

# Deep Learning-Powered Electrical Brain Signals Analysis: Advancing Neurological Diagnostics

Jiahe Li, Xin Chen, Fanqi Shen, Junru Chen, Yuxin Liu, Daoze Zhang, Zhizhang Yuan, Fang Zhao, Meng Li and Yang Yang

**Abstract**—Neurological disorders pose major global health challenges, driving advances in brain signal analysis. Scalp electroencephalography (EEG) and intracranial EEG (iEEG) are widely used for diagnosis and monitoring. However, dataset heterogeneity and task variations hinder the development of robust deep learning solutions. This review systematically examines recent advances in deep learning approaches for EEG/iEEG-based neurological diagnostics, focusing on applications across 7 neurological conditions using 46 datasets. For each condition, we review representative methods and their quantitative results, integrating performance comparisons with analyses of data usage, model design, and task-specific adaptations, while highlighting the role of pre-trained multi-task models in achieving scalable, generalizable solutions. Finally, we propose a standardized benchmark to evaluate models across diverse datasets and improve reproducibility, emphasizing how recent innovations are transforming neurological diagnostics toward intelligent, adaptable healthcare systems.

**Index Terms**—Deep learning, Neural Signal Analysis, Electroencephalography, Neurological Disorder Diagnosis

## I. INTRODUCTION

NEUROLOGICAL disorders are among the most significant global health challenges, with profound consequences for healthcare systems. According to the World Health Organization (WHO), neurological disorders affect over one-third of the global population, making them a leading cause of illness

and disability worldwide [1]. Dementia, affecting 57 million people, is a primary concern, with Alzheimer’s disease being the most common form [2]. Seizures impact more than 50 million individuals [3], while sleep disorders are widespread yet underdiagnosed [4]. Other significant disorders, including Parkinson’s disease [5], schizophrenia [6], depression [7], and ADHD [8], further exacerbate the burden, placing strain on healthcare systems. In low-income countries, where resources limit access to care, the situation is particularly dire.

Practical diagnostic tools are essential to alleviate growing global burden of neurological disorders, and electrical brain signals are indispensable. Specifically, electroencephalography is critical for understanding and diagnosing neurological disorders. Electroencephalography evaluates electrical activity in the brain and is categorized into scalp electroencephalography (EEG) and intracranial electroencephalography (iEEG). EEG is non-invasive, recording brain activity from electrodes on the scalp [9]. iEEG places electrodes into the brain (stereo-electroencephalography, SEEG) or onto brain’s surface (electrocorticography, ECoG), providing localized information [10].

The analysis of EEG/iEEG signals poses significant challenges for traditional machine learning (ML) approaches. These methods typically rely on manually engineered features that may not fully capture complex patterns in neurophysiological data, while their performance is often compromised by inherent noise and artifacts in raw recordings [11], [12]. Deep learning (DL) addresses these limitations by automatically extracting features, modeling temporal dependencies, and improving robustness against signal variability [13], [14]. The ability of DL to detect and classify neurological disorders with high accuracy has driven widespread adoption in brain signal analysis [15], [16]. This survey systematically examines the workflow of DL models in brain signal analysis, focusing on applications in diagnosing neurological disorders.

## A. General Workflow

The general workflow of EEG/iEEG analysis in neurological diagnostics is shown in Fig. 1, including three stages: signal collection, signal preprocessing, and analysis and diagnosis.

In the signal collection stage (Fig. 1.a), electrical brain activity is recorded using EEG/iEEG systems, typically across multiple channels at specific sampling rates with task-related labels. In the preprocessing stage (Fig. 1.b), techniques including denoising, filtering and normalization reduce noise and

Copyright (c) 2025 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org). Manuscript received 24 February 2025; revised 8 September 2025 and 23 October 2025; accepted 23 October 2025. Date of publication: date of current version. This work is supported by NSFC (62322606) and Zhejiang NSF (LR22F020005). (Corresponding author: Yang Yang and Meng Li.)

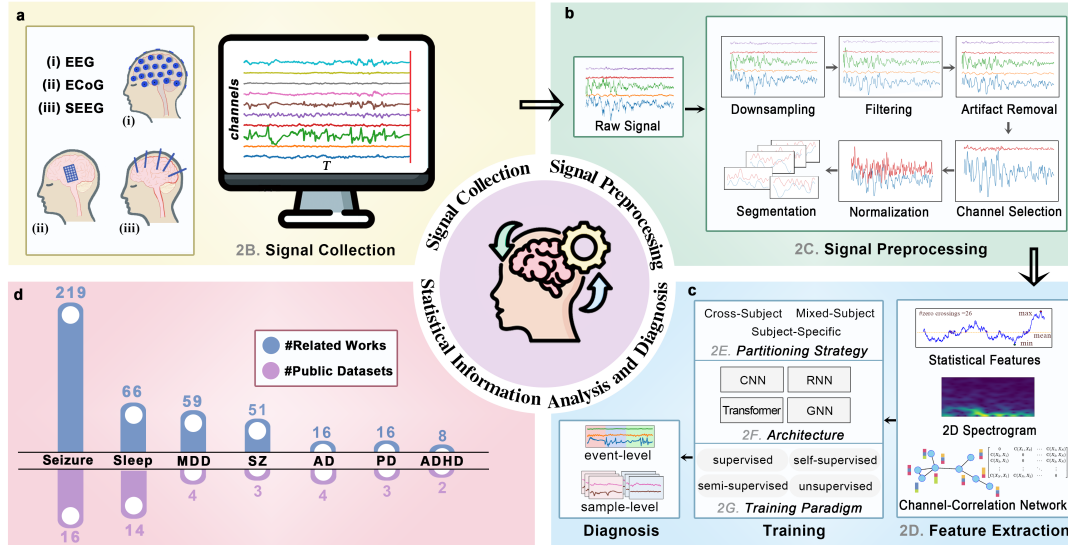
Jiahe Li, Xin Chen, Fanqi Shen, Junru Chen, Yuxin Liu, Daoze Zhang, Zhizhang Yuan and Yang Yang are with the College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310027, China (e-mail: [yangya@zju.edu.cn](mailto:yangya@zju.edu.cn)).

Fang Zhao is with the DiFint Technology (Shanghai) Co, Shanghai, 201210, China (email: [zhaofang@difint.cn](mailto:zhaofang@difint.cn)).

Meng Li is with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai, China, the School of Graduate Study, University of Chinese Academy of Sciences, Beijing, 100049, China, and The INSIDE Institute for Biological and Artificial Intelligence, Shanghai, 201210, China (email: [limeng.braindecoder@gmail.com](mailto:limeng.braindecoder@gmail.com)).

This article has supplementary downloadable material available at, provided by the authors.

Digital Object Identifier



**Fig. 1: General Workflow of Electrical Brain Signals Analysis in Neurological Diagnostics.**

- a. Signal Collection:** Acquisition of EEG/iEEG signals from patients, capturing brain electrical activity for clinical purposes.
- b. Signal Preprocessing:** A feasible workflow to process raw signals, ensuring their suitability for subsequent analysis.
- c. Analysis and Diagnosis:** Feature extraction and deep learning-based training for neurological classification.
- d. Statistical Information:** Statistical summary of resources for neurological conditions, including related work and datasets.

structure the data for feature extraction. In the analysis and diagnosis stage (Fig.1.c), preprocessed signals undergo feature extraction and classification. Traditional methods rely on manually designed features, whereas DL automatically learns diagnostically relevant patterns. Finally, the extracted features are applied to downstream tasks. Fig. 1.d highlights the distribution of related research efforts and publicly available datasets across various neurological conditions, including seizure, sleep disorders, major depressive disorder (MDD), schizophrenia (SZ), Alzheimer’s disease (AD), Parkinson’s disease (PD), and attention deficit hyperactivity disorder (ADHD).

## B. Related Studies and Our Contributions

Existing brain signal analysis surveys exhibit diverse scopes and focuses. Some focus on EEG, emphasizing their wide availability [17], [18]. Others broaden the scope to include brain signals like magnetic resonance imaging (MRI) [19], [20], which differ from EEG/iEEG in temporal resolution and preprocessing requirements. From a task perspective, some reviews focus specifically on diseases such as seizure [21], [22], providing in-depth insights into disease-specific applications. Others take a broader view, covering brain-computer interface (BCI) applications [23], [24], which focus on interaction and control, differing from neurological diagnostic tasks.

To provide a systematic perspective, we conducted a structured literature search in *PubMed*, *Science Direct*, and *Google Scholar* over the past ten years. Combinations of the terms *EEG/iEEG*, *deep learning*, and disease-related tasks were used. Studies focusing on clinically relevant diagnostics were retained, while those relying on traditional ML or non-healthcare applications were excluded. This process resulted in 450 manuscripts, from which information on publicly available datasets was extracted and cross-checked against repositories

like *PhysioNet*, *Zenodo*, and *OpenNeuro*, yielding 46 open datasets that form the empirical foundation of this survey.

Building on this systematic basis, our work establishes three contributions to advance deep learning-driven neurodiagnosis: First, we curate and analyze 46 public EEG/iEEG datasets across seven neurological conditions, establishing the most comprehensive data landscape to date while unifying fragmented methodologies by standardizing data processing, model architectures, and evaluation protocols. Besides, we identify self-supervised learning as the optimal paradigm for developing multi-task diagnostic frameworks, offering a comprehensive overview of pre-trained multi-task frameworks and their advancements. Additionally, we propose a benchmarking methodology to evaluate brain signal models across tasks, providing a foundation for scalable and versatile solutions in EEG/iEEG-based neurological diagnostics.

## II. METHODS

### A. Problem Definition

In this survey, we classify neurological diagnostic tasks into sample-level and event-level classification, both under the broader framework of classification problems. Sample-level classification involves assigning a single label to an entire signal, typically representing a specific subject or sample (e.g., Alzheimer’s disease diagnosis). Event-level classification focuses on identifying and classifying distinct temporal segments within a longer signal, thereby introducing an implicit segmentation process by associating each segment with a specific event or state (e.g., seizure detection or sleep staging).

Electrical brain signals, which capture the brain’s electrical activity over time, can be modeled as multivariate time series. Specifically, let  $\mathbf{X} \in \mathbb{R}^{C \times T}$  represent the EEG/iEEG time series, where  $C$  is the number of channels, and  $T$  is the number

TABLE I: Signal Preprocessing Techniques

| Techniques                         | Details                 | Reference |
|------------------------------------|-------------------------|-----------|
| Noise Reduction & Filtering        | FIR Filter              | [26]      |
|                                    | IIR Filter              | [27]      |
|                                    | Adaptive Filters        | [28]      |
|                                    | Manual & Custom         | [29]      |
| Artifact Removal                   | Blind Source Separation | [30]      |
|                                    | Artifact Correction     | [31]      |
| Baseline Correction & Detrending   | Baseline Correction     | [32]      |
|                                    | Baseline Removal        | [33]      |
|                                    | Detrending              | [34]      |
| Channel Processing                 | Channel Selection       | [35]      |
|                                    | Channel Mapping         | [16]      |
|                                    | Re-Referencing          | [33]      |
| Normalization & Scaling            | Z-Normalization         | [11]      |
|                                    | Quantile Normalization  | [36]      |
|                                    | Scaling & Shifting      | [16]      |
| Sampling Adjustment                | Downsampling            | [37]      |
|                                    | Resampling              | [38]      |
|                                    | Interpolation           | [39]      |
|                                    | Imputation              | [40]      |
| Segmentation                       | Windowing               | [41]      |
| Signal Alignment & Synchronization | Time Synchronization    | [42]      |
|                                    | Temporal Alignment      | [42]      |

TABLE II: Feature Extraction Techniques

| Techniques                            | Details                    | Reference |
|---------------------------------------|----------------------------|-----------|
| Data Augmentation                     | Oversampling               | [38]      |
|                                       | ELM-AE                     | [46]      |
| Signal Decomposition & Transformation | Time-Frequency Analysis    | [47]      |
|                                       | Empirical Decomposition    | [48]      |
| Spectral & Power Analysis             | Power Spectrum             | [49]      |
|                                       | Spectral Density           | [50]      |
|                                       | Partial Directed Coherence | [51]      |
| Time-Domain Features Extraction       | Statistical Measures       | [52]      |
|                                       | Amplitude & Range          | [53]      |
|                                       | Hjorth Parameters          | [54]      |
| Frequency-Domain Features Extraction  | Band Power Features        | [55]      |
|                                       | Spectral Measures          | [56]      |
| Time-Frequency Features Extraction    | Wavelet Coefficients       | [57]      |
|                                       | STFT Features              | [58]      |
|                                       | Multitaper Spectral        | [59]      |
| Other Features Extraction             | Nonlinear Features         | [60]      |
|                                       | Spatial Features           | [61]      |
|                                       | Transform-Based Features   | [62]      |
| Source Imaging                        | Conventional Methods       | [63]      |
|                                       | Deep Learning-based        | [64]      |
| Graph Analysis                        | Clustering Coefficient     | [65]      |
|                                       | Other Graph Metrics        | [66]      |

of sampling points. Each channel  $\mathbf{x}^c = \{x_1^c, x_2^c, \dots, x_T^c\}$  corresponds to the measurements from a specific source, such as an EEG electrode or a contact of an iEEG electrode.

1) *Sample-Level Classification*: In sample-level classification, the objective is to assign a single label  $y \in \mathcal{Y}$  to the entire signal  $\mathbf{X}$ . This can be formulated as:

$$y = \Phi_{\text{sample}}(\mathbf{X}; \theta), \quad y \in \mathcal{Y},$$

where  $\Phi_{\text{sample}}$  represents the DL model parameterized by  $\theta$ , and  $\mathcal{Y}$  denotes the set of classes.  $\mathbf{X}$  is treated as a unified entity, capturing sample-level or subject-level characteristics.

2) *Event-Level Classification*: In event-level classification, the goal is to classify smaller temporal segments of the signal. The signal  $\mathbf{X}$  is divided into  $K$  segments  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ , where  $\mathbf{X}_k \in \mathbb{R}^{C \times T_k}$  and  $T_k$  is the duration of the  $k$ -th segment. A classification model is applied to each segment to produce a sequence of labels  $\mathbf{Y} = \{y_1, y_2, \dots, y_K\}$ ,  $y_k \in \mathcal{Y}$ :

$$y_k = \Phi_{\text{segment}}(\mathbf{X}_k; \theta), \quad \mathbf{Y} = \bigcup_{k=1}^K \{y_k\},$$

where  $\Phi_{\text{segment}}$  denotes the DL model parameterized by  $\theta$ . This process associates each segment  $\mathbf{X}_k$  a label  $y_k$ , allowing the temporal localization of events within the signal. Event-level classification captures natural temporal dependencies between consecutive segments, reflecting the continuity of events [25].

## B. Signal Collection

EEG has evolved significantly since Hans Berger first recorded signals from the human scalp in 1924 [9]. While EEG is typically collected non-invasively with scalp electrodes placed according to the 10-20 system [43], recent studies employ higher-density configurations for enhanced

spatial resolution. EEG captures oscillations across frequency bands linked to neural states: delta (deep sleep), theta (light sleep), alpha (relaxation), beta (focus), and gamma (higher cognition) [44]. Depending on the study, participants may perform tasks or rest to elicit relevant brain activity. Resting-state EEG evaluates baseline activity, while specific tasks can highlight disease-related abnormalities [45].

iEEG involves implanting electrodes within deep or superficial structures via burr holes (SEEG) or placing grids on the cortical surface (ECoG). Compared to EEG, iEEG offers higher spatial resolution and reduced susceptibility to artifacts from scalp and eye movements. SEEG allows recording from deep and distributed regions with minimal invasiveness, while ECoG provides detailed cortical surface mapping with dense grids. However, iEEG remains affected by cardiac artifacts, electrode shifts, and other noise, making rigorous preprocessing essential for reliable clinical and research use.

## C. Signal Preprocessing

EEG/iEEG signals require low-level preprocessing to address challenges such as noise and artifact removal, normalization for consistency, and segmentation into analyzable time windows. These steps refine raw data, ensuring it accurately reflects brain activity and provides a robust foundation for analysis. Representative methods are summarized in Table I.

Noise reduction is central to this process: classical FIR/IIR filtering [26], [27] efficiently removes narrow-band artifacts like power-line interference, whereas Blind Source Separation (e.g., ICA, PCA [67]) targets ocular and muscular noise but may also suppress neural components if applied indiscriminately. Wavelet decomposition [68] offers multiscale handling of nonstationary noise, though at higher computational cost. Normalization techniques such as Z-score scaling [18]

**TABLE III:** Summary of subject-level data partitioning strategies for EEG/iEEG.

| Strategy                | Formal Definition                                                                                                                                                                       | Advantages                                       | Limitations                                                                  |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|------------------------------------------------------------------------------|
| <b>Subject-Specific</b> | $\mathcal{X}_{tr} \cup \mathcal{X}_{val} \cup \mathcal{X}_{te} = \{\mathbf{X}_k^{(i)}\}_{k=1}^{K^{(i)}}$                                                                                | Rapid prototyping<br>Useful for personalization  | Restricted clinical applicability<br>Poor transferability across individuals |
| <b>Mixed-Subject</b>    | $\mathcal{X}_{set} \subset \bigcup_{i \in \mathcal{P}} \bigcup_{k=1}^{K^{(i)}} \{\mathbf{X}_k^{(i)}\}$<br>$ \mathcal{X}_{set}  = \alpha_{set} \sum_{i=1}^N K^{(i)}$                     | Maximizes training data<br>Robust to variability | Potential risk of data leakage<br>Reduced clinical relevance                 |
| <b>Cross-Subject</b>    | $\mathcal{P} = \mathcal{P}_{tr} \cup \mathcal{P}_{val} \cup \mathcal{P}_{te}$<br>$\mathcal{X}_{set} = \bigcup_{i \in \mathcal{P}_{set}} \bigcup_{k=1}^{K^{(i)}} \{\mathbf{X}_k^{(i)}\}$ | Clinically relevant<br>Realistic deployment      | High data demand<br>Computational burden                                     |

standardize channel amplitudes, improving model stability but potentially masking inter-individual variability. Segmentation and resampling further balance efficiency and fidelity: down-sampling can reduce computational load [37], while shorter epochs facilitate localized analysis but risk fragmenting long-range dependencies. Finally, baseline correction [18], channel selection [35], and alignment [42] enhance interpretability and multimodal synchronization, though each relies on assumptions that may not hold uniformly across datasets.

#### D. Feature Extraction

Feature extraction techniques transform raw signals into structured representations by isolating salient features or reconstructing core components essential for modeling. Representative methods are summarized in Table II.

Time-domain features are straightforward and interpretable (e.g., statistical moments, Hjorth parameters [69]), but insufficient to capture complex spectral dynamics. Frequency-domain features such as power spectral density and band power [49] reveal oscillatory activity, yet assume stationarity. Time–frequency approaches address this by linking temporal and spectral information, making them effective for transient, nonstationary patterns in seizure detection and cognitive monitoring [58], [68], though at higher computational cost.

At a higher level, electrophysiological source imaging (ESI) improves spatial specificity by projecting EEG into cortical source space [63], but depends on accurate head models. Graph analysis instead quantifies network-level organization [61], offering system-wide insights while remaining sensitive to noise and thresholding. Together, these methods extend analysis from local dynamics to global connectivity, supporting applications from seizure focus localization to network alterations in Alzheimer’s disease.

#### E. Data Partitioning Strategies

Building on the definition of  $\mathbf{X}^{(i)} \in \mathbb{R}^{C \times T}$  in Section II-A, where  $\mathbf{X}^{(i)}$  represents the EEG/iEEG signal of subject  $i$ , we define notations to formalize data partitioning strategies:

- $\mathcal{P} = \{1, 2, \dots, N\}$ : The set of  $N$  subjects in the dataset.
- $\mathcal{X}_{train}, \mathcal{X}_{val}, \mathcal{X}_{test}$ : The training, validation, and testing sets.
- $\alpha_{train}, \alpha_{val}, \alpha_{test} \in (0, 1)$ : The proportion of data used for training, validation and test, and  $\alpha_{train} + \alpha_{val} + \alpha_{test} = 1$ .
- $K^{(i)}$ : The total number of temporal segments or events derived from subject  $i$ ’s data.

Previous studies have examined data partitioning strategies; for instance, Zancanaro et al. [70] compared leave-one-subject-

out, fixed splits, and pooled training in motor imagery classification. Building on these insights, we introduce a taxonomy encompassing subject-specific, mixed-subject, and cross-subject strategies, with mathematical definitions mapped to practical EEG/iEEG applications (Table III). **Subject-specific** methods are typically adopted in personalized or closed-loop systems where individual calibration is critical. **Mixed-subject** methods are common in early studies for efficient training, though they risk data leakage across sets. **Cross-subject** methods are clinically most relevant, ensuring evaluation on unseen patients and reflecting real-world deployment.

Extending subject-level partitioning strategies, dataset-level partitioning includes three approaches: **dataset-specific** (independent partitioning per dataset), **mixed-dataset** (pooling data across datasets), and **cross-dataset** (disjoint datasets for training, validation, and testing). While subject-based partitions remain the standard for evaluating patient-level relevance, dataset-based strategies are increasingly adopted in self-supervised learning to mitigate data scarcity and in multi-domain models to demonstrate cross-dataset transferability. In practice, studies combine both paradigms, using subject-based partitioning to assess patient-level performance and dataset-based partitioning to evaluate generalization, testing whether methods can achieve specialization and generalizability.

#### F. Deep Learning Architectures

Neurological data processing relies on several key architectures: **Convolutional Neural Networks (CNNs)** [13] excel at extracting spatial/spectral features through hierarchical convolutions. **Recurrent Neural Networks (RNNs)** [71] capture temporal dependencies via recurrent connections. **Transformers** [14] model long-range spatiotemporal relationships using self-attention. **Graph Neural Networks (GNNs)** [72] analyze functional connectivity in graph-structured data. **Autoencoders (AEs)** [73] learn compressed representations through encoder-decoder structures. **Generative Adversarial Networks (GANs)** [74] synthesize signals through adversarial training. **Spiking Neural Networks (SNNs)** [75] leverage spike-based computation for temporal dynamics.

#### G. Deep Learning Paradigms

Deep learning applications in neurological diagnostics fall into four paradigms: supervised, self-supervised, unsupervised, and semi-supervised learning. Each paradigm addresses specific challenges in processing brain signals by leveraging architectures tailored to data availability and task requirements.



**TABLE IV:** Public EEG/iEEG datasets for seizure detection, with **Seizures** indicating the number of episodes, **Length** the duration of each record, and **Size** the total duration of recording.

| Dataset              | Type | Subjects | Seizures  | Length         | Size                | Frequency (Hz) | Channels        |
|----------------------|------|----------|-----------|----------------|---------------------|----------------|-----------------|
| Bonn [76]            | EEG  | 10       | -         | 23.6 sec       | $\approx$ 3.3 hours | 173.61         | 1               |
| Freiburg [77]        | iEEG | 21       | 87        | 4 sec          | $\approx$ 504 hours | 256            | 128             |
| Mayo-UPenn [78]      | iEEG | 2        | 48        | 1 sec          | 583 min             | 500-5000       | 16-76           |
| CHB-MIT [79]–[81]    | EEG  | 22       | 198       | 1 hour         | $\approx$ 686 hours | 256            | 23 / 24 / 26    |
| Bern-Barcelona [82]  | iEEG | 5        | 3750      | 20 sec         | 57 hours            | 512            | 64              |
| Hauz Khas [83]       | EEG  | 10       | -         | 5.12 sec       | 87 min              | 200            | 50              |
| Melbourne [84]       | iEEG | 3        | -         | 10 min         | 81.25 hours         | 400            | 184             |
| TUSZ [85]            | EEG  | 642      | 3050      | -              | 700 hours           | 250            | 19              |
| SWEC-ETHZ [86], [87] | iEEG | 18 / 16  | 244 / 100 | 1 hour / 3 min | 2656 hours / 48 min | 512 / 1024     | 24-128 / 36-100 |
| Zenodo [88]          | EEG  | 79       | 1379      | 74 min         | $\approx$ 97 hours  | 256            | 21              |
| Mayo-Clinic [89]     | iEEG | 25       | -         | 3 sec          | 50 hours            | 5000           | 1               |
| FNUSA [89]           | iEEG | 14       | -         | 3 sec          | 7 hours             | 5000           | 1               |
| Siena [90]           | EEG  | 14       | 47        | 145-1408 min   | $\approx$ 128 hours | 512            | 27              |
| Beirut [91]          | EEG  | 6        | 35        | 1 sec          | 130 min             | 512            | 19              |
| HUP [92]             | iEEG | 58       | 208       | 300 sec        | $\approx$ 27 hours  | 500            | 52-232          |
| CCEP [93]            | iEEG | 74       | -         | -              | 89 hours            | 2048           | 48-116          |

1) **Supervised Learning:** Supervised learning is the dominant paradigm for neurological diagnostics tasks, training models to map signals  $\mathbf{X} \in \mathbb{R}^{C \times T}$  to labels  $y \in \mathcal{Y}$ .

2) **Unsupervised Learning:** Unsupervised learning is essential for uncovering intrinsic data structures in signals  $\mathbf{X}$ , enabling representation learning without relying on labels.

3) **Semi-Supervised Learning:** Semi-supervised learning combines a small set of labeled examples  $\{(x_i, \hat{y}_i)\}_{i=1}^l$ , where  $\hat{y}_i$  denotes the provided labels, with a larger set of unlabeled examples  $\{x_j\}_{j=l+1}^{l+u}$  to learn a mapping from  $\mathbf{X}$  to  $\mathcal{Y}$ .

4) **Self-Supervised Learning:** Self-supervised learning (SSL) leverages unlabeled EEG/iEEG data by constructing pretext tasks that generate pseudo-labels  $\hat{y}$  from intrinsic properties of raw signals  $\mathbf{X}$ . SSL methods fall into three main categories: contrastive, predictive, and reconstruction-based learning. **Contrastive-based methods**, such as Contrastive Predictive Coding (CPC) [26] and Transformation Contrastive Learning [94], learn by maximizing similarity between related views while minimizing it between unrelated ones, capturing distinguishing signal features. **Predictive-based learning** employs pretext tasks such as Relative Positioning and Temporal Shuffling to extract structural patterns across temporal, frequency, and spatial domains [95], [96]. By predicting transformations applied to the data, it enhances domain-specific feature learning. **Reconstruction-based learning** trains models to reconstruct masked signal segments. Methods like Masked Autoencoders (MAE) reconstruct temporal or spectral components, learning intrinsic patterns in the process [16], [97]. Studies have also explored hybrid methods, which combine elements from contrastive, predictive, and reconstruction-based approaches [26], [98].

### III. APPLICATIONS

This section reviews neurological disease diagnosis methodologies. Each subsection introduces the disease, its diagnostic tasks, and related public datasets, followed by representative studies highlighting key deep learning aspects such as data types, frequency bands, and brain regions. Summary tables

report representative studies with their reported metrics (e.g., accuracy, AUC) and dataset chance levels. These metrics are for reference only, as evaluation protocols and data selection vary across studies, which may also cause slight differences in chance levels. Technical details on preprocessing, network architectures, and training are compiled in the Appendix, covering all 450 reviewed studies.

#### A. Seizure Disorder

1) **Task Description:** Epilepsy, a neurological disorder affecting 50 million people, is characterized by recurrent seizures caused by abnormal brain activity. Seizures range from brief confusion or blanking out to severe convulsions and loss of consciousness. According to WHO, up to 70% of cases can be effectively treated with proper care. However, in low-income regions, limited resources and stigma hinder access to treatment, increasing the risk of premature death [3].

Seizure detection primarily relies on standardized EEG/iEEG datasets, summarized in Table IV. The key challenge is distinguishing seizure events from background activity, typically framed as binary classification where  $y_k \in \{0, 1\}$ . Most approaches segment EEG sequences into short windows for classification, then aggregate predictions to form event-level outcomes as  $\mathbf{Y} = \bigcup_{k=1}^K \{y_k\}$  [108], [109]. Alternatively, some methods detect cut points in continuous recordings to define segment boundaries  $\{\mathbf{X}_k\}_{k=1}^K$ , each classified independently [65]. Final event-level predictions are obtained by combining labels  $\mathbf{Y} = \bigcup_{k=1}^K \{\Phi_{\text{segment}}(\mathbf{X}_k; \theta)\}$ .

Beyond binary tasks, more fine-grained classification has been explored. Three-class settings distinguish interictal (A, between seizures), preictal (D, before onset), and ictal (E, seizure) states [101], while five-class tasks further subdivide the preictal phase into early, middle, and late stages [15]. The Temple University Seizure Corpus (TUSZ) [85] supports such studies, providing detailed annotations of pathological events (e.g., epileptiform discharges, seizure types) and non-pathological signals (e.g., background activity, artifacts).

**TABLE V:** Summary of related studies on EEG-based seizure detection with different learning paradigms, feature extraction methods, and backbones. The chance level is omitted due to inconsistent data selection criteria across studies on TUSZ.

| Dataset | Task    | Paradigm                 | Feature              | Backbone        | Splitting        | Accuracy      | AUC         |       |
|---------|---------|--------------------------|----------------------|-----------------|------------------|---------------|-------------|-------|
| Bonn    | ternary | Supervised Learning      | raw                  | CNN             | generalized      | 0.8867        | -           | [11]  |
|         |         |                          | raw                  | CNN             | generalized      | 0.9900        | -           | [99]  |
|         |         |                          | Wavelet Coefficients | CNN             | generalized      | <b>0.9940</b> | -           | [68]  |
|         |         | Unsupervised Learning    | Scalograms           | 2D-CNN          | generalized      | 0.9900        | -           | [15]  |
|         |         |                          | AE-based             | CNN             | cross-subject    | 0.9933        | -           | [100] |
|         |         | <i>chance level</i>      |                      |                 |                  |               | 0.4000      |       |
| CHB-MIT | binary  | Supervised Learning      | Spectrogram          | 2D-CNN          | subject-specific | 0.9750        | -           | [101] |
|         |         |                          | raw                  | CNN-LSTM        | cross-subject    | <b>0.9771</b> | -           | [102] |
|         |         |                          | Correlation Matrix   | GAT-Transformer | cross-subject    | 0.7315        | 0.72        | [103] |
|         |         | Self-supervised Learning | raw                  | Transformer     | cross-subject    | 0.9707        | <b>0.97</b> | [104] |
|         |         |                          | raw                  | CNN             | cross-subject    | -             | 0.88        | [105] |
|         |         | <i>chance level</i>      |                      |                 |                  |               | 0.5000      |       |
| TUSZ    | 8-class | Supervised Learning      | Spectrogram          | 2D-CNN          | cross-subject    | <b>0.8890</b> | -           | [106] |
|         | binary  |                          | Correlation Matrix   | GNN             | cross-subject    | -             | <b>0.88</b> | [107] |
|         | 4-class | Self-supervised Learning | Correlation Matrix   | GNN             | cross-subject    | -             | 0.75        | [107] |
|         | 4-class |                          | Wavelets             | Transformer     | cross-subject    | 0.7300        | -           | [108] |
|         | binary  |                          | raw                  | CNN-GCN         | cross-subject    | -             | 0.78        | [98]  |

In addition, epileptic focus localization identifies the cortical origin of pathological discharges, formulated as classifying iEEG contacts inside versus outside the epileptogenic zone [110], or reconstructing source-level activity from scalp EEG via ESI [63]. This task is clinically critical, as accurate localization guides surgical resection in drug-resistant epilepsy.

**2) Supervised Methods:** Supervised seizure detection has advanced with public datasets and progress in deep learning. Early studies relied on subject-specific or mixed-subject evaluations using short, pre-segmented EEG clips. For instance, the Bonn dataset [76] contains manually labeled seizure and non-seizure segments, enabling models to operate on fixed-length inputs. Approaches based on raw signals employ CNNs or RNNs to learn spatiotemporal features from these standardized segments [11], [99], while feature-based methods transform signals into handcrafted or derived representations, such as scalograms [15] and wavelet-based features [68], which are then used by shallow classifiers. These techniques inherently assume limited temporal context and circumvent the challenges of segmenting continuous EEG. As shown in Table V, the Bonn dataset is relatively simple and prone to overfitting, making it insufficient to represent real-world clinical scenarios.

With the adoption of long-term recordings like CHB-MIT [80], the focus shifts toward cross-subject paradigms. Unlike Bonn, CHB-MIT provides continuous recordings with multiple seizure episodes per patient, requiring models to handle variable-length inputs and detect seizure onsets in unsegmented streams. Approaches integrate temporal modeling through sliding windows [109], sequence-aware architectures such as Transformers [111], or hybrid fusion techniques [102]. Cross-subject validation becomes standard, reflecting clinical requirements that generalize across diverse conditions.

The necessity of cross-subject modeling in seizure detection stems from its critical role in ensuring clinical generalization. The invasive nature of iEEG differentiates its modeling requirements from EEG through distinct acquisition paradigms and neurophysiological characteristics, its patient-

specific recording conditions and electrode configurations lead to substantial inter-subject heterogeneity in temporal features and spatial sampling, unlike EEG's standardized scalp placement [112]. Balancing high-resolution spatiotemporal capture with robustness across patients, iEEG requires specialized methodologies to enhance generalizability while addressing its inherent complexities. Spatial modeling is essential for capturing 3D epileptogenic networks with depth electrodes. Graph-based methods model inter-channel dependencies via neuroanatomical [113] or dynamic functional connections [114], while Transformers use attention mechanisms to adapt to varying electrode configurations [115]. DMNet [116] improves domain generalization through self-comparison mechanisms.

**3) Semi- and Unsupervised Methods:** Semi-supervised and unsupervised techniques are increasingly applied in deep learning for seizure detection, particularly when labeled data is limited. A common approach incorporates clustering for event-level segmentation, allowing the model to identify and segment seizure events [65]. Another application involves using models such as Autoencoders, DBNs and GANs to automatically extract relevant features or augment datasets, thereby enhancing the model's robustness and generalizability [100], [117].

**4) Self-supervised Methods:** Self-supervised learning has emerged as an effective approach for seizure detection. Contrastive learning methods form positive and negative pairs to capture seizure-related patterns. For instance, SLAM [104] pairs an anchor with a window from a distant time point as a negative sample, while SPP-EEGNET [118] uses the absolute difference between two windows to determine pair similarity. Predictive-based methods design pretext tasks to simulate epileptic features, such as augmenting signals with amplitude or frequency changes [105] or predicting the next segment using graph-based modeling [107]. Reconstruction-based methods focus on preserving context during learning. EpilepsyNet [111] uses Pearson Correlation Coefficients to capture spatial-temporal embeddings, while Wavelet2Vec [108] reconstructs wavelet-transformed EEG patches to exploit seizure-

**TABLE VI:** Public Sleep EEG Datasets, where **Recordings** denotes the number of whole-night PSG recordings.

| Dataset               | Recordings | Frequency (Hz) | Channels |
|-----------------------|------------|----------------|----------|
| Sleep-EDF [81], [121] | 197        | 100            | 2        |
| MASS [122]            | 200        | 256            | 4-20     |
| SHHS [123], [124]     | 8362       | 125            | 2        |
| SVUH.UCD [81], [125]  | 25         | 128            | 3        |
| HMC [81], [126]       | 151        | 256            | 4        |
| PC18 [81], [127]      | 1985       | 200            | 6        |
| MIT-BIH [81], [128]   | 16         | 250            | 1        |
| DOD-O [129]           | 55         | 250            | 8        |
| DOD-H [129]           | 25         | 250            | 12       |
| ISRUC [130]           | 126        | 200            | 6        |
| MGH [131]             | 25941      | 200            | 6        |
| Piryatinska [132]     | 37         | 64             | 1        |
| DRM-SUB [133]         | 20         | 200            | 3        |
| SD-71 [134]           | 142        | 500            | 61       |

**TABLE VII:** Reported accuracies on Sleep-EDF datasets using representative models (grouped by learning paradigm).

| Learning Paradigm   | Modality | CL Strategy                 | Sleep-EDF     | Sleep-EDFx    |       |
|---------------------|----------|-----------------------------|---------------|---------------|-------|
| <b>SL</b>           | EEG      | –                           | 0.8440        | 0.8130        | [135] |
|                     | EEG+EOG  | –                           | –             | 0.8390        | [136] |
| <b>SSL</b>          | EEG      | Global Reference            | –             | <b>0.8690</b> | [137] |
|                     | EEG      | Time–spectrogram Multi-view | –             | 0.7806        | [138] |
|                     | EEG      | Time–frequency Multi-view   | 0.7160        | –             | [139] |
|                     | EEG+EOG  | Contrastive Alignment       | <b>0.8458</b> | 0.8284        | [140] |
| <i>Chance Level</i> |          |                             | 0.4207        | 0.3537        |       |

specific discharge patterns across frequency bands. EEG-CGS [66] adopts a hybrid graph-based approach, framing seizure detection as anomaly detection and integrating subgraph sampling with contrastive and reconstruction learning. As shown in Table V, SSL methods exhibit considerable performance variations across datasets. On more challenging datasets like TUSZ, performance approaches that of supervised methods, underscoring the need for larger-scale pretraining and stronger representation learning. Furthermore, four-class seizure type classification remains more difficult than detection, highlighting persistent bottlenecks in distinguishing subtypes.

SSL paradigm is also common in iEEG-based modeling. BrainNet [119] employs bidirectional contrastive predictive coding to capture temporal correlation in SEEG signals. MBrain [98] models time-varying propagation patterns and inter-channel phase delays of epileptic activity through a multi-variant contrastive-predictive learning framework, leveraging graph-based representations for spatial-temporal correlations across EEG and SEEG channels. PPI [120] accounts for regional seizure variability, employing a channel discrimination task to ensure the model captures distinct pathological patterns across brain regions rather than treating all channels uniformly.

## B. Sleep Staging

1) *Task Description*: Sleep staging is critical to understanding sleep disorders like insomnia and sleep apnea, as well as

**TABLE VIII:** Public EEG Datasets for Depression Detection, where **Exp (n)** represents the number of depressed individuals and **Ctrl (n)** represents the healthy control group.

| Dataset       | Exp (n) | Ctrl (n) | Frequency (Hz) | Channels |
|---------------|---------|----------|----------------|----------|
| HUSM [143]    | 34      | 30       | 256            | 22       |
| PRED+CT [144] | 46      | 75       | 500            | 64       |
| EDRA [145]    | 26      | 24       | 500            | 63       |
| MODMA [146]   | 24      | 29       | 250            | 128      |
|               | 26      | 29       |                | 3        |

**TABLE IX:** Reported accuracies on three MDD datasets using representative backbone architectures.

| Backbone Architecture | HUSM                | PRED+CT             | MODMA               |
|-----------------------|---------------------|---------------------|---------------------|
| CNN                   | 0.9832 [147]        | 0.9393 [148]        | 0.7400 [149]        |
| CNN-RNN               | 0.9597 [147]        | <b>0.9907</b> [150] | 0.9756 [151]        |
| GCN                   | <b>0.9844</b> [152] | 0.8317 [153]        | <b>0.9968</b> [152] |
| SNN                   | –                   | 0.9800 [154]        | –                   |
| <i>Chance Level</i>   | 0.5313              | 0.6198              | 0.5472              |

the impact on overall health. It is estimated that 20% to 41% of the global population is affected by sleep disorders, which are linked to an increased risk of obesity, cardiovascular diseases, and mental health issues [141]. Therefore, accurately identifying sleep stages is essential for addressing these concerns.

Sleep staging involves segmenting signals into 30-second epochs and classifying them into stages: awake (W), rapid eye movement (REM), and three non-REM (NREM) stages (N1, N2, N3). Wake is characterized by high-frequency  $\beta$  and  $\alpha$  waves. In N1, the transition to sleep, low-amplitude  $\theta$  waves appear. N2, light sleep, is marked by sleep spindles and K-complexes associated with sensory processing and memory consolidation. N3, or deep sleep, features slow-wave  $\delta$  activity. REM sleep, essential for emotional regulation and dreaming, is characterized by rapid, low-voltage brain activities.

Multimodal modeling is fundamental for sleep analysis, as polysomnography (PSG) integrates EEG (e.g., Fpz-Cz, Pz-Oz), Electrooculography (EOG), and Electromyography (EMG) to enhance staging accuracy. The public datasets in Table VI provide a comprehensive view of resources.

2) *Supervised methods*: Selecting biosignal modalities is critical for designing supervised learning frameworks in PSG-based sleep staging. Two primary paradigms are widely used. Single-channel EEG methods, preferred in resource-constrained settings, offer hardware simplicity, reduced cross-modal interference, and enhanced computational efficiency [142]. However, relying solely on EEG limits the detection of complementary cues—such as ocular and muscular activities—essential for identifying ambiguous sleep stages. Hybrid EEG-EOG models provide a balance between diagnostic accuracy and computational efficiency, while full multimodal designs integrating EEG, EOG, and EMG most closely emulate clinical scoring protocols [136].

3) *Self-supervised methods*: Self-supervised contrastive methods are gradually replacing traditional supervised learning, especially on large-scale EEG datasets where they demonstrate stronger generalization and robustness (Table VII). Early works explore tasks like relative positioning and temporal

**TABLE X:** Public EEG Datasets for Schizophrenia, where **Exp (n)** represents the number of schizophrenia patients and **Ctrl (n)** represents the control group.

| Dataset         | Exp (n) | Ctrl (n) | Frequency (Hz) | Channels |
|-----------------|---------|----------|----------------|----------|
| CeonRepod [155] | 14      | 14       | 250            | 19       |
| NIMH [156]      | 49      | 32       | 1024           | 64       |
| MHRC [157]      | 45      | 39       | 128            | 16       |

**TABLE XI:** Reported accuracies on three SZ datasets using representative learning strategies.

| Modeling Strategy   | CeonRepod [155]     | NIMH [156]          | MHRC [157]          |
|---------------------|---------------------|---------------------|---------------------|
| Timeseries-based    | 0.9807 [158]        | 0.9200 [159]        | <b>0.9800</b> [160] |
| 2D Representation   | <b>0.9974</b> [161] | <b>0.9635</b> [161] | 0.9740 [161]        |
| Transfer Learning   | 0.9900 [162]        | 0.9336 [163]        | 0.9773 [46]         |
| <i>Chance Level</i> | 0.5492              | 0.5962              | 0.5357              |

shuffling to extract temporal structures from multivariate signals [95], [96]. ContraWR [137] constructs contrastive pairs from distinct time windows to capture temporal dependencies, reporting notably high accuracy on Sleep-EDFx, mulEEG [138] and CoSleep [139] introduce multi-view contrastive strategies, with mulEEG focusing on cross-view consistency and CoSleep capturing temporal and spectral patterns through a dual time-frequency framework. Multimodal modeling enhances sleep staging by integrating complementary EEG, EOG, and EMG signals. Brant-X [140] tackles alignment challenges with EEG foundation models and contrastive learning, aligning EEG and EOG at local and global levels to bridge modality gaps and achieve superior performance.

### C. Depression Identification

1) *Task Description:* Depression, particularly Major Depressive Disorder (MDD), is a psychological condition affecting 5% of individuals worldwide, with a higher prevalence among women. In low- and middle-income countries, up to 75% of individuals lack adequate care due to limited resources and stigma, despite effective treatments being available [7].

Depression severity is quantified using standardized scales like the Beck Depression Inventory (BDI) to differentiate clinical depression from normal mood variations. Existing studies adopt heterogeneous classification criteria: some focus on binary discrimination (e.g., patients vs. healthy controls), while others stratify cohorts by treatment status (medicated vs. non-medicated) or severity levels (mild vs. moderate/severe). Table VIII summarizes datasets used in MDD research.

2) *Approach overview:* Depression impacts both superficial and deeper brain structures, challenging traditional handcrafted features. Acharya introduced the first end-to-end DL model for EEG-based depression detection, showing right-hemisphere signals to be more distinctive than left, consistent with clinical findings [12]. Sun et al. [152] further reported that with increasing granularity, MDD patients exhibited weakened connectivity between RF-RT and LT-LP regions; by embedding these patterns into the Multi-Granularity Graph Convolutional Network (MGGCN), clinically relevant disruptions were captured, yielding superior accuracy (Table IX).

**TABLE XII:** Public EEG Datasets for Alzheimer's Diagnosis, where **AD (n)** and **MCI (n)** represent the experimental groups, and **Ctrl (n)** represents the control group.

| Dataset      | AD (n) | MCI (n) | Ctrl (n) | Frequency (Hz) | Channels |
|--------------|--------|---------|----------|----------------|----------|
| FSA [166]    | 160    | -       | 24       | 128            | 21       |
| AD-65 [167]  | 36     | -       | 29       | 250            | 19       |
| Fiscon [168] | 49     | 37      | 14       | 1024           | 19       |
| AD-59 [169]  | 59     | 7       | 102      | 128-256        | 21       |

**TABLE XIII:** Reported accuracies on private AD datasets with feature representations (chance level in parentheses).

| Feature Representation | Accuracy                     |
|------------------------|------------------------------|
| Pearson correlation    | <b>1.0000</b> (0.5000) [170] |
| Wavelet Coherence      | 0.9230 (0.5128) [171]        |
| PSD image              | 0.9295 (0.5000) [172]        |

Spiking neural networks (SNNs) offer another direction: Shah et al. [164] used the NeuCube SNN to map EEG into a 3D reservoir aligned with the Talairach atlas, modeling spatiotemporal dynamics via STDP with interpretable connectivity visualization Sam et al. [154] integrates a 3D brain-inspired SNN with an LSTM, leveraging SNN's energy efficiency with LSTM's temporal modeling capabilities.

### D. Schizophrenia Identification

1) *Task Description:* Schizophrenia (SZ) is a psychiatric disorder affecting 24 million people worldwide, characterized by cognitive deficits, delusions, and hallucinations [6]. SZ is associated with disruptions in structural and functional brain connectivity, marked by decreased global efficiency, weakened strength, and increased clustering [165]. These abnormalities manifest in EEG signals, enabling reliable binary classification of SZ patients versus healthy controls. Publicly available datasets supporting this task are summarized in Table X.

2) *Approach overview:* EEG-based SZ diagnosis has been studied through three main strategies (Table XI). *Time-series models* work directly on raw EEG, capturing temporal dynamics with relatively simple architectures but limited spectral-spatial representation [158], [160]. *2D image-based approaches* transform EEG into spectrograms or scalograms, allowing CNNs to exploit richer spectral-spatial patterns [161]. *Transfer learning* builds on this idea by adapting pre-trained vision backbones (e.g., VGG, ResNet), whose hierarchical convolutional filters are well suited for capturing local and global patterns in spectrogram-like EEG representations, thereby achieving robust feature extraction even with limited data [162]. Reported results across representative studies show accuracies typically above 0.9 for all three strategies, suggesting that different modeling approaches can support reliable SZ classification under varied settings.

### E. Alzheimer's Disease Diagnosis

1) *Task Description:* Alzheimer's disease (AD) is a progressive neurodegenerative disorder that starts with mild memory loss and advances to severe cognitive impairment, affecting



**TABLE XIV:** Public EEG Datasets for Parkinson’s Disease Diagnosis, where **Exp (n)** represents the number of patients and **Ctrl (n)** represents the healthy control group.

| Dataset    | Exp (n) | Ctrl (n) | Frequency (Hz) | Channels |
|------------|---------|----------|----------------|----------|
| UCSD [174] | 15      | 16       | 512            | 32       |
| UNM [175]  | 27      | 27       | 500            | 64       |
| UI [176]   | 14      | 14       | 500            | 59       |

**TABLE XV:** Reported accuracies on the UCSD dataset with representative preprocessing and feature representations.

| Preprocessing   | Feature Representation | Accuracy (UCSD)     |
|-----------------|------------------------|---------------------|
| /               | Raw segments           | 0.9800 [177]        |
| Gabor Transform | Spectrograms           | 0.9946 [47]         |
| CWT             | Scalograms             | 0.9960 [178]        |
| SPWVD           | TFR                    | <b>0.9997</b> [179] |
| Chance Level    | –                      | 0.6522              |

daily life. While medical interventions can improve quality of life, a definitive cure remains elusive [2]. Alzheimer’s disease (AD) progresses through three stages: preclinical, mild cognitive impairment (MCI), and Alzheimer’s dementia. Classification tasks typically distinguish MCI or Alzheimer’s dementia from healthy controls. EEG abnormalities, such as slowed brain rhythms and desynchronization, serve as biomarkers for AD-related neurodegeneration [173]. Table XII summarizes publicly available datasets.

2) *Approach overview:* EEG abnormalities in Alzheimer’s disease, such as disrupted functional connectivity and altered brain rhythms, provide critical insights into the neurological changes. Representative strategies are summarized in Table XIII, noting that results on private datasets are not strictly comparable. For instance, Alves et al. [170] employed Pearson correlation to construct connectivity matrices, achieving near-perfect discrimination between AD and healthy controls. Shan et al. [171] explored six functional connectivity measures for constructing adjacency matrices, reporting that wavelet coherence yielded the best performance for capturing spatial–temporal dependencies. Beyond connectivity, 2D spectral representations—such as PSD-based images—have been employed to enable feature learning for AD classification [172].

## F. Parkinson’s Disease Diagnosis

1) *Task Description:* Parkinson’s disease (PD) is a progressive neurodegenerative disorder marked by motor (tremors, rigidity, bradykinesia) and non-motor symptoms (depression, sleep disturbances, cognitive decline). In 2019, over 8.5 million people lived with PD [5]. EEG is used in PD research for noise resistance and sensitivity to neurological changes, such as slowed cortical oscillations and increased low-frequency power [180]. Most studies focus on supervised learning for binary classification, with some incorporating transfer learning. Table XIV summarizes publicly available datasets.

2) *Approach overview:* Transforming raw EEG signals into 2D representations is a well-established approach for PD classification (Table XV). Time–frequency transformations such as Gabor Transform and CWT have been widely adopted: spectrograms [47] and scalograms [178] capture temporal–spectral

**TABLE XVI:** Public EEG Datasets for ADHD Identification, where **Exp (n)** represents the number of ADHD patients and **Ctrl (n)** represents the healthy control group.

| Dataset        | Exp (n) | Ctrl (n) | Frequency (Hz) | Channels |
|----------------|---------|----------|----------------|----------|
| ADHD-79 [181]  | 37      | 42       | 256            | 2        |
| ADHD-121 [182] | 61      | 60       | 128            | 19       |

**TABLE XVII:** Reported accuracies across frequency bands on the ADHD-121 dataset [183].

| Theta  | Alpha  | Beta   | Gamma  | Full          | chance level |
|--------|--------|--------|--------|---------------|--------------|
| 0.9374 | 0.9724 | 0.9825 | 0.9725 | <b>0.9975</b> | 0.5825       |

dynamics more effectively than raw waveforms. More recently, advanced representations such as the Smoothed Pseudo Wigner–Ville Distribution (SPWVD) [179] generate high-resolution time–frequency maps, allowing CNNs to exploit fine-grained signal structure. Collectively, these approaches illustrate a methodological shift from direct time-series analysis to progressively richer 2D representations, each achieving performance substantially above chance level.

## G. ADHD Identification

1) *Task Description:* Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder affecting around 3.1% of individuals aged 10–14 and 2.4% of those aged 15–19 [8]. It is categorized into three subtypes: Inattentive (ADHD-I), Hyperactive-Impulsive (ADHD-H), and Combined (ADHD-C) [184]. EEG is widely used alongside neuroimaging and physiological measures for ADHD diagnosis. However, deep learning remains underexplored, with most existing approaches relying on supervised learning and feature-based classification. Research focuses on binary classification tasks, and Table XVI lists two publicly available datasets.

2) *Approach overview:* Studies on ADHD diagnosis report elevated  $\theta$  and reduced  $\beta$  in children with ADHD, aligning with medical findings [185]. More recent work compared models trained on individual bands with those using the full denoised range, showing that integrated inputs achieved the best results (Table XVII). This indicates that although  $\theta$  and  $\beta$  remain the most consistent group-level markers, combining multiband better captures individual variability and cross-band interactions, providing richer features for DL models [183].

## IV. UNIVERSAL PRE-TRAINED MODELS

In recent years, SSL has revolutionized EEG/iEEG analysis in neurological diagnosis. Emerging methods focus on generalizable SSL frameworks that integrate heterogeneous datasets during pre-training, overcoming the limitations of task- and dataset-specific models and enabling seamless adaptation to multiple downstream tasks. These innovations bring us closer to the development of universal neurodiagnostic models capable of addressing challenges across diverse clinical settings.

Table XVIII summarizes pre-trained SSL frameworks for multi-task neurodiagnosis, organized by the SSL paradigms to align with their technical evolution analyzed in this section. While some frameworks extend to broader time-series data,

**TABLE XVIII:** Summary of pre-trained SSL frameworks for multi-task neurodiagnosis, highlighting datasets and tasks, with paradigms such as Contrastive Learning (CL), Contrastive Predictive Coding (CPC), and Masked Autoencoding (MAE)

| Work                 | SSL Paradigm          | Backbone        | Data Type | Partitioning     | pre-training Dataset                                   | Downstream Tasks                  |
|----------------------|-----------------------|-----------------|-----------|------------------|--------------------------------------------------------|-----------------------------------|
| Banville et al. [26] | CPC                   | CNN             | EEG       | dataset-specific | TUSZ, PC18                                             | Seizure, Sleep                    |
| MBrain [98]          | CPC                   | CNN+LSTM+GNN    | EEG, iEEG | dataset-specific | TUSZ, private                                          | Seizure, etc.                     |
| TS-TCC [186]         | CPC                   | CNN+Transformer | EEG       | cross-dataset    | Bonn, Sleep-EDF, etc.                                  | Seizure, Sleep, etc.              |
| SeqCLR [94]          | CL                    | CNN+GRU         | EEG       | mixed-dataset    | TUSZ, Sleep-EDF, ISRUC, etc.                           | Seizure, Sleep, etc.              |
| TF-C [187]           | CL                    | CNN             | EEG       | cross-dataset    | Sleep-EDF, etc.                                        | Seizure, Sleep, etc.              |
| BIOT [188]           | CL                    | Transformer     | EEG, etc. | cross-dataset    | SHHS, etc.                                             | Seizure, etc.                     |
| Jo et al. [189]      | Predictive            | CNN             | EEG       | mixed-dataset    | CHB-MIT, Sleep-EDF                                     | Seizure, Sleep                    |
| neuro2vec [97]       | MAE                   | CNN+Transformer | EEG       | cross-dataset    | Bonn, Sleep-EDF, etc.                                  | Seizure, Sleep                    |
| CRT [190]            | MAE                   | Transformer     | EEG       | dataset-specific | Sleep-EDF, etc.                                        | Sleep, etc.                       |
| NeuroBERT [191]      | MAE                   | Transformer     | EEG, etc. | dataset-specific | Bonn, SleepEDF, etc.                                   | Seizure, Sleep, etc.              |
| BENDR [16]           | CPC+MAE               | CNN+Transformer | EEG       | cross-dataset    | TUEG                                                   | Sleep, etc.                       |
| CBRAMOD [192]        | MAE                   | Transformer     | EEG       | cross-dataset    | TUEG                                                   | Seizure, Sleep, MDD               |
| Brant [112]          | MAE                   | Transformer     | iEEG      | cross-dataset    | private                                                | Seizure, etc.                     |
| Brainwave [193]      | MAE                   | Transformer     | EEG, iEEG | cross-dataset    | TUEG, Siena, CCEP, Sleep-EDF, NIMH, FSA, private, etc. | Seizure, Sleep, MDD, SZ, AD, ADHD |
| EEGFormer [194]      | VQ+MAE                | Transformer     | EEG       | cross-dataset    | TUEG                                                   | Seizure, etc.                     |
| LaBraM [195]         | VQ+MAE                | Transformer     | EEG       | cross-dataset    | TUEG, Siena, etc.                                      | Seizure, etc.                     |
| NeuroLM [196]        | VQ+MAE<br>+Predictive | Transformer     | EEG       | cross-dataset    | TUEG, Siena, etc.                                      | Seizure, Sleep, etc.              |

such as BCI signals and motion sensor data, we focus on datasets and tasks directly relevant to neurological applications. Below, we explore these frameworks, examining their contributions to unified pre-training strategies, multitask adaptability, and their potential to impact real-world applications.

#### A. Contrastive- and Predictive- Based Learning

*a) Contrastive Predictive Coding:* Early SSL approaches in EEG/iEEG analysis primarily followed the Contrastive Predictive Coding (CPC) paradigm [98], which learns representations by predicting signal segments through contrastive learning. While models employed generic architectures across neurophysiological tasks, they fail to achieve true cross-task generalization and were thus trained separately on task-specific datasets, limiting their clinical applicability. CPC variants like TS-TCC [186] introduce a one-to-one feature transfer, enabling feature migration across tasks such as human activity recognition, sleep staging, and seizure detection, paving the way for broader multi-domain generalization.

Building on the foundational principles of CPC, two distinct approaches have emerged: contrastive learning (CL) and predictive-based variants. CL retains CPC's contrastive framework but emphasizes explicit instance-level discrimination through hand-crafted augmentations for positive/negative pairs, instead of CPC's autoregressive future state prediction. Predictive variants inherit CPC's structure but replace its auto-learned latent contexts with manually defined features.

*b) Contrastive-Based learning:* SeqCLR [94], inspired by SimCLR, employs contrastive learning to EEG data, enhancing similarity between augmented views of the same channel through domain-specific transformations. Adopting a mixed-dataset training approach, it unifies diverse EEG datasets for robust representation learning. TF-C [187] incorporates dual time-frequency contrastive learning with a cross-domain consistency loss to align embeddings across temporal and spectral

representations. It further examines cross-dataset generalization, training on a source dataset and evaluating transferability to multiple targets, highlighting the potential of cross-task feature sharing for universal neural signal models. BIOT [188] integrates contrastive learning, unifying multimodal biosignals (e.g., EEG, ECG) via tokenization and linear attention to learn invariant physiological patterns for cross-task generalization.

*c) Predictive-Based Learning:* Jo et al. [189] proposes a channel-aware predictive-based framework, which leverages stopped band prediction for spectral feature learning and employs temporal trend identification to capture dynamic patterns. By integrating mix-dataset pretraining, it enhances generalization through cross-domain feature fusion. However, the pretraining scale remains limited.

#### B. Reconstruction-Based Learning

*a) Masked Autoencoding:* The paradigm shift from CPC to masked reconstruction in SSL aims for higher data efficiency and scalability, inspired by cross-domain advances like masked language modeling in NLP (e.g., BERT [197]), with MAE's generative approach enhancing classification performance while avoiding complex negative sampling.

Neuro2vec [97] extends masked reconstruction by integrating EEG-specific spatiotemporal recovery and spectral component prediction into a unified framework, utilizing a CNN-ViT hybrid architecture for patch embedding and reconstruction. CRT [190] further introduces multi-domain reconstruction through cross-domain synchronization of temporal and spectral features, replacing conventional masking with adaptive input dropping to preserve data distribution integrity, thereby improving robustness in physiological signal modeling. NeuroBERT [191] introduces Fourier Inversion Prediction (FIP), reconstructing masked signals by predicting their Fourier amplitude and phase, then applying an inverse Fourier transform. The spectral-based prediction framework inherently matches the physiological nature of EEG signals.

*b) Large-Scale Continuous-Reconstruction Models:* Transformer architectures are increasingly applied in neurodiagnostics, leveraging their scalability and attention mechanisms to capture global dependencies in irregular neural signals. BERT-style pretraining, particularly masked reconstruction, enhances neurodiagnostic classification by enforcing robust contextual learning of latent bioelectrical patterns, which is crucial for distinguishing subtle neurological signatures. Their parallelizable training and tokenized time-frequency representations pave the way for scalable foundation models, driving large-scale pretraining in neural signal analysis.

Inspired by Bert, BENDR [16] integrates CPC with MAE-inspired reconstruction for temporal feature learning. Pre-trained on the Temple University Hospital EEG Corpus—a diverse dataset containing 1.5 TB of raw clinical EEG from over 10,000 subjects—BENDR represents the emergence of large-scale pretraining for neurodiagnostics, showcasing the cross-subject scalability of transformers. It demonstrates how foundation models can unify heterogeneous neural signal paradigms, advancing generalized, scalable EEG analysis. CBRAMOD [192] introduces a criss-cross transformer framework to explicitly model EEG’s spatial-temporal heterogeneity. Using patch-based masked reconstruction, it separately processes spatial and temporal patches through parallel attention, preserving the structural dependencies to EEG.

Brant [112] and Brainwave [193] represent a unified effort to establish foundation models for neural signal analysis. Brant focuses on SEEG signals, employing a masked autoencoding framework with dual Transformer encoders to capture temporal dependencies and spatial correlations, enabling seizure detection and signal forecasting. Brainwave pioneers large-scale pretraining with an unprecedented multimodal corpus of over 40,000 hours of EEG/iEEG data from 16,000 subjects, marking a significant milestone in neural signal foundation models. Its pre-training strategy follows a masked modeling paradigm that randomly masks time-frequency patches of neural signals, and the model is trained to reconstruct the missing regions. To enhance generalizability across neural data types, Brainwave employs a shared encoder for both EEG and iEEG, coupled with modality-specific reconstruction decoders. These innovations position Brainwave as the first comprehensive foundation model unifying EEG/iEEG analysis, with transformative implications for neuroscience research.

*c) Large-Scale Discrete-Reconstruction Models:* Vector Quantized Variational Autoencoder (VQ-VAE) is a powerful framework for learning discrete representations of continuous data by mapping inputs to a predefined codebook, which has been widely adopted in domains like speech and image processing [198]. By tokenizing raw data into discrete codes, this approach enhances cross-subject generalization while preserving interpretable spatiotemporal patterns.

LaBraM [195] trains its discrete codebook by reconstructing spectral magnitudes and phases of EEG segments, then pre-trains with a symmetric masking task that predicts masked code indices bidirectionally. NeuroLM [196] extends this approach by introducing VQ Temporal-Frequency Prediction, aligning EEG tokens with textual representations through adversarial training. After tokenization, it employs autoregressive

modeling, enabling an LLM to predict the next EEG token analogous to language modeling. EEGFormer [194] focuses on reconstructing raw waveforms for codebook training, followed by BERT-style masked signal reconstruction pretraining. These methods demonstrate how VQ-based tokenization adapts to EEG modeling—whether prioritizing spectral synchrony (LaBraM), fusing time-frequency features (NeuroLM), or preserving temporal fidelity (EEGFormer).

### C. BrainBenchmark

The development of universal pre-trained frameworks represents a transformative advancement in healthcare, enabling the integration of heterogeneous datasets and generalization across diverse diagnostic tasks. However, existing studies—whether supervised or self-supervised—often adopt inconsistent dataset usage, validation splits, and evaluation metrics. These inconsistencies make it difficult to fairly compare different paradigms and accurately assess progress in the field. To address this issue, we have established an open benchmark that provides a unified evaluation standard and toolset for the community. It currently includes 8 representative models and 9 public datasets, with support for flexible model integration and dataset expansion. Our goal is to encourage researchers to adopt this common framework for consistent, reproducible benchmarking and to lower the barrier for integrating new methods. The implementation is publicly available at <https://github.com/ZJU-BrainNet/BrainBenchmark>, and we hope it will serve as a foundation for advancing universal pre-trained frameworks in EEG/iEEG analysis.

## V. CONCLUSION

This survey systematically reviews 448 studies and 46 public datasets to advance deep learning-driven analysis of EEG/iEEG signals across seven neurological diagnostic tasks: seizure detection, sleep staging and disorder, major depressive disorder, schizophrenia, Alzheimer’s disease, Parkinson’s disease, and ADHD. Our work establishes three foundational contributions: First, we unify fragmented methodologies across neurological conditions by standardizing data processing, model architectures, and evaluation protocols. Second, we identify self-supervised learning as the most promising paradigm for multi-task neurodiagnosis, providing a comprehensive overview of pre-trained SSL frameworks and their advancements. Third, we introduce BrainBenchmark, a unified platform that standardizes evaluations and integrates neurological datasets with diverse models, aiming to improve comparability and reproducibility across studies.

Looking back, the pursuit of universal models capable of learning from diverse, multimodal data reflects the field’s growing ambition, laying the groundwork for a new era of intelligent and adaptable healthcare systems. Over the past decades, significant progress has established a strong foundation for neurological diagnostics based on electrical brain signals. Key contributions include advances in signal pre-processing, curating large-scale, well-annotated datasets, and developing deep learning architectures for specific tasks. The integration of self-supervised pretraining marks a paradigm



shift, enabling models to extract rich and meaningful representations from vast amounts of unlabeled, heterogeneous data.

Looking forward, the ultimate goal is to develop genuinely universal and adaptable frameworks capable of transcending individual tasks and datasets to address a broader range of neurological disorders. These advancements will pave the way for intelligent diagnostic tools that deliver precise, efficient, and accessible healthcare solutions globally, driving transformative progress in biomedical research and clinical applications.

## REFERENCES

- [1] World Health Organization, "Over 1 in 3 people affected by neurological conditions, the leading cause of illness and disability worldwide," 2024.
- [2] —, "Dementia," 2025.
- [3] —, "Epilepsy," 2024.
- [4] —, "Mental disorders," 2025.
- [5] —, "Parkinson disease," 2023.
- [6] —, "Schizophrenia," 2024.
- [7] —, "Depression," 2024.
- [8] —, "Adolescent mental health," 2021.
- [9] H. Berger, "Über das elektroencephalogramm des menschen," *Archiv für psychiatrie und nervenkrankheiten*, vol. 87, no. 1, pp. 527–570, 1929.
- [10] G. Ramantani and et al., "Correlation of invasive eeg and scalp eeg," *Seizure*, vol. 41, pp. 196–200, 2016.
- [11] U. R. Acharya and et al., "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.
- [12] —, "Automated eeg-based screening of depression using deep convolutional neural network," *Computer methods and programs in biomedicine*, vol. 161, pp. 103–113, 2018.
- [13] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The handbook of brain theory and neural networks*, 1998, pp. 255–258.
- [14] A. Vaswani and et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [15] Ö. Türk and M. S. Özerdem, "Epilepsy detection by using scalogram based convolutional neural network from eeg signals," *Brain sciences*, vol. 9, no. 5, p. 115, 2019.
- [16] D. Kostas and et al., "Bendr: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data," *Frontiers in Human Neuroscience*, vol. 15, p. 653659, 2021.
- [17] Y. Roy and et al., "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051001, 2019.
- [18] N. S. Amer and S. B. Belhaoui, "Eeg signal processing for medical diagnosis, healthcare, and monitoring: A comprehensive review," *IEEE Access*, 2023.
- [19] X. Zhang and et al., "A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers," *Journal of neural engineering*, vol. 18, no. 3, p. 031002, 2021.
- [20] P. Khan and et al., "Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances," *Ieee Access*, vol. 9, pp. 37 622–37 655, 2021.
- [21] A. Shoeibi and et al., "Epileptic seizures detection using deep learning techniques: a review," *International journal of environmental research and public health*, vol. 18, no. 11, p. 5780, 2021.
- [22] J. Rahul and et al., "A systematic review of eeg based automated schizophrenia classification through machine learning and deep learning," *Frontiers in Human Neuroscience*, vol. 18, p. 1347082, 2024.
- [23] K. M. Hossain and et al., "Status of deep learning for eeg-based brain-computer interface applications," *Frontiers in computational neuroscience*, vol. 16, p. 1006763, 2023.
- [24] W. Weng and et al., "Self-supervised learning for electroencephalogram: A systematic survey," *arXiv preprint arXiv:2401.05446*, 2024.
- [25] J. Chen and et al., "Con4m: Context-aware consistency learning framework for segmented time series classification," *The Thirty-Eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [26] H. Banville and et al., "Uncovering the structure of clinical eeg signals with self-supervised learning," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046020, 2021.
- [27] S. L. Oh and et al., "A deep learning approach for parkinson's disease diagnosis from eeg signals," *Neural Computing and Applications*, vol. 32, pp. 10927–10933, 2020.
- [28] C. Chatzichristos and et al., "Epileptic seizure detection in eeg via fusion of multi-view attention-gated u-net deep neural networks," in *2020 IEEE Signal Processing in Medicine and Biology Symposium*, 2020, pp. 1–7.
- [29] B. Ay and et al., "Automated depression detection using deep representation and sequence learning with eeg signals," *Journal of medical systems*, vol. 43, p. 205, 2019.
- [30] N. Grover and et al., "Schizo-net: A novel schizophrenia diagnosis framework using late fusion multimodal deep learning on electroencephalogram-based brain connectivity indices," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 464–473, 2023.
- [31] M. Moghaddari and et al., "Diagnose adhd disorder in children using convolutional neural network based on continuous mental task eeg," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105738, 2020.
- [32] J. Sun and et al., "A hybrid deep neural network for classification of schizophrenia using eeg data," *Scientific Reports*, vol. 11, no. 1, p. 4706, 2021.
- [33] A. Nouri and Z. Tabanfar, "Detection of adhd disorder in children using layer-wise relevance propagation and convolutional neural network: An eeg analysis," *Frontiers in Biomedical Technologies*, vol. 11, no. 1, pp. 14–21, 2024.
- [34] A. Antoniadis and et al., "Deep learning for epileptic intracranial eeg data," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing*, 2016, pp. 1–6.
- [35] T. Wen and Z. Zhang, "Deep convolution neural network and autoencoders-based unsupervised feature learning of eeg signals," *IEEE Access*, vol. 6, pp. 25 399–25 410, 2018.
- [36] W. Ko and H.-I. Suk, "Eeg-oriented self-supervised learning and cluster-aware adaptation," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 4143–4147.
- [37] Z. Mousavi and et al., "Deep convolutional neural network for classification of sleep stages from single-channel eeg signals," *Journal of neuroscience methods*, vol. 324, p. 108312, 2019.
- [38] J. Zhang and et al., "Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel eeg," *Computer Methods and Programs in Biomedicine*, vol. 183, p. 105089, 2020.
- [39] S. Siuly and et al., "Exploring deep residual network based features for automatic schizophrenia detection from eeg," *Physics in Engineering and Science Medicine*, vol. 46, pp. 561–574, 2023.
- [40] G. Sharma and et al., "Dephnn: a novel hybrid neural network for electroencephalogram (eeg)-based screening of depression," *Biomedical signal processing and control*, vol. 66, p. 102393, 2021.
- [41] A. Seal and et al., "Depmnet: A deep convolution neural network framework for detecting depression using eeg," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [42] S. Iwama and et al., "Two common issues in synchronized multimodal recordings with eeg: Jitter and latency," *Neuroscience Research*, 2024.
- [43] H. H. Jasper, "Ten-twenty electrode system of the international federation," *Electroencephalogr Clin Neurophysiol*, vol. 10, pp. 371–375, 1958.
- [44] G. Buzsáki and A. Draguhn, "Neuronal oscillations in cortical networks," *Science*, vol. 304, no. 5679, pp. 1926–1929, 2004.
- [45] J. Jeong, "Eeg dynamics in patients with alzheimer's disease," *Clinical neurophysiology*, vol. 115, no. 7, pp. 1490–1505, 2004.
- [46] N. Sobahi and et al., "A new signal to image mapping procedure and convolutional neural networks for efficient schizophrenia detection in eeg recordings," *IEEE Sensors Journal*, vol. 22, no. 8, pp. 7913–7919, 2022.
- [47] H. W. Loh and et al., "Gaborpdnet: Gabor transformation and deep neural network for parkinson's disease detection using eeg signals," *Electronics*, vol. 10, no. 14, p. 1740, 2021.
- [48] A. Zülfiqar and A. Mehmet, "Empirical mode decomposition and convolutional neural network-based approach for diagnosing psychotic disorders from eeg signals," *Applied Intelligence*, vol. 52, no. 11, pp. 12 103–12 115, 2022.
- [49] X. Li and et al., "Eeg-based mild depression recognition using convolutional neural network," *Medical & biological engineering & computing*, vol. 57, pp. 1341–1352, 2019.



- [50] A. M. Karim and et al., "A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 148–159, 2019.
- [51] D. M. Khan and et al., "Automated diagnosis of major depressive disorder using brain effective connectivity and 3d convolutional neural network," *Ieee Access*, vol. 9, pp. 8835–8846, 2021.
- [52] J. Zhu and et al., "Multimodal mild depression recognition based on eeg-em synchronization acquisition network," *IEEE Access*, vol. 7, pp. 28 196–28 210, 2019.
- [53] J. Liu and B. Woodson, "Deep learning classification for epilepsy detection using a single channel electroencephalography (eeg)," in *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*, 2019, pp. 23–26.
- [54] X. Li and et al., "Depression recognition using machine learning methods with different feature generation strategies," *Artificial intelligence in medicine*, vol. 99, p. 101696, 2019.
- [55] I. Bhattacharjee, "Epileptic seizure detection using multicolumn convolutional neural network," in *2020 7th International Conference on Computing for Sustainable Global Development*, 2020, pp. 58–63.
- [56] M. Tosun, "Effects of spectral features of eeg signals recorded with different channels and recording statuses on adhd classification with deep learning," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 3, pp. 693–702, 2021.
- [57] Z. Aslan and M. Akin, "A deep learning approach in automated detection of schizophrenia using scalogram images of eeg signals," *Physical and Engineering Sciences in Medicine*, vol. 45, no. 1, pp. 83–96, 2022.
- [58] G. Choi and et al., "A novel multi-scale 3d cnn with deep neural network for epileptic seizure detection," in *2019 IEEE International Conference on Consumer Electronics*, 2019, pp. 1–2.
- [59] A. Vilamala and et al., "Deep convolutional neural networks for interpretable analysis of eeg sleep stage scoring," in *2017 IEEE 27th international workshop on machine learning for signal processing*, 2017, pp. 1–6.
- [60] J. Xu and et al., "Epileptic seizure detection based on feature extraction and cnn- bigru network with attention mechanism," in *International Conference on Intelligent Computing*. Springer, 2023, pp. 308–319.
- [61] C.-R. Phang and et al., "A multi-domain connectome convolutional neural network for identifying schizophrenia from eeg connectivity patterns," *IEEE journal of biomedical and health informatics*, vol. 24, no. 5, pp. 1333–1343, 2019.
- [62] D.-W. Ko and J.-J. Yang, "Eeg-based schizophrenia diagnosis through time series image conversion and deep learning," *Electronics*, vol. 11, no. 14, 2022.
- [63] A. Sohrabpour and et al., "Noninvasive electromagnetic source imaging of spatiotemporally distributed epileptogenic brain sources," *Nature communications*, vol. 11, no. 1, p. 1946, 2020.
- [64] R. Sun and et al., "Deep neural networks constrained by neural mass models improve electrophysiological source imaging of spatiotemporal brain dynamics," *Proceedings of the National Academy of Sciences*, vol. 119, no. 31, p. e2201128119, 2022.
- [65] Q. Zhan and W. Hu, "An epilepsy detection method using multiview clustering algorithm and deep features," *Computational and Mathematical Methods in Medicine*, vol. 2020, p. 5128729, 2020.
- [66] T. K. K. Ho and N. Armanfard, "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, 2023, pp. 7866–7874.
- [67] K. A. Robbins and et al., "How sensitive are eeg results to preprocessing methods: A benchmarking study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 5, pp. 1081–1090, 2020.
- [68] R. Akut, "Wavelet based deep learning approach for epilepsy detection," *Health information science and systems*, vol. 7, no. 1, p. 8, 2019.
- [69] S.-H. Oh and et al., "A novel eeg feature extraction method using hjorth parameter," *International Journal of Electronics and Electrical Engineering*, vol. 2, no. 2, pp. 106–110, 2014.
- [70] A. Zancanaro and et al., "Cnn-based approaches for cross-subject classification in motor imagery: From the state-of-the-art to dynamicnet," in *2021 IEEE conference on computational intelligence in bioinformatics and computational biology*, 2021, pp. 1–7.
- [71] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
- [72] F. Scarselli and et al., "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009.
- [73] G. E. Hinton and R. Zemel, "Autoencoders, minimum description length and helmholtz free energy," in *Advances in neural information processing systems*, vol. 6, 1993.
- [74] I. Goodfellow and et al., "Generative adversarial nets," in *Advances in neural information processing systems*, vol. 27, 2014.
- [75] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural networks*, vol. 10, no. 9, pp. 1659–1671, 1997.
- [76] R. G. Andrzejak and et al., "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*, vol. 64, no. 6, p. 061907, 2001.
- [77] M. Ihle and et al., "Epilepsiae – A European epilepsy database," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 127–138, 2012.
- [78] bbrinkm and et al., "Upenn and mayo clinic's seizure detection challenge," 2014.
- [79] J. Guttg, "CHB-MIT Scalp EEG Database," 2010.
- [80] A. Shueb, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
- [81] A. Goldberger and et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [82] R. G. Andrzejak and et al., "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients," *Physical Review E*, vol. 86, p. 046206, 2012.
- [83] P. Swami and et al., "Eeg epilepsy datasets," 2016.
- [84] L. Kuhlmann and et al., "Melbourne university aes/mathworks/nih seizure prediction," 2016.
- [85] V. Shah and et al., "The Temple University Hospital Seizure Detection Corpus," *Frontiers in Neuroinformatics*, vol. 12, p. 83, 2018.
- [86] A. Burrello and et al., "One-shot learning for ieeg seizure detection using end-to-end binary operations: Local binary patterns with hyperdimensional computing," in *Proceedings of the IEEE Biomedical Circuits and Systems Conference*, 2018.
- [87] —, "Hyperdimensional computing with local binary patterns: One-shot learning of seizure onset and identification of ictogenic brain regions using short-time ieeg recordings," *IEEE Transactions on Biomedical Engineering*, 2020.
- [88] N. Stevenson and et al., "A dataset of neonatal EEG recordings with seizure annotations," *Scientific Data*, vol. 6, p. 190039, 2019.
- [89] P. Nejedly and et al., "Multicenter intracranial eeg dataset for classification of graphoelements and artifactual signals," *Scientific Data*, vol. 7, no. 1, p. 179, 2020.
- [90] P. Detti and et al., "Eeg synchronization analysis for seizure prediction: A study on data of noninvasive recordings," *Processes*, vol. 8, no. 7, p. 846, 2020.
- [91] W. Nasreddine, "Epileptic eeg dataset," 2021.
- [92] J. M. Bernabei and et al., "Hup ieeg epilepsy dataset," 2022.
- [93] D. van Blooij and et al., "Ccep ecog dataset across age 4-51," 2023.
- [94] M. N. Mohsenvand and et al., "Contrastive representation learning for electroencephalogram classification," in *Machine Learning for Health*, 2020, pp. 238–253.
- [95] H. Banville and et al., "Self-supervised representation learning from electroencephalography signals," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing*, 2019, pp. 1–6.
- [96] A. v. d. Oord and et al., "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.
- [97] D. Wu and et al., "neuro2vec: Masked fourier spectrum prediction for neurophysiological representation learning," *arXiv preprint arXiv:2204.12440*, 2022.
- [98] D. Cai and et al., "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*, 2023, pp. 130–141.
- [99] I. Ullah and et al., "An automated system for epilepsy detection using eeg brain signals based on deep learning approach," *Expert Systems with Applications*, vol. 107, pp. 61–71, 2018.
- [100] A. M. Abdelhameed and et al., "Epileptic seizure detection using deep convolutional autoencoder," in *2018 IEEE international workshop on signal processing systems*, 2018, pp. 223–228.
- [101] M. Zhou and et al., "Epileptic seizure detection based on eeg signals and cnn," *Frontiers in neuroinformatics*, vol. 12, p. 95, 2018.
- [102] A. K. Dutta and et al., "Deep learning-based multi-head self-attention model for human epilepsy identification from eeg signal for biomedical traits," *Multimedia Tools and Applications*, p. 80201, 2024.

- [103] Y. Zhao, , and et al., "Hybrid attention network for epileptic eeg classification," *International Journal of Neural Systems*, vol. 33, no. 06, p. 2350031, 2023.
- [104] T. Xiao and et al., "Self-supervised learning with attention mechanism for eeg-based seizure detection," *Biomedical Signal Processing and Control*, vol. 87, p. 105464, 2024.
- [105] Y. Zheng and et al., "Task-oriented self-supervised learning for anomaly detection in electroencephalography," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022, pp. 193–203.
- [106] S. Raghu and et al., "Eeg based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Networks*, vol. 124, pp. 202–212, 2020.
- [107] S. Tang and et al., "Self-supervised graph neural networks for improved electroencephalographic seizure analysis," *10th International Conference on Learning Representations*, 2022.
- [108] R. Peng and et al., "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*, 2023, pp. 1–5.
- [109] X. Xu and et al., "Patient-specific method for predicting epileptic seizures based on drsn-gru," *Biomedical Signal Processing and Control*, vol. 81, p. 104449, 2023.
- [110] L. Sui and et al., "Localization of epileptic foci by using convolutional neural network based on ieeg," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 2019, pp. 331–339.
- [111] O. S. Lih and et al., "Epilepsynet: Novel automated detection of epilepsy using transformer model with eeg signals from 121 patient population," *Computers in Biology and Medicine*, vol. 164, p. 107312, 2023.
- [112] D. Zhang and et al., "Brant: Foundation model for intracranial neural signal," in *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [113] J. Guo and et al., "Detecting high frequency oscillations for stereoelectroencephalography in epilepsy via hypergraph learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 587–596, 2021.
- [114] A. Rahmani and et al., "A meta-gnn approach to personalized seizure detection and classification," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.
- [115] Y. Sun and et al., "Continuous seizure detection based on transformer and long-term ieeg," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5418–5427, 2022.
- [116] S. Tu and et al., "Dmnet: Self-comparison driven model for subject-independent seizure detection," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- [117] S. You and et al., "Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear eeg using an anomaly-detecting generative adversarial network," *Computer Methods and Programs in Biomedicine*, vol. 193, p. 105472, 2020.
- [118] X. Li and V. Metsis, "Spp-eegnet: An input-agnostic self-supervised eeg representation model for inter-dataset transfer learning," in *International Conference on Computing and Information Technology*, 2022, pp. 173–182.
- [119] J. Chen and et al., "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*. ACM, 2022, pp. 2741–2751.
- [120] Z. Yuan and et al., "Ppi: Pretraining brain signal model for patient-independent seizure detection," in *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [121] B. Kemp and et al., "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000.
- [122] C. O'Reilly and et al., "Montreal Archive of Sleep Studies: An open-access resource for instrument benchmarking and exploratory research," *Journal of Sleep Research*, vol. 23, no. 6, pp. 628–635, 2014.
- [123] S. F. Quan and et al., "The Sleep Heart Health Study: Design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997.
- [124] G. Q. Zhang and et al., "The National Sleep Research Resource: Towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351–1358, 2018.
- [125] "St. Vincent's University Hospital / University College Dublin Sleep Apnea Database," 2007.
- [126] D. Alvarez-Estevéz and R. Rijsman, "Haaglanden medisch centrum sleep staging database," 2022.
- [127] M. M. Ghassemi and et al., "You Snooze, You Win: The PhysioNet/Computing in Cardiology Challenge 2018," in *2018 Computing in Cardiology Conference*, 2018, pp. 1–4.
- [128] Y. Ichimaru and G. B. Moody, "Development of the polysomnographic database on CD-ROM," *Psychiatry and Clinical Neurosciences*, vol. 53, pp. 175–177, 1999.
- [129] A. Guillot and et al., "Dreem open datasets: Multi-scored sleep datasets to compare human and automated sleep staging," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 1955–1965, 2020.
- [130] S. Khalighi and et al., "ISRUC-Sleep: A comprehensive public dataset for sleep researchers," *Computer Methods and Programs in Biomedicine*, vol. 124, pp. 180–192, 2016.
- [131] S. Biswal and et al., "Expert-level sleep scoring with deep neural networks," *Journal of the American Medical Informatics Association*, vol. 25, no. 12, pp. 1643–1650, 2018.
- [132] A. Piryatinska and et al., "Automated detection of neonate eeg sleep stages," *Computer methods and programs in biomedicine*, vol. 95, no. 1, pp. 31–46, 2009.
- [133] S. Devuyst, "The dreams databases and assessment algorithm," 2005.
- [134] C. Xiang and et al., "A resting-state eeg dataset for sleep deprivation."
- [135] E. Eldele and et al., "An attention-based deep learning approach for sleep stage classification with single-channel eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 809–818, 2021.
- [136] K. H and et al., "Accurate deep learning-based sleep staging in a clinical population with suspected obstructive sleep apnea," *IEEE journal of biomedical and health informatics*, vol. 24(7), pp. 2073–2081, 2019.
- [137] C. Yang and et al., "Self-supervised electroencephalogram representation learning for automatic sleep staging: model development and evaluation study," *JMIR AI*, vol. 2, no. 1, p. e46769, 2023.
- [138] V. Kumar and et al., "muleeg: a multi-view representation learning on eeg signals," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022, pp. 398–407.
- [139] J. Ye and et al., "Cosleep: A multi-view representation learning framework for self-supervised learning of sleep stage classification," *IEEE Signal Processing Letters*, vol. 29, pp. 189–193, 2022.
- [140] D. Zhang and et al., "Brant-x: A unified physiological signal alignment framework," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 4155–4166.
- [141] The Recovery Village, "Sleep disorders statistics," 2023.
- [142] O. Tsalis and et al., "Automatic sleep stage scoring with single-channel eeg using convolutional neural networks," *arXiv preprint arXiv:1610.01683*, 2016.
- [143] W. Mumtaz, "MDD Patients and Healthy Controls EEG Data," 2016.
- [144] J. Cavanagh and et al., "The patient repository for eeg data + computational tools (pred+ct)," *Frontiers in Neuroinformatics*, vol. 11, p. 67, 2017.
- [145] L. Yang and et al., "Automatic feature learning model combining functional connectivity network and graph regularization for depression detection," *Biomedical Signal Processing and Control*, vol. 82, p. 104520, 2023.
- [146] H. Cai and et al., "A multi-modal open dataset for mental-disorder analysis," *Scientific Data*, vol. 9, no. 1, p. 178, 2022.
- [147] W. Mumtaz and A. Qayyum, "A deep learning framework for automatic diagnosis of unipolar depression," *International Journal of Medical Informatics*, vol. 132, p. 103983, 2019.
- [148] L. Li and et al., "An eeg-based marker of functional connectivity: detection of major depressive disorder," *Cognitive Neurodynamics*, vol. 18, no. 4, pp. 1671–1687, 2024.
- [149] B. Wang and et al., "Depression signal correlation identification from different eeg channels based on cnn feature extraction," *Psychiatry Research: Neuroimaging*, vol. 328, p. 111582, 2023.
- [150] P. P. Thoduparambil and et al., "Eeg-based deep learning model for the automatic detection of clinical depression," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 4, pp. 1349–1360, 2020.
- [151] L. Yang and et al., "A gated temporal-separable attention network for eeg-based depression recognition," *Computers in Biology and Medicine*, vol. 157, p. 106782, 2023.
- [152] X. Sun and et al., "Multi-granularity graph convolution network for major depressive disorder recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 559–569, 2023.
- [153] Z. Zhang and et al., "A novel eeg-based graph convolution network for depression detection: incorporating secondary subject partitioning and attention mechanism," *Expert Systems with Applications*, vol. 239, p. 122356, 2024.

- [154] A. Sam and et al., "Depression identification using eeg signals via a hybrid of lstm and spiking neural networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 4725–4737, 2023.
- [155] E. Olejarczyk and W. Jernajczyk, "Graph-based analysis of brain connectivity in schizophrenia," *PloS one*, vol. 12, no. 11, 2017.
- [156] J. M. Ford and et al., "Did i do that? abnormal predictive processes in schizophrenia when button pressing to deliver a tone," *Schizophrenia bulletin*, vol. 40, no. 4, pp. 804–812, 2014.
- [157] S. Borisov and et al., "Analysis of eeg structural synchrony in adolescents with schizophrenic disorders," *Human Physiology*, vol. 31, pp. 255–261, 2005.
- [158] S. L. Oh and et al., "Deep convolutional neural network model for automated diagnosis of schizophrenia using eeg signals," *Applied Sciences*, vol. 9, no. 14, p. 2870, 2019.
- [159] Z. Guo and et al., "Deep neural network classification of eeg data in schizophrenia," in *2021 IEEE 10th Data Driven Control and Learning Systems Conference*, 2021, pp. 1322–1327.
- [160] R. Supakar and et al., "A deep learning based model using rnn-lstm for the detection of schizophrenia from eeg data," *Computers in Biology and Medicine*, vol. 151, p. 106225, 2022.
- [161] S. K. Khare and et al., "Schizonet: a robust and accurate marginau-hill time-frequency distribution based deep neural network model for schizophrenia detection using eeg signals," *Physiological Measurement*, vol. 44, no. 3, p. 035005, 2023.
- [162] A. Shalhaf and et al., "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from eeg signals," *Physical and Engineering Sciences in Medicine*, vol. 43, pp. 1229–1239, 2020.
- [163] S. K. Khare and et al., "Spwvd-cnn for automated detection of schizophrenia patients using eeg signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.
- [164] D. Shah and et al., "Deep learning of eeg data in the neucube brain-inspired spiking neural network architecture for a better understanding of depression," in *Neural Information Processing: 26th International Conference*, 2019, pp. 195–206.
- [165] A. Zalesky and et al., "Disrupted axonal fiber connectivity in schizophrenia," *Biological psychiatry*, vol. 69, no. 1, pp. 80–89, 2011.
- [166] M. L. Vicchietti and et al., "Data from: Computational methods of eeg signals analysis for alzheimer's disease classification," 2023.
- [167] A. Miltiadous and et al., "A dataset of 88 eeg recordings from: Alzheimer's disease, frontotemporal dementia and healthy subjects," 2023.
- [168] G. Fiscon and et al., "In alzheimer's disease patients classification through eeg signals processing," in *Computational Intelligence & Data Mining*, 2014, pp. 105–112.
- [169] M. Cejnek and et al., "Novelty detection-based approach for alzheimer's disease and mild cognitive impairment diagnosis from eeg," *Medical & Biological Engineering & Computing*, vol. 59, no. 11, pp. 2287–2296, 2021.
- [170] A. C. L and et al., "Eeg functional connectivity and deep learning for automatic diagnosis of brain disorders: Alzheimer's disease and schizophrenia," *Journal of Physics: complexity*, vol. 3(2), p. 025001, 2022.
- [171] X. Shan and et al., "Spatial-temporal graph convolutional network for alzheimer classification based on brain functional connectivity imaging of electroencephalogram," *Human Brain Mapping*, vol. 43, no. 17, pp. 5194–5209, 2022.
- [172] C. Ieracitano and et al., "A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings," *Neurocomputing*, vol. 323, pp. 96–107, 2019.
- [173] D. Labate and et al., "Eeg complexity modifications and altered compressibility in mild cognitive impairment and alzheimer's disease," in *Recent Advances of Neural Network Models and Applications: Proceedings of the 23rd Workshop of the Italian Neural Networks Society*, 2014, pp. 163–173.
- [174] A. P. Rockhill and et al., "Uc san diego resting state eeg data from patients with parkinson's disease," 2021.
- [175] J. F. Cavanagh and et al., "Diminished eeg habituation to novel events effectively classifies parkinson's patients," *Clinical Neurophysiology*, vol. 129, no. 2, pp. 409–418, 2018.
- [176] A. Singh and et al., "Frontal theta and beta oscillations during lower-limb movement in parkinson's disease," *Clinical Neurophysiology*, vol. 131, pp. 694–702, 2020.
- [177] M. Shaban, "Automated screening of parkinson's disease using deep learning based electroencephalography," in *2021 10th international IEEE/EMBS conference on neural engineering*, 2021, pp. 158–161.
- [178] M. Shaban and A. W. Amara, "Resting-state electroencephalography based deep-learning for the detection of parkinson's disease," *Plos one*, vol. 17, no. 2, p. e0263159, 2022.
- [179] S. K. Khare and et al., "Pdcnnnet: An automatic framework for the detection of parkinson's disease using eeg signals," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17017–17024, 2021.
- [180] A. Morita and et al., "Relationship between slowing of the eeg and cognitive impairment in parkinson disease," *Journal of Clinical Neurophysiology*, vol. 28, no. 4, pp. 384–387, 2011.
- [181] G. S. Bajestani and et al., "A dataset of eeg signals from adults with adhd and healthy controls: Resting state, cognitive function, and sound listening paradigm," 2023.
- [182] M. Nasrabadi and et al., "Eeg data for adhd / control children," 2020.
- [183] M. Bakhtyari and S. Mirzaei, "Adhd detection using dynamic connectivity patterns of eeg data and convlstm with attention framework," *Biomedical Signal Processing and Control*, vol. 76, p. 103708, 2022.
- [184] National Institute of Mental Health. (2023) Attention-deficit/hyperactivity disorder.
- [185] H. Chen and et al., "Use of deep learning to detect personalized spatial-frequency abnormalities in eegs of children with adhd," *Journal of neural engineering*, vol. 16, no. 6, p. 066046, 2019.
- [186] E. Eldele and et al., "Time-series representation learning via temporal and contextual contrasting," *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 2021.
- [187] X. Zhang and et al., "Self-supervised contrastive pre-training for time series via time-frequency consistency," *Advances in Neural Information Processing Systems*, vol. 35, pp. 3988–4003, 2022.
- [188] C. Yang and et al., "Biot: Biosignal transformer for cross-data learning in the wild," in *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [189] S. Jo and et al., "Channel-aware self-supervised learning for eeg-based bci," in *2023 11th International Winter Conference on Brain-Computer Interface (BCI)*, 2023, pp. 1–4.
- [190] W. Zhang and et al., "Self-supervised time series representation learning via cross reconstruction transformer," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [191] D. Wu and et al., "Neuro-bert: Rethinking masked autoencoding for self-supervised neurological pretraining," *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [192] J. Wang and et al., "Cbramod: A criss-cross brain foundation model for eeg decoding," *The Thirteenth International Conference on Learning Representations*, 2025.
- [193] Z. Yuan and et al., "Brainwave: A brain signal foundation model for clinical applications," 2024.
- [194] Y. Chen and et al., "Eegformer: Towards transferable and interpretable large-scale eeg foundation model," in *AAAI 2024 Spring Symposium on Clinical Foundation Models*, 2024.
- [195] W.-B. Jiang and et al., "Large brain model for learning generic representations with tremendous eeg data in bci," *The Twelfth International Conference on Learning Representations*, 2024.
- [196] W. Jiang and et al., "Neurolm: A universal multi-task foundation model for bridging the gap between language and eeg signals," *The Thirteenth International Conference on Learning Representations*, 2025.
- [197] J. Devlin, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [198] A. Van Den Oord and et al., "Neural discrete representation learning," in *Advances in neural information processing systems*, vol. 30, 2017.