An Improved Self-Supervised Framework on EEG Signals for Seizure Detection and Classification

Jia Han[1,a], Xuande Zhang[1,b], Long Xu[2,c], Zhijie Gao[3,d], Xin Huang[2,e,*]

[1]School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, 710021, China
[2]Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, 315211, China
[3]Department of Neurology, Capital Center for Children's Health, Capital Medical University, Beijing, 100045, China
[a]Jea_Han@163.com, [b]love_truth@126.com, [c]xulong1@nbu.edu.cn, [d]gaozhijie20082008@163.com, [e]huangxin@nbu.edu.cn
**Corresponding author*

Abstract: *Supervised learning approaches for deep learning-based epilepsy detection from Electroencephalogram (EEG) signals face significant limitations, including poor generalization across patients and a heavy reliance on large-scale labeled datasets. The acquisition of such datasets is highly labor-intensive, which in turn restricts the practical deployment of these methods. Here, we propose a self-supervised learning (SSL) framework to reduce this dependency for seizure detection and classification. Our method combines a time-frequency data augmentation module with a representation-level reconstruction task, guided by a novel semantic-subsequence-preserving (SSP) masking strategy, to learn semantic representations from unlabeled EEG. When evaluated on 5,499 public EEG recordings, our model achieves an AUROC of 0.848 for detection and a weighted F1-score of 0.900 for classification. This demonstrates the ability of our SSL approach to deliver high performance with minimal labeled data, offering a promising path toward more scalable and accurate clinical diagnostic tools.*

Keywords: *self-supervised learning; Electroencephalogram; seizure detection; seizure classification*

1. Introduction

Epilepsy, caused by abnormal discharges of neurons in the brain, has become one of the most common neurological disorders worldwide[1]. Due to differences in the origin and propagation of abnormal brain electrical activity, the clinical manifestations of epilepsy are diverse and complex. These variations can lead to significant challenges in diagnosis and treatment. In clinical practice, diagnosis is primarily based on EEG recordings to assess brain activity during seizures[2].The interpretation and analysis of EEG signals mainly rely on visual inspection and manual annotation by medical professionals. However, the unpredictable onset and duration of seizures make it labor-intensive to extract relevant segments from large volumes of EEG data. This process is also highly dependent on the subjective judgment of experts.

In recent years, substantial advancements have been made in automated seizure detection and classification, offering clinicians efficient tools for assisted diagnosis[3-5]. With the increasing adoption of deep learning in biomedical fields, its superiority over traditional machine learning methods in seizure-related tasks has been well-documented. Liu et al. [6] proposed a hybrid architecture that integrates convolutional neural networks (CNNs) with long short-term memory (LSTM) networks to improve epilepsy classification. Similarly, Priyasad et al.[7] developed a deep learning framework leveraging attention-based data fusion to optimize seizure type identification. Further advancing the field, Arshia et al.[8] introduced a residual state update mechanism (REST), which combines graph neural networks (GNNs) with recurrent structures to enable real-time EEG signal analysis. More recently, models such as the Transformer [9-12] and Mamba [13-16] have also demonstrated success in epileptic seizure detection and classification tasks.

While effective, these supervised learning methods depend heavily on large-scale labeled datasets to achieve robust generalization. However, high-quality annotated EEG data remain scarce in practice. To address this limitation and leverage abundant unlabeled data, researchers have increasingly turned to self-supervised learning (SSL) for epilepsy diagnosis.

Recent work has explored SSL techniques for EEG analysis. Tang et al. [17] adopted SSL-based pre-training for graph neural networks to improve epilepsy detection and classification. Lam et al. [18] proposed a masked prediction task to reconstruct intracranial EEG spectrograms, enabling label-free representation learning. Han et al. [19] introduced a contrastive learning framework with a channel-abnormality detection pretext task. Collectively, these studies highlight the potential of SSL paradigms to enhance EEG-based models via task-specific pre-training. Nevertheless, their representation capacity is often insufficient for complex epilepsy diagnosis tasks, leaving room for further improvement.

Moreover, the reconstruction of raw EEG signals may amplify inherent recording noise. Compared to conventional self-prediction methods that directly reconstruct raw EEG signals [20-21], reconstructing more abstract representations in the latent space has demonstrated superior effectiveness - an approach already validated in image and text representation learning [22-24] and recently adapted to EEG data [25].

To address these challenges, we propose a novel self-supervised learning (SSL) framework incorporating advanced data augmentation techniques. Specifically, we introduce a time-frequency mixed augmentation strategy to enhance the temporal stability of learned semantic representations. Furthermore, by leveraging a masked prediction task in the latent space, our model captures intrinsic data correlations while generating highly informative representations. This approach significantly reduces the dependency on labeled data while maintaining robust performance.

2. Methods

2.1. Overall Model Architecture

The self-supervised model proposed in this study is illustrated in Figure 1. Each EEG sample $X_i = \{x_1, x_2, \dots, x_L\}$ from the dataset D is transformed into a latent representation $R_i \in R^{d_e}$ of dimension d_e, where d_e denotes the embedding size. The network consists of three main components: a data augmentation module, an encoder module, and reconstruction module. The overall architecture of the proposed model is illustrated in Figure 1. The process begins by applying data augmentation to the input signals. These augmented signals are then fed into an encoder to generate effective representations. Subsequently, a portion of these representations is masked, and the model is trained to reconstruct the masked parts as a pretext task.

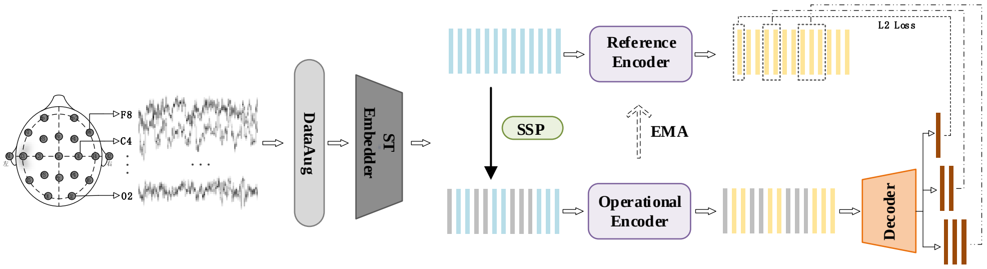

Figure 1. The overall architecture of the proposed model.

2.2. Data Augmentation Module

In time-series representation learning, designing data augmentation strategies that balance semantic consistency with model robustness remains a critical challenge. Conventional augmentation approaches often rely on task-specific priors, which can easily disrupt the intrinsic temporal dependencies in the data. Our method builds upon the FT-Aug framework proposed by Liu et al.[26], which constructs augmented views via frequency mixing and overlapping cropping to preserve temporal semantics. Instead of employing an explicit contrastive loss, we implement an implicit contrastive learning mechanism through a dual-path masked reconstruction task. By applying distinct augmentation perturbations to the same time segment and performing masked reconstruction separately, the encoder is guided to learn perturbation-invariant feature representations. An overview of the proposed data augmentation framework is illustrated in Figure 2.

The proposed framework consists of two key components:

Overlapping Cropping: To enable effective dual-path learning, it is essential to ensure strict semantic consistency between the two input views. We design an overlapping cropping algorithm: for an input sequence X_i, a target length T and a start point t are randomly determined, defining a core interval $[t, t+T]$. Two segments both covering this core interval are then generated by introducing controlled random offsets:

Original segment X_{orig}: obtained by directly cropping the raw signal of X_i within $[t, t+T]$.

Augmented segment X_{aug}: It is derived by applying frequency mixing to a segment of X_i encompassing the interval$[t, t+T]$, and subsequently truncating the result to the core interval$[t, t+T]$.

Frequency Mixing: To preserve the macroscopic structure of the time series during augmentation, we adopt a frequency-domain mixing strategy. This operation exchanges partial frequency components between randomly selected samples within the same batch to construct semantically consistent hard samples. Specifically, for a candidate segment X_i , it is first transformed into the frequency domain via the Fast Fourier Transform (FFT) to obtain its spectrum F. A random proportion of non-dominant frequency components is then replaced with the corresponding components from another randomly selected sample $X_j$in the same batch, yielding a mixed spectrum F'. An inverse FFT is applied to reconstruct the augmented segment X_{aug} in the time domain.

Notably, the frequency mixing operation is performed on a segment longer than the target core interval. This design allows the augmentation process to leverage richer spectral context, thereby generating more diverse and challenging augmented examples. The resulting augmented segment is then cropped to the same core interval $[t, t+T]$ as the original view, forming the augmented view X_{aug} . This approach ensures strict semantic alignment while significantly enhancing augmentation diversity and regularization strength, effectively encouraging the model to learn robust features that are invariant to frequency-domain perturbations.

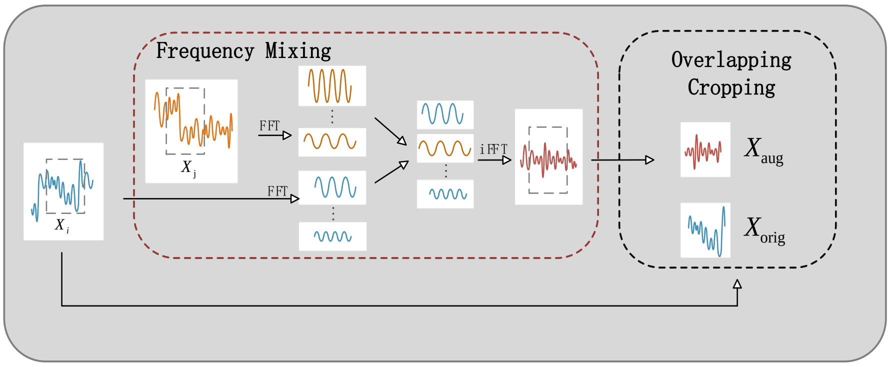

Figure 2. Data Augmentation: Cropping and Frequency Mixing.

2.3. Patch Embedding and Masking

Following the data augmentation, the Patch Embedding and Masking module transforms the processed EEG signals into a structured sequence of embedded patches and prepares them for self-supervised pre-training. This module consists of two key operations: spatial-temporal embedding which projects the input into patch representations, and strategic masking which constructs the pretext task objectives.

The ST Embedder: Building upon established methodologies—particularly the approach of Mohammadi Foumani et al.[25] ,which enhances the signal-to-noise ratio (SNR) by leveraging systematic spatial differences between neural signals and noise. Neural activities typically exhibit structured topographical distributions, whereas noise tends to manifest stochastic spatial patterns[27]. We maintain a depth-wise convolution layer followed by spatial filtering to ensure effective denoising while preserving semantically meaningful neural features. Furthermore, we introduce a multi-scale convolutional module to capture temporal dynamics across varying receptive fields. This design enables the extraction of comprehensive temporal representations that incorporate both fine-grained variations and broader contextual information.

For each EEG sample X_k, a depth-wise convolutional layer is applied across channels to extract spatial dependencies, while linear spatial filters are applied to amplify the signal-to-noise ratio. The filtered output is then processed by the multi-scale temporal module. Finally, after adding positional encoding to each patch, the output of this network consists of embedded EEG patches $S_x = \{S_{x1}, \dots, S_{xl}\}$, where $S_{xi} \in R^{d_x}$,d_x are the embedding dimensions, and l is the number of fragments.

Masking Strategy: To address the challenge of high amplitude variability in raw EEG signals, we employ a representation-level masked prediction task using the Semantic Subsequence Preserving (SSP) method introduced by Mohammadi Foumani et al. [25]. This task is conducted in the latent space, which encourages the model to learn robust and discriminative features while mitigating the influence of noise. The SSP strategy ensures the masking process preserves semantically continuous subsequences, thereby improving the model's efficacy for downstream tasks.

The SSP method is particularly suited to this objective due to its strategy of preserving semantically continuous subsequences during masking. Instead of randomly selecting patches to remove, it actively determines which time steps to retain in a block-wise manner. This design ensures that the encoder receives input with coherent contextual structure, which helps prevent the model from being misled by high-amplitude but semantically sparse segments and guides it to learn representations based on meaningful neurophysiological contexts.

Formally, for the preprocessed EEG patches$S_x = \{S_{x1}, \dots, S_{xl}\}$, we construct a preserved block set $B = \{B_k\}_{k=1}^{\beta}$ according to a masking ratio ρ and block count β. The width of each block B_k is given by:

$$B = \lceil ((1-\rho) \times l)/\beta \rceil \tag{1}$$

The visible subset P is formed by concatenating these preserved blocks selected S_x from via the SSP strategy. Then passed as input to the encoder, enabling it to learn from structured and semantically informative subsequences.

2.4. Encoder-Decoder Architecture

This encoder-decoder architecture is designed for EEG signal representation learning. It employs a Transformer encoder with multi-head attention and a decoder with cross-attention mechanisms to achieve effective self-supervised pre-training.

The encoder is built on a shared Transformer architecture with multi-head attention, which allows parallel attention computations across multiple representation subspaces to capture diverse high-level features. It consists of two components: a Reference Encoder and an Operational Encoder. The Reference Encoder operates on the complete set of EEG patches, whereas the Operational Encoder handles the masked patches. A momentum-based strategy is employed for updating both encoders: the Operational Encoder is directly optimized via gradient descent, whereas the Reference Encoder is updated via an Exponential Moving Average (EMA) strategy[28-29], ensuring stable and robust representation learning.

Specifically, the complete set of EEG patches $S_x = \{S_{x1}, \dots, S_{xl}\}$obtained from the ST Embedder is processed by the Reference Encoder $f_{\hat{\theta}}$ which transforms it into the patch-level representation $= \{y_1, \dots, y_l\}$:

$$y = f_{\hat{\theta}}(S_x) \tag{2}$$

Where $y_i \in R^{d_e}$ and d_e are the embedding dimensions of the Transformer.

Concurrently, a visible subset P is obtained from S_x is encoded by the Operational Encoder f_θ into a latent representation :

$$r = f_\theta(P) \tag{3}$$

The Reference Encoder acts as a momentum encoder, updated through the EMA synchronization mechanism $\hat{\theta} = \tau\hat{\theta} + (1-\tau)\theta$. Its lagged updated parameters $\hat{\theta}$ provide a consistent learning objective for the prediction task. After pre-training, only the Operational Encoder is used for downstream tasks.

The decoder generates predicted representations for the masked regions based on the encoded visible context. Given the masked positions $I_m = \{1, \dots, l\} \backslash B$, where B are SSP preserved blocks, we randomly sample M target blocks $\{B_i\}_{i=1}^{M}$. For each target block B_i, the predictor g_φ takes as input the latent representation r from the Operational Encoder and a set of mask tokens $m(i) = \{m_j\}_{j \in B_i}$ for each patch we wish to predict and outputs a patch-level prediction $\hat{y}_i$:

$$\hat{y}_i = g_\varphi(r, m(i)) \tag{4}$$

The model is optimized by minimizing the L2 loss between the predicted representation $\hat{y}_i$ and the corresponding true target representation y_i from the Reference Encoder. The loss function for the masked prediction task is defined as:

$$L_{mask} = \frac{1}{|M|} \sum_{i=1}^{M} ||\hat{y}_i - y_i||_2^2 \tag{5}$$

2.5. Downstream Task Fine-tuning

To quantitatively evaluate the quality of the representations learned through self-supervised pre-training, we adopt a linear evaluation protocol, as illustrated in the downstream phase of Figure 3. The pre-trained encoder weights are frozen. Specifically, we extract fixed feature vectors for all training and test set samples using the frozen encoder. A logistic regression (LR) classifier is then trained on these fixed representations to perform the downstream seizure detection and classification task. This protocol effectively isolates and assesses the linear separability and discriminative power of the learned features, providing a clear benchmark of their utility for the target clinical application.

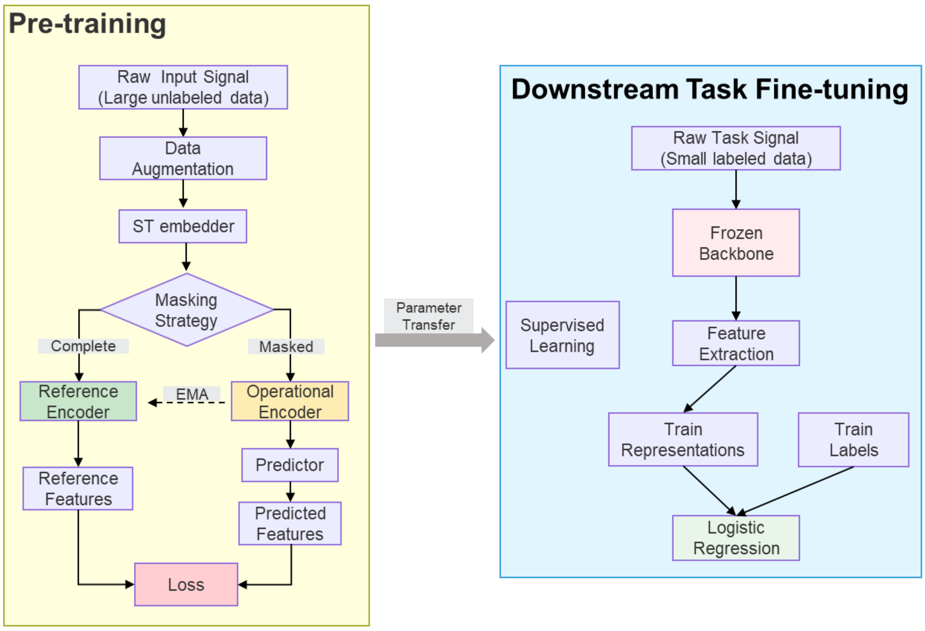

Figure 3. Illustration of the Self-Supervised Pre-training and Downstream Fine-tuning Pipeline.

3. Results

3.1. Experimental Setup

Experimental environment: NVIDIA 3090 GB GPU, Python 3.9, PyTorch 1.12.1, CUDA 11.3

Dataset segmentation: We randomly split the official TUSZ train set by patients into train and validation sets by 90/10 for model pre-training and fine-tuning, respectively, and we hold-out the official TUSZ test set for model evaluation. The pre-train, train, and test sets consist of distinct patients.

Training parameters: The model was trained for 200 epochs using the RAdam optimizer with a base learning rate of 1e-3 and batch size of 128.

3.2. Dataset and Preprocessing

The Temple University Hospital EEG Epilepsy Seizure Corpus (TUSZ) version 1.5.2 was employed in this study [30] . During self-supervised pre-training, the data were sampled at 200 Hz with a segment length of 12 seconds. The epilepsy types in the dataset were reclassified based on previous studies [17,31] . Due to the clinical distinction between simple partial (SP) and complex partial (CP) seizures—based solely on consciousness during the event, as they are not discernible from EEG signals alone [31] —we merged them with focal non-specific (FN) seizures into a combined focal (CF) seizure class. Given the extremely limited sample size (only three instances of myoclonic seizures in the TUSZ corpus), this class was excluded from analysis. The epilepsy seizure types were thus categorized into four classes: CF, generalized nonspecific (GN), absence (AB), and CT seizures.

3.3. Baseline Methods

To comprehensively evaluate the effectiveness of the proposed model, four state-of-the-art self-supervised learning methods were selected as baselines. A brief description of each method is provided below:

EEG-GNN-SSL (Corr-DCRNN)[17]: A graph neural network (GNN) based model that utilizes a diffusion convolutional recurrent neural network (DCRNN) on graphs constructed from the cross-correlation of EEG data at each time step.

REST (DS)[8]: Combines graph neural networks and recurrent structures for epilepsy detection and classification.

WGTS[32]: A recurrent graph neural network that constructs graphs using weighted graph time series.

EEG2Rep[25]: A framework for learning EEG representations using a reconstruction task.

All models were pre-trained on the same dataset and evaluated on seizure detection and classification tasks. The performance was measured using the AUROC and the Weighted F1-score.

Table 1. Comparison with state-of-the-art self-supervised methods.

Model	Seizure Detection AUROC	Seizure Classification Weighted F1-Score
EEG-GNN-SSL (Corr-DCRNN)	**0.850**	0.749
REST (DS)	0.706	0.792
WGTS	0.762	0.847
Eeg2rep	0.810	0.843
Ours	0.848	**0.900**

Table 1 summarizes the comparative performance of different self-supervised learning methods. As shown, the proposed model achieves seizure detection performance comparable to the best-performing baselines and surpasses all baselines in seizure classification.

3.4. Impact of Pre-training

To assess the impact of self-supervised pre-training on model performance, we compared networks trained with and without pre-training for epilepsy detection and classification. As presented in Table 2, the network with self-supervised pre-training consistently outperforms the non-pre-trained model across all evaluation metrics. Notably, substantial improvements are observed in overall seizure detection AUROC (+16.16%) and weighted F1-score for 4-class classification (+26.23%). Performance gains are also evident across individual seizure types, including CF, GN, AB, and CT.

Table 2. Effect of self-supervised pre-training on seizure detection and classification performance.

	Seizure Detection AUROC	4-class Weighted F1-Score	CF AUROC	GN AUROC	AB AUROC	CT AUROC
w/ pre-training	0.730	0.713	0.769	0.617	0.939	0.909
w/o pre-training	**0.848**	**0.900**	**0.904**	**0.885**	**0.983**	**0.954**
Improvement	+16.16	+26.23%	+17.56%	+43.44%	+4.69%	+4.95%

3.5. Ablation Study

To validate the effectiveness of our data augmentation module, ablation experiments were conducted comparing different augmentation strategies: no augmentation, traditional Gaussian noise injection, and the proposed time-frequency augmentation method. As shown in Table 3, our augmentation approach outperforms the traditional noise injection technique and significantly improves model performance in both seizure detection and classification tasks.

Table 3. Performance comparison of different data augmentation methods on seizure detection and classification.

	Seizure Detection AUROC	Seizure Classification Weighted F1-Score
No Augmentation	0.808	0.861
Gaussian Noise	0.799	0.862
ours	0.848	0.900

Furthermore, we investigated the impact of different masking strategies and masking ratios on model performance during pre-training. Specifically, the Semantic Subsequence Preservation (SSP) masking strategy was compared against random masking, using five masking rates: 10%, 20%, 50%, 75%, and 90%. Results in Table 4 indicate that for the SSP method, moderate masking ratios (around 50%) yield the best performance, whereas excessively high or low masking ratios degrade effectiveness. Conversely, under random masking, increasing the masking ratio generally improves performance.

Table 4. Masking strategies and ratios in pre-training.

	Seizure Detection AUROC		Seizure Classification Weighted F1-Score	
Mask Ratio	SSP	RANDOM	SSP	RANDOM
10%	0.798	0.790	0.861	0.764
20%	0.789	0.791	0.859	0.800
50%	0.848	0.792	0.900	0.844
75%	0.833	0.794	0.872	0.873
90%	0.815	0.824	0.856	0.899

4. Discussion

The self-supervised network proposed in this paper offers an effective self-supervised model for epilepsy diagnosis. The main contribution of this study is the development of a framework that integrates multiple representations to enhance task performance, generating more robust representations for various epilepsy types via mixed-frequency augmentation, and mitigating the dependency on labeled data through self-supervised pre-training. Experiments on the TUSZ dataset validate the effectiveness of the proposed method, and comparisons with other models pre-trained on the same dataset indicate superior performance of the proposed framework. However, our experiments revealed an important limitation: detection performance consistently underperforms relative to classification accuracy across all evaluated models. Future research will focus on developing enhanced representation learning techniques to bridge this performance gap while maintaining the framework's strong classification capabilities. Specifically, we plan to investigate task-specific attention mechanisms and hierarchical feature aggregation approaches to improve detection sensitivity without compromising classification accuracy

References

[1] Jiang L, Fan Q, Ren J, Dong F, Jiang T, Liu J. An improved BECT spike detection method with functional brain network features based on PLV. Front. Neurosci. 2023, 17, 1150668.
[2] Slater JD, Benbadis S, Verrier RL. The brain-heart connection: Value of concurrent ECG and EEG recordings in epilepsy management. Epilepsy Behav Rep. 2024, 28, 100726.
[3] Vidyaratne, L.S.; Iftekharuddin, K.M. Real-Time Epileptic Seizure Detection Using EEG. IEEE Trans. Neural Syst. Rehabil. Eng. 2017, 25, 2146-2156.
[4] Tripathi PM, Kumar A, Kumar M, Komaragiri RS. Automatic seizure detection and classification using super-resolution superlet transform and deep neural network - a preprocessing-less method. Comput. Methods Programs Biomed. 2023, 240, 107680.
[5] Kantipudi MVVP, Kumar NSP, Aluvalu R, et al. An improved GBSO-TAENN-based EEG signal classification model for epileptic seizure detection. Sci Rep. 2024, 14, 843.
[6] Liu S, Wang J, Li S, Cai L. Multi-dimensional hybrid bilinear CNN-LSTM models for epileptic seizure detection and prediction using EEG signals. J Neural Eng. 2024, 21, 066045.
[7] Priyasad D, Fernando T, Denman S, Sridharan S, Fookes C. Interpretable seizure classification using unprocessed EEG with multi-channel attentive feature fusion. IEEE Sens J. 2021, 21, 19186-97.
[8] Afzal, A.; Chrysos, G.; Cevher, V.; Shoaran, M. REST: Efficient and accelerated EEG seizure analysis through residual state updates. In Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024; pp. 271–290.
[9] Lih OS, Jahmunah V, Palmer EE, Barua PD, Dogan S, Tuncer T, et al. EpilepsyNet: novel automated detection of epilepsy using transformer model with EEG signals from 121 patient population. Comput Biol Med. 2023, 164, 107312.
[10] Zhu R, Pan WX, Liu JX, et al. Epileptic seizure prediction via multidimensional transformer and recurrent neural network fusion. J Transl Med. 2024, 22, 895.
[11] Hu S, et al. Exploring the applicability of transfer learning and feature engineering in epilepsy prediction using hybrid transformer model. IEEE Trans. Neural Syst. Rehabil. Eng. 2023, 31, 1321-32.
[12] Holguin-Garcia SA, Guevara-Navarro E, Daza-Chica AE, et al. A comparative study of CNN-capsule-net, CNN-transformer encoder, and traditional machine learning algorithms to classify epileptic seizure. BMC Med Inform Decis Mak. 2024, 24, 60.
[13] Lu, G.; Peng, J.; Huang, B.; Gao, C.; Stefanov, T.; Hao, Y.; et al. SlimSeiz: Efficient Channel-Adaptive Seizure Prediction Using a Mamba-Enhanced Network. In Proceedings of the 2025 IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Korea, 2025; pp. 1–5.
[14] Deng Q, Wang Q, Liu W, Xue Y. EEG VMamba: vision Mamba for seizure prediction based on EEG. In Proceedings of the 2024 6th International Conference on Video, Signal and Image Processing; 2025; New York, NY, USA. p. 119-24.
[15] Li, Z. Research on EEG Acquisition System of Smart Internet of Things with Enhanced Mamba. In Proceedings of the 5th International Conference on Signal Processing and Machine Learning (CONF-SPML 2025), Hybrid Conference, Portsmouth, UK, 2025; pp. 190–195.
[16] Wang, J.; Zhao, S.; Luo, Z.; Zhou, Y.; Li, S.; Pan, G. EEGMamba: An EEG Foundation Model with Mamba. Neural Netw. 2025, 192, 107816. https://doi.org/10.1016/j.neunet.2025.107816.
[17] Tang, S.; Dunnmon, J.; Saab, K.K.; Zhang, X.; Huang, Q.; Dubost, F.; Rubin, D.; Lee-Messer, C. Self-supervised graph neural networks for improved electroencephalographic seizure analysis. In Proceedings of the 10th International Conference on Learning Representations (ICLR), Virtual, 25–29 April 2022.
[18] Van Lam; Oliugbo, C.; Parida, A.; Linguraru, M.G.; Anwar, S.M. Self-Supervised Learning for Seizure Classification Using ECoG Spectrograms. In Proceedings of the SPIE Medical Imaging 2024: Computer-Aided Diagnosis, San Diego, CA, USA, 2024; 12927, 129272J.
[19] Han H, Fan H, Huang X, Han C. Self-supervised multi-transformation learning for time series anomaly detection. Expert Syst Appl. 2024, 253, 124339.
[20] Kostas D, Aroca-Ouellette S, Rudzicz F. BENDR: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. Front Hum Neurosci. 2021, 15, 653659.
[21] Chien, H.-Y.S.; Goh, H.; Sandino, C.M.; Cheng, J.Y. MAEEG: Masked auto-encoder for EEG representation learning. In Proceedings of the NeurIPS 2022 Workshop on Learning from Time Series for Health, New Orleans, LA, USA, 2 December 2022.
[22] Chen, X.; Ding, M.; Wang, X.; Yan, Z.; Wong, K.; Xu, Q.; Han, Y.; Zhang, P.; Li, H.; Han, X.; et al. Context Autoencoder for Self-supervised Representation Learning. Int. J. Comput. Vis. 2024, 132, 208–223.
[23] Misra, I.; van der Maaten, L. Self-supervised learning of pretext-invariant representations.

In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 6706–6716.

[24] Goyal, P.; Mahajan, D.K.; Gupta, A.K.; Misra, I. Scaling and benchmarking self-supervised visual representation learning. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 6390–6399.

[25] Mohammadi Foumani, N.; Mackellar, G.; Ghane, S.; Irtza, S.; Nguyen, N.; Salehi, M. EEG2Rep: Enhancing self-supervised EEG representation through informative masked inputs. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), Barcelona, Spain, 25–29 August 2024; pp. 5544–5555.

[26] Liu, J.; Chen, S. TimesURL: Self-Supervised Contrastive Learning for Universal Time Series Representation Learning. Proc. AAAI Conf. Artif. Intell. 2024, 38, 13918-13926.

[27] Chambon S, Galtier MN, Arnal PJ, Wainrib G, Gramfort A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. IEEE Trans Neural Syst Rehabil Eng. 2018, 26, 758-69.

[28] Baevski, A.; Hsu, W.-N.; Xu, Q.; Babu, A.; Gu, J.; Auli, M. data2vec: A general framework for self-supervised learning in speech, vision and language. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, 17–23 July 2022; pp. 1298–1312.

[29] Zhou, J.; Wei, C.; Wang, H.; Shen, W.; Xie, C.; Yuille, A.; Kong, T. Image BERT pre-training with online tokenizer. In Proceedings of the 10th International Conference on Learning Representations, Virtual, 25–29 April 2022.

[30] Shah V, von Weltin E, Lopez S, McHugh JR, Veloso L, Golmohammadi M, et al. The Temple University Hospital seizure detection corpus [dataset]. Front Neuroinform. 2018, 12, 83.

[31] Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, et al. Operational classification of seizure types by the International League Against Epilepsy: position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017, 58, 522-30.

[32] Cappelletti, W.; Xie, Y.; Frossard, P. Learning Self-Supervised Dynamic Networks for Seizure Analysis. In Proceedings of the ICLR 2024 Workshop on Learning from Time Series for Health, Vienna, Austria, 2024.