both convolutional layers and Transformers layers. They first use a small number of convolutional layers to preliminarily extract the features of time series signals and transform their dimensions, and then use a large number of Transformers layers to further capture the correlation between different sequence patches and better represent time series signals.

Since the input of the Transformer layer is tokens, and the time series data is a continuous value, the foundation model needs to convert the time series data into patches before subsequent calculations can be performed. A common approach is to split the original data by a fixed window size and a fixed strides. Specifically, given a neural signal $\boldsymbol{x} \in \mathbb{R}^{N \times C}$, where N is the number of timestamps and C is the number of electrode channels, we divide x with window size M and stride S to generate a set of patches $\boldsymbol{p} \in \mathbb{R}^{N_p \times C \times M}$, where $N_p = \lfloor \frac{N-M}{S} \rfloor + 1$ is the number of patches in each channel. After obtaining the segmented patches, additional position or frequency encoding information is usually added to them to help the model learn better. Some researchers [11] also map each patch to a fixed codebook in order to make the foundation model have a fixed vocabulary like a large language model. Specifically, it first represents the patch and then utilizes quantizer to quantize all the patch representations into the neural codebook embeddings. The codebook looks up the nearest neighbor of each patch in the neural codebook.

The parameter size of the existing foundation models in the EEG field is usually between tens and hundreds of millions, which is still relatively small compared to the parameters of large language models. This may be because the amount of EEG data is still much smaller than text data. However, we believe that with the continuous development of the field, the scale of the foundation model will continue to increase, and its capabilities will continue to increase.

4.2.2 *Training Methods*

In order for the model to learn useful knowledge from massive amounts of unlabeled data, it is essential to design an effective training method. A good training method is like a good teacher, which can make the learning process more efficient.

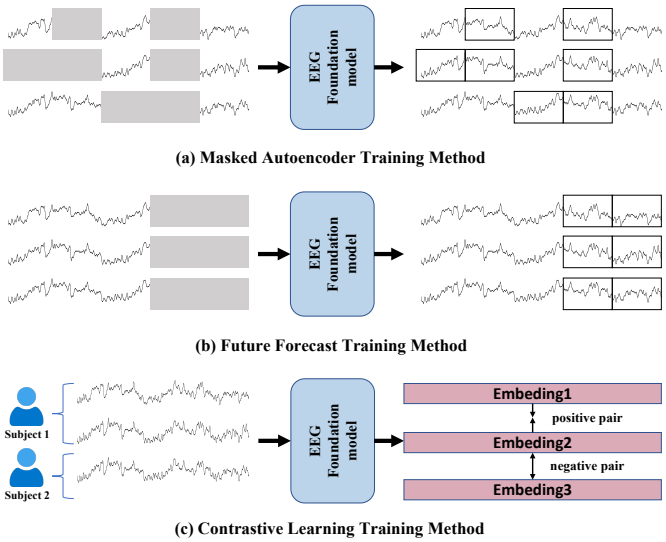


Figure 4. The different training methods of EEG foundation models.

Existing foundation models are all pre-trained using self-supervised methods. One of the mainstream approaches is to use masked autoencoder as a pre-training task [11, 81, 87]. Masked autoencoder has been proven to be a simple and effective method in many fields, which trains model to reconstruct the whole input given its partial observation (as shown in Figure 4(a)). In this way, the foundation model can be forced to infer the whole from partial information, so that the model can learn powerful representation capabilities.

There is another pre-training method that is similar to masked autoencoder, which can be understood as

masking only the latter part of the input (as shown in Firuge 4(b)). During the training process, the model predicts the future situation based on the historical content of the time series data [82]. Its goal is actually the same as the short-term or long-term prediction in the downstream task. Therefore, the foundation model pre-trained by this method usually has strong predictive ability, which can capture regularities from historical time series data.

Another type of work uses contrastive learning to train the foundation model. The core idea is to learn how to effectively distinguish similar (positive) and dissimilar (negative) data points by comparing data samples, so as to optimize the data representation or feature vector. This method can help the model capture the intrinsic structure and relationship between data, thereby improving its generalization ability on downstream tasks. For example, Salar et al. [83] constructed positive and negative pairs at the participant level. Specifically, the positive pairs are selected as augmented views of two different segments from the same participant, while the segments from different subjects are regarded as negative samples (as shown in Figure 4(c)). Through this training method, the model can not only acquire strong representation capabilities, but also enhance its generalization ability on different subjects.

Using various pre-training methods, the foundation model can acquire enough knowledge from a large amount of unlabeled data. Therefore, it only needs to be fine-tuned with a small amount of data to be well adapted to various downstream tasks. It can even have zero-shot capabilities like a large language model. This makes it possible to build a universal EEG foundation model.

4.2.3 Discussion

The emergence of foundation models for EEG data processing marks a significant advancement in the field of EEG signal analysis. By leveraging the principles and techniques of large-scale pretrained models, researchers can increasingly overcome the limitations posed by traditionally small and costly annotated EEG datasets. The ability of these foundation models to extract meaningful patterns from large amounts of unlabeled EEG data opens new avenues for improving diagnostic and therapeutic applications.

Despite these advancements, several challenges remain. Although the parameter size of existing EEG foundation models has significantly improved compared to before, it is still smaller than that of large language models because the amount of EEG data is far less than the amount of text data. This disparity highlights the need for more extensive and standardized EEG datasets, potentially through collaborative data-sharing initiatives or the integration of synthetic data generation techniques. Meanwhile, ethical considerations surrounding the use of EEG data must also be addressed. Issues of privacy, data security, and informed consent are paramount, especially as these models become more integrated into clinical workflows. Ensuring that these models are developed and implemented with a strong ethical framework will be crucial for their acceptance and success in the medical community.

4.3 LLMs-based Methods

Large Language Models (LLMs) [93–95] have revolutionized the field of natural language processing (NLP) by demonstrating remarkable capabilities in understanding, generating, and translating human language. The application of LLMs in EEG analysis represents a novel and innovative approach to interpreting complex brain signals. Unlike traditional machine learning methods, LLMs can be fine-tuned with relatively small amounts of task-specific data, making them particularly well-suited for the analysis of EEG data, which can be challenging to annotate and label.

The integration of LLMs into EEG analysis can take two forms: **Single-tower Models**: These approaches use LLMs as feature extractors for EEG data sets, which are of a single modality, implicitly leveraging the semantic knowledge that these models contain. Here, LLMs can be fine-tuned to classify different neurological states or forecast outcomes based on EEG data with Parameter Efficient Fine-Tuning (PEFT) techniques [129], such as LoRA [101] or soft prompt [130]. Their proficiency in handling sequential data makes them particularly adept at time-series analysis. **Dual-tower Models**: These approaches deals with multi-modal data, where EEG is paired with text using LLMs through knowledge distillation [131] or cross-modal contrastive learning [78]. What's more, there has been significant progress in adapting LLMs for general time series analysis[10, 12, 132]. For those familiar with the field, it is well understood that EEG data is a type of time series data. Given this, we are confident that the advancements made in general time series analysis can be successfully applied to EEG data analysis in the near future. Consequently, we intend to provide a brief overview of some mainstream methods

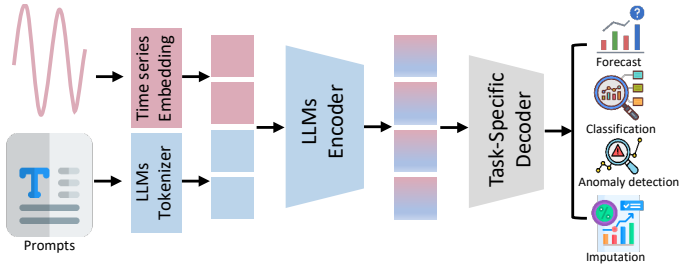
Table 4. Summary of LLMs-based Methods for EEG Analysis.

Method	Task	Language model	Training	Datasets	Metric
Victor[96]	Prediction	BERT[91]	Cross-entropy	American Epilepsy Society[97]	AUC, Accuracy
PromptCast[98]	Forecast	T5[99], BART[134], etc.	Template-Based Prompting	PISA	MAE, MSE
TEMPO[135]		GPT2[104] , T5[99], LLaMA[94], etc.	STL[100], LoRA[101]		
LLM4TS[103]		GPT2[104]	Autoregressive	**LTSF**[102]	
Time-LLM[105]		LLaMA[94]	Reprogramming, Prompt-as-Prefix		
S^2IP-LLM[106]		GPT2[104]	Partial fine-tune		
GPT4TS[107]	Classification, Forecast, etc.	GPT2[104]	Partial fine-tune	UEA[108]	Accuracy
TEST[109]	Classification	GPT2[104]	Contrastive learning		
Zhang[110]	Eye-tracking	GPT-3.5's and GPT-4's APIs	LLM agent	ZuCo[111]	AUC, Accuracy
EEG-To-Text [133]	Sentiment Classification	BART[134]	Reconstruction, cross-entropy	ZuCo[111]	AUC, Accuracy
MTAM[112]	Analysis, Relation Detection	Transformer	CCA, WD	K-EmoCon[113], ZuCo[111]	Precision, Recall,F1-score, Accuracy
METS[114]	Clinical Diagnosis	ClinicalBert[115]	Contrastive Learning	PTB-XL[116],MIT-BIH[117]	Precision, Recall, F1-score, Accuracy
GPT4MTS[118]	Forecast	BERT[91], GPT2[104]	Partial fine-tune	GDELT	MAE, MSE
ESI[119]	Diagnosis	BioLinkBERT[120]	Contrastive Learning, RAG	CSX[121], PTB-XL[116], MIT-BIH[117]	AUC, Accuracy
InstructTime[122]	Classification	GPT2[104]	VQ-VAE[123], Full fine-tuning	EEG[42], ECG[124], HAR[41], FD[125]	F1-score, Accuracy
CrossTimeNet[126]	Classification	BERT[91]	VQ-VAE[123]	EEG[42], ECG[124], HAR[41]	F1-score, Accuracy
CALF[136]	Forecast	GPT2[104]	Distill Knowledge	**LTSF**[102]	MAE, MSE
EEG-GPT[127]	Classification	Vinci GPT-3	prompt-completion API	TUH EEG Corpus[37]	AUC-ROC
K-Link[128]	Forecast	CLIP-Text[78]	Contrasting Learning	**LTSF**[102]	MAE, MSE

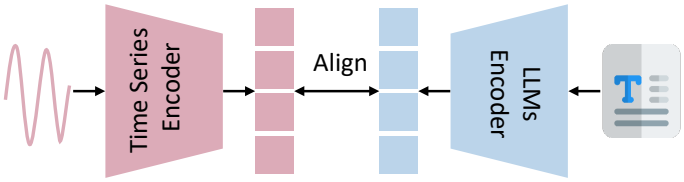
LTSF contains ETTh1/h2/m1/m2, Weather, Electricity, Traffic

currently utilized in general time series analysis. All of the methods are summarized in Table 4.

4.3.1 *Sinlge-tower Models*



(a) Single-tower methods.



(b) Dual-tower methods.

Figure 5. Two types of LLMs-based methods.

These approaches use LLMs as the backbone, harnessing the models' inherent semantic understanding(as shown in Figure 5(a)). Some works adapts them for time-series forecasting tasks. Victor et.al [96] first employs the Kolmogorov-Chaitin algorithm to convert EEG data into a text-like format, and then constructs a machine-learning model based on language models to predict epilepsy. PromptCast [98] introduces an innovative "codeless" approach to time series forecasting, offering a fresh perspective that moves away from the sole emphasis on creating complex architectures. TEMPO [135] concentrats exclusively on time series forecasting while integrating additional intricate elements such as time series decomposition and soft prompts. LLM4TS [103] proposes a two-stage fine-tuning framework for time-series forecasting, addresses challenges in incorporating LLMs with time-series data. Time-LLM [105] reprograms time series by incorporating the source data modality and utilizing natural language-based prompting, which unlocks the potential of LLMs as efficient time series machines. S^2IP-LLM [106] leverages LLMs by aligning their semantic space with time series embeddings to enhance time series forecasting through semantic space-informed prompt learning. The vast majority of existing research in the field has been centered on time-series forecasting tasks. This focus may stem from the inherent similarities between the autoregressive processes of LLMs and the forecasting nature of time-series prediction models. In other words, the resemblance lies in the fact that both types of models rely on historical data (or context) to make predictions about future data points (or words in the case of LLMs). In addition to forecasting, a few works have adapted LLMs for time-series classification. GPT4TS [107] presents a unified framework with frozening the self-attention and feedforward layers of the residual blocks in the LLMs and fine-tuning the layer norm layer. TEST [109] converts time-series data into a format suitable for pre-trained LLMs by employing a three-level contrast approach, which includes instance-wise, feature-wise, and text-prototype-aligned contrasts. Zhang et al. [110] utilize LLMs to generate labels that guide a new reading embedding representation for EEG, enabling the prediction of human reading comprehension at the word level. In summary, recent studies reflect a burgeoning interest in harnessing

the capabilities of LLMs for time-series analysis by integrating them into the architecture in ways that capitalize on the inherent strengths of LLMs.

4.3.2 *Dual-tower Models*

In addition to methods that focus solely on time series data, there have been significant efforts to develop multi-modal applications(as shown in Figure 5(b)). EEG-To-Text [133] presents a novel framework using LLMs to extend brain-to-text decoding to open vocabulary and achieve zero-shot sentiment classification. MTAM [112] uses a multimodal transformer alignment model to investigate the correlation between EEG data and language, enabling the observation of synchronized representations across these modalities and utilizing these aligned representations for various downstream tasks. METS [114] employs a trainable ECG encoder alongside a frozen language model to embed paired ECG signals and automatically generated clinical reports separately through multimodal contrastive learning. GPT4MTS [118] introduces a multimodal time series dataset for news impact forecasting and proposes a prompt-based LLM framework that leverages both numerical values and textual information. ESI [119] integrates a retrieval-augmented generation (RAG) pipeline to obtain external medical knowledge, thereby enriching textual descriptions. InstructTime [122] formulates the classification of time series as a multimodal understanding task, treating both task-specific instructions and raw time series data as multimodal inputs, with label information represented in text form. CrossTimeNet [126] designs a time series tokenization module that effectively converts raw time series data into a sequence of discrete tokens based on a reconstruction optimization process. CALF [136] develops a cross-modal match module to align cross-modal input distributions between textual and temporal data, further bridging the modality distribution gap in both feature and output spaces. EEG-GPT [127] offers intermediate reasoning steps and coordinate EEG tools across different scales, providing a transparent, interpretable, step-by-step analysis that enhances trustworthiness in clinical application. K-Link [128] proposes a framework that enriches a signal-derived graph by integrating a knowledge-link graph, which is constructed using LLMs, through the process of graph alignment. In summary, these efforts underscore the potential of integrating time series methods with the capabilities of LLMs to develop more robust and informative models. This is achieved through techniques that utilize a dual-tower architecture, such as cross-modal contrastive learning and knowledge distillation processes.

4.3.3 *Discussion*

The application of LLMs to EEG and other time series data modalities offers a promising approach, bridging the gap between advancements in natural language processing and time series analysis. By leveraging the inherent strengths of LLMs in semantic understanding and sequence processing, these models can unify various EEG tasks, as many of these tasks—like neurological state classification and signal forecasting—can be framed similarly to tasks in NLP. This positions LLMs as a competitive choice for building more generalizable models for EEG data analysis.

However, it is important to acknowledge the inherent limitations of current approaches that leverage LLMs as backbones for EEG analysis. LLMs may encounter difficulties with the unique characteristics of EEG data, including the requirement to capture fine-grained temporal patterns and the dynamic nature of evolving brain signals. Additional challenges arise as LLMs attempt to model the intricate dependencies within EEG data or to fully account for the broader topological relationships that may become relevant when integrating EEG with other data sources, such as clinical notes or external physiological signals. These limitations underscore the need for continued research and innovation in the application of LLM-based models to EEG data.

There remains active debate over whether LLMs are truly effective for time series analysis [137, 138]. Greater theoretical support is needed—for instance, in single-tower structures, where it has been suggested that the self-attention module functions analogously to principal component analysis (PCA) [139], and in dual-tower structures, which offer a probabilistic perspective that supports cross-modal fine-tuning techniques [136].

Future research directions should focus on enhancing the ability of LLMs to more efficiently process and understand the temporal and structural information embedded in EEG data. Given that LLMs were not originally designed to directly handle time series, new modeling techniques—such as structured embeddings and task-specific adaptations—are required to bridge the gap between natural language prompts and the detailed temporal patterns present in EEG. Insights

Table 5. Summary of EEG-To-Modality Generation Models.

Modality	Method	Encoder	Decoder	Pretrained	Dataset	Eval Metric
Image	Brain2Image[143]	LSTM	VAE	Classification	Spampinato[144]	IS
	ThoughtViz[145]	CNN	GAN	Classification	Kumar[146]	IS & Accuracy
	EEG2Image[147]	LSTM	DCGAN	Constrastive learning	Kumar[146]	IS
	EEGStyleGAN-ADA[148]	LSTM	SyleGAN-ADA	Constrastive learning	Spampinato[144] Kumar[146] Kaneshiro[149]	IS & FID & KID
	DreamDiffusion[150]	VQ	LDM	MAE	Spampinato[144]	Accuracy
	NeuroImagen[151]	Saliency Map, BLIP	LDM	Map	Spampinato[144]	IS & Accuracy & SSIM
Text	EEG-To-Text[133]	Transformer	BART[134]	Map	ZuCo[111]	BLEU-N & ROUGE-1
	EEG2Text[152]	Convolutional Transformer		MAE	ZuCo[111] Image-EEG[153]	
	E2T-PTR[154]	Multi-stream Transformer		MAE	ZuCo[111]	
	DeWave[155]	VQ-VAE		-	ZuCo[111]	
Others	ETCAS[156]	-	Dual-DualGAN	-	Privated data	Accuracy & PCC & MCD
	NDMusic[157]	-	BiLSTM	-	MusicAffect	Rank accuracy

from a recent survey [140] may provide valuable guidance. These developments will likely help unlock the full potential of LLMs in this domain and drive further advancements in EEG analysis through a language model-based approach.

5 Generative-based EEG Analysis

In this section, we will delve into innovative generative applications that utilize EEG data to produce images or text, providing novel approaches to the visualization and understanding of brain activity.In this section, we explore the performance of EEG analysis methods on multi-modal generation tasks. Previous works have proved that EEG signal contain abundant semantics. It's intuitively that we can reconstruct the semantics information from EEG signal instead of just catch their representation from raw data with the help of generative model such as GANs [141], Diffusion Models [142] and Transformers based models. All of the methods are presented in Table 5.

5.1 Image Generation

EEG-Image generation tasks typically follow the Map-Train-Finetune paradigm, which ensures high semantic fidelity but poses challenges in training and fine-tuning. As shown in Figure 6, the EEG-to-Image generation task involves three phases: data collection, model training, and testing. During the data collection phase, paired EEG signals and corresponding images are recorded while the subject views an image. This paired data is then used to jointly train the EEG encoder and image generator. In the testing phase, the trained model generates images directly from EEG signals. Brain2Image [143] addresses these challenges by dividing the EEG-Image generation task into two distinct phases. In the first phase, Brain2Image encodes EEG signals into a lower-dimensional feature vector for conditioning in image generation. Specifically, a standard LSTM layer followed by a nonlinear layer is trained to classify the EEG signals, serving as the encoder.

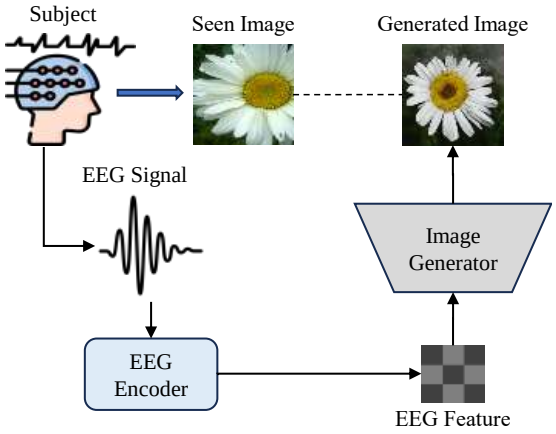


Figure 6. EEG based image generation task pipeline.

An additional fully-connected layer is then added to ensure the learned EEG feature vector follows a Gaussian distribution, as required by Variational Autoencoders (VAEs). In the second phase, for each EEG sequence provided to the encoder, Brain2Image uses the encoder's output to train the VAE's decoder to generate images corresponding to what the subject is observing at that precise moment. Compared to Brain2Image, ThoughtViz [145] employs a 1D-CNN followed by a 2D-CNN for EEG classification as an encoder. Building on the traditional GAN architecture, ThoughtViz introduces a pre-trained classifier to classify the samples generated by the generator. The generator loss in ThoughtViz incorporates both the discriminative loss from the discriminator and the classification loss from the classifier. Unlike training the EEG encoder through a supervised classification task, EEG2Image [147] and EEGStyleGAN-ADA [148] employ a triplet loss-based contrastive learning approach in their proposed frameworks for EEG feature learning. The triplet loss function aims to minimize the distance between data points with the same labels while maximizing the distance between data points with different labels. This approach prevents the EEG encoder from compressing the representations into small, indistinct clusters. EEG2Image utilizes a Conditional DCGAN [158]

architecture with hinge loss for stable training, whereas EEGStyleGAN-ADA employs StyleGAN-ADA [159] with adaptive discriminator augmentation. This augmentation helps the discriminator effectively learn from limited data by augmenting real images during training.

With the powerful generative capabilities of Diffusion Models, an increasing number of researchers are applying these models to the EEG-Image generation task. DreamDiffusion [150], for instance, collects a large-scale unlabeled EEG dataset from the MOABB [160] platform and uses the MAE method for brain pretraining. During the fine-tuning stage, DreamDiffusion employs a projection layer to align brain latent representations with CLIP-Image semantic information. NeuroImagen [151], on the other hand, uses detail and semantic extractors to map EEG signals to pixel and CLIP-Text priors, which are then decoded by a pretrained Stable Diffusion model following the image-to-image pipeline.

5.2 Text Generation

Unlike EEG-image generation, EEG-text generation is a sequence-to-sequence process. As shown in Figure 7, the EEG-to-Text generation task involves collecting word-level EEG signals while the subject views text (e.g., "He likes apple"). Eye-tracking may also be utilized to align EEG signals with specific words. These word-level EEG signals are then processed by an EEG-to-Text model, which decodes the signals and generates the corresponding text. Inspired by machine translation applications using pretrained BART [134], Wang et al. [133] consider the human brain as a unique type of encoder. They treat each EEG feature sequence as an encoded sentence by the human brain and then train an additional encoder to map the brain's embeddings to the embeddings from the pretrained BART model. Instead of using the word-level EEG features crafted based on the eye-tracking data like [133], EEG2Text [152] directly use the sentence-level EEG signals as input to the model. Specifically, EEG2Text leverages EEG pre-training to enhance the learning of semantics from EEG signals and proposes a multiview transformer to model the EEG signal processing by different spatial regions of the brain. Wang et al. [154] introduced CET-MAE, a model that combines contrastive learning and masked signal modeling via a multi-stream encoder. It effectively learns EEG and text representations by balancing self-reconstructed latent embeddings with aligned text and EEG features. They also

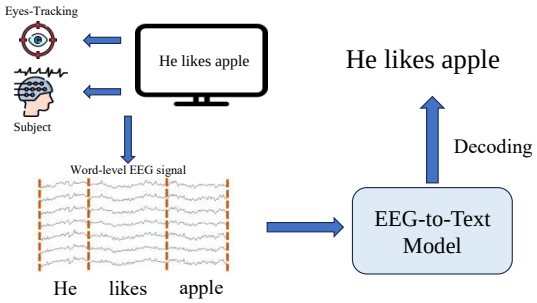


Figure 7. EEG based text generation task pipeline.

propose an EEG-to-Text decoding framework using Pretrained Transferable Representations, leveraging LLMs for language understanding and generation, and fully utilizing the pre-trained representations from CET-MAE. To address significant distribution variances in EEG waves across individuals and rectify order mismatches between raw wave sequences and text, DeWave [155] uses a vector quantized variational encoder. This encoder transforms EEG waves into a discrete codex, linking them to tokens based on proximity to codex book entries. DeWave is the first to introduce discrete encoding into EEG signal representation, benefiting both word-level EEG features and raw EEG wave translation.

5.3 Others

In addition to image and text generation, many other EEG-to-modality generation tasks deserve attention. ETCAS [156], an end-to-end GAN model tailored for EEG-based sound generation tasks, introduces a Dual-DualGAN to directly map EEG signals to speech signals. NDMusic [157] adopts an end-toend bidirectional LSTM (BiLSTM) architecture to establish a direct mapping from fMRI-informed EEG signals to music signals.

5.4 Discussion

The advancements in EEG-based generation tasks, spanning image, text, and even audio outputs, highlight the growing potential of generative models in decoding brain activity into multi-modal representations. The evolution of methods from simpler Map-Train-Finetune paradigms to more advanced approaches like contrastive learning and transformer-based architectures illustrates a robust progression toward higher semantic fidelity and model adaptability. Works such as Brain2Image [143], ThoughtViz [145], EEG2Image [147], and EEGStyleGAN-ADA [148] demonstrate the success of diverse model architectures—including GANs, StyleGANs, and VAEs—particularly in leveraging

novel feature extraction techniques that preserve the temporal and semantic richness of EEG data. Likewise, diffusion models, such as DreamDiffusion [150] and NeuroImagen [151], signify a leap in the generative quality and capacity to incorporate external semantic information, revealing promising directions for highly detailed image reconstruction.

In the realm of EEG-text generation, models inspired by sequence-to-sequence frameworks in NLP, like BART [134] and multi-view transformers, enable more coherent mapping from EEG signals to language representations. Innovations such as CET-MAE [154] and DeWave [155] further address challenges related to individual variability and sequence alignment, showcasing effective strategies to bridge the distinct characteristics of EEG signals and natural language representations. These frameworks mark significant progress toward the seamless integration of pre-trained language models and EEG features, opening new avenues for interpretable and accurate text generation.

Future research should aim to address several critical challenges: (1) improving the robustness of EEG-based generation models across diverse data sources and populations; (2) enhancing data efficiency through unsupervised or few-shot learning approaches to mitigate the need for large labeled datasets; and (3) refining alignment techniques for cross-modal integration with clinical and physiological data. Additionally, continued exploration of discrete and structured representations, as in DeWave, could prove transformative for other EEG-based tasks by establishing a consistent framework for handling the complexity of EEG signals. These efforts will be vital in pushing the boundaries of brain decoding technologies and in developing universally applicable, robust EEG-based generative models for various modalities.

6 Datasets and Metrics

The analysis of spatio-temporal EEG data relies heavily on the availability of high-quality datasets and robust evaluation metrics. This section provides an overview of the most widely used datasets and the key metrics employed to assess the performance of various EEG analysis models.

6.1 Datasets

6.1.1 *Publicly Available EEG Datasets*

Several publicly available EEG datasets have been instrumental in advancing the field. These datasets vary in their focus, including different cognitive tasks, subject demographics, and recording conditions.

Discriminative EEG Task Dataset: These datasets are typically employed for tasks that involve distinguishing between different cognitive states or mental activities, such as classifying brain signals associated with motor imagery, attention, or emotional responses. Some of the most notable datasets include:

- **BCI Competition IV [161]**: This dataset comprises multiple sub-datasets, each designed for specific brain-computer interface (BCI) challenges. It includes motor imagery tasks and event-related potentials (ERPs) recorded from healthy subjects.
- **TUH EEG Corpus[37]**: The Temple University Hospital EEG Corpus is one of the largest publicly available EEG datasets. It contains EEG signals collected from 14,987 subjects, with more than 40 different channel configurations and different recording duration, including normal and abnormal samples, making it suitable for both research and clinical applications.
- **DEAP (Database for Emotion Analysis using Physiological Signals[74])**: This dataset includes EEG and other physiological signals recorded while subjects watched music videos. It is widely used for emotion recognition and affective computing studies.
- **CHB-MIT Scalp EEG Database[52]**: This dataset contains EEG recordings from pediatric subjects with intractable seizures. It is commonly used for seizure detection and prediction research.
- **SEED (SJTU Emotion EEG Dataset)[38]**: The SEED dataset includes EEG recordings from subjects experiencing emotional stimuli, such as movie clips. It is used to study emotional recognition and related applications.
- **ISRUC-S3 dataset[40]**: This dataset contains 10 healthy subjects. Each recording contains 6 EEG channels, 2 EOG channels, 3 EMG channels, and 1 ECG channel. It is widely used for sleep stage classification studies.
- **MASS-SS3 dataset[50]**: This dataset contains 62 healthy subjects. Each recording contains 20 EEG channels, 2 EOG channels, 3 EMG channels, and 1 ECG channel. It is widely used for sleep stage classification studies.

Generative EEG Task Dataset: These datasets are

typically used for tasks that involve the generation of images, sentences, and other signals. For the image generative task, Spampinato et al. [144], Kumar et al. [146], and Kaneshiro et al. [149] obtain image semantics from EEG by employing EEG data recorded while subjects looked at images on a screen. The classical dataset constructed for the generative EEG task is shown in Table 6.

Table 6. EEG-Image Dataset for Image Generation.

Item \ Dataset	Spampinato [144]	Kumar [146]	Kaneshiro [149]
Classes	40	30	6
Subjects	6	23	10
Channels	128	14	128
Quantity	2000	30	72
Frequency (Hz)	1000	2048	1000
Time(s)	0.5	10	0.5
Pause(s)	10	20	0.75

- Spampinato et al[144] employed a subset of ImageNet containing 40 classes of easily recognizable objects for visual stimuli, using a 128-channel cap (actiCAP 128Ch), Brainvision DAQs and amplifiers for the EEG data acquisition. Sampling frequency and data resolution were set, respectively, to 1000 Hz and 16 bits. During the recording process, 2,000 images (50 from each class) were shown in bursts for 0.5 seconds each. A burst lasts for 25 seconds, followed by a 10-second pause where a black image was shown for a total running time of 1,400 seconds (23 minutes and 20 seconds).
- Kumar et al[146] prepared a slide presentation that consisted of 20 text and 10 non-text items in 3 categories of object to the subjects, namely digits, characters and object images, each slide was showed for 10 seconds, then recording the EEG data via a wireless neuro-headset Emotiv EPOC+ at a frequency of 2048Hz and there was a 20 seconds gap between 2 record.
- Kaneshiro et al[149] used 72 images from 6 categories of real objects as visual stimuli, acquired the EEG data via 128-channel EGI HCGSN 110 nets in the frequency of 1000 Hz. Each image was displayed for 0.5 seconds, and there was a 0.75 second interval between each image.
- **ZuCo**[111] contains EEG and eyetracking data from 12 healthy adult native English speakers engaged in natural English text reading for 4 - 6 hours. This dataset covers two standard reading tasks and a taskspecific reading task, offering EEG and eye-tracking data for 21,629 words across 1,107 sentences and 154,173 fixations.

6.1.2 *Private EEG Datasets*

In addition to publicly available datasets, researchers often collect private EEG datasets tailored to specific research questions or applications. These datasets may focus on particular cognitive tasks, clinical conditions, or subject populations. Specifically, private data also forms the basis of foundation models, and while its importance has been highlighted in Table 4.2. Collecting custom datasets allows for greater control over experimental conditions and data quality, but it also requires significant resources and expertise.

- BrainBERT[87] collected stereo electroencephalogram (SEEG) data from 10 subjects(5 male, 5 female; aged 4-19, with a mean age of 11.9 and a standard deviation of 4.6) over 26 sessions, who are pharmacologically intractable epilepsy patients.
- BrainNet [73] collected 796 GB of SEEG data from a first-class hospital. The subjects suffering from epilepsy undergo a surgical procedure to implant 4 to 10 invasive electrodes, with 52 to 126 channels, in their brain. In total, the dataset contains 526 hours of 256Hz to 1024Hz recordings.
- MBrain [2] collected 550 GB of SEEG data from a first-class hospital. The subjects suffering from epilepsy undergo a surgical procedure to implant 4 to 10 invasive electrodes, with 52 to 124 channels, in their brain. In total, the dataset contains 470 hours of 1000Hz to 2000Hz recordings.
- Brant [81] collected 1.01 TB of SEEG data from a first-class hospital. The subjects undergo a surgical procedure to implant 4 to 11 invasive electrodes, each with 52 to 153 channels, in their brain. The dataset contains 2528 hours of 1000Hz recordings with more than 1 trillion timestamps. In addition, it also collected 29.39 GB and 43 hours of epilepsy labeled data for fine-tuning of specific downstream tasks.
- LaBraM [11] further collected 342.23 hours of data from more than 140 subjects through the ESI neural scanning system.

6.2 Metrics

Evaluating the performance of EEG analysis models involves several key metrics, which are crucial for

comparing different approaches and understanding their effectiveness. The most commonly used metrics include:

- **Accuracy**[1]: The proportion of correctly classified instances among the total instances. It is a fundamental metric for classification tasks but may be misleading for imbalanced datasets.
- **Precision and Recall**[2]: Precision is the proportion of true positive results among the predicted positives, while recall is the proportion of true positive results among the actual positives. These metrics are particularly useful for tasks with imbalanced classes.
- **F1 Score**[3]: The harmonic mean of precision and recall, providing a single metric that balances both concerns. It is especially useful when the dataset has imbalanced classes.
- **F2 Score**[3]: The harmonic mean of precision and recall, giving twice as much weight to recall. It is particularly useful in applications such as epilepsy detection, where missing positive instances (epileptic events) can be fatal.
- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC)**[162]: This metric evaluates the ability of a model to distinguish between classes, considering both the true positive rate and the false positive rate. It is widely used for binary classification tasks.
- **Mean Squared Error (MSE)**[4]: Used for regression tasks, MSE measures the average squared difference between predicted and actual values. Lower MSE indicates better model performance.
- **Mean Absolute Error (MAE)**[5]: Another metric for regression tasks, MAE measures the average absolute difference between predicted and actual values. It is less sensitive to outliers compared to MSE.
- **Cohen's Kappa**[6]: A statistical measure of inter-rater agreement for categorical items, which takes into account the possibility of agreement occurring by chance. It is useful for evaluating the reliability of classifications.
- **Inception Score (IS)**[163]: A metric used to evaluate the performance of generative models, such as Generative Adversarial Networks (GANs), by assessing the quality and diversity of the generated images. It calculates the classification probabilities of the generated images using a pre-trained Inception network, and measures both how distinct and realistic the generated images are. Higher scores indicate better performance in terms of generating high-quality and diverse images.
- **Frechet Inception Distance (FID)**[164]: A metric for evaluating the quality of generated images by comparing the feature distributions of these images to those of real images. Lower FID scores indicate more realistic and diverse generated images.
- **Kernel Inception Distance (KID)**[165]: A more robust measure of image quality in generative models than FID, KID compares the similarity of feature distributions between generated and real images using a kernel method. It provides a more nuanced assessment by considering both the mean and covariance of the feature distributions, making it sensitive to both the style and content of the images. Lower KID scores suggest better image generation performance.
- **Structural Similarity Index (SSIM)**[166]: A metric for assessing the visual similarity between two images. It evaluates the similarity by comparing the luminance, contrast, and structure of the images. The SSIM index ranges from 0 to 1, with values closer to 1 indicating higher similarity. It is commonly used to measure the effectiveness of image processing techniques like enhancement, compression, and super-resolution.
- **BLEU-N**[167]: A metric used to evaluate the quality of machine-translated text. It measures the correspondence between a machine's translations and human translations by comparing n-gram overlaps. Higher BLEU-N scores indicate better translation accuracy and fluency. BLEU stands for Bilingual Evaluation Understudy.
- **ROUGE-1**[168]: A metric used to evaluate the quality of automatic summarization and machine translation. It focuses on the overlap of unigrams (single words) between a generated summary or translation and a set of reference summaries or translations. Higher ROUGE-1 scores indicate a

[1] https://en.wikipedia.org/wiki/Accuracy_and_precision
[2] https://en.wikipedia.org/wiki/Precision_and_recall
[3] https://en.wikipedia.org/wiki/F-score
[4] https://en.wikipedia.org/wiki/Mean_squared_error
[5] https://en.wikipedia.org/wiki/Mean_absolute_error
[6] https://en.wikipedia.org/wiki/Cohen's_kappa

better match between the generated text and the reference texts.

- **Pearson Correlation Coefficient (PCC)**[7]: a statistical measure that expresses the linear correlation between two variables. It ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation), with 0 indicating no correlation. PCC is commonly used in finance and economics to assess the strength and direction of the relationship between variables.
- **Melcepstral distance**[169]: A measure used in audio processing to evaluate the similarity between two sound signals, often employed in speech recognition and audio analysis. It's calculated based on the Mel-cepstral coefficients derived from the Fourier transform of the audio. Lower melcepstral distances indicate more similar sounds.

In summary, the availability of diverse and high-quality datasets, combined with robust evaluation metrics, is essential for advancing spatio-temporal EEG data analysis. These resources enable researchers to develop, compare, and refine models, ultimately leading to more accurate and insightful interpretations of brain activity.

7 Concludes and Future Directions

Conclusion: In conclusion, this paper has reviewed the current advancements in EEG analysis, focusing on three key areas: representation learning, discriminative-based methods, and generative-based methods. These areas collectively enhance the precision, interpretability, and application scope of EEG signal analysis, addressing significant challenges and paving the way for future research.

- **Learning Useful Representation from EEG Signals**: The first step in understanding EEG signals is representation learning, where we automatically extract important information. Self-supervised learning techniques are effective in this process, helping us create strong representations of EEG signals. These representations improve our ability to interpret the data accurately and handle large amounts of brain signal data efficiently.
- **Identifying Patterns in EEG Signals**: Discriminative methods are crucial for recognizing different patterns or categories within EEG signals. Using advanced techniques like Graph Neural Networks (GNNs) and foundation models, we can gain deeper insights into brain activity by capturing these patterns effectively. Understanding these patterns is essential for deciphering complex neural processes.
- **Generating New Insights from EEG Signals**: Generative methods focus on generating new types of data from EEG signals. Techniques like diffusion models allow us to create images or text based on EEG data, providing innovative ways to visualize and understand brain activity. These generative methods also have applications in generating AI-generated content.

Future Directions: Looking ahead, several promising directions for future research in EEG signal analysis and understanding can be identified:

- **Enhanced Integration of Self-Supervised and Semi-Supervised Learning**: Further exploration into the integration of self-supervised and semi-supervised learning techniques could yield even more robust and generalized representations. This will enable better handling of diverse and complex EEG data with minimal labeled data, driving improvements in accuracy and efficiency.
- **Development of Advanced Network Architectures**: Continued innovation in network architectures, such as the refinement and combination of Mamba [92, 170], KAN [171], and MoE models [172], is essential. These advancements should focus on improving training efficiency and inference speed, particularly for deployment on mobile and edge devices. Research into optimizing these architectures for real-time analysis and low-power consumption is also crucial.
- **Expansion of Multimodal Generative Techniques**: Expanding the capabilities of multimodal generative techniques to include more diverse forms of data, such as tactile or olfactory signals, could open new avenues for EEG applications. Additionally, improving the quality and realism of generated outputs, whether they be images, text, or speech, will enhance their utility in practical scenarios, particularly for assisting individuals with disabilities.
- **Addressing Constrained Conditions in Brain Signals**: Variable missing [173], class-incremental

[7] https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

[174], and source-free domain adaptation [175] are constrained conditions in brain signal analysis that present significant challenges but also offer important research opportunities. Addressing these issues can enhance the accuracy and stability of analyses, leading to broad impacts in practical applications.

- **Interdisciplinary Collaboration and Real-World Applications**: Encouraging interdisciplinary collaboration between neuroscientists, computer scientists, and clinicians will be vital for translating these technological advancements into real-world applications. This includes the development of user-friendly interfaces and tools for clinical use, as well as ensuring the ethical and responsible deployment of these technologies.
- **Establishing a Unified Evaluation Benchmark**: As the volume of EEG data, task variety, and computational capabilities increase, establishing a comprehensive and standardized evaluation system becomes crucial. Similar to the challenges observed in general time-series analysis [176, 177], there is currently no unified benchmark or complete dataset for consistent comparisons in EEG-based methods. To address this, we advocate for the future development of unified evaluation metrics and standardized datasets, which would enable more comprehensive and fair comparisons between different methods, and help assess their practical utility more accurately.

By focusing on these future directions, the field of EEG signal analysis can continue to advance, providing deeper insights into brain function and enabling more effective applications in both clinical and non-clinical settings.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This work was supported by NSFC under grant 62136002 and 62477014, Ministry of Education Research Joint Fund Project under grant 8091B042239, and Shanghai Trusted Industry Internet Software Collaborative Innovation Center.

References

[1] David, O., Blauwblomme, T., Job, A. S., Chabardès, S., Hoffmann, D., Minotti, L., & Kahane, P. (2011). Imaging the seizure onset zone with stereo-electroencephalography. *Brain*, 134(10), 2898–2911. [CrossRef]

[2] Cai, D., Chen, J., Yang, Y., Liu, T., & Li, Y. (2023, August). MBrain: A Multi-channel Self-Supervised Learning Framework for Brain Signals. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 130–141). [CrossRef]

[3] Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, 16(3), 031001.

[4] Hosseini, M. P., Hosseini, A., & Ahi, K. (2020). A review on machine learning for EEG signal processing in bioengineering. *IEEE reviews in biomedical engineering*, 14, 204–218. [CrossRef]

[5] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.

[6] Jiang, X., Bian, G. B., & Tian, Z. (2019). Removal of artifacts from EEG signals: a review. *Sensors*, 19(5), 987. [CrossRef]

[7] Zhang, X., Yao, L., Wang, X., Monaghan, J., Mcalpine, D., & Zhang, Y. (2019). A survey on deep learning based brain computer interface: Recent advances and new frontiers. *arXiv preprint arXiv:1905.04149*, 66.

[8] Zhang, K., Wen, Q., Zhang, C., Cai, R., Jin, M., Liu, Y., ... & Pan, S. (2024). Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. [CrossRef]

[9] Jin, M., Koh, H. Y., Wen, Q., Zambon, D., Alippi, C., Webb, G. I., ... & Pan, S. (2024). A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10466-10485. [CrossRef]

[10] Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., ... & Wen, Q. (2024, August). Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 6555-6565). [CrossRef]

[11] Jiang, W. B., Zhao, L. M., & Lu, B. L. (2024). Large brain model for learning generic representations with tremendous EEG data in BCI. *arXiv preprint arXiv:2405.18765*. [CrossRef]

[12] Zhang, X., Chowdhury, R. R., Gupta, R. K., & Shang, J. (2024). Large language models for time series: A survey. *arXiv preprint arXiv:2402.01801*. [CrossRef]

[13] Jin, M., Wen, Q., Liang, Y., Zhang, C., Xue, S., Wang, X., ... & Xiong, H. (2023). Large models for time series and spatio-temporal data: A survey and outlook. *arXiv preprint arXiv:2310.10196*. [CrossRef]

[14] Yang, Y., Jin, M., Wen, H., Zhang, C., Liang, Y., Ma, L., ... & Wen, Q. (2024). A survey on diffusion models for

time series and spatio-temporal data. *arXiv preprint arXiv:2404.18886*. [CrossRef]

[15] Zhang, Z., Sun, Y., Wang, Z., Nie, Y., Ma, X., Sun, P., & Li, R. (2024). Large language models for mobility in transportation systems: A survey on forecasting tasks. *arXiv preprint arXiv:2405.02357*. [CrossRef]

[16] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2022). Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125*. [CrossRef]

[17] Liu, K., Xiao, A., Zhang, X., Lu, S., & Shao, L. (2023). Fac: 3d representation learning via foreground aware feature contrast. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9476-9485).

[18] Gao, T., Yao, X., & Chen, D. (2021). Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*. [CrossRef]

[19] Mohsenvand, M. N., Izadi, M. R., & Maes, P. (2020, November). Contrastive representation learning for electroencephalogram classification. In *Machine Learning for Health* (pp. 238-253). PMLR.

[20] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PMLR.

[21] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C. K., Li, X., & Guan, C. (2021). Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*. [CrossRef]

[22] Jiang, X., Zhao, J., Du, B., & Yuan, Z. (2021, July). Self-supervised contrastive learning for EEG-based sleep staging. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE. [CrossRef]

[23] Kumar, V., Reddy, L., Kumar Sharma, S., Dadi, K., Yarra, C., Bapi, R. S., & Rajendran, S. (2022, September). mulEEG: a multi-view representation learning on EEG signals. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 398-407). Cham: Springer Nature Switzerland.

[24] Chuang, C. Y., Robinson, J., Lin, Y. C., Torralba, A., & Jegelka, S. (2020). Debiased contrastive learning. *Advances in neural information processing systems*, 33, 8765-8775.

[25] Robinson, J., Chuang, C. Y., Sra, S., & Jegelka, S. (2020). Contrastive learning with hard negative samples. *arXiv preprint arXiv:2010.04592*. [CrossRef]

[26] Yang, C., Xiao, C., Westover, M. B., & Sun, J. (2023). Self-supervised electroencephalogram representation learning for automatic sleep staging: model development and evaluation study. *JMIR AI*, 2(1), e46769. [CrossRef]

[27] Wang, Y., Han, Y., Wang, H., & Zhang, X. (2024). Contrast everything: A hierarchical contrastive framework for medical time-series. *Advances in Neural Information Processing Systems*, 36.

[28] Zhang, H., Wang, J., Xiao, Q., Deng, J., & Lin, Y. (2021). Sleeppriorcl: Contrastive representation learning with prior knowledge-based positive mining and adaptive temperature for sleep staging. *arXiv preprint arXiv:2110.09966*. [CrossRef]

[29] Weng, W., Gu, Y., Zhang, Q., Huang, Y., Miao, C., & Chen, Y. (2023). A Knowledge-Driven Cross-view Contrastive Learning for EEG Representation. *arXiv preprint arXiv:2310.03747*. [CrossRef]

[30] Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. [CrossRef]

[31] Kostas, D., Aroca-Ouellette, S., & Rudzicz, F. (2021). BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience*, 15, 653659. [CrossRef]

[32] Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33, 12449-12460.

[33] Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

[34] Chien, H. Y. S., Goh, H., Sandino, C. M., & Cheng, J. Y. (2022). Maeeg: Masked auto-encoder for eeg representation learning. *arXiv preprint arXiv:2211.02625*. [CrossRef]

[35] Peng, R., Zhao, C., Xu, Y., Jiang, J., Kuang, G., Shao, J., & Wu, D. (2023, June). Wavelet2vec: a filter bank masked autoencoder for EEG-based seizure subtype classification. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. [CrossRef]

[36] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. [CrossRef]

[37] Obeid, I., & Picone, J. (2016). The temple university hospital EEG data corpus. *Frontiers in neuroscience*, 10, 196. [CrossRef]

[38] Zheng, W. L., Zhu, J. Y., & Lu, B. L. (2017). Identifying stable patterns over time for emotion recognition from EEG. *IEEE transactions on affective computing*, 10(3), 417-429. [CrossRef]

[39] Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A., & Oberye, J. J. (2000). Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering*, 47(9), 1185-1194. [CrossRef]

[40] Khalighi, S., Sousa, T., Santos, J. M., & Nunes, U. (2016). ISRUC-Sleep: A comprehensive public dataset for sleep researchers. *Computer methods and programs in biomedicine*, 124, 180-192. [CrossRef]

[41] Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013, April). A public domain dataset for human activity recognition using smartphones. In *Esann* (Vol. 3, p. 3). [CrossRef]

[42] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6), 061907. [CrossRef]

[43] Lessmeier, C., Kimotho, J. K., Zimmer, D., & Sextro, W. (2016, July). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *PHM Society European Conference* (Vol. 3, No. 1). [CrossRef]

[44] Guillot, A., Sauvet, F., During, E. H., & Thorey, V. (2020). Dreem open datasets: Multi-scored sleep datasets to compare human and automated sleep staging. *IEEE transactions on neural systems and rehabilitation engineering*, 28(9), 1955-1965. [CrossRef]

[45] Zhang, G. Q., Cui, L., Mueller, R., Tao, S., Kim, M., Rueschman, M., ... & Redline, S. (2018). The National Sleep Research Resource: towards a sleep data commons. *Journal of the American Medical Informatics Association*, 25(10), 1351-1358. [CrossRef]

[46] Biswal, S., Sun, H., Goparaju, B., Westover, M. B., Sun, J., & Bianchi, M. T. (2018). Expert-level sleep scoring with deep neural networks. *Journal of the American Medical Informatics Association*, 25(12), 1643-1650. [CrossRef]

[47] Escudero, J., Abásolo, D., Hornero, R., Espino, P., & López, M. (2006). Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy. *Physiological measurement*, 27(11), 1091.

[48] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23), e215-e220. [CrossRef]

[49] Van Dijk, H., Van Wingen, G., Denys, D., Olbrich, S., Van Ruth, R., & Arns, M. (2022). The two decades brainclinics research archive for insights in neurophysiology (TDBRAIN) database. *Scientific data*, 9(1), 333.

[50] O'reilly, C., Gosselin, N., Carrier, J., & Nielsen, T. (2014). Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research. *Journal of sleep research*, 23(6), 628-635. [CrossRef]

[51] Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on biomedical engineering*, 51(6), 1034-1043. [CrossRef]

[52] Shoeb, A. H. (2009). *Application of machine learning to epileptic seizure onset detection and treatment* (Doctoral dissertation, Massachusetts Institute of Technology).

[53] Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., ... & Blankertz, B. (2012). Review of the BCI competition IV. *Frontiers in neuroscience*, 6, 55. [CrossRef]

[54] Margaux, P., Emmanuel, M., Sébastien, D., Olivier, B., & Jérémie, M. (2012). Objective and Subjective Evaluation of Online Error Correction during P300-Based Spelling. *Advances in Human-Computer Interaction*, 2012(1), 578295. [CrossRef]

[55] Peng, R., Zhao, C., Jiang, J., Kuang, G., Cui, Y., Xu, Y., ... & Wu, D. (2022). TIE-EEGNet: Temporal information enhanced EEGNet for seizure subtype classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30, 2567-2576. [CrossRef]

[56] Loshchilov, I. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*. [CrossRef]

[57] Park, H. J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, 342(6158), 1238411. [CrossRef]

[58] Jia, Z., Lin, Y., Wang, J., Zhou, R., Ning, X., He, Y., & Zhao, Y. (2020, July). GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In *Ijcai* (Vol. 2021, pp. 1324-1330).

[59] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29.

[60] Wang, Y., Xu, Y., Yang, J., Wu, M., Li, X., Xie, L., & Chen, Z. (2024, March). Graph-Aware Contrasting for Multivariate Time-Series Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 14, pp. 15725-15734). [CrossRef]

[61] Cai, W., Liang, Y., Liu, X., Feng, J., & Wu, Y. (2024, March). Msgnet: Learning multi-scale inter-series correlations for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 10, pp. 11141-11149). [CrossRef]

[62] Deng, A., & Hooi, B. (2021, May). Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 5, pp. 4027-4035). [CrossRef]

[63] Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. D. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex*, 15(9), 1332-1342. [CrossRef]

[64] Pearson, K., & Lee, A. (1903). On the laws of inheritance in man: I. Inheritance of physical characters. *Biometrika*, 2(4), 357-462. [CrossRef]

[65] Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment*, 2005(09), P09008.

[66] Aydore, S., Pantazis, D., & Leahy, R. M. (2013). A note on the phase locking value and its properties. *Neuroimage*, 74, 231-244. [CrossRef]

[67] Tang, S., Dunnmon, J. A., Saab, K., Zhang, X., Huang, Q., Dubost, F., ... & Lee-Messer, C. (2021). Self-supervised graph neural networks for improved electroencephalographic seizure analysis. *arXiv preprint arXiv:2104.08336*. [CrossRef]

[68] Ho, T. K. K., & Armanfard, N. (2023, June). Self-supervised learning for anomalous channel detection in EEG graphs: Application to seizure analysis. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 7, pp. 7866-7874).

[69] Jia, Z., Lin, Y., Wang, J., Ning, X., He, Y., Zhou, R., ... & Li-wei, H. L. (2021). Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, 1977-1986. [CrossRef]

[70] Li, R., Wang, Y., & Lu, B. L. (2021, October). A multi-domain adaptive graph convolutional network for EEG-based emotion recognition. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 5565-5573). [CrossRef]

[71] Wang, J., Ning, X., Shi, W., & Lin, Y. (2023, April). A Bayesian Graph Neural Network for EEG Classification—A Win-Win on Performance and Interpretability. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)* (pp. 2126-2139). IEEE. [CrossRef]

[72] Jia, Z., Lin, Y., Wang, J., Feng, Z., Xie, X., & Chen, C. (2021, October). HetEmotionNet: two-stream heterogeneous graph recurrent neural network for multi-modal emotion recognition. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 1047-1056). [CrossRef]

[73] Chen, J., Yang, Y., Yu, T., Fan, Y., Mo, X., & Yang, C. (2022, August). Brainnet: Epileptic wave detection from seeg with hierarchical graph diffusion learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 2741-2751). [CrossRef]

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

[75] Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1), 42-55. [CrossRef]

[76] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. [CrossRef]

[77] Brown, T. B. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

[78] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PMLR.

[79] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4015-4026).

[80] Wagh, N., & Varatharajah, Y. (2020, November). Eeg-gcnn: Augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network. In *Machine Learning for Health* (pp. 367-378). PMLR.

[81] Zhang, D., Yuan, Z., Yang, Y., Chen, J., Wang, J., & Li, Y. (2024). Brant: Foundation model for intracranial neural signal. *Advances in Neural Information Processing Systems*, 36.

[82] Cui, W., Jeong, W., Thölke, P., Medani, T., Jerbi, K., Joshi, A. A., & Leahy, R. M. (2024, May). Neuro-GPT: Towards a foundation model for EEG. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE. [CrossRef]

[83] Abbaspourazad, S., Elachqar, O., Miller, A. C., Emrani, S., Nallasamy, U., & Shapiro, I. (2023). Large-scale training of foundation models for wearable biosignals. *arXiv preprint arXiv:2312.05409*. [CrossRef]

[84] Zhang, D., Yuan, Z., Chen, J., Chen, K., & Yang, Y. (2024, August). 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). [CrossRef]

[85] Yuan, Z., Zhang, D., Chen, J., Gu, G., & Yang, Y. (2024). Brant-2: Foundation Model for Brain Signals. *arXiv preprint arXiv:2402.10251*. [CrossRef]

[86] Chen, Y., Ren, K., Song, K., Wang, Y., Wang, Y., Li, D., & Qiu, L. (2024). EEGFormer: Towards transferable and interpretable large-scale EEG foundation model. *arXiv preprint arXiv:2401.10278*. [CrossRef]

[87] Wang, C., Subramaniam, V., Yaari, A. U., Kreiman, G., Katz, B., Cases, I., & Barbu, A. (2023). BrainBERT: Self-supervised representation learning for intracranial recordings. *arXiv preprint arXiv:2302.14367*. [CrossRef]

[88] Apple Heart & Movement Study – Study site for information and progress updates for AH&MS. https://appleheartandmovementstudy.bwh.harvard.edu/

[89] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

[90] Zaremba, W. (2014). Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*. [CrossRef]

[91] Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

[92] Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*. [CrossRef]

[93] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*. [CrossRef]

[94] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*. [CrossRef]

[95] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*. [CrossRef]

[96] Iapascurta, V., & Fiodorov, I. (2023, September). NLP Tools for Epileptic Seizure Prediction Using EEG Data: A Comparative Study of Three ML Models. In *International Conference on Nanotechnologies and Biomedical Engineering* (pp. 170-180). Cham: Springer Nature Switzerland.

[97] bbrinkm, & Will Cukierski. (2014). *American Epilepsy Society Seizure Prediction Challenge*. https://kaggle.com/competitions/seizure-prediction.

[98] Xue, H., & Salim, F. D. (2023). Promptcast: A new prompt-based learning paradigm for time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*. [CrossRef]

[99] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140), 1-67.

[100] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. off. Stat*, 6(1), 3-73.

[101] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*. [CrossRef]

[102] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34, 22419-22430.

[103] Chang, C., Peng, W. C., & Chen, T. F. (2023). Llm4ts: Two-stage fine-tuning for time-series forecasting with pre-trained llms. *arXiv preprint arXiv:2308.08469*. [CrossRef]

[104] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.

[105] Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., ... & Wen, Q. (2023). Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*. [CrossRef]

[106] Pan, Z., Jiang, Y., Garg, S., Schneider, A., Nevmyvaka, Y., & Song, D. (2024). $ S^ 2$ IP-LLM: Semantic Space Informed Prompt Learning with LLM for Time Series Forecasting. In *Forty-first International Conference on Machine Learning*.

[107] Zhou, T., Niu, P., Sun, L., & Jin, R. (2023). One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36, 43322-43355.

[108] Bagnall, A., Dau, H. A., Lines, J., Flynn, M., Large, J., Bostrom, A., ... & Keogh, E. (2018). The UEA multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*. [CrossRef]

[109] Sun, C., Li, H., Li, Y., & Hong, S. (2023). TEST: Text prototype aligned embedding to activate LLM's ability for time series. *arXiv preprint arXiv:2308.08241*. [CrossRef]

[110] Zhang, Y., Yang, S., Cauwenberghs, G., & Jung, T. P. (2024). From Word Embedding to Reading Embedding Using Large Language Model, EEG and Eye-tracking. *arXiv preprint arXiv:2401.15681*. [CrossRef]

[111] Hollenstein, N., Rotsztejn, J., Troendle, M., Pedroni, A., Zhang, C., & Langer, N. (2018). ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Scientific data*, 5(1), 1-13.

[112] Qiu, J., Han, W., Zhu, J., Xu, M., Weber, D., Li, B., & Zhao, D. (2023, December). Can brain signals reveal inner alignment with human languages?. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 1789-1804). [CrossRef]

[113] Park, C. Y., Cha, N., Kang, S., Kim, A., Khandoker, A. H., Hadjileontiadis, L., ... & Lee, U. (2020). K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data*, 7(1), 293.

[114] Li, J., Liu, C., Cheng, S., Arcucci, R., & Hong, S. (2024, January). Frozen language model helps ecg zero-shot learning. In *Medical Imaging with Deep Learning* (pp. 402-415). PMLR.

[115] Alsentzer, E., Murphy, J. R., Boag, W., Weng, W. H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. *arXiv preprint arXiv:1904.03323*. [CrossRef]

[116] Wagner, P., Strodthoff, N., Bousseljot, R. D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific data*, 7(1), 1-15.

[117] Moody, G. B., & Mark, R. G. (2001). The impact of

the MIT-BIH arrhythmia database. *IEEE engineering in medicine and biology magazine*, 20(3), 45-50. [CrossRef]

[118] Jia, F., Wang, K., Zheng, Y., Cao, D., & Liu, Y. (2024, March). GPT4MTS: Prompt-based Large Language Model for Multimodal Time-series Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 21, pp. 23343-23351). [CrossRef]

[119] Yu, H., Guo, P., & Sano, A. (2024). ECG Semantic Integrator (ESI): A Foundation ECG Model Pretrained with LLM-Enhanced Cardiological Text. *arXiv preprint arXiv:2405.19366*. [CrossRef]

[120] Yasunaga, M., Leskovec, J., & Liang, P. (2022). Linkbert: Pretraining language models with document links. *arXiv preprint arXiv:2203.15827*. [CrossRef]

[121] Zheng, J., Chu, H., Struppa, D., Zhang, J., Yacoub, S. M., El-Askary, H., ... & Rakovski, C. (2020). Optimal multi-stage arrhythmia classification approach. *Scientific reports*, 10(1), 2898.

[122] Cheng, M., Chen, Y., Liu, Q., Liu, Z., & Luo, Y. (2024). Advancing Time Series Classification with Multimodal Language Modeling. *arXiv preprint arXiv:2403.12371*. [CrossRef]

[123] Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 12.

[124] Cheng, M., Liu, Q., Liu, Z., Zhang, H., Zhang, R., & Chen, E. (2023). Timemae: Self-supervised representations of time series with decoupled masked autoencoders. *arXiv preprint arXiv:2303.00320*. [CrossRef]

[125] Liu, M., Ren, S., Ma, S., Jiao, J., Chen, Y., Wang, Z., & Song, W. (2021). Gated transformer networks for multivariate time series classification. *arXiv preprint arXiv:2103.14438*. [CrossRef]

[126] Cheng, M., Tao, X., Liu, Q., Zhang, H., Chen, Y., & Lei, C. (2024). Learning Transferable Time Series Classifier with Cross-Domain Pre-training from Language Model. *arXiv preprint arXiv:2403.12372*. [CrossRef]

[127] Kim, J. W., Alaa, A., & Bernardo, D. (2024). EEG-GPT: exploring capabilities of large language models for EEG classification and interpretation. *arXiv preprint arXiv:2401.18006*. [CrossRef]

[128] Wang, Y., Jin, R., Wu, M., Li, X., Xie, L., & Chen, Z. (2024). K-Link: Knowledge-Link Graph from LLMs for Enhanced Representation Learning in Multivariate Time-Series Data. *arXiv preprint arXiv:2403.03645*. [CrossRef]

[129] Han, Z., Gao, C., Liu, J., Zhang, J., & Zhang, S. Q. (2024). Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608*. [CrossRef]

[130] Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691*. [CrossRef]

[131] Hinton, G. (2015). Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531*. [CrossRef]

[132] Jiang, Y., Pan, Z., Zhang, X., Garg, S., Schneider, A., Nevmyvaka, Y., & Song, D. (2024). Empowering time series analysis with large language models: A survey. *arXiv preprint arXiv:2402.03182*. [CrossRef]

[133] Wang, Z., & Ji, H. (2022, June). Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 5, pp. 5350-5358). [CrossRef]

[134] Lewis, M. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*. [CrossRef]

[135] Cao, D., Jia, F., Arik, S. O., Pfister, T., Zheng, Y., Ye, W., & Liu, Y. (2023). Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*. [CrossRef]

[136] Liu, P., Guo, H., Dai, T., Li, N., Bao, J., Ren, X., ... & Xia, S. T. (2024). Taming Pre-trained LLMs for Generalised Time Series Forecasting via Cross-modal Knowledge Distillation. *arXiv preprint arXiv:2403.07300*. [CrossRef]

[137] Tan, M., Merrill, M. A., Gupta, V., Althoff, T., & Hartvigsen, T. (2024, June). Are language models actually useful for time series forecasting?. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

[138] Zheng, L. N., Dong, C. G., Zhang, W. E., Yue, L., Xu, M., Maennel, O., & Chen, W. (2024). Revisited Large Language Model for Time Series Analysis through Modality Alignment. *arXiv preprint arXiv:2410.12326*. [CrossRef]

[139] Zhou, T., Niu, P., Wang, X., Sun, L., & Jin, R. (2023). One fits all: Universal time series analysis by pretrained lm and specially designed adaptors. *arXiv preprint arXiv:2311.14782*. [CrossRef]

[140] Li, T., Kong, L., Yang, X., Wang, B., & Xu, J. (2024). Bridging Modalities: A Survey of Cross-Modal Image-Text Retrieval. *Chinese Journal of Information Fusion*, 1(1), 79-92. [CrossRef]

[141] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144. [CrossRef]

[142] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33, 6840-6851.

[143] Kavasidis, I., Palazzo, S., Spampinato, C., Giordano, D., & Shah, M. (2017, October). Brain2image:

Converting brain signals into images. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 1809-1817). [CrossRef]

[144] Spampinato, C., Palazzo, S., Kavasidis, I., Giordano, D., Souly, N., & Shah, M. (2017). Deep learning human mind for automated visual classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6809-6817).

[145] Tirupattur, P., Rawat, Y. S., Spampinato, C., & Shah, M. (2018, October). Thoughtviz: Visualizing human thoughts using generative adversarial network. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 950-958). [CrossRef]

[146] Kumar, P., Saini, R., Roy, P. P., Sahu, P. K., & Dogra, D. P. (2018). Envisioned speech recognition using EEG sensors. *Personal and Ubiquitous Computing*, 22, 185-199.

[147] Singh, P., Pandey, P., Miyapuram, K., & Raman, S. (2023, June). EEG2IMAGE: image reconstruction from EEG brain signals. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. [CrossRef]

[148] Singh, P., Dalal, D., Vashishtha, G., Miyapuram, K., & Raman, S. (2024). Learning Robust Deep Visual Representations from EEG Brain Recordings. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 7553-7562).

[149] Kaneshiro, B., Perreau Guimaraes, M., Kim, H. S., Norcia, A. M., & Suppes, P. (2015). A representational similarity analysis of the dynamics of object processing using single-trial EEG classification. *Plos one*, 10(8), e0135697. [CrossRef]

[150] Bai, Y., Wang, X., Cao, Y. P., Ge, Y., Yuan, C., & Shan, Y. (2023). Dreamdiffusion: Generating high-quality images from brain eeg signals. *arXiv preprint arXiv:2306.16934*. [CrossRef]

[151] Lan, Y. T., Ren, K., Wang, Y., Zheng, W. L., Li, D., Lu, B. L., & Qiu, L. (2023). Seeing through the brain: image reconstruction of visual perception from human brain signals. *arXiv preprint arXiv:2308.02510*. [CrossRef]

[152] Liu, H., Hajialigol, D., Antony, B., Han, A., & Wang, X. (2024). EEG2TEXT: Open Vocabulary EEG-to-Text Decoding with EEG Pre-Training and Multi-View Transformer. *arXiv preprint arXiv:2405.02165*. [CrossRef]

[153] Gifford, A. T., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, 264, 119754. [CrossRef]

[154] Wang, J., Song, Z., Ma, Z., Qiu, X., Zhang, M., & Zhang, Z. (2024). Enhancing EEG-to-Text Decoding through Transferable Representations from Pre-trained Contrastive EEG-Text Masked Autoencoder. *arXiv preprint arXiv:2402.17433*. [CrossRef]

[155] Duan, Y., Chau, C., Wang, Z., Wang, Y. K., & Lin, C. T. (2024). Dewave: Discrete encoding of eeg waves for eeg to text translation. *Advances in Neural Information Processing Systems*, 36.

[156] Guo, Y., Liu, T., Zhang, X., Wang, A., & Wang, W. (2023). End-to-end translation of human neural activity to speech with a dual–dual generative adversarial network. *Knowledge-Based Systems*, 277, 110837. [CrossRef]

[157] Daly, I. (2023). Neural decoding of music from the EEG. Scientific Reports, 13(1), 624.

[158] Radford, A. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*. [CrossRef]

[159] Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33, 12104-12114.

[160] Jayaram, V., & Barachant, A. (2018). MOABB: trustworthy algorithm benchmarking for BCIs. *Journal of neural engineering*, 15(6), 066011.

[161] Blankertz, B., Dornhege, G., Krauledat, M., Müller, K. R., & Curio, G. (2007). The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2), 539-550. [CrossRef]

[162] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145-1159. [CrossRef]

[163] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. *Advances in neural information processing systems*, 29.

[164] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

[165] Bińkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2018). Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*. [CrossRef]

[166] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600-612. [CrossRef]

[167] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).

[168] Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out* (pp. 74-81).

[169] Kubichek, R. (1993, May). Mel-cepstral distance measure for objective speech quality assessment.

In *Proceedings of IEEE pacific rim conference on communications computers and signal processing* (Vol. 1, pp. 125-128). IEEE. [CrossRef]

[170] Dao, T., & Gu, A. (2024). Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. *arXiv preprint arXiv:2405.21060*. [CrossRef]

[171] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., ... & Tegmark, M. (2024). Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*. [CrossRef]

[172] Ni, R., Lin, Z., Wang, S., & Fanti, G. (2024, April). Mixture-of-Linear-Experts for Long-term Time Series Forecasting. In *International Conference on Artificial Intelligence and Statistics* (pp. 4672-4680). PMLR.

[173] Yu, C., Wang, F., Shao, Z., Qian, T., Zhang, Z., Wei, W., & Xu, Y. (2024, August). Ginar: An end-to-end multivariate time series forecasting model suitable for variable missing. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 3989-4000). [CrossRef]

[174] Qiao, Z., Pham, Q., Cao, Z., Le, H. H., Suganthan, P. N., Jiang, X., & Savitha, R. (2024). Class-incremental learning for time series: Benchmark and evaluation. *arXiv preprint arXiv:2402.12035*. [CrossRef]

[175] Ragab, M., Eldele, E., Wu, M., Foo, C. S., Li, X., & Chen, Z. (2023, August). Source-free domain adaptation with temporal imputation for time series data. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 1989-1998). [CrossRef]

[176] Qiu, X., Hu, J., Zhou, L., Wu, X., Du, J., Zhang, B., ... & Yang, B. (2024). Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *arXiv preprint arXiv:2403.20150*. [CrossRef]

[177] Wang, Y., Wu, H., Dong, J., Liu, Y., Long, M., & Wang, J. (2024). Deep time series models: A comprehensive survey and benchmark. *arXiv preprint arXiv:2407.13278*. [CrossRef]

[178] Savran, A., Ciftci, K., Chanel, G., Cruz_Mota, J., Viet, L. H., Sankur, B., ... & Rombaut, M. (2006). Emotion detection in the loop from brain signals and facial images. In *eINTERFACE'06-SIMILAR NoE Summer Workshop on Multimodal Interfaces*.

[179] Trujillo, L. T., Stanfield, C. T., & Vela, R. D. (2017). The effect of electroencephalogram (EEG) reference choice on information-theoretic measures of the complexity and integration of EEG signals. *Frontiers in neuroscience*, 11, 425. [CrossRef]

Pengfei Wang is currently a PhD student at East China Normal University. His main research interests include multi-modal learning, time-series analysis, and healthcare applications. (Email: pfwang@stu.ecnu.edu.cn)

Huanran Zheng is currently a PhD student at East China Normal University. His main research interests include natural language processing, machine translation, large language models and brain-computer interface. (Email: hrzheng@stu.ecnu.edu.cn)

Silong Dai is currently a master student at East China Normal University. His main research interests include brain-signals analysis and AI-generated content. (Email: 51265901134@stu.ecnu.edu.cn)

Yiqiao Wang is currently a master student at East China Normal University. Her main research interests include sequence modeling and physiological signals classification. (Email: yqwang_01@stu.ecnu.edu.cn)

Xiaotian Gu is currently a master student at East China Normal University. His main research interests include time-series forecasting and brain-signals applications. (Email: 51265901064@stu.ecnu.edu.cn)

Yuanbin Wu received the Ph.D. from Fudan University in 2012. He is currently an Associate Professor in the School of Computer Science and Technology at East China Normal University. His research interests include natural language processing and machine learning (structured prediction, online learning). He is to build efficient and effective models for computing human languages. More details about his research can be found at https://faculty.ecnu.edu.cn/_s16/wyb/main.psp. (Email: ybwu@cs.ecnu.edu.cn)

Xiaoling Wang received the Ph.D. from Southeast University in 2003. She is currently a professor in the School of Computer Science and Technology at East China Normal University. Her research interests include web data management, web service technology, data mining, information retrieval, distributed graph data processing technology, knowledge graphs, sequential recommendation and sequential data analysis, and explainability techniques. More details about her research can be found at https://faculty.ecnu.edu.cn/_s16/wxl2/main.psp. (Email: xlwang@cs.ecnu.edu.cn).