*applied sciences*

MDPI

*Review*

# Trends and Limitations in Transformer-Based BCI Research

**Maximilian Achim Pfeffer [1], Johnny Kwok Wai Wong [2] and Sai Ho Ling [1,*]**

[1] Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia; maximilianachim.pfeffer@student.uts.edu.au
[2] Faculty of Design, Architecture and Building, University of Technology Sydney, Ultimo, NSW 2007, Australia; johnny.wong@uts.edu.au
* Correspondence: steve.ling@uts.edu.au

## Abstract

Transformer-based models have accelerated EEG motor imagery (MI) decoding by using self-attention to capture long-range temporal structures while complementing spatial inductive biases. This systematic survey of Scopus-indexed works from 2020 to 2025 indicates that reported advances are concentrated in offline, protocol-heterogeneous settings; inconsistent preprocessing, non-standard data splits, and sparse efficiency frequently reporting cloud claims of generalization and real-time suitability. Under session- and subject-aware evaluation on the BCIC IV 2a/2b dataset, typical performance clusters are in the high-80% range for binary MI and the mid-70% range for multi-class tasks with gains of roughly 5–10 percentage points achieved by strong hybrids (CNN/TCN–Transformer; hierarchical attention) rather than by extreme figures often driven by leakage-prone protocols. In parallel, transformer-driven denoising—particularly diffusion–transformer hybrids—yields strong signal-level metrics but remains weakly linked to task benefit; denoise → decode validation is rarely standardized despite being the most relevant proxy when artifact-free ground truth is unavailable. Three priorities emerge for translation: protocol discipline (fixed train/test partitions, transparent preprocessing, mandatory reporting of parameters, FLOPs, per-trial latency, and acquisition-to-feedback delay); task relevance (shared denoise → decode benchmarks for MI and related paradigms); and adaptivity at scale (self-supervised pretraining on heterogeneous EEG corpora and resource-aware co-optimization of preprocessing and hybrid transformer topologies). Evidence from subject-adjusting evolutionary pipelines that jointly tune preprocessing, attention depth, and CNN–Transformer fusion demonstrates reproducible inter-subject gains over established baselines under controlled protocols. Implementing these practices positions transformer-driven BCIs to move beyond inflated offline estimates toward reliable, real-time neurointerfaces with concrete clinical and assistive relevance.

**Keywords:** brain-computer interfaces; artificial intelligence; EEG; transformers; signal processing; self-attention; diffusion; deep learning; neural decoding; noise removal



Academic Editor: Alexander N. Pisarchik

Received: 12 September 2025
Revised: 9 October 2025
Accepted: 16 October 2025
Published: 17 October 2025

**Citation:** Pfeffer, M.A.; Wong, J.K.W.; Ling, S.H. Trends and Limitations in Transformer-Based BCI Research. *Appl. Sci.* **2025**, *15*, 11150. https://doi.org/10.3390/app152011150



## 1. Introduction

Brain–computer interfaces (BCIs) based on electroencephalography (EEG) have seen significant progress in recent years, particularly in motor imagery (MI) classification, which enables users to control external systems through imagined movements. The success of these systems relies on accurate neural decoding, yet EEG signals remain inherently challenging due to their low signal-to-noise ratio, non-stationarity, and considerable inter-subject variability. Traditional machine learning approaches, including feature-based

classifiers and shallow neural networks, have demonstrated limited performance in dynamic or real-world settings, motivating a shift toward deep learning architectures capable of more robust feature extraction and generalization.

Deep learning, particularly convolutional neural networks (CNNs) and recurrent models such as long short-term memory (LSTM) networks, has improved EEG-based MI classification by better capturing spatial and temporal structure. However, these models often struggle with long-range dependencies in EEG sequences and require extensive dataset-specific tuning. Transformer-based architectures, originally developed for natural language processing, have emerged as a powerful alternative through their ability to model global temporal dependencies via self-attention mechanisms. When appropriately adapted to EEG, these architectures dynamically emphasize task-relevant temporal segments and suppress background noise, outperforming many CNN–LSTM baselines in controlled MI benchmarks.

Despite their rapid rise, transformer-based EEG models remain methodologically fragmented and largely confined to offline validation. Most studies continue to rely on static datasets and subject-specific tuning with limited exploration of real-time decoding, cross-session adaptability, or latency constraints. Moreover, the computational overhead of multi-head attention, which scales quadratically with sequence length, raises practical challenges for deployment in wearable or embedded systems. Cross-subject generalization likewise remains unresolved, as current transformer-based models are seldom benchmarked under subject-agnostic conditions. These factors underscore the gap between research-grade architectures and deployable neural decoding systems.

To provide a structured assessment of progress in this rapidly evolving domain, this review systematically analyzes transformer-based MI–BCI research, integrating both quantitative and qualitative perspectives. A comprehensive literature review was conducted using Scopus indexing, filtering works based on EEG-, BCI-, and transformer-related terms to ensure coverage of all major methodological directions since 2020. The analysis highlights patterns in dataset usage, architectural evolution, and reported performance while identifying key methodological limitations, including the absence of standardized evaluation protocols, inconsistent efficiency reporting, and a lack of real-time validation frameworks. Additionally, the review incorporates transformer-driven EEG denoising research—an area of accelerating growth—to demonstrate how attention mechanisms and diffusion-based hybrids are converging toward unified signal enhancement and decoding frameworks. This parallel development illustrates that transformer paradigms in EEG are not merely improving classification but redefining the preprocessing and denoising foundations that determine BCI reliability itself.

The present review is particularly timely, as transformer architectures have reached a stage of maturity in other domains such as vision and speech, yet EEG-based research continues to develop without clear methodological consensus or standardized validation practices. A coherent synthesis of these efforts is now essential to consolidate fragmented approaches, benchmark progress, and clarify the direction of future work. By integrating transformer-based classification and denoising into a single analytical framework, this review provides both an evidence-based overview of current achievements and a forward-looking perspective on how these methods can be translated into reproducible, real-time, and adaptive BCI systems.

## 2. Methodology

For this study, a systematic review of published work was conducted using scientific database indexing using Scopus [1], which is a major database of research literature and web sources that includes abstracts, citations, and full text documents, and more.

To assess and distill trends from the literature, a focus on recent years was set. To be exact, no literature before 2020 was deemed eligible for the scope of this review, and all works until published before 21st September 2025 were accessed. The graphics displayed in this work may neglect 2025 data to visualize trends more clearly without steep drop-off, yet all works including those of September 2025 of each query scope were numerically and sematically included in this review. As for Scopus indexing, advanced query strings were used, all of which utilized the search fields Title, Abstract, and Keywords of publications. All works that are presented in the following sections satisfied the conditions ((“EEG” AND “BCI”) OR “Electroencephalography” AND “BCI”) AND PUBYEAR > 2019 AND PUBYEAR < 2026.

For the meta-analysis as approached as shown in Figure 1, all works were included if indexed according to Scopus search criteria, unless the following applied:

- The work was not published in a journal or conference.
- The Results section of the original article is not presented clearly or in a manner that is sufficiently transparent.
- The work is a literature review, or any type of work other than original, and hence not a primary source of presented data.

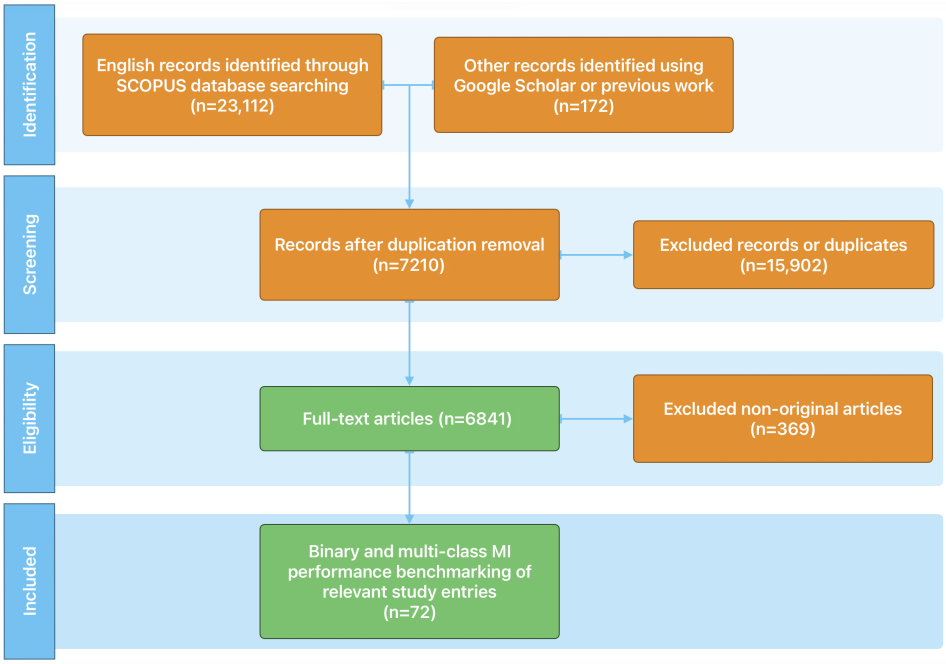


**Figure 1.** Systematic screening and selection of relevant studies and database documents, delineating the body of eligible works and benchmarked models forming the basis for comparative EEG-BCI performance assessment.

Furthermore, the following subcategories were defined and subsequently assessed after compiling the metadata of all resulting publications concerning general works in EEG signal processing, which are categorically structured by the following:

1. Works by mental task (paradigm-based);
2. Works by dataset selection (metric-based);
3. Works and technologies by signal processing application (i.e., signal classification vs. signal denoising).

For the qualitative assessment of works by dataset selection, we further specified the niche application current transformer-based deep learning by further limiting search

criteria appending AND Transformers to the query string, given their importance and trending popularity among data scientists and contemporary research across domains [2–5].

## 3. Results

### 3.1. Meta Analysis

As per Table 1, documents addressing or mentioning the complications with noise-level and problematic signal-to-noise ratio (SNR), or documents that are aimed at directly denoising the EEG signal, are resulting in a count of 5167, with 2788 documents resulting within the subgroup of the "EEG" AND "BCI" AND "signal processing" query.

For works indexed in the query for general Signal Processing, 6841 entries were identified as shown in Table 2 with MI-based BCI-EEG works appearing in 2602 documents. Queries for EEG-signal proccesing works with the search keys ML and CNN comprised a total number of 1421 and 1597 hits, respectively.

**Table 1.** Scopus search queries results, selectively summarized and sorted. Left: Comparison by BCI approach and subcategory. Right: Comparison by EEG-BCI denoising paradigms.

| BCI Approach and Subcategories | Count | Denoising Study Document Query | Count |
|---|---|---|---|
| General Signal Processing | 2788 | No Exclusion Criteria | 5167 |
| General Motor Imagery | 2602 | Transformers | 146 |
| CNNs | 951 | Diffusion | 78 |
| Transformers | 205 | BCI and Diffusion | 36 |
| CNNs and MI | 511 | GANs | 45 |
| Transformers and MI | 54 | Self-Attention | 65 |

**Table 2.** Scopus search queries results, selectively summarized and sorted. Left: Comparison by feature paradigm search. Right: Comparison by data processing methodologies for feature extraction.

| By BCI Feature Paradigm | Count | By Signal Processing Methodology | Count |
|---|---|---|---|
| Motor Imagery | 2602 | No Exclusion Criteria | 6841 |
| P300 | 388 | Traditional ML | 1421 |
| Error-related Potential(s) | 314 | Deep Learning | 1597 |
| Imagined/Inner Speech | 138 | | |
| Mental Workload Estimation | 107 | | |

Out of 54 works that correspond to the query string TITLE-ABS-KEY (("EEG" AND "BCI" AND "transformer" AND "MI") OR ("EEG" AND "BCI" AND "transformer" AND "motor imagery")) AND PUBYEAR > 2020 AND PUBYEAR < 2026, a total of 44 works would ultimately be used to compute and assess dataset utilization in the research field's subcategory of transformer-based MI-BCI research, the extensive list of which is included in Appendix A Table A1. Overall, the BCI Competition IV 2a dataset [6] is, with a wide margin, the most widely adopted dataset to evaluate model performance, which is followed by subset 2b of the BCICIV competition. Another substantial albeit much smaller proportion of only 15.91% of all studies in the scope of selected works utilized the PhysioNet dataset [7].

The overall trends of the meta-analysis are summarized in Figure 2. As per subplot (a), traditional ML has been the most prominent methodology employed until 2020 when more modern methodologies (in particular, DL such as CNNs) took over the majority share. For the period from 2022 to 2023, we can note a decrease in the upward trend of EEG-BCI signal processing works with a significant decrease of overall output within the umbrella domain that year. However, despite this overall drop of resaerch efforts in this domain, the transformer-based subdomain of the receding research field keeps emerging and extending not only their share but also continuing to increase their total research outputs. As per Figure 2 subplot (a), transformer-based BCI outputs are also the only

subcategory whose reseach outout volume keeps increasing for five years consecutively since 2020, with CNNs slumping to a plateau during 2022 and never recovering, eventually declining in terms of output volume after 2023. While transformer's total share of research in the domain as depicted in subplot (a) is still comparatively small, one may predict a break-even point in the near future, at which they will emerge as the predominant and de facto DL methodology in the realm of EEG-BCI signal processing. To be exact, should this trend continue, transformers will replace CNN-based feature extraction for BCIs in the next 1–3 years, as both CNN-based BCI research and CNN-based MI-decoding are receding strongly, with transformer-based signal BCI output already having passed general CNN-BCI output in 2025 (as of September). Riemannian manifold-based processing continues to denote a negligible share of total research efforts in this domain as depicted per subplot (b) of Figure 2 with a now receeding research interest. As per subplot (c), MI continues to present the vast majority within the distribution mental tasks for the development of EEG-based BCI research and development.

One major trend is clearly depicted in Figure 2 in subplot (d), which shows annualized outputs as a proxy for relative reserach interests in EEG signal denoising and artifact removal, the transformer-based application of which recorded an outstanding growth for the past three years (since mid-2022). Whilst relatively understudied so far, and there being relatively low total hits when compared to MI-based decoding and other research fields of the EEG-BCI domain, the transformer-based application has seen the sharpest upward trend out of all the subdomains and application fields covered in this review study, progressing from only one work in 2021 all the way to presenting aboout half (48.2%) of the Scopus-indexed denoising studies presented here.

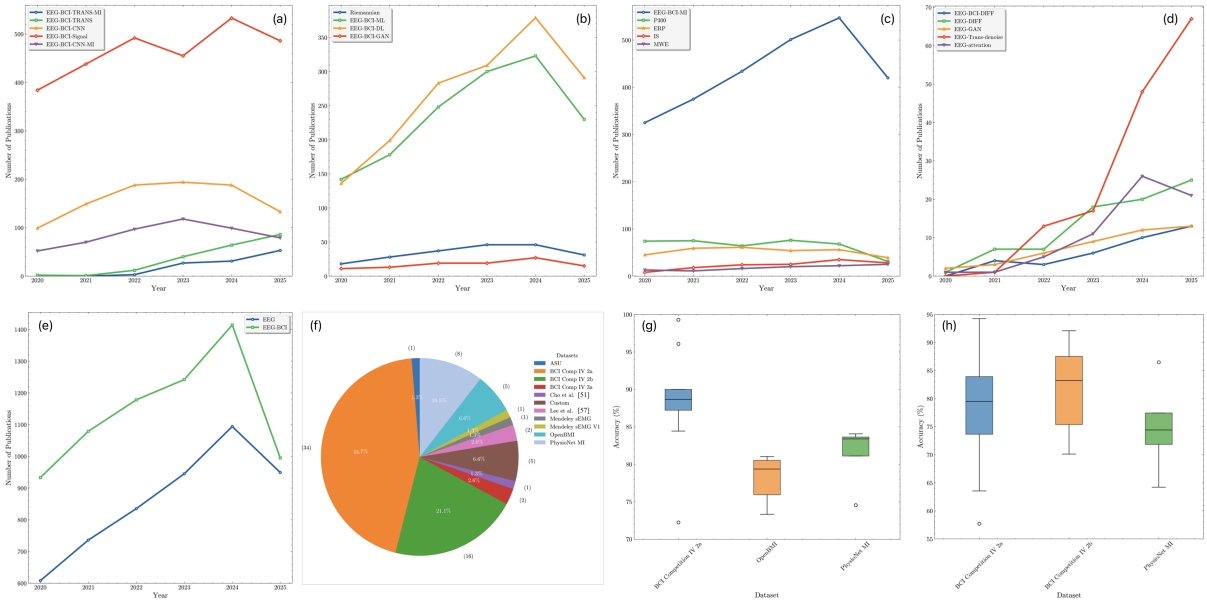


**Figure 2.** Overview of trends identified in meta-analysis. (**a**) Annual publication output broken down by keywords combined with EEG and BCI, showcasing a plateau and eventual decline of research interest in CNN-based processing, whilst transformer-based approaches steadily increase for 5 years consecutively. (**b**) Annualized trends in selected signal processing categories. (**c**) Annualized reserach outputs sorted by BCI paradigm (mental task used to evoke potentials). (**d**) Trends in EEG signal denoising and artifact removal, showing an explosive interest in transformer-based denoising since 2023. (**e**) Growing overall research outputs referencing EEG and EEG-BCI systems. (**f**) Benchmarking dataset distribution across pubications in the MI-BCI domain [8,9]. (**g**,**h**) Reported benchmarking performances of proposed transformer-based models in binary and multi-class MI-BCI systems, respectively.

As depicted in subplots (g) and (h), transformer-based MI classifiers trained and evaluated on BCI Competition IV 2a typically achieve mean validation accuracies around the high-80% range for binary tasks and mid-70% for multi-class settings. These averages establish a realistic reference point for interpreting the highest-reported results—such as CNN–Transformer fusion, hierarchical attention, or deep ensemble variants—which exceed this baseline mainly under subject-specific or binary-class conditions, or when extended time-window augmentation is applied. When normalized to multi-class, cross-subject evaluations, the practical gains of these top architectures correspond to relative improvements of roughly 5–10 percentage points rather than orders of magnitude. This contextual framing clarifies that while peak performances demonstrate the potential of attention-based models, the field's typical accuracy distribution remains concentrated around these central values. However, qualitative limitations of such assumptions and benchmarking are discussed in Section 4.

As for the most prominent BCI paradigm (Motor Imagery), it is must be mentioned that after compiling all transformer-based EEG–BCI publications retrieved through the updated Scopus query set, the consolidated findings in subplot (f) of Figure 2, which denote and summarize various datasets and their utilization across the complete corpus of transformer-based motor-imagery studies published since 2020, there is a remaining focus on the now almost 20-year old BCI Competition IV dataset of 2008 [6]. The distilled data clearly show how the BCI Competition IV 2a dataset remains the predominant benchmark, accounting for the clear majority of experimental validations, followed by its subset 2b and, with markedly lower frequency, the PhysioNet MI dataset. All remaining public datasets together contribute only a small fraction of total works, while a separate Custom category captures studies employing proprietary or non-public recordings that are not directly comparable to standard benchmarks.

On an additional note, the reviewed works report accuracies in diverse formats and often did not include repeated-trial statistics or variance measures; therefore, it is not feasible to compute aggregate confidence intervals or effect sizes across studies. Accordingly, this review reports the mean accuracies as published by the respective authors while emphasizing the need for future EEG-BCI research to provide mean $\pm$ SD or confidence intervals and apply statistical tests when comparing models on identical datasets. Table 3 compiles representative results from both pure transformer and hybrid CNN–Transformer models across widely used motor imagery datasets, including BCIC-IV 2a/2b, PhysioNet MI, and OpenBMI. Each entry lists reported performance metrics and the corresponding evaluation protocol. Although such aggregation provides a useful overview of the field, the heterogeneity of preprocessing steps, subject partitioning strategies, and evaluation criteria across studies means that the values should be interpreted descriptively rather than as directly comparable benchmarks.

**Table 3.** Representative transformer-based and hybrid CNN–Transformer models evaluated on major MI EEG datasets. The table summarizes reported accuracies and evaluation protocols, illustrating typical performance ranges across datasets and architectural paradigms. While hybrid models often achieve higher accuracies under subject-dependent conditions, methodological inconsistencies across preprocessing, partitioning, and evaluation settings preclude direct quantitative comparison.

| # | Study (First Author) | Model | Type | n | Acc [%] | Evaluation Protocol |
|---|---|---|---|---|---|---|
| | | *Continued on next page* | | | | |
| **BCI Competition IV 2a** | | | | | | |
| 1 | Deny (2023) [10] | Hierarchical Transformer | Pure | 2 | 90.00 | subject-dependent |
| 2 | Chaudhary (2024) [11] | Two-stage Transformer | Pure | 2 | 88.50 | subject-dependent |
| 3 | Keutayeva (2024) [12] | Compact Conv. Transformer | Pure | 4 | 70.12 | subject-independent |
| 4 | Liu (2022) [13] | CNN–Transformer Fusion | Hybrid | 2 | 99.29 | non-standard split |
| 5 | Mehtiyev (2023) [14] | DeepEnsemble (ViT + CNN) | Hybrid | 2 | 96.07 | subject-dependent |
| 6 | Hameed (2024) [15] | Transformer + ICA | Hybrid | 2 | 88.75 | subject-dependent |
| 7 | Li (2024) [16] | Transformer + M-LLE + SW-CNN | Hybrid | 2 | 84.44 | subject-dependent |
| 8 | Shi (2024) [17] | EEG-VTTCNet (ViT + TCN) | Hybrid | 4 | 84.58 | subject-dependent |
| 9 | Nguyen (2024) [18] | EEG-TCNTransformer | Hybrid | 4 | 83.41 | subject-dependent |
| 10 | Zhao (2024) [19] | CTNet | Hybrid | 4 | 82.52 | subject-dependent |
| 11 | Zhang (2023) [20] | Local + Global Conformer/Transformer | Hybrid | 4 | 80.20 | subject-dependent |
| 12 | Shi (2024) [21] | Swin-CANet | Hybrid | 4 | 78.78 | subject-dependent |
| 13 | Zare (2024) [22] | Integrated Transformer–CNN | Hybrid | 4 | 75.30 | subject-dependent |
| 14 | SCTrans (2024) [23] | SCTrans | Hybrid | 4 | 68.61 | subject-dependent |
| **BCI Competition IV 2b** | | | | | | |
| 15 | Luo (2023) [24] | Shallow Mirror Transformer | Pure | 3 | 77.36 | subject-independent |
| 16 | Song (2023) [25] | Global Adaptive Transformer | Pure | 3 | 92.08 | subject-dependent |
| 17 | Keutayeva (2024) [12] | Compact Conv. Transformer | Pure | 3 | 70.12 | subject-independent |
| 18 | Shi (2024) [17] | EEG-VTTCNet (ViT + TCN) | Hybrid | 3 | 90.94 | subject-dependent |
| 19 | Song (2023) [26] | EEG Conformer | Hybrid | 3 | 84.63 | subject-dependent |
| 20 | Zhao (2024) [19] | CTNet | Hybrid | 3 | 76.27 | subject-dependent |
| **PhysioNet MI** | | | | | | |
| 21 | Keutayeva (2023) [27] | Attention-based Transformer | Pure | 3 | 86.47 | subject-independent |
| 22 | Xie (2022) [28] | Transformer + DL (2-class) | Hybrid | 2 | 83.31 | subject-dependent |
| 23 | Xie (2022) [28] | Transformer + DL (3-class) | Hybrid | 3 | 74.44 | subject-dependent |
| 24 | Xie (2022) [28] | Transformer + DL (4-class) | Hybrid | 4 | 64.22 | subject-dependent |
| 25 | Luo (2024) [29] | MI-MBFT (Multi-branch Transformer) | Hybrid | 2 | 84.07 | subject-dependent |
| 26 | Ajali (2024) [30] | Optimization + DL | Hybrid | 2 | 74.54 | subject-dependent |
| **OpenBMI** | | | | | | |
| 27 | Luo (2023) [24] | Shallow Mirror Transformer | Pure | 2 | 80.54 | subject-independent |
| 28 | Liu (2024) [31] | MSVTNet (ViT) | Pure | 2 | 75.93 | subject-dependent |
| 29 | Zhang (2023) [20] | Local + Global Conformer/Transformer | Hybrid | 2 | 81.04 | subject-dependent |
| 30 | SCTrans (2024) [23] | SCTrans | Hybrid | 2 | 73.33 | subject-dependent |
| **Other / Custom or Additional Benchmarks** | | | | | | |
| 31 | Lee (2023) [32] | Continual Transformer (online) | Pure | 2 | 77.00 | online/real-time |
| 32 | Ahn (2023) [33] | Multiscale Conv-Transformer | Hybrid | 2 | 72.00 | subject-dependent |
| 33 | Li (2024) [34] | Transformer + EEGNet (metaverse) | Hybrid | 3 | 71.31 | online/real-time |
| 34 | Liu (2024) [35] | Multimodal Transformer (exoskeleton) | Hybrid | 4 | 91.25 | online/real-time |

### *3.2. Qualitative Assessment*

Overall, a trend is depicted in which more emphasis is given to DL-based approaches in neural decoding than when compared to traditional ML-based decoders and classifiers with DL substantially increasing computational complexity and resource demand whilst simultaneously being able to draw improved features from complex EEG recordings [36]. This trend has been observed over the past few years, resulting in steadily improved max performances globally across methodologies [5,37].

#### 3.2.1. Proposed Frameworks Improvements for Task-Related Processing

One outstanding performance was derived by the introduction of hierarchical transformer (HT) models as proposed by [10]. Reporting a validation accuracy of 90%, this approach demonstrates the potential of transformer architectures to overcome long-standing challenges in EEG-based MI decoding. This is largely due to the inherent variability of EEG signals, the presence of noise and artifacts, and the difficulty in capturing long-range temporal dependencies. The success of this study stems from its ability to effectively extract and selectively weight task-relevant EEG features while suppressing irrelevant signals, which is an area where conventional CNNs and recurrent architectures fall short [10,38].

The key innovation of this model is its hierarchical attention framework, which refines MI feature extraction across multiple temporal scales. Unlike traditional CNN-based approaches that apply convolutional filters to extract features from fixed temporal windows, or LSTMs that model sequential dependencies with recurrent processing, this model explicitly learns which time intervals within an EEG trial contain the most relevant motor imagery patterns. This is achieved through a two-stage transformer architecture: a low-level transformer (LLT) and a high-level transformer (HLT). The LLT operates on short, overlapping time segments, treating each segment as an independent input token. Through self-attention, it captures localized temporal features while preserving spatial information across EEG channels. This is crucial because MI patterns are not uniformly distributed across a trial; some periods contain clear task-related activations, while others are dominated by noise or non-task-related brain activity. By encoding short-term dependencies, the LLT ensures that relevant patterns are effectively extracted without excessive temporal smoothing. Once these localized representations are obtained, the HLT takes over, processing the sequence of extracted features to assign adaptive attention weights. This allows the model to prioritize segments where strong MI-related activity occurs, effectively filtering out irrelevant portions of the EEG trial. Unlike LSTMs, which struggle with long-range dependencies due to vanishing gradients, the transformer's self-attention mechanism enables direct global interactions between distant time intervals, significantly improving feature selection. The hierarchical design also prevents information loss, which is a common problem in deep CNN pipelines where successive pooling and convolution layers progressively discard temporal details.

Another factor contributing to the model's high accuracy is its training methodology. Instead of jointly training the LLT and HLT, the authors use a staged training approach. The LLT is first optimized independently, ensuring the robust extraction of short-term EEG features. Only after this step is the HLT trained, learning to refine and combine these extracted features for final classification. This hierarchical optimization prevents gradient interference between the two levels of processing, which is a common issue when multiple transformer layers are stacked directly.

The DeepEnsemble model as proposed in [14] introduces a novel ensemble learning framework by integrating multiple deep learning architectures to enhance accuracy and robustness. The proposed method leverages the complementary strengths of multi-layer perceptrons (MLPs), CNNs, vision transformers (ViTs), and a hybrid CNN-XGBoost model. By combining these architectures through a soft voting ensemble strategy, the model significantly improves generalization across subjects, mitigating inter-subject variability, which remains a persistent challenge in MI-based BCI research.

For the ensemble models, each deep learning model within the ensemble is trained independently before being integrated into the final classifier. The MLP baseline is a fully connected architecture that flattens the EEG feature space into a structured network of 20 hidden layers, each containing 100 neurons, allowing for efficient pattern learning. The CNN model extracts local spatial–temporal patterns through three convolutional layers

with varying kernel sizes (3 × 3, 1 × 1, 3 × 3), which is followed by max-pooling and dense layers with ReLU activation functions. To enhance the CNN's feature representation, the hybrid CNN-XGBoost model replaces the final dense layers with gradient boosting decision trees, improving classification stability by leveraging XGBoost's ability to model complex nonlinear relationships in the EEG data. The ViT component, inspired by recent advancements in self-attention architectures, restructures EEG signals into patch-embedded image representations, enabling a more global feature extraction approach without the constraints of convolutional filtering. This component allows the model to capture long-range dependencies between EEG segments, complementing the local feature extraction performed by CNN layers.

The final classification decision is obtained through soft voting ensemble learning, in which probability outputs from all models are combined to create a more generalized decision boundary across subjects. This ensemble method ensures that weaknesses inherent in individual models—such as the CNN's susceptibility to local feature bias or the ViT's reliance on extensive data—are offset by the strengths of the other architectures. The experimental results demonstrate that DeepEnsemble achieves a classification accuracy of 96.07% on the best-performing subject, surpassing individual deep learning models and conventional single-network approaches. This improvement is particularly pronounced for subjects with smaller training datasets, where the diverse ensemble architecture prevents overfitting and enhances performance stability.

Compared to the hierarchical transformer (HT) model, which achieved 90% accuracy using a structured multi-scale attention framework, DeepEnsemble adopts a fundamentally different paradigm by aggregating multiple independent classifiers rather than refining feature extraction through attention-based weighting. While the HT model effectively prioritizes time-dependent EEG features by dynamically re-weighting intervals of high MI activity, DeepEnsemble bypasses explicit temporal attention mechanisms and instead relies on the diversity of architectures to improve classification robustness. The HT model excels at structured feature extraction, particularly in scenarios where precise temporal segmentation is critical, whereas DeepEnsemble thrives on architectural redundancy, leveraging multiple perspectives on feature representation to ensure higher accuracy across diverse subjects. However, while DeepEnsemble benefits from ensemble diversity, it comes at the cost of increased computational complexity and training overhead, as multiple models must be trained and optimized separately before integration. The HT model, in contrast, achieves high accuracy with a single transformer-based pipeline, making it more computationally efficient while still achieving competitive performance. These distinctions highlight a broader trade-off in MI EEG classification: feature refinement through structured attention mechanisms versus robustness through model diversity, with each approach offering distinct advantages depending on dataset constraints and computational resources.

The by far greatest performance was reported by Liu et al. [13], which significantly improved MI-EEG classification accuracy by integrating CNNs with Transformer-based feature extraction blocks. Their approach involves a structured pipeline, starting with rigorous preprocessing to refine the raw EEG signals. Artifacts were removed, and dimensionality is reduced through bandpass filtering and principal component analysis (PCA), resulting in a more compact feature representation in the time–space–frequency domain. Following preprocessing, local temporal-frequency features are extracted through 1D-CNN layers. The CNN applies convolutional and pooling operations to efficiently capture localized patterns in the temporal domain while simultaneously reducing the length of the EEG time series. This step not only helps mitigate overfitting by reducing model complexity but also creates a refined, lower-dimensional feature representation for subsequent processing. Compared to the conventional CNN-LSTM approaches, which utilize long short-term memory (LSTM)

networks to capture temporal dependencies, the CNN-Transformer architecture leverages multi-head self-attention mechanisms to extract richer, more abstract representations from the EEG time series [13].

This Transformer component of the model further enhanced feature extraction by leveraging position encoding and multi-head attention. Unlike LSTM networks, which process sequential data step by step, the Transformer architecture allows for parallel computation, significantly improving training efficiency. The position-encoding mechanism ensured that temporal relationships were preserved despite the parallelization, while the attention mechanism facilitated the capture of dependencies between distant time steps. The encoder processes CNN-derived feature vectors by applying multi-head self-attention layers, which is followed by residual connections and layer normalization to stabilize training and enhance representational capacity.

As for the decoding stage, the transformer outputs are combined with the original CNN-derived features, ensuring that both low-level and high-level temporal representations are integrated. A fully connected layer with a softmax activation function is used to generate classification probabilities, ultimately producing highly accurate predictions for motor imagery tasks. The researchers validate their model using the most common dataset (BCI Competition IV 2a), demonstrating that the CNN–Transformer model achieved an average accuracy of 99.29% and a kappa value of 98.43%, significantly outperforming the CNN-LSTM baseline by 3.72% and 7.68%, respectively.

The superior performance of this approach can be attributed to several key factors. First, the combination of CNNs and transformers allow for a more effective decomposition of EEG signals into meaningful temporal and frequency components, leveraging CNNs for localized feature extraction and transformers for long-range dependencies. Second, the parallelized nature of the transformer architecture enables more efficient training and inference compared to sequential models like LSTMs. Third, the preprocessing steps ensure that high-quality, noise-reduced input data were provided to the network, enhancing the robustness of feature extraction. Finally, the careful tuning of hyperparameters, including the choice of kernel sizes, filter counts, dropout rates, and attention heads, contributes to the model's stability and accuracy.

By swapping a recurrent LSTM structure for a transformer-based attention mechanism, the researchers address the inherent limitations of sequential processing while preserving the ability to capture long-term dependencies in EEG signals. Their findings not only suggest that hybrid CNN–Transformer architectures offer a promising direction for improving MI-EEG classification accuracy, as has been proposed by previous researchers in the past [5], but also provide substantial evidence for the superiority of transformer-based neural decoding as an effective foundation for real-world BCI applications.

While the CNN–Transformer fusion network demonstrated exceptional accuracy through its hierarchical feature extraction strategy, DeepEnsemble takes a fundamentally different approach by integrating multiple independent classifiers to achieve robustness and generalizability. The CNN–Transformer model relies on a structured pipeline, where CNN layers extract localized time-frequency features, followed by a transformer module that refines these features through self-attention and position encoding. This hierarchical attention mechanism allows for the precise weighting of relevant EEG time intervals, improving classification accuracy by prioritizing salient MI-related features while suppressing noise. The architecture's reliance on a single, highly optimized deep learning pipeline ensures computational efficiency and a well-defined feature extraction process.

In contrast, DeepEnsemble eschews hierarchical feature learning in favor of architectural diversity, combining MLPs, CNNs, vision transformers, and gradient boosting (XGBoost) classifiers in an ensemble framework. Rather than progressively refining EEG

features through structured attention mechanisms, DeepEnsemble aggregates diverse feature representations from multiple independent classifiers, which are combined through soft voting to achieve a more generalized decision boundary. This ensemble approach mitigates the bias–variance trade-off, effectively balancing localized feature extraction (CNN), long-range dependencies (ViT), and structured decision making (XGBoost). However, while DeepEnsemble achieves slightly higher accuracy in some subjects, it comes at the cost of increased computational complexity and higher training overhead, as multiple models must be independently trained and optimized before integration.

The key trade-off between these approaches lies in their philosophy of feature extraction and classification robustness. The CNN–Transformer model excels in structured, hierarchical feature learning, making it ideal for datasets where temporal segmentation and long-range EEG dependencies are crucial. In contrast, DeepEnsemble leverages classifier diversity, making it more adaptive to inter-subject variability and small training datasets. While CNN–Transformer achieves state-of-the-art accuracy with a single, well-optimized deep learning pipeline, DeepEnsemble's multi-model fusion strategy offers greater resilience to individual model weaknesses at the expense of efficiency. These distinctions highlight two divergent yet effective paths for improving MI-EEG classification: refining structured temporal feature extraction versus leveraging classifier diversity for robustness.

#### 3.2.2. Classification Performance

Today, the highest-reported validation accuracies on the BCI Competition IV 2a dataset are reported using CNN–Transformer fusion networks with 99.29% [13], deep learning ensemble models using vision transformer modules with 96.07% [14], and approaches that employ hierarchical transformer modules with 90.00% [10] validation accuracy, all of which are significantly higher than the average accuracies recorded in all selected papers as per subplot (d) of Figure 2. This presents a significant improvement as most prior deep learning models applied to this dataset struggle to reach this level of performance, typically plateauing below 80% [5].

However, some of these numbers seem too good to be true. One recurring problem in MI–EEG benchmarking is the use of non-standard evaluation protocols that inadvertently leak subject information between training and test sets, inflating reported accuracies far beyond what is typically attainable under subject- or session-independent testing [39,40]. For BCI Competition IV–2a, the standard and defensible practice is to respect the dataset's two-session structure (train on one session and test on the other, or perform leave-one-session/subject-out evaluation), as recommended in the competition documentation and widely adopted in recent transformer baselines [41–43]. Under such protocols, contemporary CNN/Transformer or hybrid models report subject-dependent means in the mid-80% range and subject-independent means around the mid-70% range, e.g., 85.15% on IV–2a with CIACNet, 82.95% with MSCFormer, and 74.48% in subject-independent testing in a recent transformer study [44–46]. Results in this regime are also consistent with other recent augmentation or GAN-assisted pipelines on IV–2a, reporting around 80–81% [47]. By contrast, claims of $\geq$99% accuracy on IV–2a typically arise when trial segments from the same subjects are randomly split across folds, when both sessions are pooled and windowed before splitting, or when hyperparameters are tuned on data that later appear in evaluation—each of which constitutes information leakage that substantially overestimates out-of-subject performance [39,40]. As one illustrative example, an EasyChair preprint reports 99.29% on IV–2a using a CNN–Transformer fusion network, but the paper does not adhere to session-preserving or subject-independent evaluation; under the community's standard protocols, such numbers would be outliers by a very large margin relative to recently published baselines [13,44–46]. Taken together, these observations emphasize that

meaningful comparison on IV–2a requires session- and subject-aware splits; otherwise, leakage can yield headline figures that do not translate to real-world generalization [39,41,42].

#### 3.2.3. Artifact Removal

Table 4 summarizes representative transformer and hybrid transformer architectures applied to EEG denoising across datasets and artifact types. The aggregation highlights how performance evaluation remains inconsistent—correlation coefficients, relative errors, and SNR improvements are variably reported often without standardized baselines or matched signal conditions. Despite these discrepancies, the compilation evidences the rapid diversification of transformer-based denoising paradigms, ranging from diffusion-conditioned models to recurrent–attention hybrids.

As summarized in Table 4, the highest-performing transformer-based EEG denoising frameworks are consistently reported for hybrid diffusion–transformer architectures and attention-enhanced recurrent hybrids. Among these, Huang et al. (2024) introduced the EEG-DFUS framework, which achieved correlation coefficients of up to 0.989 on EEGDenoiseNet and 0.992 on SSED for muscular and ocular artifact suppression, respectively [48]. EEG-DFUS integrates a Denoising Diffusion Implicit Model (DDIM) with transformer-driven attention refinement. In this configuration, the diffusion module iteratively refines signal estimates through progressive denoising steps that learn the conditional distribution of clean EEG given its corrupted form, while transformer-based attention blocks serve as contextual priors that emphasize physiologically plausible spatio-temporal patterns. The model uses cross-attention between diffusion steps and intermediate transformer embeddings, enabling the selective reconstruction of phase-coherent neural components while suppressing structured non-cortical noise. This hierarchical fusion of diffusion inference and transformer attention results in unprecedented reconstruction fidelity, setting a new benchmark for EEG denoising.

Yin et al. (2024) presented a pure transformer-based denoiser on the SSED dataset that achieved $CC = 0.988 \pm 0.003$ and $SNR = 17.73 \pm 1.24$ dB [8]. Their design emphasizes dual-stage cross-attention operating jointly on spatial and temporal dimensions, using separate transformer encoder branches for channel-wise and temporal feature representation. The architecture employs symmetric encoder–decoder blocks with skip connections to preserve fine-grained signal detail during reconstruction. Unlike diffusion-based methods, this transformer relies entirely on self-attention to learn structured noise correlations across electrodes, optimizing a reconstruction loss composed of relative root mean square error (RRMSE) and cosine similarity. The resulting denoiser demonstrates that a sufficiently expressive transformer can match or exceed diffusion-conditioned hybrids when equipped with artifact-specific attention heads and balanced feature normalization across layers.

Bellamkonda et al. (2025) proposed a CNN–Transformer hybrid [49] that achieved $CC = 0.9212$ and an SNR improvement of approximately 35 dB on EEGDenoiseNet, representing one of the most efficient non-diffusion-based approaches to date. The model employs convolutional blocks for local spatial filtering and short-term temporal smoothing, feeding their feature maps into a multi-head self-attention transformer encoder. This design captures both localized artifact morphologies and long-range dependencies in the EEG time series. The transformer module's attention maps allow for the selective reweighting of temporally coherent neural activity, while the CNN front-end ensures robustness to transient high-amplitude disturbances. The architecture's relatively low parameter count and strong reconstruction fidelity make it particularly suitable for online or low-latency denoising applications.

**Table 4.** Representative transformer and hybrid transformer architectures for EEG denoising organized by dataset and artifact type. The compilation delineates performance dispersion across contemporary studies, reflecting the methodological heterogeneity that characterizes current transformer-driven artifact suppression research.

| # | Study | Dataset | Artifact Type | Architecture | Symbiont | CC | MSE | RRMSE | tRRMSE | sRRMSE | RE | PRD | SNR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **EEGDenoiseNet** | | | | | | | | | | | | | |
| 1 | Bellamkonda et al. [49] | EEGDenoiseNet | muscular | hybrid | CNN | 0.9212 | NR | 0.353 | NR | NR | NR | 2.412 | 35.39 |
| 2 | Tiwari et al. [50] | EEGDenoiseNet | not specified | hybrid | LSTM/GRU | – | NR | NR | NR | NR | NR | NR | NR |
| 3 | Xiaorang et al. [51] | EEGDenoiseNet | muscular | pure | – | 0.732 | NR | NR | NR | 0.677 | 0.626 | NR | NR |
| 4 | Xiaorang et al. [51] | EEGDenoiseNet | ocular | pure | – | 0.868 | NR | NR | NR | 0.497 | 0.491 | NR | NR |
| 5 | Wang et al. [52] | EEGDenoiseNet | muscular [1] | hybrid | GRU | 0.844–0.982 | NR | NR | NR | NR | NR | NR | 10.06–35.13 |
| 6 | Wang et al. [52] | EEGDenoiseNet | ocular [1] | hybrid | GRU | 0.922–0.987 | NR | NR | NR | NR | NR | NR | 19.92–39.93 |
| 7 | Huang et al. [48] | EEGDenoiseNet | muscular | hybrid | DDIM | 0.989 | NR | NR | NR | 0.171 | 0.154 | NR | NR |
| 8 | Huang et al. [48] | EEGDenoiseNet | ocular | hybrid | DDIM | 0.983 | NR | NR | NR | 0.182 | 0.188 | NR | NR |
| **SSED** | | | | | | | | | | | | | |
| 9 | Yin et al. [8] | SSED | ocular | pure | – | $0.978 \pm 0.007$ | – | $0.156 \pm 0.016$ | $0.164 \pm 0.019$ | $0.163 \pm 0.013$ | – | – | $16.914 \pm 0.948$ |
| 10 | Yin et al. [8] | SSED | muscular | pure | – | $0.988 \pm 0.003$ | – | $0.135 \pm 0.020$ | NR | NR | – | – | $17.73 \pm 1.236$ |
| 11 | Huang et al. [48] | SSED | ocular | hybrid | DDIM | 0.992 | NR | NR | NR | 0.121 | 0.127 | NR | NR |
| **BCIC IV 2a / 2b** | | | | | | | | | | | | | |
| 12 | Chen et al. [53] | BCIC IV 2a | non-specific | hybrid | DDIM | $58.4 \pm 1.3$ | NR | NR | NR | NR | NR | NR | NR |
| 13 | Yin et al. [8] | BCIC IV 2a | non-specific | pure | – | NR | – | – | – | – | – | – | – |
| 14 | Yin et al. [8] | BCIC IV 2b | non-specific | pure | – | NR | – | – | – | – | – | – | – |
| **Other / Miscellaneous** | | | | | | | | | | | | | |
| 15 | Tiwari et al. [50] | VEP | not specified | hybrid | LSTM/GRU | 0.9513 | 0.033 | NR | NR | NR | NR | NR | 10.56 |
| 16 | Tiwari et al. [50] | MNIST | not specified | hybrid | LSTM/GRU | 0.813 | 0.0286 | NR | NR | NR | NR | NR | NR |
| 17 | Yin et al. [8] | MNE Sample | non-specific | pure | – | $91.34 \pm 3.87$ | NR | NR | NR | NR | NR | NR | NR |
| 18 | Alzahab et al. [54] | AMIGOS | non-specific | pure | – | $93.04 \pm 2.72$ | 0.9665 | 0.0004 | 0.0192 | NR | NR | NR | NR |
| 19 | Chen et al. [53] | DEAP | non-specific | hybrid | DDIM | $58.3 \pm 1.4$ | NR | NR | NR | NR | NR | NR | NR |

[1] The results of this study were reported over a range from −7 dB to 2 dB.

Overall, these results underline two converging methodological directions: (i) diffusion-conditioned hybrids that model noise evolution across scales while leveraging transformer attention for contextual recovery, and (ii) pure attention-based transformers that, when sufficiently deep and regularized, achieve near-diffusion performance without stochastic refinement. Despite these advances, cross-study comparisons remain difficult due to inconsistent metrics (CC, RRMSE, PRD, SNR) and the limited reporting of downstream decoding improvements, emphasizing the need for unified evaluation protocols linking denoising quality to BCI task performance.

## 4. Discussion

The advancements in transformer-based architectures for MI-BCI classification present a significant departure from traditional deep learning models, yet their integration into real-world systems remains a challenge due to fundamental technical and methodological limitations. While the reviewed works demonstrate that self-attention mechanisms improve temporal feature extraction, many aspects of these architectures are still underexplored, particularly in the context of cross-subject generalization, real-time applicability, and computational feasibility. In other words, attention improves what models can represent, but it does not by itself solve the procedural issues of evaluation rigor, domain shift across users/sessions, or the engineering constraints of closed-loop systems.

Developing industry-level or real-time classification applications with online decoding is crucial in MI research because it ensures that signal processing and decoding methods are not only optimized for controlled offline datasets but also validated in dynamic, real-world conditions where real-time adaptability, user interaction, and system robustness are essential for practical BCI deployment. In the domain of transformer-based neural decoding for MI BCIs, however, only a few papers (a total of four, which equals less than 9% of the publications investigated as part of this systematic review) have demonstrated an actual application for online (herein referred to as live) applications using transformer-based or hybrid-based BCIs only [32,34,35,55]. Notably, even these demonstrations often report partial pipeline latencies or omit I/O and feedback timing, underscoring the gap between offline accuracy and deployable closed-loop control.

A critical bottleneck in EEG-based MI classification is the balance between model complexity and real-time performance. Transformer architectures, particularly those that rely on multi-head attention mechanisms and large embedding dimensions, exhibit quadratic computational complexity with respect to input sequence length. This makes them significantly more demanding than CNN or hybrid CNN-LSTM architectures, which operate with fixed receptive fields and localized feature extraction. While hierarchical transformer models attempt to mitigate these challenges by reducing sequence length at lower processing levels, they remain inherently heavier than traditional architectures. Few works explicitly quantify the inference time of transformer-based MI models, which raises concerns about their practical viability in real-time BCI applications, where decoding must occur within milliseconds. Where inference times are reported, differences in windowing, overlap, and pre/postprocessing pipelines complicate fair comparison. Standardized efficiency reporting (parameters, FLOPs, per-trial latency on specified hardware, and end-to-end acquisition-to-feedback delay) is therefore essential to substantiate real-time claims. Future research must address latency reduction and develop hardware-efficient adaptations of these architectures to enable their use in real-world neurotechnology.

Beyond computational constraints, the lack of any systematic exploration of pretraining strategies in transformer-based MI classification is another overlooked factor. Unlike in natural language processing (NLP) or computer vision, where transformers benefit from large-scale pretraining on vast datasets, EEG-based models are often trained from scratch

on small, subject-dependent datasets. Given the limited availability of large, high-quality EEG datasets, transformers in MI applications do not leverage the full potential of transfer learning. There is a clear gap in research regarding the use of self-supervised learning (SSL), contrastive learning, or domain adaptation techniques that could allow transformers to learn generalizable EEG features across subjects and tasks. The reliance on fully supervised learning may contribute to overfitting on specific datasets, limiting the scalability and robustness of these models outside laboratory conditions. Recent modality-agnostic SSL objectives (e.g., masked prediction or temporal contrast) are natural candidates for EEG, but they require careful adaptation to non-stationary rhythms and subject/session variability; reporting label-efficiency curves after SSL pretraining would make claimed benefits tangible.

Another crucial but often neglected issue is the role of inductive biases in model design. CNNs, due to their localized receptive fields, inherently capture spatially structured relationships in EEG data, making them particularly effective for motor imagery classification, where specific electrode regions encode task-related signals. Transformers, on the other hand, are data-driven and require significantly larger datasets to learn meaningful relationships. The reviewed studies demonstrate that hybrid models combining CNNs with transformers outperform standalone transformers, yet there is little consensus on optimal architectural design choices for these hybrid approaches. Whether spatial embeddings should be learned explicitly within transformer layers or whether CNNs should be strictly confined to low-level feature extraction remains an open question. Addressing these design choices through ablation studies and architecture search methods would provide a clearer framework for constructing optimized transformer-based EEG models. In practice, the strongest hybrids exploit CNNs or shallow temporal convolutions for inductive bias and downsampling, reserving attention for global context; however, standardized ablations (with fixed splits/preprocessing) are still rare, making it hard to separate architectural merit from protocol variance.

The cross-subject variability problem also remains unresolved despite the reported improvements in transformer-based models. EEG signals exhibit significant inter-subject differences due to variations in neurophysiology, electrode placement, and signal-to-noise characteristics. While ensemble models such as DeepEnsemble address this issue by aggregating predictions from multiple classifiers, this approach is computationally prohibitive and does not fundamentally resolve the domain shift problem across subjects. Hierarchical transformer models attempt to overcome this by dynamically weighting time intervals, but they still require subject-specific fine tuning. There is a need for research into adaptive, subject-independent learning strategies, such as meta-learning or domain adversarial training, which could improve transformer generalization without requiring individual recalibration for new users. Calibration-efficient adapters (lightweight, few-shot layers) and conditional normalization tied to subject/session statistics are promising, but they must be evaluated on truly held-out subjects and sessions to avoid optimistic bias.

While transformer-based MI classifiers are achieving record-breaking accuracies, their practical deployment remains limited by computational demands, the absence of large-scale pretraining strategies, unresolved architectural design questions, and inadequate cross-subject generalization. Addressing these issues will require a multi-faceted approach, incorporating efficient model compression, transfer learning techniques, adaptive learning strategies, and standardized multi-dataset evaluations to ensure that the next generation of MI-BCI systems can function beyond laboratory constraints and into real-world applications. To address the here described challenges and gaps, exploring evolution-based, hyper-individualized, and automated MI classification strategies using particle swarm optimization (PSO) and genetic algorithms (GAs) may provide a viable solution. Such

optimization frameworks could enable the dynamic adaptation of transformer-based BCI architectures, optimizing hyperparameters, attention mechanisms, and feature extraction strategies based on subject-specific EEG characteristics. Such an approach could improve cross-subject generalization and session variability, mitigating the constraints of static, pretrained models that struggle in real-world applications. Meanwhile, subject-self-adjusting pipelines that have begun to show concrete benefits utilize a GA-driven framework that jointly searches CNN–Transformer hybrid topology [56], attention configuration, and training hyperparameters together with preprocessing choices (e.g., band selection, windowing, normalization), and they demonstrate improved inter-subject performance across multiple datasets, surpassing strong baselines such as EEG-Conformer and related hybrids under protocol-controlled evaluation. This line of work supports the view that (i) automated design can recover performant inductive biases tailored to each cohort, and (ii) optimizing the full pipeline—including data curation steps—is as important as optimizing the network alone. Further research into reproducible neuroevolution and a resource-aware architecture search (with explicit latency/FLOP constraints) may therefore be pivotal for making transformer-based MI BCIs both accurate and deployable.

In summary, the current evidence positions transformers and their hybrids as the most promising successors to purely convolutional pipelines for MI decoding, provided that future studies pair accuracy advances with protocol discipline, efficiency reporting, and calibration-efficient generalization. In the next section, we turn to denoising—an area of explosive recent growth—where comparable issues of metric heterogeneity and proxy-task validation currently limit cumulative progress despite strong signal-level gains.

The emerging body of transformer-driven EEG denoising research is expanding rapidly yet remains methodologically fragmented. Reported metrics across studies display wide dispersion, reflecting divergent preprocessing, artifact generation, and validation protocols rather than genuine performance differentials. This variability undermines comparability and reproducibility, as studies often differ in channel montages, sampling rates, baseline corrections, and normalization procedures—parameters that substantially affect quantitative measures such as CC, RMSE, or SNR. In several cases, datasets are resampled or filtered differently from their reference implementations, producing results that are not directly transferable across publications. The lack of uniform reporting of data conditioning steps and evaluation metrics constitutes one of the most pressing gaps in the field.

Establishing comparability will require community-wide conventions for reference signal generation, standardized artifact taxonomies (ocular, muscular, motion, mixed), and unified evaluation protocols that report at least CC, RMSE, and task-relevant downstream performance. These conventions should also define how synthetic artifacts are generated and embedded into clean EEG signals—whether linearly, additively, or via physiologically informed simulation. Further, a minimal benchmark specification should require that denoised outputs be tested not only at the signal level but also at the decoding level, using fixed pipelines for motor imagery or emotion recognition tasks, thereby linking noise suppression to functional impact. Public benchmarks—analogous to EEGDenoiseNet but extended to multi-artifact, multi-subject, and real-recording conditions—would enable systematic cross-architecture comparison, a longitudinal tracking of progress, and statistically meaningful meta-analysis.

EEGDenoiseNet currently serves as the de facto substrate for most transformer-based denoising studies, yet its synthetic and single-artifact design constrains ecological validity. The dominance of this dataset risks overfitting algorithmic advances to an idealized problem formulation that excludes the complexities of real EEG. Without demonstrating measurable improvement in functional decoding tasks (e.g., motor imagery classification or cognitive workload detection) following denoising, these studies remain academic proofs

of concept. Integrating proxy performance evaluation—where improvements in decoding accuracy or kappa are quantified after denoising—should become a standard practice. This would clarify whether denoisers enhance information transfer or merely reconstruct visually plausible waveforms without downstream utility.

Future work should prioritize (i) the creation of open, real EEG datasets with annotated multi-artifact contamination and synchronized auxiliary modalities (EOG, EMG, motion sensors); (ii) unified preprocessing pipelines that include fixed resampling, reference, and filtering stages; (iii) a shared evaluation protocol that publishes per-artifact and mixed-artifact metrics under identical conditions; and (iv) standardized proxy-task evaluation scripts that assess improvements in motor or cognitive decoding. Collectively, these steps would establish a reproducible foundation for transformer-based denoising research and allow algorithmic improvements to be attributed to genuine model capability rather than dataset or preprocessing variance.

Overall, the reviewed evidence suggests that EEG denoising research stands at an inflection point—conceptually advanced but procedurally immature. Transformer-based models have demonstrated exceptional representational power, yet the absence of shared data standards, consistent metrics, and functional validation prevents cumulative progress. Convergent efforts toward standardized benchmarks, transparent preprocessing, and unified proxy-task assessment will be decisive for transforming current methodological diversity into coherent, reproducible advancement in transformer-driven artifact suppression.

### *4.1. Computational Efficiency and Real-Time Considerations*

While computational efficiency is fundamental for practical online BCIs, most transformer-based MI studies do not yet report standardized metrics of model cost or latency. Across the surveyed works, details such as parameter counts, FLOPs, single-trial inference time, device specifications, and energy per inference are rarely provided, preventing consistent quantitative comparison across studies. In this review, accuracy is therefore treated as an offline metric unless a complete end-to-end latency measurement is available. To improve reproducibility and practical assessment, future publications should report at least the following: parameter count and FLOPs; single-trial inference time with explicit hardware and software configuration; inclusion or exclusion of preprocessing; and closed-loop latency from signal acquisition to feedback, following established timing procedures [57]. A few compact CNN-based pipelines have reported per-window inference times in the sub-millisecond to few-millisecond range on GPUs, though these values typically omit preprocessing and I/O overhead and cannot be directly compared to larger attention-based models [9]. For real-time motor imagery control, total loop delays of approximately 100–300 ms are generally targeted to maintain responsiveness, while emotion-recognition applications often tolerate longer analysis windows, where energy efficiency becomes the primary constraint [5,57,58]. Until such efficiency metrics are systematically and transparently reported, claims of real-time suitability should be interpreted with caution, as latency remains one of the least quantified yet most decisive factors in transformer-based BCI development.

### *4.2. Research Directions and Key Observations*

Transformer-based EEG decoding and denoising have evolved into a technically sophisticated but methodologically fragmented domain. While recent models demonstrate state-of-the-art performance in controlled evaluations, the field still lacks the experimental standardization and translational focus necessary for cumulative progress. The next phase of MI-BCI research must transition from ad hoc innovation toward reproducible, efficiency-aware, and functionally validated frameworks.

Key challenges, open gaps, and future priorities are outlined below:

- **Standardization and reproducibility:** Transformer-based EEG studies vary dramatically in preprocessing pipelines, channel configurations, and evaluation splits. To establish comparability, future research should adopt shared benchmark datasets with fixed train/test partitions, transparent preprocessing scripts, and mandatory reporting of key metrics (accuracy, F1, $\kappa$, CC, RMSE, SNR). A community-agreed benchmarking protocol that integrates both classification and denoising pipelines would enable cumulative progress rather than isolated performance reports.
- **Efficiency and real-time applicability:** The computational footprint of transformer models remains prohibitive for live BCI use. Future architectures should prioritize efficient attention mechanisms (linear, kernelized, or windowed attention), model compression, and adaptive windowing. Inference-time reporting, hardware benchmarking, and real-time latency validation should become standardized requirements for all BCI transformer publications.
- **Pretraining and transferability:** Most EEG transformers are trained from scratch, forfeiting the benefits of pretraining common in NLP and vision. Future work must leverage self-supervised or contrastive pretraining across large, diverse EEG corpora, enabling generalizable embeddings transferable to MI, ERP, or cognitive tasks. Cross-task transfer and domain adaptation frameworks should be benchmarked on fixed cross-subject splits to quantify real generalization rather than overfitting to dataset idiosyncrasies.
- **Automated and adaptive architecture optimization:** Manually designed hybrids are reaching diminishing returns. Recent work using genetic and evolutionary optimization to co-tune CNN–Transformer topologies, attention depth, and preprocessing parameters demonstrates clear gains in cross-subject robustness. Extending this paradigm with multi-objective optimization—balancing accuracy, latency, and stability—can accelerate the discovery of individualized and hardware-aware architectures.
- **Denoising integration and task relevance:** The field of transformer-based EEG denoising is expanding rapidly yet remains disconnected from downstream performance evaluation. Most studies report signal-level metrics (CC, RMSE, SNR) without validating how denoising affects decoding accuracy or information transfer. Future pipelines must adopt proxy-task evaluations, training denoising and classification jointly or sequentially, ensuring that signal cleaning translates to functional improvement in cognitive or motor decoding.
- **Dataset realism and artifact taxonomy:** EEGDenoiseNet remains the de facto benchmark for transformer-based EEG denoising, yet its reliance on synthetic, single-artifact data fundamentally limits ecological validity. A growing portion of recent denoising research is therefore methodologically detached from real-world benefit, as improvements in correlation or error metrics on synthetic signals do not necessarily translate to enhanced decoding in practical BCI tasks. To ensure translational relevance, future studies must couple denoising evaluation with task-driven benchmarks—for instance, retraining or testing MI classifiers on denoised BCIC IV 2a data as a proxy for assessing functional impact. In realistic settings where no artifact-free ground truth exists, performance improvement in downstream decoding (e.g., motor imagery accuracy or kappa) should serve as the principal validation metric. Benchmark updates should thus include real, multi-artifact recordings with synchronized EOG, EMG, and motion channels, standardized artifact taxonomies, and shared denoise–decode baselines to allow a reproducible and functionally meaningful comparison across architectures.
- **Cross-subject and adaptive generalization:** Inter-subject domain shifts continue to degrade performance in both MI classification and denoising. Meta-learning, con-

ditional normalization, and domain-adversarial training offer promising adaptation mechanisms but lack standardized evaluation. Future benchmarks should require subject-agnostic validation, reporting both absolute accuracy and adaptation cost per new user.

- **Integrative and hybrid paradigms:** Diffusion–transformer hybrids, temporal variational models, and cross-modal fusion architectures (EEG–EOG, EEG–fNIRS) are emerging but lack principled justification for their added complexity. Future studies should evaluate hybrid gains using ablation-based efficiency metrics, ensuring that each added mechanism contributes measurable benefit to denoising quality or classification robustness.
- **From offline to closed-loop BCIs:** A critical step toward translation lies in online validation. Offline pipelines must evolve into real-time adaptive systems with latency-aware inference, dynamic feedback, and cross-session persistence. Integrating lightweight transformer variants (e.g., Performer, Longformer) into embedded hardware or wearable platforms will mark the transition from research prototypes to deployable neurotechnologies.

In summary, the next generation of transformer-driven EEG research must converge on standardization, real-time efficiency, and task relevance. Models must be evaluated not only for numerical accuracy but for their contribution to interpretable, reproducible, and deployable BCI systems. By integrating denoising, classification, and adaptation within unified, benchmarked pipelines, transformer-based neurointerfaces can progress from conceptual innovation toward practical, real-world impact.

## 5. Conclusions

Transformer-based MI-BCI research has progressed from proof-of-concept to competitive performance on standard benchmarks with hybrids (e.g., CNN/TCN–Transformer) and hierarchical attention modules delivering state-of-the-art accuracies under subject-dependent settings. Yet our synthesis shows that headline gains still rely disproportionately on offline protocols, non-standard splits, and heterogeneous preprocessing, which together obscure real generalization and deployability. The evidence across Sections 2 and 3 indicates that while top systems can exceed typical baselines by 5–10 percentage points, the median operating regime on IV–2a/2b remains well below extreme claims when session- and subject-aware evaluation is enforced.

A central lesson of this review is that attention improves what models can represent, but reproducible progress depends on how we evaluate and deploy these models. Three priorities emerge. First, protocol discipline: session/subject-preserving splits, transparent preprocessing, and mandatory efficiency reporting (parameters, FLOPs, per-window latency and end-to-end loop delay) must accompany accuracy. Second, task relevance: denoising advances should be validated through denoise $\rightarrow$ decode pipelines (e.g., MI on BCIC IV 2a) when clean ground truth is unavailable, making downstream information transfer the principal success metric. Third, adaptivity at scale: cross-subject robustness improves when models couple inductive biases (compact convolutions for spatial structure) with attention for global context, and when the entire pipeline (filters, windowing, normalization, and network topology) is co-optimized; emerging evolutionary/automated design shows that such co-adaptation yields tangible, protocol-controlled gains over strong baselines.

Looking forward, the path from accuracy to impact requires the following: (i) standardized, multi-dataset benchmarks spanning MI and denoising with shared train/test partitions and proxy-task scripts; (ii) self-supervised and transfer learning across large, heterogeneous EEG corpora to reduce label and calibration burden; (iii) resource-aware architecture search that jointly optimizes accuracy, latency, and stability for embedded

hardware; and (iv) online validation with closed-loop timing to demonstrate sustained performance beyond the lab. With these elements in place, transformer-driven BCIs can transition from sophisticated offline analyzers to deployable neurointerfaces that are efficient, adaptive, and reproducible across users, sessions, and recording conditions.

**Author Contributions:** Conceptualization, M.A.P.; methodology, M.A.P.; software, M.A.P.; validation, M.A.P.; formal analysis, M.A.P.; investigation, S.H.L.; data curation, M.A.P.; writing—original draft preparation, M.A.P.; writing—review and editing, M.A.P. and S.H.L.; supervision, J.K.W.W. and S.H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data of the meta-analysis can be made available upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| BCI | Brain–Computer Interface |
| CNN | Convolutional Neural Network |
| DDIM | Denoising Diffusion Implicit Models |
| DL | Deep Learning |
| EEG | Electroencephalography |
| GAN | Generative Adversarial Network |
| GRU | Gated Recurrent Unit |
| HLT | High-Level Transformer |
| HT | Hierarchical Transformer |
| LLT | Low-Level Transformer |
| LSTM | Long-Short Term Memory |
| MI | Motor Imagery |
| ML | Machine Learning |
| MLP | Multi-Layer Perceptron |
| NLP | Natural Language Processing |
| SNR | Signal-to-Noise Ratio |
| TCN | Temporal Convolution Network |
| XGBoost | eXtreme Gradient Boosting |

## Appendix A

**Table A1.** List of studies as primary sources of data sorted by dataset name (ascending), n_classes (ascending), and reported validation accuracies (descending).

| # | Study Name | Dataset | n | Acc [%] | Year | Country |
|---|---|---|---|---|---|---|
| | *Continued on next page* | | | | | |
| 1 | Multiscale Convolutional Transformer for EEG Classification of Mental Imagery in Different Modalities [33] | Arizona State University Dataset | 2 | 72.0 | 2023 | South Korea |
| 2 | A hybrid network using transformer with modified locally linear embedding and sliding window convolution for EEG decoding [16] | BCI Competition IV 2a | 2 | 84.44 | 2024 | China |
| 3 | A two-stage transformer based network for motor imagery classification [11] | BCI Competition IV 2a | 2 | 88.5 | 2024 | India |
| 4 | Three-stage transfer learning for motor imagery EEG recognition [59] | BCI Competition IV 2a | 2 | 72.24 | 2024 | China |

**Table A1.** *Cont.*

| # | Study Name | Dataset | n | Acc [%] | Year | Country |
|---|---|---|---|---|---|---|
| 5 | Temporal–spatial transformer based motor imagery classification for BCI using independent component analysis [15] | BCI Competition IV 2a | 2 | 88.75 | 2024 | Saudi Arabia |
| 6 | Hierarchical Transformer for Motor Imagery-Based Brain Computer Interface [10] | BCI Competition IV 2a | 2 | 90.0 | 2023 | South Korea |
| 7 | DeepEnsemble: A Novel Brain Wave Classification in MI-BCI using Ensemble of Deep Learners [14] | BCI Competition IV 2a | 2 | 96.07 | 2023 | Canada |
| 8 | EEG classification algorithm of motor imagery based on CNN-Transformer fusion network [13] | BCI Competition IV 2a | 2 | 99.29 | 2022 | China |
| 9 | Compact convolutional transformer for subject-independent motor imagery EEG-based BCIs [12] | BCI Competition IV 2a | 4 | 70.12 | 2024 | Kazakhstan |
| 10 | CTNet: a convolutional transformer network for EEG-based motor imagery classification [19] | BCI Competition IV 2a | 4 | 82.52 | 2024 | China |
| 11 | EEG-VTTCNet: A loss joint training model based on the vision transformer and the temporal convolution network for EEG-based motor imagery classification [17] | BCI Competition IV 2a | 4 | 84.58 | 2024 | China |
| 12 | EEG-TCNTransformer: A Temporal Convolutional Transformer for Motor Imagery Brain–Computer Interfaces [18] | BCI Competition IV 2a | 4 | 83.41 | 2024 | Australia |
| 13 | Swin-CANet: A Novel Integration of Swin Transformer with Channel Attention for Enhanced Motor Imagery Classification [21] | BCI Competition IV 2a | 4 | 78.78 | 2024 | China |
| 14 | BDAN-SPD: A Brain Decoding Adversarial Network Guided by Spatiotemporal Pattern Differences for Cross-Subject MI-BCI | BCI Competition IV 2a | 4 | 77.49 | 2024 | China |
| 15 | Temporal Focal Modulation Networks for EEG-Based Cross-Subject Motor Imagery Classification [60] | BCI Competition IV 2a | 4 | 84.57 | 2024 | Tunisia |
| 16 | MSVTNet: Multi-Scale Vision Transformer Neural Network for EEG-Based Motor Imagery Decoding [31] | BCI Competition IV 2a | 4 | 82.56 | 2024 | China |
| 17 | EEG Motor Imagery Classification using Integrated Transformer-CNN for Assistive Technology Control [22] | BCI Competition IV 2a | 4 | 75.3 | 2024 | United States |
| 18 | SCTrans: Motor Imagery EEG Classification Method based on CNN-Transformer Structure [23] | BCI Competition IV 2a | 4 | 68.61 | 2024 | China |
| 19 | Deep temporal networks for EEG-based motor imagery recognition [61] | BCI Competition IV 2a | 4 | 84.0 | 2024 | India |
| 20 | Classification Algorithm for Electroencephalogram-based Motor Imagery Using Hybrid Neural Network with Spatio-temporal Convolution and Multi-head Attention Mechanism [62] | BCI Competition IV 2a | 4 | 83.3 | 2023 | China |
| 21 | A shallow mirror transformer for subject-independent motor imagery BCI [24] | BCI Competition IV 2a | 4 | 70.41 | 2023 | China |
| 22 | Research on Motor Imagery EEG Classification Method based on Improved Transformer [63] | BCI Competition IV 2a | 4 | 94.24 | 2023 | China |
| 23 | Transformer-Based Network with Optimization for Cross-Subject Motor Imagery Identification [64] | BCI Competition IV 2a | 4 | 63.56 | 2023 | China |
| 24 | A Spatial-Temporal Transformer based on Domain Generalization for Motor Imagery Classification [65] | BCI Competition IV 2a | 4 | 57.705 | 2023 | China |
| 25 | Front-End Replication Dynamic Window (FRDW) for Online Motor Imagery Classification [55] | BCI Competition IV 2a | 4 | 66.51 | 2023 | China |
| 26 | Exploring the Potential of Attention Mechanism-Based Deep Learning for Robust Subject-Independent Motor-Imagery Based BCIs [27] | BCI Competition IV 2a | 4 | 74.73 | 2023 | Kazakhstan |
| 27 | Local and global convolutional transformer-based motor imagery EEG classification [20] | BCI Competition IV 2a | 4 | 80.2 | 2023 | China |
| 28 | Global Adaptive Transformer for Cross-Subject Enhanced EEG Classification [25] | BCI Competition IV 2a | 4 | 76.58 | 2023 | China |
| 29 | A Channel Selection Method for Motor Imagery EEG Based on Fisher Score of OVR-CSP [66] | BCI Competition IV 2a | 4 | 85.54 | 2023 | China |
| 30 | EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization [26] | BCI Competition IV 2a | 4 | 78.66 | 2023 | China |
| 31 | Excellent fine-tuning: From specific-subject classification to cross-task classification for motor imagery [67] | BCI Competition IV 2a | 4 | 86.11 | 2023 | China |
| 32 | A novel hybrid CNN-Transformer model for EEG Motor Imagery classification [68] | BCI Competition IV 2a | 4 | 83.91 | 2022 | China |
| 33 | Three-stage transfer learning for motor imagery EEG recognition [59] | BCI Competition IV 2b | 2 | 69.29 | 2024 | China |
| 34 | ConTraNet: A hybrid network for improving the classification of EEG and EMG signals with limited training data [69] | BCI Competition IV 2b | 2 | 83.61 | 2024 | Germany |
| 35 | Compact convolutional transformer for subject-independent motor imagery EEG-based BCIs [12] | BCI Competition IV 2b | 3 | 70.12 | 2024 | Kazakhstan |
| 36 | CTNet: a convolutional transformer network for EEG-based motor imagery classification [19] | BCI Competition IV 2b | 3 | 76.27 | 2024 | China |

**Table A1.** *Cont.*

| # | Study Name | Dataset | n | Acc [%] | Year | Country |
|---|---|---|---|---|---|---|
| 37 | EEG-VTTCNet: A loss joint training model based on the vision transformer and the temporal convolution network for EEG-based motor imagery classification [17] | BCI Competition IV 2b | 3 | 90.94 | 2024 | China |
| 38 | A two-stage transformer based network for motor imagery classification | BCI Competition IV 2b | 3 | 88.3 | 2024 | India |
| 39 | BDAN-SPD: A Brain Decoding Adversarial Network Guided by Spatiotemporal Pattern Differences for Cross-Subject MI-BCI [70] | BCI Competition IV 2b | 3 | 85.19 | 2024 | China |
| 40 | Temporal Focal Modulation Networks for EEG-Based Cross-Subject Motor Imagery Classification [60] | BCI Competition IV 2b | 3 | 82.22 | 2024 | Tunisia |
| 41 | MSVTNet: Multi-Scale Vision Transformer Neural Network for EEG-Based Motor Imagery Decoding [31] | BCI Competition IV 2b | 3 | 70.3 | 2024 | China |
| 42 | Temporal–spatial transformer based motor imagery classification for BCI using independent component analysis | BCI Competition IV 2b | 3 | 84.2 | 2024 | Saudi Arabia |
| 43 | A shallow mirror transformer for subject-independent motor imagery BCI [24] | BCI Competition IV 2b | 3 | 77.36 | 2023 | China |
| 44 | A Spatial-Temporal Transformer based on Domain Generalization for Motor Imagery Classification [65] | BCI Competition IV 2b | 3 | 75.089 | 2023 | China |
| 45 | Exploring the Potential of Attention Mechanism-Based Deep Learning for Robust Subject-Independent Motor-Imagery Based BCIs [27] | BCI Competition IV 2b | 3 | 72.0 | 2023 | Kazakhstan |
| 46 | Global Adaptive Transformer for Cross-Subject Enhanced EEG Classification [25] | BCI Competition IV 2b | 3 | 92.08 | 2023 | China |
| 47 | EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization [26] | BCI Competition IV 2b | 3 | 84.63 | 2023 | China |
| 48 | Excellent fine-tuning: From specific-subject classification to cross-task classification for motor imagery [67] | BCI Competition IV 2b | 3 | 88.39 | 2023 | China |
| 49 | MI-MBFT: Superior Motor Imagery Decoding of Raw EEG Data Based on a Multibranch and Fusion Transformer Framework [29] | BCI Competition IV 3a | 2 | 94.64 | 2024 | China |
| 50 | Deep temporal networks for EEG-based motor imagery recognition [61] | BCI Competition IV 3a | 2 | 99.7 | 2024 | India |
| 51 | Hierarchical Transformer for Motor Imagery-Based Brain Computer Interface [10] | Cho Dataset | - | 84.6 | 2023 | South Korea |
| 52 | Continual Learning of a Transformer-Based Deep Learning Classifier Using an Initial Model from Action Observation EEG Data to Online Motor Imagery Classification [32] | Custom | 2 | 77.0 | 2023 | Taiwan |
| 53 | A Novel Algorithmic Structure of EEG Channel Attention Combined With Swin Transformer for Motor Patterns Classification [71] | Custom | 2 | 87.67 | 2023 | China |
| 54 | Utilizing the Transformer Architecture Combined with EEGNet to Achieve Real-Time Manipulation of EEG in the Metaverse [34] | Custom | 3 | 71.31 | 2024 | China |
| 55 | Multimodal brain-controlled system for rehabilitation training: Combining asynchronous online brain–computer interface and exoskeleton [35] | Custom | 4 | 91.25 | 2024 | China |
| 56 | Multiscale Convolutional Transformer for EEG Classification of Mental Imagery in Different Modalities [33] | Custom | 4 | 62.0 | 2023 | South Korea |
| 57 | Hierarchical Transformer for Brain Computer Interface | Lee Dataset | 2 | 81.3 | 2023 | South Korea |
| 58 | Hierarchical Transformer for Motor Imagery-Based Brain Computer Interface [10] | Lee Dataset | 2 | 82.1 | 2023 | South Korea |
| 59 | ConTraNet: A hybrid network for improving the classification of EEG and EMG signals with limited training data [69] | Mendeley sEMG | 10 | 77.15 | 2024 | Germany |
| 60 | ConTraNet: A hybrid network for improving the classification of EEG and EMG signals with limited training data [69] | Mendeley sEMG V1 | 7 | 85.0 | 2024 | Germany |
| 61 | BDAN-SPD: A Brain Decoding Adversarial Network Guided by Spatiotemporal Pattern Differences for Cross-Subject MI-BCI [70] | OpenBMI | 2 | 79.37 | 2024 | China |
| 62 | MSVTNet: Multi-Scale Vision Transformer Neural Network for EEG-Based Motor Imagery Decoding [31] | OpenBMI | 2 | 75.93 | 2024 | China |
| 63 | SCTrans: Motor Imagery EEG Classification Method based on CNN-Transformer Structure [23] | OpenBMI | 2 | 73.33 | 2024 | China |
| 64 | A shallow mirror transformer for subject-independent motor imagery BCI [24] | OpenBMI | 2 | 80.54 | 2023 | China |
| 65 | Local and global convolutional transformer-based motor imagery EEG classification [20] | OpenBMI | 2 | 81.04 | 2023 | China |

**Table A1.** *Cont.*

| # | Study Name | Dataset | n | Acc [%] | Year | Country |
|---|---|---|---|---|---|---|
| 66 | MI-MBFT: Superior Motor Imagery Decoding of Raw EEG Data Based on a Multibranch and Fusion Transformer Framework [29] | PhysioNet MI | 2 | 84.07 | 2024 | China |
| 67 | Study of an Optimization Tool Avoided Bias for Brain-Computer Interfaces Using a Hybrid Deep Learning Model [30] | PhysioNet MI | 2 | 74.54 | 2024 | Spain |
| 68 | Hierarchical Transformer for Motor Imagery-Based Brain Computer Interface [10] | PhysioNet MI | 2 | 83.5 | 2023 | South Korea |
| 69 | A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification [28] | PhysioNet MI | 2 | 83.31 | 2022 | China |
| 70 | ConTraNet: A hybrid network for improving the classification of EEG and EMG signals with limited training data [69] | PhysioNet MI | 3 | 74.38 | 2024 | Germany |
| 71 | Exploring the Potential of Attention Mechanism-Based Deep Learning for Robust Subject-Independent Motor-Imagery Based BCIs [27] | PhysioNet MI | 3 | 86.47 | 2023 | Kazakhstan |
| 72 | A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification [28] | PhysioNet MI | 3 | 74.44 | 2022 | China |
| 73 | A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification [28] | PhysioNet MI | 4 | 64.22 | 2022 | China |
| 74 | Motor Imagery and Mental Arithmetic Classification Based on Transformer Deep Learning Network [72] | Shin, Blankertz | 2 | 88.67 | 2024 | China |
| 75 | Classification of EEG signals based on CNN-Transformer model [73] | Shin, Blankertz | 2 | 87.23 | 2023 | China |

## References


1. Burnham, J.F. Scopus database: A review. *Biomed. Digit. Libr.* **2006**, *3*, 1. [CrossRef] [PubMed]
2. Fantozzi, P.; Naldi, M. The explainability of transformers: Current status and directions. *Computers* **2024**, *13*, 92. [CrossRef]
3. Vidyasagar, K.C.; Kumar, K.R.; Sai, G.A.; Ruchita, M.; Saikia, M.J. Signal to image conversion and convolutional neural networks for physiological signal processing: A review. *IEEE Access* **2024**, *12*, 66726–66764. [CrossRef]
4. Su, L.; Zuo, X.; Li, R.; Wang, X.; Zhao, H.; Huang, B. A systematic review for transformer-based long-term series forecasting. *Artif. Intell. Rev.* **2025**, *58*, 80. [CrossRef]
5. Pfeffer, M.A.; Ling, S.S.H.; Wong, J.K.W. Exploring the Frontier: Transformer-Based Models in EEG Signal Analysis for Brain-Computer Interfaces. *Comput. Biol. Med.* **2024**, *178*, 108705. [CrossRef]
6. BCI Competition IV. 2008. Available online: http://www.bbci.de/competition/iv (accessed on 2 September 2025).
7. Schalk, G.; McFarland, D.J.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1034–1043. [CrossRef]
8. Yin, J.; Liu, A.; Wang, L.; Qian, R.; Chen, X. Integrating spatial and temporal features for enhanced artifact removal in multi-channel EEG recordings. *J. Neural Eng.* **2024**, *21*, 056018. [CrossRef]
9. Ouahidi, Y.E.; Mohammadi, P. A Strong and Simple Deep Learning Baseline for BCI Classification from EEG. *arXiv* **2023**, arXiv:2309.07159.
10. Deny, P.; Cheon, S.; Son, H.; Choi, K.W. Hierarchical transformer for motor imagery-based brain computer interface. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 5459–5470. [CrossRef]
11. Chaudhary, P.; Dhankhar, N.; Singhal, A.; Rana, K. A two-stage transformer based network for motor imagery classification. *Med. Eng. Phys.* **2024**, *128*, 104154. [CrossRef]
12. Keutayeva, A.; Fakhrutdinov, N.; Abibullaev, B. Compact convolutional transformer for subject-independent motor imagery EEG-based BCIs. *Sci. Rep.* **2024**, *14*, 25775. [CrossRef] [PubMed]
13. Liu, H.; Liu, Y.; Wang, Y.; Liu, B.; Bao, X. EEG classification algorithm of motor imagery based on CNN-Transformer fusion network. In Proceedings of the 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Wuhan, China, 9–11 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1302–1309.
14. Mehtiyev, A.; Al-Najjar, A.; Sadreazami, H.; Amini, M. Deepensemble: A novel brain wave classification in MI-BCI using ensemble of deep learners. In Proceedings of the 2023 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 6–8 January 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
15. Hameed, A.; Fourati, R.; Ammar, B.; Ksibi, A.; Alluhaidan, A.S.; Ayed, M.B.; Khleaf, H.K. Temporal–spatial transformer based motor imagery classification for BCI using independent component analysis. *Biomed. Signal Process. Control* **2024**, *87*, 105359. [CrossRef]

16. Li, K.; Chen, P.; Chen, Q.; Li, X. A hybrid network using transformer with modified locally linear embedding and sliding window convolution for EEG decoding. *J. Neural Eng.* **2024**, *21*, 066049. [CrossRef] [PubMed]
17. Shi, X.; Li, B.; Wang, W.; Qin, Y.; Wang, H.; Wang, X. EEG-VTTCNet: A loss joint training model based on the vision transformer and the temporal convolution network for EEG-based motor imagery classification. *Neuroscience* **2024**, *556*, 42–51. [CrossRef]
18. Nguyen, A.H.P.; Oyefisayo, O.; Pfeffer, M.A.; Ling, S.H. EEG-TCNTransformer: A Temporal Convolutional Transformer for Motor Imagery Brain–Computer Interfaces. *Signals* **2024**, *5*, 605–632. [CrossRef]
19. Zhao, W.; Jiang, X.; Zhang, B.; Xiao, S.; Weng, S. CTNet: A convolutional transformer network for EEG-based motor imagery classification. *Sci. Rep.* **2024**, *14*, 20237. [CrossRef]
20. Zhang, J.; Li, K.; Yang, B.; Han, X. Local and global convolutional transformer-based motor imagery EEG classification. *Front. Neurosci.* **2023**, *17*, 1219988. [CrossRef]
21. Shi, Y.; Wang, M. Swin-CANet: A Novel Integration of Swin Transformer with Channel Attention for Enhanced Motor Imagery Classification. In Proceedings of the 2024 IEEE 4th International Conference on Software Engineering and Artificial Intelligence (SEAI), Xiamen, China, 21–23 June 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 57–62.
22. Zare, S.; Sun, Y. EEG Motor Imagery Classification using Integrated Transformer-CNN for Assistive Technology Control. In Proceedings of the 2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Wilmington, DE, USA, 19–21 June 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 189–190.
23. Sun, B.; Wang, Q.; Li, S.; Deng, Q. SCTrans: Motor Imagery EEG Classification Method based on CNN-Transformer Structure. In Proceedings of the 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, 29–31 March 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 2001–2004.
24. Luo, J.; Wang, Y.; Xia, S.; Lu, N.; Ren, X.; Shi, Z.; Hei, X. A shallow mirror transformer for subject-independent motor imagery BCI. *Comput. Biol. Med.* **2023**, *164*, 107254. [CrossRef]
25. Song, Y.; Zheng, Q.; Wang, Q.; Gao, X.; Heng, P.A. Global adaptive transformer for cross-subject enhanced EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 2767–2777. [CrossRef]
26. Song, Y.; Zheng, Q.; Liu, B.; Gao, X. EEG conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *31*, 710–719. [CrossRef]
27. Keutayeva, A.; Abibullaev, B. Exploring the potential of attention mechanism-based deep learning for robust subject-independent motor-imagery based BCIs. *IEEE Access* **2023**, *11*, 107562–107580. [CrossRef]
28. Xie, J.; Zhang, J.; Sun, J.; Ma, Z.; Qin, L.; Li, G.; Zhou, H.; Zhan, Y. A transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2126–2136. [CrossRef] [PubMed]
29. Luo, J.; Cheng, Q.; Wang, H.; Du, Q.; Wang, Y.; Li, Y. MI-MBFT: Superior Motor Imagery Decoding of Raw EEG Data Based on a Multi-Branch and Fusion Transformer Framework. *IEEE Sens. J.* **2024**, *24*, 34879–34891. [CrossRef]
30. Ajali-Hernández, N.I.; Travieso-González, C.M.; Bermudo-Mora, N.; Reino-Cacho, P.; Rodríguez-Saucedo, S. Study of an Optimization Tool Avoided Bias for Brain-Computer Interfaces Using a Hybrid Deep Learning Model. *IRBM* **2024**, *45*, 100836. [CrossRef]
31. Liu, K.; Yang, T.; Yu, Z.; Yi, W.; Yu, H.; Wang, G.; Wu, W. MSVTNet: Multi-Scale Vision Transformer Neural Network for EEG-Based Motor Imagery Decoding. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 7126–7137. [CrossRef]
32. Lee, P.L.; Chen, S.H.; Chang, T.C.; Lee, W.K.; Hsu, H.T.; Chang, H.H. Continual learning of a transformer-based deep learning classifier using an initial model from action observation EEG data to online motor imagery classification. *Bioengineering* **2023**, *10*, 186. [CrossRef]
33. Ahn, H.J.; Lee, D.H.; Jeong, J.H.; Lee, S.W. Multiscale convolutional transformer for EEG classification of mental imagery in different modalities. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *31*, 646–656. [CrossRef]
34. Li, P.L.; Yuan, J.J. Utilizing the Transformer Architecture Combined with EEGNet to Achieve Real-Time Manipulation of EEG in the Metaverse. In Proceedings of the 2024 International Conference on System Science and Engineering (ICSSE), Hsinchu, Taiwan, 26–28 June 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–8.
35. Liu, L.; Li, J.; Ouyang, R.; Zhou, D.; Fan, C.; Liang, W.; Li, F.; Lv, Z.; Wu, X. Multimodal brain-controlled system for rehabilitation training: Combining asynchronous online brain–computer interface and exoskeleton. *J. Neurosci. Methods* **2024**, *406*, 110132. [CrossRef]
36. Hossain, K.M.; Islam, M.A.; Hossain, S.; Nijholt, A.; Ahad, M.A.R. Status of deep learning for EEG-based brain–computer interface applications. *Front. Comput. Neurosci.* **2023**, *16*, 1006763. [CrossRef]
37. Elashmawi, W.H.; Ayman, A.; Antoun, M.; Mohamed, H.; Mohamed, S.E.; Amr, H.; Talaat, Y.; Ali, A. A Comprehensive Review on Brain–Computer Interface (BCI)-Based Machine and Deep Learning Algorithms for Stroke Rehabilitation. *Appl. Sci.* **2024**, *14*, 6347. [CrossRef]
38. Khademi, Z.; Ebrahimi, F.; Kordy, H.M. A review of critical challenges in MI-BCI: From conventional to deep learning methods. *J. Neurosci. Methods* **2023**, *383*, 109736. [CrossRef] [PubMed]

39. Brookshire, G.; Kasper, J.; Blauch, N.M.; Wu, Y.C.; Glatt, R.; Merrill, D.A.; Gerrol, S.; Yoder, K.J.; Quirk, C.; Lucero, C. Data leakage in deep learning studies of translational EEG. *Front. Neurosci.* **2024**, *18*, 1373515. [CrossRef] [PubMed]
40. Varoquaux, G. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage* **2017**, *145*, 166–179. [CrossRef]
41. Brunner, C.; Leeb, R.; Müller-Putz, G.P.; Schlögl, A.; Pfurtscheller, G. *BCI Competition 2008—Graz Data Set A (Data Set 2a)*; Technical report; Graz University of Technology: Graz, Austria, 2008.
42. Tangermann, M.; Müller, K.R.; Aertsen, A.; Birbaumer, N.; Braun, C.; Brunner, C.; Leeb, R.; Mehring, C.; Miller, K.J.; Müller-Putz, G.R.; et al. Review of the BCI Competition IV. *Front. Neurosci.* **2012**, *6*, 55. [CrossRef]
43. Zhang, C.; Liu, Y.; Wu, X. TFANet: A temporal fusion attention neural network for motor imagery decoding. *Front. Neurosci.* **2025**, *19*, 1635588. [CrossRef]
44. Liao, W.; Miao, Z.; Liang, S.; Zhang, L.; Li, C. A composite improved attention convolutional network for motor imagery EEG classification. *Front. Neurosci.* **2025**, *19*, 1543508. [CrossRef]
45. Zhao, W.; Zhang, B.; Zhou, H.; Wei, D.; Huang, C.; Lan, Q. Multi-scale convolutional transformer network for motor imagery brain–computer interface. *Sci. Rep.* **2025**, *15*, 96611. [CrossRef]
46. Liao, W.; Liu, H.; Wang, W. Advancing BCI with a transformer-based model for motor imagery decoding. *Sci. Rep.* **2025**, *15*, 06364. [CrossRef]
47. Song, J.; Zhai, Q.; Wang, C.; Liu, J. EEGGAN-Net: Enhancing EEG signal classification through data augmentation based on generative adversarial networks. *Front. Hum. Neurosci.* **2024**, *18*, 1430086. [CrossRef]
48. Huang, X.; Li, C.; Liu, A.; Qian, R.; Chen, X. EEGDfus: A conditional diffusion model for fine-grained EEG denoising. *IEEE J. Biomed. Health Inform.* **2024**, *29*, 2557–2569. [CrossRef]
49. Bellamkonda, N.L.; Goru, H.K.; Solasuttu, B.; Gangu, V.R. A Hybrid Residual CNN and Multi-Head Self-Attention Network for Denoising Muscle Artifacts in EEG Signals. In Proceedings of the 2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 9–11 July 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 21–27.
50. Tiwari, N.; Anwar, S. BiGRU-TFA: An Attention-Enhanced Model for EEG Signal Reconstruction Using Temporal and Frequency Features. *IEEE Sens. J.* **2025**, *25*, 27077–27085. [CrossRef]
51. Pu, X.; Yi, P.; Chen, K.; Ma, Z.; Zhao, D.; Ren, Y. EEGDnet: Fusing non-local and local self-similarity for EEG signal denoising with transformer. *Comput. Biol. Med.* **2022**, *151*, 106248. [CrossRef]
52. Wang, W.; Li, B.; Wang, H. A novel end-to-end network based on a bidirectional GRU and a self-attention mechanism for denoising of electroencephalography signals. *Neuroscience* **2022**, *505*, 10–20. [CrossRef]
53. Chen, J.; Pi, D.; Jiang, X.; Gao, F.; Wang, B.; Chen, Y. EEGCiD: EEG Condensation Into Diffusion Model. *IEEE Trans. Autom. Sci. Eng.* **2024**, *22*, 8502–8518. [CrossRef]
54. Alzahab, N.A.; Alshawa, N. Automatic Reconstruction of Noisy Electroencephalography (EEG) Channels with Transformer-Based Architectures for Sustainable Systems. In Proceedings of the 2024 10th International Conference on Computing, Engineering and Design (ICCED), Jeddah, Saudi Arabia, 11–12 December 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
55. Chen, X.; An, J.; Wu, H.; Li, S.; Liu, B.; Wu, D. Front-end Replication Dynamic Window (FRDW) for Online Motor Imagery Classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 3906–3914. [CrossRef]
56. Pfeffer, M.A.; Nguyen, A.H.P.; Kim, K.; Wong, J.K.W.; Ling, S.H. Evolving optimized transformer-hybrid systems for robust BCI signal processing using genetic algorithms. *Biomed. Signal Process. Control* **2025**, *108*, 107883. [CrossRef]
57. Wilson, J.A.; Mellinger, J.; Schalk, G.; Williams, J. A Procedure for Measuring Latencies in Brain–Computer Interfaces. *Front. Neurosci.* **2010**, *4*, 306. [CrossRef] [PubMed]
58. LaRocco, J.; Le, M.; Paeng, D.H. Optimizing Computer–Brain Interface Parameters for Non-Invasive Applications. *Front. Neuroinformatics* **2020**, *14*, 1. [CrossRef]
59. Li, J.; She, Q.; Meng, M.; Du, S.; Zhang, Y. Three-stage transfer learning for motor imagery EEG recognition. *Med. Biol. Eng. Comput.* **2024**, *62*, 1689–1701. [CrossRef]
60. Hameed, A.; Fourati, R.; Ammar, B.; Sanchez-Medina, J.; Ltifi, H. Temporal Focal Modulation Networks for EEG-Based Cross-Subject Motor Imagery Classification. In Proceedings of the International Conference on Computational Collective Intelligence, Leipzig, Germany, 9–11 September 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 445–457.
61. Sharma, N.; Upadhyay, A.; Sharma, M.; Singhal, A. Deep temporal networks for EEG-based motor imagery recognition. *Sci. Rep.* **2023**, *13*, 18813. [CrossRef]
62. Shi, X.; Li, B.; Wang, W.; Qin, Y.; Wang, H.; Wang, X. Classification algorithm for electroencephalogram-based motor imagery using hybrid neural network with spatio-temporal convolution and multi-head attention mechanism. *Neuroscience* **2023**, *527*, 64–73. [CrossRef] [PubMed]
63. Liu, Y.; Liu, Z.; Huang, L. Research on Motor Imagery EEG Classification Method based on Improved Transformer. In Proceedings of the Fifth International Conference on Image Processing and Intelligent Control (IPIC 2025), Qingdao, China, 9–11 May 2025; Volume 13782, pp. 195–201.

64. Tan, X.; Wang, D.; Chen, J.; Xu, M. Transformer-based network with optimization for cross-subject motor imagery identification. *Bioengineering* **2023**, *10*, 609. [CrossRef] [PubMed]
65. Liu, S.; An, L.; Zhang, C.; Jia, Z. A spatial-temporal transformer based on domain generalization for motor imagery classification. In Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Honolulu, HI, USA, 1–4 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 3789–3794.
66. Mu, W.; Wang, J.; Wang, L.; Wang, P.; Han, J.; Niu, L.; Bin, J.; Liu, L.; Zhang, J.; Jia, J.; et al. A channel selection method for motor imagery EEG based on Fisher score of OVR-CSP. In Proceedings of the 2023 11th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Republic of Korea, 20–22 February 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–4.
67. Jia, X.; Song, Y.; Xie, L. Excellent fine-tuning: From specific-subject classification to cross-task classification for motor imagery. *Biomed. Signal Process. Control* **2023**, *79*, 104051. [CrossRef]
68. Ma, Y.; Song, Y.; Gao, F. A novel hybrid CNN-transformer model for EEG motor imagery classification. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–8.
69. Ali, O.; Saif-ur Rehman, M.; Glasmachers, T.; Iossifidis, I.; Klaes, C. ConTraNet: A hybrid network for improving the classification of EEG and EMG signals with limited training data. *Comput. Biol. Med.* **2024**, *168*, 107649. [CrossRef]
70. Wei, F.; Xu, X.; Li, X.; Wu, X. BDAN-SPD: A brain decoding adversarial network guided by spatiotemporal pattern differences for cross-subject MI-BCI. *IEEE Trans. Ind. Inform.* **2024**, *20*, 14321–14329. [CrossRef]
71. Wang, H.; Cao, L.; Huang, C.; Jia, J.; Dong, Y.; Fan, C.; De Albuquerque, V.H.C. A novel algorithmic structure of EEG Channel Attention combined with Swin Transformer for motor patterns classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 3132–3141. [CrossRef]
72. Ye, Y.; Tong, J.; Yang, S.; Chang, Y.; Du, S. Motor Imagery and Mental Arithmetic Classification Based on Transformer Deep Learning Network. In Proceedings of the 2024 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 357–362.
73. Liu, J.; Dong, E.; Tong, J.; Yang, S.; Du, S. Classification of EEG signals based on CNN-Transformer model. In Proceedings of the 2023 IEEE International Conference on Mechatronics and Automation (ICMA), Harbin, China, 6–9 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 2095–2099.