Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 11
TPCEE 2025

Deep Learning Architectures for EEG: from CNN to Transformers

Xuanzheng He

Department of Computer Science, New York University, New York, USA

xh2596@nyu.edu

Abstract. This review provides a comprehensive overview of the advancements in deep learning for electroencephalography (EEG) analysis. EEG offers a non-invasive means to monitor brain activity, with applications in brain-computer interfaces (BCIs), clinical diagnostics, emotion recognition, and sleep monitoring. We trace the historical progression from early convolutional neural networks (CNNs) in the mid-2010s to the recent emergence of transformer-based models. Key CNN architectures, such as EEGNet, are examined for their efficiency in feature extraction from raw EEG data. Transformer and hybrid models are discussed, highlighting their ability to model long-range dependencies and achieve state-of-the-art performance across diverse tasks like motor imagery, seizure detection, and sleep staging. The review also addresses self-supervised and unsupervised learning methods to mitigate data scarcity. Challenges including inter-subject variability, noise artifacts, and model interpretability are analyzed, alongside future directions such as transfer learning, synthetic data generation, domain adaptation, and multi-modal fusion. By synthesizing these developments, this work aims to guide future research toward more generalizable and practical EEG deep learning systems.

Keywords: Electroencephalography (EEG); convolutional neural networks (CNNs); transformer; brain-computer interfaces (BCIs).

1. Introduction

Electroencephalography (EEG) provides a non-invasive window into brain activity by recording electrical potentials from the scalp. Its high temporal resolution and portability have enabled applications ranging from brain–computer interfaces (BCIs) to clinical diagnostics, emotion analysis, and sleep monitoring. Classical EEG analysis relied on hand-crafted features and conventional classifiers, but in the past decade deep learning approaches have become dominant. [1] Notably, deep neural networks remove much of the need for manual feature engineering, directly learning representations from raw or minimally processed EEG data. The literature now includes CNN-based models, recurrent and temporal architectures, and more recently transformer-based networks, each applied to tasks such as motor imagery BCIs, seizure detection, emotion recognition, sleep staging, and other cognitive and clinical applications. [2] For example, EEGNet was proposed as a compact CNN that generalized across multiple BCI paradigms even with limited training data. [3] Transformer-based models like EEGformer and Patched Brain Transformer have demonstrated high performance on SSVEP, emotion, and depression EEG tasks. [13] This review traces the development of deep learning for EEG, examining CNN architectures in detail, then surveying transformer and hybrid models. We highlight recent self-supervised and unsupervised representation methods, and discuss challenges such as data scarcity, inter-subject variability, and noise. We conclude with future research directions to guide continued progress.

2. Historical Perspective of Deep Learning in EEG

Deep learning for EEG gained momentum in the mid-2010s as researchers adapted convolutional and recurrent neural networks to brain signals. [1] Early influential work systematically evaluated CNN designs for raw EEG decoding and visualization. [4] Their 2017 study considered shallow (2-layer), deep (5-layer), and ResNet (31-layer) CNNs, showing that end-to-end ConvNets can approach or surpass classical methods in tasks like motor imagery BCI. [4] Lawhern et al. introduced EEGNet, a

compact CNN exploiting depthwise and separable convolutions tailored to EEG. [3] EEGNet generalized across four diverse BCI paradigms (P300, error potentials, MRCP, SMR), achieving high accuracy even with limited data. [3] By 2019 reviews reported that over half of EEG deep-learning studies used CNNs, with the remainder split between RNNs and other architectures. [2] The volume of studies grew rapidly: between 2010–2018, one review identified 154 papers applying DL to EEG across applications (BCI, epilepsy, sleep, emotion, etc.). [2] They noted a clear trend away from subject-specific models toward inter-subject and cross-paradigm approaches. [2] More recently, around 2020–2023, transformer architectures (successful in NLP and vision) were introduced for EEG. [12] The adaptation was nontrivial due to EEG's multi-channel time series nature. [12] Early transformer-based EEG models built on prior advances in speech (e.g. wav2vec2.0) and vision (ViT). [12] For instance, Brain Encoder-Decoder Representations (BENDR) extended wav2vec2.0 to multichannel EEG by learning latent embeddings via CNN and then feeding masked sequences to a Transformer encoder. [12] The first pure transformer decoders for EEG emerged shortly after. [12] Wan et al. proposed EEGformer, using 1D CNNs followed by three specialized Transformer sub-blocks (temporal, synchronous, regional) to classify SSVEP, emotion, and depression EEG. [13] Their results showed state-of-art performance across these domains, demonstrating the feasibility of end-to-end Transformer models on raw EEG. [13] Klein et al. introduced the Patched Brain Transformer (PBT), a ViT adaptation that tokenizes EEG channels into patches and achieved superior MI-BCI classification after supervised pretraining. [14] Thus, in under a decade EEG DL has progressed from shallow CNNs on single tasks to sophisticated hybrid architectures and large-scale transformer models pre-trained on diverse data. [1]

3. Applications of Deep Learning in EEG

Deep learning has been applied to essentially every major EEG application domain. [1] In BCIs for motor imagery and SSVEP, CNNs and LSTM networks have achieved human-level performance.[22] For example, EEGNet and related ConvNets have been used to classify motor imagery (MI) tasks across subjects. [3] Hybrid models combining CNNs, LSTMs, and Transformers have further improved MI-EEG decoding. [12] Zhao et al. proposed CTNet, which concatenates an EEGNet-like convolutional extractor with a Transformer encoder, achieving over 88% accuracy on subject-specific MI datasets. [6] Similarly, Shi et al. introduced EEG-VTTCNet, a vision-transformer + temporal convolution network, reaching 84.6% and 90.9% accuracy on the BCI Competition IV-2a/2b MI datasets. [7] These models suggest that adding self-attention can capture global dependencies that CNNs alone miss. [12] In clinical diagnosis, EEG DL systems target epilepsy, neurological disorders, and other conditions. [2] EEG is the gold standard for epilepsy detection, and learning-based methods aim to improve seizure forecasting and artifact rejection. [19] A recent survey noted that CNNs and RNNs now "revolutionize seizure prediction by directly learning from raw EEG". [16] Hybrid architectures like CNN-LSTM capture both spectral and temporal patterns, advancing ictal detection accuracy. [8] Deep learning also aids neurological disorder recognition: for instance, Wan et al. mention applications such as early glaucoma diagnosis and depression detection using EEGTransformer models. [13] Affective computing is another major EEG domain. [2] EEG-based emotion recognition has seen many CNN approaches, often using spectrogram inputs.[1] Cheng et al. combined a multi-scale dynamic CNN with a gated Transformer encoder (MSDCGTNet) to classify emotions from raw EEG. [5] By extracting spatial-spectral features via CNN and then capturing global context with self-attention, they achieved high accuracy on DEAP, SEED, and SEED-IV datasets. [5] This illustrates how combining convolutional feature extraction with transformers can effectively learn complex, task-specific EEG patterns for emotion. [12] Sleep staging is a well-studied EEG task for sleep medicine. [2] Transformer-based models have recently shown promise here as well. [12] Mostafaei et al. introduced a Transformer encoder-decoder with cross-modality attention that ingests multiple EEG channels and other physiological signals to classify five sleep stages. [11] Their model, trained on the large SHHS dataset, reached 91.3% accuracy and outperformed prior approaches by effectively fusing multimodal information through attention mechanisms. [11] Another recent Transformer model by Wan et al. focused on sleep-related motor imagery but also

suggests the adaptability of ViT-like models to variable channel configurations. [13] Overall, transformer architectures are being rapidly explored for all EEG tasks, often surpassing pure CNN baselines. [12]

4. CNN-Based Models for EEG

Convolutional neural networks (CNNs) have been the workhorse of EEG deep learning. [1] They naturally exploit the spatial (electrode) and temporal structure of EEG. [4] Early CNN models often took time-series or time-frequency EEG segments as 2D inputs (channels×time) and applied 2D convolutions across them. [4] Others employed 1D temporal convolutions followed by spatial filters. [4] Many architectures also integrate neuroscientific priors: for example, Lawhern et al. embedded depthwise and separable convolutions in EEGNet to implement bandpass filtering and spatial filtering akin to common EEG features. [3] The result was a compact network that required few parameters yet generalized across visual and motor BCI paradigms. [3] Deep EEG CNNs vary in depth and design. [1] Schirrmeister et al. compared a shallow 2-layer CNN, a 5-layer CNN, and a 31-layer residual network on several motor-task datasets. [4] They showed that adding layers and regularization can improve accuracy but that even shallow networks are competitive on some EEG tasks. [4] More recent CNN designs have borrowed from computer vision: for instance, the Inception-style and ResNet-style blocks appear in models like "IENet" and WaveNet-inspired CNNs. [20] Some work focuses on efficiency: for example, Waytowich et al. developed compact CNNs for asynchronous SSVEP detection, and others have proposed algorithms to prune or quantize EEG CNNs for deployment on portable devices. [22] Typical CNN pipelines begin with minimal preprocessing (often bandpass filtering), then feed either raw multichannel EEG or its spectral representations into the network. [4] For example, a time–frequency image (spectrogram or wavelet coefficients) can be constructed per channel and stacked, turning EEG into an "image" over channels and frequency, suitable for 2D ConvNets. [4] Alternatively, one-dimensional convolutions can slide across each channel's waveform to capture temporal features, followed by 2D convolutions across channels to capture spatial relationships. [4] A key advantage of CNNs is their ability to learn end-to-end from raw EEG, as demonstrated in many studies. [2] For example, Schirrmeister et al. visualized the spatial filters learned by their networks, showing correspondence with known EEG rhythms. [4] CNNs can implicitly discover features like event-related potentials or oscillatory band power that might be missed by fixed filters. [4] In practice, a plethora of CNN variants have been proposed. [2] Deep ConvNets (multi-layer, high-dimensional) and Shallow ConvNets (fewer layers) were both explored by Schirrmeister et al. [4] EEGNet (Lawhern et al.) used very few layers with separable convolutions to reduce parameters. [3] Other models like IENet and FBCNet integrated filter-bank ideas explicitly. [1] More recently, convolutional architectures have been combined with attention or other mechanisms: for example, Cheng's MSDCGTNet employed a 1D multi-scale CNN, and then applied a gated multi-head self-attention layer on its output. [5] This shows that even within the "CNN" category, modern models often hybridize multiple ideas to better capture EEG's complexity. [12] When evaluating CNN models, studies have emphasized both within-subject and cross-subject performance. [2] EEGNet, for instance, achieved robust within-subject accuracy on P300, ERN, MRCP, and SMR tasks, and also showed generalization across paradigms. [3] Its low-parameter design helps prevent overfitting on small EEG datasets. [3] Similarly, data augmentation (e.g. cropping trials, adding noise) is often used to enlarge datasets for CNN training, recognizing that CNNs generally require more data than classical methods. [4] In terms of citations, CNN-based EEG models dominate the literature reflecting the maturity of this approach. [2] Representative high-impact examples include Lawhern et al. (EEGNet), Schirrmeister et al. (DeepConvNet), and many recent CNN+LSTM hybrids (e.g. EEG-VTTCNet, CTNet) which still rely on convolutional front-ends. [3]

5. Transformer-Based and Hybrid Models for EEG

Transformers have gained prominence in EEG analysis only recently, but studies show they can capture long-range dependencies and flexible feature interactions that CNNs may miss. [12] Transformer architectures rely on self-attention to weigh relationships between all positions in the input sequence, making them powerful for complex temporal-spatial patterns. [21] In EEG, various transformer flavors have been explored. [12]

5.1. Pure Transformers (Time-Series Transformers)

Some works apply 1D Transformers directly to raw or windowed EEG time-series. [12] For example, Klein et al. (2025) proposed the Patched Brain Transformer, essentially a Vision Transformer (ViT) adapted to EEG: it tokenizes each channel's time segment into patches, embeds them, and feeds them to a standard Transformer encoder. [14] PBT achieved state-of-art imagined-movement decoding after supervised pre-training and data augmentation. [14] These models show that Transformers can be effective even without convolutional front-ends, provided enough data and proper tokenization. [14]

5.2. CNN-Transformer Hybrids

To leverage CNNs' local feature extraction and Transformers' global modeling, many hybrid architectures have been proposed. [12] The EEGformer (Wan et al.) is one example: it begins with a 1D CNN to extract channel-wise features, then applies three sequential Transformer sub-layers (capturing temporal, synchronous, and regional EEG features). [13] Another example is Zhao et al.'s CTNet: it uses an EEGNet-like CNN for feature extraction followed by a Transformer encoder, yielding >88% MI classification accuracy. [6] Shi et al.'s EEG-VTTCNet blended a shared CNN feature extractor with a vision transformer and a temporal convolutional network, achieving top performance on BCI datasets. [18] These hybrids typically train end-to-end, letting the CNN layers and Transformer layers jointly optimize for the task. [12] The combination often outperforms either alone, suggesting that the Transformer can capture global patterns (such as inter-channel synchrony) that complement the CNN's local filters. [12]

5.3. Vision-Transformer (ViT) Approaches

Some researchers have directly employed ViT models, often with domain adaptation. [12] Yang et al. (2023) fine-tuned an ImageNet-pretrained ViT on EEG regression tasks, finding significant gains over unpretrained models. [15] This indicates that even vision pretraining encodes features (edges, textures) useful for EEG spectrogram-like data.[15] Other studies design novel EEG-specific tokenizers for ViT. [12] For instance, the Flexible Patched Brain Transformer tokenizes each channel's waveform into temporal patches, akin to image patches in ViT, enabling cross-channel attention. [14] Relatedly, Deng et al. developed a 3D CNN+Swin-Transformer for MI (Sensors 2025) by converting EEG into 3D "images", showing that vision transformer concepts can extend to EEG. [12]

5.4. Temporal Convolution + Transformer (TCN-Transformer)

A few works combine Temporal Convolutional Networks (TCNs) with Transformers. [12] The EEG-VTTCNet introduced by Shi et al. includes a TCN branch that extracts temporal features in parallel with the ViT encoder. [7] Others, like ConTraNet (Ali et al., 2024), use a CNN to learn spatial patterns and a Transformer (or LSTM) to model temporal evolution. [10] These models highlight the modularity of deep architectures: one can mix convolution, attention, and even recurrence. [12]

5.5. Overall Analysis

Overall, recent transformer-based studies have spanned BCI tasks, emotion, seizure, and sleep. [12] For example, Ding et al. used a dense CNN-Transformer ("EEG-Deformer") for BCI (IEEE JBHI

2024) and reported strong MI accuracy. [5] Cheng et al. used a gated multi-head Transformer for emotion recognition. [5] Vu et al. applied a Graph Attention Transformer to sleep EEG (FlexSleepTransformer, Nat. Sci Reports 2025) to handle variable channels. [12] Notably, these works often emphasize scalability and pretraining: Transformers for EEG are usually larger and require more data. [12] The transformer review by Vafaei and Hosseini points out that "only the size of the CNN tokenizer for these transformers exceeds the size of previously used CNN-based models". [12] Nonetheless, the performance gains can be substantial. [12] For instance, EEGformer achieved near-perfect accuracy on a depressive EEG dataset, outperforming multiple CNN baselines. [13] Transformer models also excel in interpretability efforts. [12] Self-attention weights can be visualized to show which timepoints or channels the model attends to. [21] Cross-modal transformers (like Mostafaei's sleep model) further allow insight into how different physiological signals contribute to decisions. [11] In sum, transformer-based EEG architectures represent a new frontier: they push state-of-art in many domains and facilitate new capabilities like multimodal fusion and unsupervised pretraining. [12]

6. Self-Supervised and Unsupervised Learning

The limited availability of labeled EEG data has spurred interest in self-supervised (SSL) and unsupervised methods. [1] EEG datasets are often small (typically 10–100 subjects with a few hundred trials each), so SSL can leverage unlabeled recordings to pretrain models. [12] Weng et al. survey categorizes EEG SSL methods into predictive (e.g. predicting signal transformations), generative (autoencoding/masked reconstruction), contrastive (distance-based), and graph-based approaches. [12] For instance, a predictive task might mask random EEG segments and train a model to reconstruct them, forcing it to learn salient features. [12] Contrastive SSL has also been applied: by treating EEG segments from the same subject as positive pairs and different subjects as negatives, a network learns subject-invariant embeddings (useful for emotion or MI classification). [12] Weng et al. note that SSL methods have already been applied to downstream tasks like emotion recognition and motor imagery. [12] Practical SSL approaches for EEG include models inspired by wav2vec, BERT, and CPC. [12] For example, Kostas et al. used a CPC-like objective in a CNN-Transformer model (BENDR) to learn EEG embeddings. [12] They pretrained on large unlabeled EEG corpora and then fine-tuned for classification, analogously to NLP and speech. [12] This "foundation model" paradigm is promising: Klein et al. found that supervised pretraining on large combined datasets significantly outperformed self-supervised pretraining for their ViT model, but the trend suggests more effort into efficient SSL. [14] In addition to SSL, truly unsupervised techniques have been explored. [17] Autoencoders and variational autoencoders (VAEs) can learn latent representations of EEG without labels, which can be later used for tasks or clustering. [17] For example, Suryawati et al. applied autoencoders to epilepsy EEG: they used bottleneck features from an autoencoder as input to a classifier, reducing dimensionality and improving seizure detection. [17] Generative adversarial networks (GANs) have been used to synthesize realistic EEG segments, both to augment training data and to learn discriminative features (the discriminator's features). [17] One study notes that "autoencoder and generative adversarial networks (GAN) are examples of unsupervised deep learning methods" in EEG. [17] While GAN-based EEG generation is still nascent, early work shows it can help mitigate data scarcity by creating pseudo-samples. [17] Overall, self-supervised and unsupervised learning in EEG is an active frontier. [12] They address a core challenge: that expert labels are scarce and costly. [2] SSL approaches have so far demonstrated improved representations for standard EEG tasks. [12] As these methods mature, we expect EEG deep models to increasingly leverage unlabelled datasets, possibly through large-scale consortium data or continuous recordings. [12]

7. Challenges and Future

Despite progress, EEG deep learning faces several persistent challenges. [1] Data scarcity is paramount: as noted, typical EEG datasets are small and non-standardized. [2] Unlike image or text

corpora with millions of examples, EEG datasets often comprise only tens of subjects. [12] This scarcity makes large models prone to overfitting. [12] Even when data exist, they are heterogeneous: different recording systems, electrode montages, and preprocessing steps. [12] This leads to inter-subject and inter-session variability: EEG signals vary greatly across people and even within the same person over time. [12] As Vafaei and Hosseini note, "variability between participants and even between recording sessions" is a fundamental difficulty for EEG decoding. [12] Consequently, a model trained on one cohort often fails to generalize to new subjects. [12] Wan et al. highlight that collecting large volumes of subject-specific data is expensive and impractical, limiting real-world scalability. [13] Signal quality and noise are also significant issues. [2] EEG has a low signal-to-noise ratio: artifacts (eye blinks, muscle activity) and environmental noise can dominate the true neural signal. [12] Label noise can arise from ambiguous or inconsistent experimental conditions. [12] Klein et al. point out that "hardware discrepancies, electrode placements, and environmental factors introduce noise" which degrades robustness. [14] Deep models can unintentionally learn to rely on noise patterns, which hurts generalization. [12] Mitigating this requires careful preprocessing and possibly noise-robust model training, but these solutions are not yet standardized across studies. [12] Another challenge is reproducibility and benchmarking. [2] Roy et al. lament that many EEG deep learning papers lack open data and code, making replication difficult. [2] As a result, reported improvements are hard to validate. [2] There is a shortage of widely-accepted benchmarks: EEG datasets vary in size, protocols, and tasks, so it is challenging to compare models fairly. [12] Vafaei and Hosseini (2025) reinforce that "public availability of datasets influences the frequency of certain studies". [12] In practice, a few public datasets (e.g. BCI Competitions, DEAP, TUH) dominate the field, and models tuned to these may not generalize broadly. [12] Establishing standardized benchmarks or shared pretraining corpora would help overcome this. [12] Finally, deep learning models in EEG often act as "black boxes," raising interpretability concerns. [2] Clinicians especially demand models whose decisions can be understood. [12] Although some works apply attention visualizations or ablation (e.g. EEGNet filters or t-SNE of embeddings), a generally accepted interpretability framework is lacking. [12] The dimensional complexity of EEG (time, space, frequency) makes it non-trivial to attribute a model's decision to neurophysiologically meaningful features. [12] This remains an open challenge, particularly for clinical adoption. [12] Looking forward, several avenues promise to advance EEG deep learning. [12] Transfer learning and pretraining are likely central. [12] As Vafaei and Hosseini discuss, leveraging pretrained models – either pretrained on large EEG corpora or even on related domains – can ameliorate limited labeled data. [12] For example, Yang et al. (2023) showed that starting from an ImageNet-pretrained ViT significantly improved EEG regression accuracy over a randomly-initialized model. [15] Similarly, several studies have applied transfer learning strategies (e.g. dual-transfer or domain adaptation) to EEG with success. [8] Future work should explore large-scale self-supervised pretraining on unlabeled EEG or multimodal brain data, akin to foundation models in NLP and vision. [12] Creating publicly-shared pretrained EEG models (Transformers or CNNs) could become an infrastructure goal, analogous to BERT or GPT. [12] Data augmentation and synthetic data generation will also be important. [17] The transformer review highlights augmentation (adding noise, scaling, mixing trials, etc.) as a key strategy to increase training diversity. [12] More sophisticated approaches (e.g. using GANs to generate realistic EEG) are nascent but promising. [17] Synthetic EEG generation could populate the sparse corners of state space (rare events, underrepresented tasks). [12] Research on physiologically-informed augmentation (e.g. simulating electrode shifts) may help generalization to new subjects. [12] Domain adaptation and personalization methods are another direction. [12] Rather than training a one-size-fits-all model, recent work has looked at adapting global models to individuals. [12] For instance, contrastive self-supervised learning has been used to align representations across subjects, improving cross-subject emotion recognition. [8] Techniques like adversarial domain adaptation or meta-learning could further mitigate inter-subject shifts. [12] Federated learning – training models across multiple centers without sharing raw data – may address privacy and personalization simultaneously. [12] Multi-modal and hybrid architectures will continue to expand. [12] EEG often comes with other signals (e.g. eye-tracking, ECG, fNIRS). [12] Models that can jointly process

multiple modalities (as in Mostafaei's sleep model) could yield richer insights and better performance. [11] Within EEG itself, graph neural networks (GNNs) or spatio-temporal attention models could further capture the complex sensor relationships. [9] Additionally, integrating physiological knowledge (e.g. head models, source localization) into deep architectures remains largely unexplored and could improve both accuracy and interpretability. [12] Finally, the community must address standardization and collaboration. [2] Creating large, open EEG datasets (with diverse tasks and consistent preprocessing) will boost reproducibility. [12] Shared challenges (similar to ImageNet) could catalyze progress. [20] Publishing code and models is essential. [2] As noted by Roy et al., reproducibility was low in 2019, and this must improve. [2] Conferences and journals should encourage (or require) data/code sharing for EEG DL studies. [12]

8. Conclusion

In conclusion, deep learning for EEG has already transformed many applications from BCIs to clinical diagnostics. The shift from CNNs to Transformers reflects the field's rapid evolution. Recent works demonstrate that hybrid architectures and self-/unsupervised methods can further push performance. Nonetheless, the challenges of data limitations, variability, and interpretability demand creative solutions. Future research combining large-scale pretraining, domain adaptation, and multi-modal integration holds promise to overcome these hurdles. By building on the wealth of existing work and addressing its shortcomings, the field of EEG deep learning can continue advancing toward robust, generalizable, and clinically useful systems.

References

[1] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," J. Neural Eng., vol. 16, no. 3, p. 031001, 2019.

[2] Y. Roy et al., "Deep learning-based electroencephalography analysis: a systematic review," J. Neural Eng., vol. 16, no. 3, p. 031001, 2019.

[3] V. J. Lawhern et al., "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," J. Neural Eng., vol. 15, no. 5, 2018, Art. no. 056013.

[4] RT. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," J. Neural Eng., vol. 14, no. 2, 2017, Art. no. 026003.

[5] M. X. Ding et al., "EEG-based emotion recognition using multi-scale dynamic CNN and gated transformer," Sci. Rep., vol. 14, Art. 31319, 2024.

[6] W. Zhao et al., "CTNet: a convolutional transformer network for EEG-based motor imagery classification," Sci. Rep., vol. 14, Art. 20237, 2024.

[7] X. Shi et al., "EEG-VTTCNet: a loss joint training model based on the vision transformer and the temporal convolution network for EEG-based motor imagery classification," Neuroscience, vol. 556, pp. 42–51, 2024.

[8] W. Lu et al., "CNN+LSTM hybrid deep transfer learning for cross-subject EEG emotion recognition," Front. Hum. Neurosci., vol. 17, 2023, Art. 13280241.

[9] G. Greiner and Y. Zhang, "Multi-modal EEG NEO-FFI with trained attention layer (MENTAL) for mental disorder prediction," Brain Inform., vol. 11, Art. 26, 2024.

[10] O. Ali et al., "ConTraNet: a hybrid network for improving the classification of EEG and EMG signals with limited training data," Comput. Biol. Med., vol. 168, 2024, Art. 107649.

[11] SH. Mostafaei et al., "A novel deep learning model based on transformer encoder-decoder and cross-modal attention for classification of sleep stages," J. Biomed. Inform., vol. 157, 2024, Art. 104689.

[12] E. Vafaei and M. Hosseini, "Transformers in EEG analysis: a review of architectures and applications in motor imagery, seizure, and emotion classification," Sensors, vol. 25, no. 5, 2025, Art. 1293.

[13] Z. Wan et al., "EEGformer: a transformer–based brain activity classification method using EEG signals," Front. Neurosci., vol. 17, 2023, Art. 1148855.

[14] T. Klein et al., "Flexible Patched Brain Transformer model for EEG decoding," Sci. Rep., vol. 15, Art. 10935, 2025.

[15] R. Yang et al., "ViT2EEG: Leveraging hybrid pretrained vision transformers for EEG data," Proc. ACM KDD Workshop MLBI, 2023.

[16] Y. Saadoon, M. Khalil, and D. Battikh, "Machine and deep learning-based seizure prediction: a scoping review," Appl. Sci., vol. 15, no. 11, Art. 6279, 2025.

[17] E. Suryawati et al., “Unsupervised feature learning-based encoder and adversarial networks,” J. Big Data, vol. 8, no. 1, Art. 148, 2021.

[18] X. Shi et al., “EEG-VTTCNet: Vision Transformer and temporal convolution network for MI-EEG classification,” Neuroscience, vol. 556, pp. 42–51, 2024.

[19] A. Shoeibi et al., “Epileptic seizures detection using deep learning techniques: a review,” Int. J. Environ. Res. Public Health, vol. 18, Art. 5780, 2021.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Adv. Neural Inf. Process. Syst., vol. 25, pp. 1097–1105, 2012.

[21] A. Vaswani et al., “Attention is all you need,” Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 5998–6008.

[22] Z. Tayeb et al., “Validating deep neural networks for online decoding of motor imagery movements from EEG signals,” Sensors, vol. 19, no. 1, Art. 210, 2019.