International Journal of
Network Dynamics and Intelligence (IJNDI)

Scilight

Article

Check for updates

Fed-VQCS: Vector-Quantized Content-Style State-Space Disentanglement for Federated EEG Generalization

Jun Li [1], Jiping Cui [1], Rui Yang [1,2,*], Hanjing Cheng [3], and Jinling Liang [4]

[1] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
[2] State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing 210093, China
[3] School of Electronic & Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China
[4] School of Mathematics, Southeast University, Nanjing 210096, China
* Correspondence: R.Yang@xjtlu.edu.cn
Received: 30 July 2026; Revised: 31 August 2026; Accepted: 2 November 2026; Published: 9 November 2026

Abstract: Federated learning provides a natural framework for decentralized EEG analysis, but learning shared and transferable representations remains challenging because task-relevant neural patterns are often entangled with subject-dependent characteristics in local EEG data. Existing federated EEG methods mainly focus on backbone design, aggregation strategies, or personalized optimization, while representation-level control of what information is globally shared remains insufficiently explored. In this work, we propose Fed-VQCS, a vector-quantized content-style state-space disentanglement framework for robust federated EEG learning. Fed-VQCS introduces a plug-and-play Vector-Quantized Content-Style State-Space (VQCS) module that can be inserted after different EEG feature extractors. Given a latent EEG representation, a bidirectional state-space semantic extractor first produces a continuous task-semantic representation, which is further mapped to globally shared discrete content prototypes through vector quantization. A complementary residual pathway captures variation outside the continuous semantic estimate and is adversarially constrained to suppress task-discriminative information. During federated training, the shared model parameters and content codebook are aggregated globally, while the residual pathway is evaluated within each client's local training pass. Experiments on BCI Competition IV-2a and PhysioNet, repeated over five random seeds, show consistent improvements in accuracy and Cohen's κ across four EEG backbones. Comparisons with representative federated learning methods, sensitivity analyses, ablation studies, low-data adaptation, and representation probing further support the effectiveness and robustness of the proposed framework under heterogeneous subject distributions.

Keywords: brain–computer interface; electroencephalography; federated learning; representation disentanglement; cross-subject generalization

1. Introduction

Brain–computer interfaces (BCIs) based on electroencephalography (EEG) have attracted increasing attention in motor imagery decoding, cognitive monitoring, neurorehabilitation, and assistive interaction [1–3]. Recent advances in deep learning have significantly improved EEG decoding by learning discriminative neural representations from data [4–6]. However, EEG data are often distributed across different subjects, devices, hospitals, or recording environments [7]. Directly centralizing such data is difficult due to privacy concerns, data ownership restrictions, and practical limitations in cross-institutional data sharing [8–12].

Federated learning (FL) provides a promising paradigm for distributed EEG modeling, as it enables multiple clients to collaboratively train a global model without directly sharing raw EEG signals [8,13–15]. In a typical federated EEG scenario, each client may correspond to an individual subject, a local device, or an institution. Clients update their models locally and communicate model-related information with the server for global aggregation. This training paradigm is naturally suitable for privacy-sensitive EEG applications and large-scale BCI deployment [16–18].

Despite its potential, federated EEG learning still suffers from severe performance degradation under heterogeneous client distributions. A major reason is inter-subject variability, which is intrinsic to EEG signals. Neural responses may vary significantly across individuals due to differences in physiology, brain activation patterns, electrode placement, signal quality, and recording conditions, making robust learning important under noisy or outlier-contaminated observations [19,20]. As a result, local EEG data collected from different clients are usually not independently and identically distributed (non-IID). When standard FL algorithms aggregate local models trained on such subject-dependent data, the global model can be biased by client-associated patterns and may generalize poorly under heterogeneous subject distributions.

Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license(https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

We argue that a key factor behind this problem is the representation-level entanglement between task-relevant neural patterns and subject-associated variations. Ideally, EEG representations used for global learning should emphasize transferable content information related to the target task, such as motor imagery intention or cognitive state, while reducing sensitivity to client-associated nuisance factors. In practice, however, features learned by conventional EEG models may mix task-discriminative information with variations arising from individual physiology, electrode placement, recording conditions, session effects, and signal quality. When models are optimized from such entangled representations, client-associated variation can influence global parameter updates and produce fragmented feature distributions or inconsistent decision boundaries across clients. Therefore, improving federated EEG generalization requires not only better aggregation strategies, but also mechanisms that explicitly control the information represented by the shared semantic pathway [21].

Existing studies have attempted to address client heterogeneity in federated EEG learning through stronger EEG backbones, robust aggregation rules, personalized federated optimization, normalization strategies, or domain adaptation techniques [22, 23]. Although these methods can alleviate non-IID effects, many primarily modify optimization, normalization, or personalization behavior rather than explicitly constraining the information carried by globally shared latent representations. The problem of how to concentrate transferable task information in a shared semantic pathway while relegating complementary client-associated variation to a non-shared residual pathway remains insufficiently explored. This motivates us to investigate a representation disentanglement strategy for federated EEG generalization [24].

In this work, we propose **Fed-VQCS**, a vector-quantized content-style state-space disentanglement framework for robust federated EEG learning. Instead of designing a task-specific EEG backbone, Fed-VQCS introduces a plug-and-play **Vector-Quantized Content-Style State-Space (VQCS)** module that can be inserted after different EEG feature extractors. Given latent EEG features produced by a backbone network, the VQCS module first learns a continuous task-semantic representation and then quantizes it into discrete content prototypes. A complementary residual representation is defined relative to the continuous semantic estimate and is trained to capture variation outside the task-semantic pathway. We refer to this residual as "style", encompassing subject-associated and acquisition-related variation that is complementary to the shared task content.

Specifically, the continuous dynamics branch adopts bidirectional state-space modeling to capture temporally meaningful task semantics. The vector-quantized content branch imposes a discrete information bottleneck by mapping the continuous semantic representation into compact prototypes shared through federated optimization. The residual pathway is computed before quantization, preventing codebook approximation error from being directly absorbed into the residual definition, and an adversarial objective further suppresses task-discriminative information in that pathway. During federated training, the backbone, Bi-Mamba module, codebook, and shared classifier are globally optimized, while the residual pathway provides local regularization for representation learning.

Figure 1 illustrates the motivation of the proposed framework. Conventional federated EEG learning can yield entangled representations in which task-relevant information coexists with strong client-associated variation. Fed-VQCS instead concentrates task-relevant semantics in a quantized shared content space while assigning complementary non-task variation to a residual pathway. This design promotes more consistent cross-client feature alignment and improves robustness under heterogeneous subject-as-client distributions.

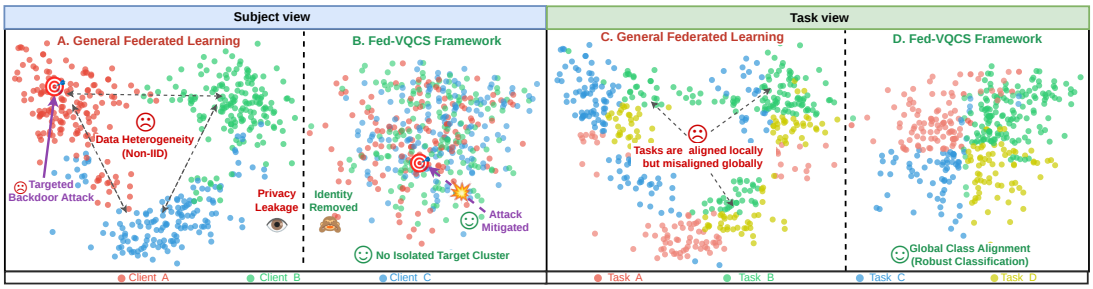


Figure 1. Motivation of Fed-VQCS from subject and task views. Conventional federated EEG learning may preserve strongly subject-associated clusters under non-IID data, while Fed-VQCS reduces subject-associated variation in the shared content representation through a complementary residual pathway. From the task view, conventional FL may align classes locally but remain globally misaligned across clients, whereas Fed-VQCS learns shared content prototypes to promote more consistent global class alignment and robust federated EEG classification.

The main contributions of this work are summarized as follows:

- We identify representation-level entanglement between transferable task semantics and client-associated

nuisance variation as an important factor limiting robust federated EEG learning, and formulate a shared-content/complementary-residual view of subject heterogeneity.
- We propose Fed-VQCS, a plug-and-play vector-quantized content-style state-space framework that combines bidirectional semantic modeling, a discrete content bottleneck, and adversarial suppression of task information in the residual pathway.
- We validate Fed-VQCS on BCI Competition IV-2a and PhysioNet across four EEG backbones and five random seeds, and further evaluate it against representative federated learning strategies using statistical significance tests, sensitivity analysis, ablation studies, low-data adaptation, and task/subject probing.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed Fed-VQCS framework. Section 4 reports experimental settings and results. Section 5 concludes this paper.

2. Related Work

The rapid development of deep learning has significantly transformed EEG decoding, enabling models to extract complex spatiotemporal patterns from non-invasive recordings. Early compact architectures such as EEGNet introduced efficient temporal–spatial filtering strategies, demonstrating competitive performance across various BCI paradigms [25]. Building on this foundation, task-specific models emerged to address particular signal properties. MI-EEGNet adapted convolutional layers to motor imagery tasks to better capture sensorimotor rhythms [26], while TSception exploited multi-scale temporal kernels and asymmetric spatial filters to enhance emotion recognition and cross-dataset robustness [27]. More recently, transformer-based architectures have been applied to EEG to model long-range dependencies and multi-channel interactions. These approaches, including AETF and related encoder designs, achieved notable improvements in SSVEP and MI decoding under controlled conditions [28]. In addition, sequence modeling techniques such as state-space models provide an efficient alternative for capturing long-range temporal dependencies in neural signals [29]. Despite this progress, most existing EEG backbones are designed under centralized or controlled acquisition settings, and their learned representations may still entangle task-relevant neural patterns with subject-specific variations, limiting their applicability to real-world, multi-site BCI deployments.

To address the limitations of centralized training, federated learning (FL) has emerged as a natural solution for decentralized EEG analysis [30]. In early explorations, privacy-preserving FL frameworks demonstrated the feasibility of collaborative training across sleep centers without sharing raw signals [31]. Subsequent studies extended this direction with semi-supervised knowledge sharing to reduce manual labeling effort [32], and edge-oriented frameworks that leveraged meta-learning to accelerate convergence on low-resource devices [33]. These works represent important steps toward practical federated BCI systems. However, they also exposed fundamental trade-offs between accuracy, communication cost, and robustness, especially under non-IID distributions where clients may correspond to different subjects, devices, recording protocols, or institutions.

As federated EEG learning scaled to more realistic scenarios, a second wave of work began to explore richer training strategies for handling client heterogeneity. Hierarchical FL introduced multi-level aggregation for drowsiness detection with EEG/EOG, improving performance across organizations but increasing orchestration complexity [34]. Personalized federated frameworks such as FLEEG combined global encoders with local adaptation to address device and subject heterogeneity, achieving improved accuracy under heterogeneous settings [35]. These developments highlight both the maturity and fragility of current EEG–FL solutions: they can improve local adaptation or aggregation robustness, but the shared model updates are still commonly learned from entangled representations in which task-relevant and subject-specific information may be mixed.

Recent research has therefore shifted attention from aggregation schemes to representation learning itself, aiming to control what information is shared across clients [36]. Hardware-assisted split learning explored transmitting intermediate activations rather than raw signals to reduce exposure, introducing explicit privacy metrics to quantify residual leakage [37]. Parallel to this, multi-source decentralized transfer learning demonstrated that subjects can collaborate through model predictions rather than raw data, reducing central dependencies and improving privacy [38]. These directions mark an important shift from surface-level defenses to architectural approaches. Nevertheless, they still lack explicit mechanisms for separating transferable task information from client-specific variations in the latent representation space.

Another important challenge arises from how heterogeneous client distributions interact with optimization dynamics. Studies on batch normalization in FL revealed that inconsistent statistics across clients severely degrade convergence, and proposed FIXBN to stabilize training under non-IID conditions [39]. Hybrid normalization approaches, such as HBN, further blended local and global statistics to improve robustness with minimal overhead [40]. Meanwhile, test-time personalization strategies allowed models to adapt to client-specific distributions without labeled target data,

significantly improving out-of-domain generalization [41]. Complementary evidence from centralized baselines such as U-Sleep shows that even highly accurate models degrade under cross-dataset mismatch, underscoring the need for federated methods that explicitly address distribution shifts at the representation level [42,43].

Prototype-based and discretized representation learning provides another promising perspective for improving representation stability under heterogeneous distributions. By mapping continuous features into compact codebook entries or semantic prototypes, vector quantization can impose a discrete information bottleneck and encourage representations to capture more stable and transferable factors [44,45]. This property is particularly relevant to federated EEG learning, where globally shared representations should focus on task-relevant neural content while suppressing subject-dependent variations. However, existing federated EEG studies rarely exploit vector-quantized prototypes to explicitly define a shared content space, nor do they combine discrete content modeling with local residual representations for subject-specific variability.

Taken together, these advances trace a clear trajectory. Initial work focused on centralized deep models that established strong EEG decoding baselines; FL then enabled collaborative learning but introduced new heterogeneity, robustness, and privacy-related concerns; more recent studies began probing how the structure of shared representations affects these problems. However, representation-level control of transferable task content versus client-associated nuisance variation remains insufficiently explored in federated EEG learning. Fed-VQCS addresses this gap by mapping a continuous task-semantic representation into globally shared vector-quantized content prototypes and using a complementary residual as a locally evaluated regularization pathway that discourages task information from remaining outside the shared content representation.

3. Methodology

This section presents Fed-VQCS, a vector-quantized content-style disentanglement framework for federated EEG decoding. The key idea is to learn a globally shareable task-semantic pathway while constraining complementary residual variation so that task-discriminative information is preferentially concentrated in the shared content representation. Given an EEG trial, Fed-VQCS first extracts a latent EEG representation using a backbone encoder and then employs a Bi-Mamba semantic extractor to model bidirectional temporal dependencies and produce a continuous semantic representation. A vector-quantized layer maps this representation to shared discrete content prototypes. The residual pathway is defined with respect to the continuous semantic estimate before quantization and is trained as a complementary style representation. Here, "style" denotes subject-associated and acquisition-related variation outside the task-semantic estimate. As illustrated in Figure 2, Fed-VQCS contains a backbone EEG encoder, a Bi-Mamba semantic dynamics branch, a vector-quantized content branch, and a complementary residual pathway. The shared classifier imposes task supervision on the semantic representations and adversarially suppresses task-discriminative information in the residual. The main notations used throughout this section are summarized in Table 1.

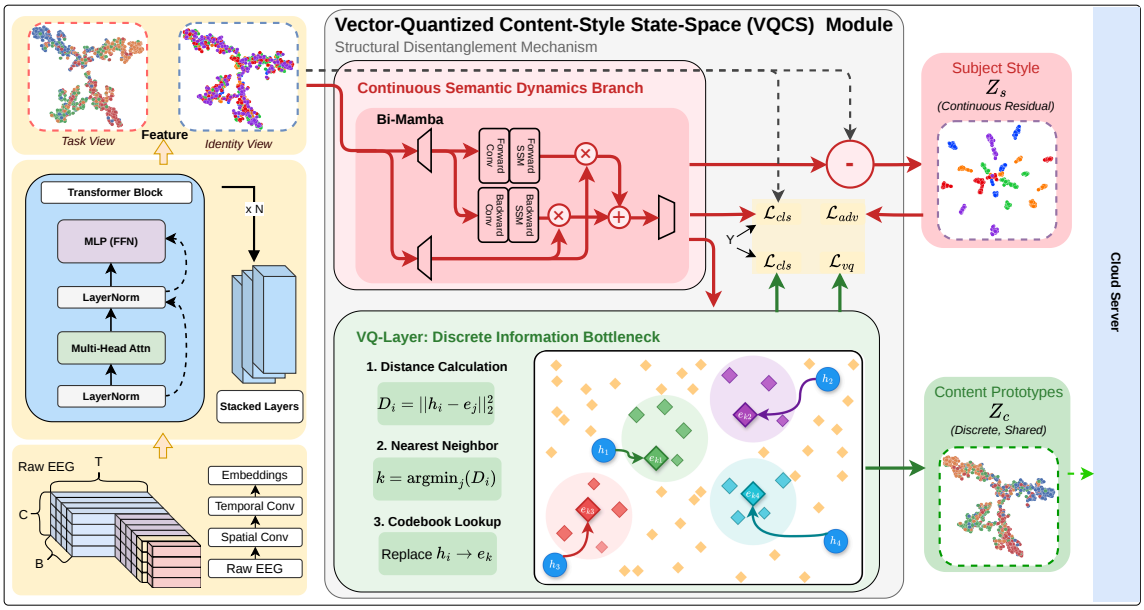


Figure 2. Overview of the proposed Fed-VQCS framework. The VQCS module maps the latent EEG representation H_k to shared vector-quantized content prototypes Z_c and a complementary residual representation Z_s. During federated training, the shared model parameters and content codebook are optimized through global aggregation, while Z_s is computed and constrained within each client's local forward pass.

Table 1. Main notations used in the methodology.

Notation	Description
a_t	Assigned codebook index.
d	Latent feature dimension.
$\mathcal{D}_k$	Local dataset of client k.
$\mathcal{E}$	Learnable content codebook.
e_m	The m-th content prototype.
$g(\cdot)$	Shared classifier.
H	Backbone latent EEG representation.
K, k	Number of clients and client index.
K_c	Number of codebook prototypes.
$\mathcal{L}_{\text{style}}$	Adversarial style loss.
$\mathcal{L}_{\text{task}}$	Task supervision loss.
$\mathcal{L}_{\text{VQ}}$	Vector quantization loss.
$\mathcal{M}_f$, $\mathcal{M}_b$	Forward and backward Mamba operators.
N_k	Number of local samples.
P_k	Data distribution of client k.
$\mathcal{Q}(\cdot)$	Vector quantization operator.
$\mathcal{S}_r$	Selected clients at round r.
$\text{sg}[\cdot]$	Stop-gradient operation.
T'	Latent sequence length.
x_i^k	EEG trial with C channels and T samples.
y_i^k	Class label among M task classes.
Z_c	Quantized content representation.
$\tilde{Z}_c$	Continuous semantic representation.
Z_s	Complementary residual (style) representation.
$z_{c,t}$	Quantized content token.
$\tilde{z}_{c,t}$	The t-th semantic token.
$\lambda_c, \lambda_{\text{VQ}}, \lambda_{\text{adv}}$	Task, VQ, and adversarial loss trade-off coefficients.
Θ, Θ_g	Representation and shared global parameters.
$\Psi(\cdot,\cdot)$	Bidirectional fusion function.

3.1. Problem Formulation and Framework Overview

We consider a federated EEG learning scenario with K clients. The k-th client owns a private local dataset $\mathcal{D}_k = \{(x_i^k, y_i^k)\}_{i=1}^{N_k}$, where $x_i^k \in \mathbb{R}^{C\times T}$ denotes an EEG trial with C channels and T temporal samples, and $y_i^k \in \{1,\ldots,M\}$ denotes the corresponding class label. Due to inter-subject variability, electrode placement differences, and recording-dependent noise, EEG data collected from different clients are usually non-IID, namely $P_k(x,y) \neq P_j(x,y)$ for $k \neq j$. The goal is to learn a global EEG decoder without exchanging raw EEG signals.

In conventional federated EEG learning, local models are optimized on subject-dependent data and then aggregated into a global model. This procedure may be problematic when latent features contain both task-relevant neural patterns and strong client-associated nuisance variation. Aggregating parameters learned from such entangled representations can propagate subject-dependent bias into the global model. Fed-VQCS addresses this issue by explicitly allocating task semantics to a shared content pathway while introducing a complementary residual pathway that is adversarially discouraged from preserving task-discriminative information.

Let H be a latent EEG representation extracted by a backbone encoder. Fed-VQCS decomposes H as

$$H = \tilde{Z}_c + Z_s, \qquad Z_c = \mathcal{Q}(\tilde{Z}_c), \qquad Z_s = H - \tilde{Z}_c. \tag{1}$$

Here, $\tilde{Z}_c$ is the continuous semantic representation produced by the Bi-Mamba extractor, Z_c is the vector-quantized content representation, $\mathcal{Q}(\cdot)$ denotes codebook quantization, and Z_s is the complementary residual representation. The residual is defined before quantization, i.e., from $H - \tilde{Z}_c$ rather than $H - Z_c$, so that codebook approximation error is not directly folded into the residual definition. We use "style" to describe this complementary variation relative to the task-semantic estimate. It can include individual physiological characteristics as well as session-, device-, electrode-, noise-, and modeling-related effects. The adversarial objective in Section 3.5 suppresses task predictability in Z_s, while Section 4.7 empirically examines its subject association through representation probing.

3.2. Backbone Feature Extraction

The backbone encoder provides the initial latent representation for subsequent semantic purification. In EEG decoding, convolutional front-ends are widely used because they naturally match the structure of EEG trials. Temporal convolutions act as learnable frequency- and rhythm-sensitive filters along the time dimension, while spatial convolutions integrate information across electrodes and capture channel-specific activation patterns. Such temporal-spatial filtering is particularly important for motor imagery and cognitive-state decoding, where discriminative information is often reflected in both rhythmic temporal responses and spatially distributed cortical patterns.

Fed-VQCS does not depend on a specific backbone architecture. The front-end can be instantiated by compact temporal-spatial CNNs, hierarchical convolutional encoders, or convolutional tokenizers followed by attention-based contextual blocks. We write this stage abstractly as

$$H_i^k = f_{\text{base}}(x_i^k; \theta_{\text{base}}) \in \mathbb{R}^{T' \times d}, \tag{2}$$

where T' denotes the length of the latent sequence and d denotes the embedding dimension. This abstraction emphasizes the functional role of the backbone rather than its implementation details: it transforms raw EEG signals into a compact sequence of latent neural descriptors.

Although H_i^k is expected to contain discriminative information for the target task, it is not necessarily suitable for direct global sharing. Because the backbone is trained on local subject-dependent data, H_i^k may still encode nuisance factors such as amplitude scale, local signal quality, electrode placement variation, and individual response patterns. Therefore, the following modules operate on H_i^k to extract the semantic component that is more appropriate for federated aggregation.

3.3. Bi-Mamba Semantic Modeling

The purpose of the Bi-Mamba semantic extractor is to derive a continuous semantic representation $\tilde{Z}_c$ from the backbone feature H. Unlike local convolutional filters, state-space sequence models are designed to capture long-range temporal dependencies with linear-complexity sequence processing. A standard continuous state-space model can be written as

$$\dot{s}(t) = As(t) + Bu(t), \qquad y(t) = Cs(t), \tag{3}$$

where $u(t)$ is the input signal, $s(t)$ is the hidden state, and $y(t)$ is the output. After discretization, the recurrent form becomes

$$s_t = \bar{A}s_{t-1} + \bar{B}u_t, \qquad y_t = \bar{C}s_t. \tag{4}$$

Mamba extends this state-space formulation with input-selective parameterization, where the transition and projection terms are dynamically conditioned on the current token. This selectivity allows the model to adaptively retain task-relevant temporal patterns while suppressing less informative fluctuations.

For EEG trials, discriminative evidence may depend on both preceding and succeeding temporal contexts within the same trial. We therefore employ a bidirectional Mamba formulation. Given $H = [h_1, \ldots, h_{T'}]$, the forward Mamba operator summarizes information from the past-to-future direction, while the backward operator processes the reversed sequence to incorporate future-to-past context. The continuous semantic representation is obtained as

$$\tilde{Z}_c = \mathrm{LN}(H + \Psi(\mathcal{M}_f(H),\, \mathrm{rev}(\mathcal{M}_b(\mathrm{rev}(H))))\,, \tag{5}$$

where $\mathcal{M}_f$ and $\mathcal{M}_b$ denote the forward and backward Mamba operators, $\mathrm{rev}(\cdot)$ reverses the temporal order, $\Psi(\cdot,\cdot)$ denotes learnable bidirectional fusion, and $\mathrm{LN}(\cdot)$ denotes layer normalization. The residual connection preserves the original latent feature structure while enabling the Bi-Mamba module to refine it with bidirectional temporal semantics.

The resulting $\tilde{Z}_c$ is a continuous semantic representation designed to capture temporally meaningful task-related structure from H. Because it is continuous and learned from heterogeneous clients, it may still retain client-associated details. Fed-VQCS therefore further constrains $\tilde{Z}_c$ through vector quantization, converting it into a discrete prototype-based content representation. Accordingly, the role of Bi-Mamba is semantic refinement rather than direct style extraction; the complementary residual is computed only after this semantic estimate has been formed.

3.4. Vector-Quantized Content Prototypes and Residual Style

The vector-quantized content branch imposes a discrete semantic bottleneck on $\tilde{Z}_c$. Let $\mathcal{E} = \{e_m\}_{m=1}^{K_c}$ denote a learnable codebook with K_c content prototypes, where $e_m \in \mathbb{R}^d$. For each semantic token $\tilde{z}_{c,t}$ in $\tilde{Z}_c$, the nearest prototype is selected according to

$$a_t = \arg\min_m \|\mathrm{sg}[\tilde{z}_{c,t}] - e_m\|_2^2, \qquad z_{c,t} = e_{a_t}, \tag{6}$$

where $\mathrm{sg}[\cdot]$ denotes the stop-gradient operator. In implementation, the straight-through estimator is used so that the forward pass uses the selected prototype while gradients are passed to the continuous semantic encoder. The quantized sequence is denoted as $Z_c = [z_{c,1}, \ldots, z_{c,T'}]$.

The quantization loss follows the standard commitment mechanism:

$$\mathcal{L}_{\mathrm{VQ}} = \mathbb{E}_x \left[\left\| \mathrm{sg}[\tilde{Z}_c] - Z_c \right\|_2^2 + \beta \left\| \tilde{Z}_c - \mathrm{sg}[Z_c] \right\|_2^2 \right], \tag{7}$$

where β controls the commitment strength. The first term moves the selected prototypes toward the continuous semantic embeddings, while the second term encourages the semantic extractor to commit to the selected prototypes. By forcing task-related semantics to be expressed through a finite set of shared prototypes, the VQ-layer restricts representational degrees of freedom and discourages the content pathway from encoding arbitrary client-associated details.

The style component is defined as the residual of the continuous semantic extractor:

$$Z_s = H - \tilde{Z}_c. \tag{8}$$

This definition is central to Fed-VQCS. Instead of constructing the residual from the quantized content representation, we treat $\tilde{Z}_c$ as the continuous task-semantic estimate and Z_s as the complementary variation not represented by that estimate. The discrete content Z_c forms the shared predictive pathway, whereas Z_s provides a locally computed adversarial regularization signal. The resulting decomposition encourages task information to concentrate in the shared content pathway while retaining complementary variation in the residual pathway.

3.5. Shared-Classifier-Guided Federated Optimization

Fed-VQCS introduces a shared classifier $g(\cdot)$ to enforce task-style separation within a unified decision space. The classifier is applied to the discrete content representation Z_c and the continuous semantic representation $\tilde{Z}_c$. We define the task objective as

$$\mathcal{L}_{\mathrm{task}} = \mathbb{E}_{(x,y)}[-\log g_y(Z_c)] + \lambda_c \, \mathbb{E}_{(x,y)}[-\log g_y(\tilde{Z}_c)]. \tag{9}$$

The first term provides discrete semantic supervision for the quantized content prototypes, while the second term provides continuous semantic supervision for the Bi-Mamba extractor. When backbone-level regularization is used, an auxiliary term $\lambda_h \mathbb{E}_{(x,y)}[-\log g_y(H)]$ can be added to Equation (9) using the same classifier. This auxiliary constraint preserves the task-discriminative structure of the backbone feature without introducing a separate classifier.

Meanwhile, the style component is explicitly suppressed through an adversarial objective:

$$\mathcal{L}_{\mathrm{style}} = \mathbb{E}_{(x,y)}[-\log g_y(Z_s)]. \tag{10}$$

The classifier is optimized to predict motor-imagery labels from both the quantized content representation and the complementary residual representation:

$$\min_g \mathcal{L}_{\mathrm{task}} + \lambda_{\mathrm{adv}} \mathcal{L}_{\mathrm{style}}. \tag{11}$$

The representation learning modules are optimized with an adversarial objective, where the complementary residual representation is discouraged from preserving task-discriminative information:

$$\min_{\Theta} \mathcal{L}_{\mathrm{task}} + \lambda_{\mathrm{VQ}} \mathcal{L}_{\mathrm{VQ}} - \lambda_{\mathrm{adv}} \mathcal{L}_{\mathrm{style}}, \tag{12}$$

where Θ denotes the representation learning parameters, including the backbone, Bi-Mamba semantic extractor, and VQ-layer. This min–max optimization creates an asymmetric allocation of task information: the classifier and representation learner jointly increase the predictive utility of Z_c and $\tilde{Z}_c$, while the negative adversarial term reduces the task predictability of Z_s. Conceptually, the objective encourages $I(Z_c; y) > I(Z_s; y)$, and the resulting separation is evaluated empirically through the ablation and probing studies.

In federated training, the server maintains the globally shared semantic parameters

$$\Theta_g = \{\theta_{\text{base}}, \theta_{\text{BiMamba}}, \mathcal{E}, g\}. \tag{13}$$

At communication round r, the server broadcasts Θ_g^r to a selected client subset $\mathcal{S}_r$. Each client optimizes the local objective using its private EEG data and uploads only the updated shared parameters. The server performs weighted aggregation:

$$\Theta_g^{r+1} = \sum_{k \in \mathcal{S}_r} \frac{N_k}{\sum_{j \in \mathcal{S}_r} N_j} \Theta_{g,k}^{r+1}. \tag{14}$$

The residual representation Z_s is computed and constrained within each client's local training pass. Accordingly, "local style" denotes a residual activation that is evaluated and regularized locally, while the backbone, Bi-Mamba, codebook, and classifier parameters participate in global aggregation. The residual branch shapes the shared representation through this local adversarial regularization.

During inference, an EEG trial is encoded into H, transformed into $\tilde{Z}_c$ by the Bi-Mamba semantic extractor, and quantized into Z_c. The final prediction is produced by the shared classifier:

$$\hat{y} = \arg\max_m g_m(Z_c). \tag{15}$$

Although H and $\tilde{Z}_c$ are used for supervision during training, the final decision relies on Z_c, since it is explicitly designed to represent transferable task-relevant content in a shared discrete prototype space.

4. Experiments

In this section, we evaluate the proposed Fed-VQCS framework on two representative motor imagery EEG benchmarks. The experiments are designed to answer six questions: (1) whether Fed-VQCS consistently improves federated EEG decoding across different backbones; (2) whether the gains remain statistically stable across repeated random seeds; (3) whether Fed-VQCS is competitive with representative personalization-, normalization-, and disentanglement-based federated strategies; (4) whether it remains useful when only limited calibration samples are available for target clients; (5) how sensitive the method is to key VQCS hyperparameters; and (6) whether the learned shared-content/complementary-residual decomposition exhibits the intended task-versus-subject information allocation. We first describe the datasets, preprocessing pipeline, federated evaluation protocol, implementation details, and comparison methods, and then report the corresponding results.

4.1. Experimental Setup

We evaluate Fed-VQCS on two widely used motor imagery EEG benchmarks: the PhysioNet EEG Motor Movement/Imagery dataset [46] and the BCI Competition IV-2a dataset [47]. These datasets provide complementary federated settings: PhysioNet offers a large subject-as-client population with substantial inter-subject variability, whereas BCI IV-2a provides a smaller but well-established four-class motor imagery benchmark. In all experiments, one subject corresponds to one federated client. The present study was approved by the University Research Ethics Review Panel of Xi'an Jiaotong-Liverpool University (Proposal No. 00100000900202508201604 45; approval date: 25 August 2025).

The PhysioNet EEG Motor Movement/Imagery dataset contains recordings from 109 subjects acquired with 64 EEG channels at 160 Hz. We retain only subjects for whom all motor-imagery runs required for four-class construction are readable, contain valid event annotations, and provide complete channel data. Recordings failing any of these integrity checks are excluded, resulting in 104 valid subjects in the federated client pool. The four target classes are left-fist, right-fist, both-fists, and both-feet imagery, constructed from the corresponding motor-imagery runs and official event annotations.

The BCI Competition IV-2a dataset consists of 9 subjects performing left-hand, right-hand, both-feet, and tongue motor imagery. Signals were acquired from 22 EEG channels at 250 Hz, with two sessions recorded on different days. Each subject is treated as one client, and the official session-wise training/evaluation partition is used in the experiments. The main characteristics of the two datasets are summarized in Table 2.

4.1.1. Preprocessing and Data Splits

EEG signals are band-pass filtered to 4–40 Hz using a fourth-order zero-phase Butterworth filter. For BCI IV-2a, we retain the 22 EEG channels and segment each trial from 2.0 to 6.0 s after cue onset. For PhysioNet, we use the motor-imagery runs and extract a 4.0 s segment from each annotated imagery onset. Each channel is standardized using the mean and standard deviation estimated only from the corresponding local training data. For

PhysioNet, trials of each client are stratified into 70%/10%/20% training/validation/test splits. For BCI IV-2a, we use the official predefined training/evaluation partition: the official training set (A0xT) serves as the local federated training data for each subject, and the corresponding official evaluation set (A0xE) is used for testing. Normalization statistics for BCI IV-2a are estimated exclusively from A0xT and then applied to A0xE.

Table 2. Summary of the datasets used in the experiments.

Dataset	Subjects	Channels	Classes	Sampling
PhysioNet MI	104	64	4	160 Hz
BCI IV-2a	9	22	4	250 Hz

4.1.2. Federated Protocol

At round r, the server samples participating clients uniformly without replacement, broadcasts the current shared parameters, and aggregates returned updates with sample-count weighting as in Equation (14). For PhysioNet, the participation ratio is 20% per round; for BCI IV-2a, all 9 clients (100%) participate because of the small client population. Each selected client performs 5 local epochs before communication. Training runs for 120 rounds on PhysioNet and 150 rounds on BCI IV-2a. Raw EEG signals, intermediate residual activations, and client labels are never transmitted to the server; only the shared model parameters defined in Equation (13) are communicated.

4.1.3. Evaluation Regimes

In the standard federated regime, the final global model is evaluated on the dataset-specific test data: the held-out 20% test split for PhysioNet and the official A0xE evaluation set for BCI IV-2a. We additionally conduct a low-data transfer/adaptation evaluation starting from the trained global model, where target evaluation trials are not used during the original federated training stage. The 0-shot setting directly evaluates the trained global model without using any labeled target-domain trial for adaptation. In the 5-shot and 10-shot settings, "shot" denotes *labeled target trials per class per target client*: five or ten trials from each class are sampled uniformly without replacement for local adaptation, respectively, and performance is then evaluated on the remaining target trials. The adaptation samples are re-sampled independently for each random seed and do not overlap with the trials used for evaluation.

4.1.4. Comparison Methods

We first evaluate the plug-and-play effect of VQCS with four representative EEG backbones: EEGNet [25], DeepConvNet [48], EEG Conformer [49], and CTNet [50]. We then fix CTNet as the common backbone and compare Fed-VQCS with representative federated strategies covering personalization (FLEEG [35]), normalization (FIXBN [39] and HBN [40]), and representation disentanglement (DFL [24]). Wherever a method permits backbone substitution, its method-specific training mechanism is retained while the EEG feature extractor and data protocol are kept identical. The main implementation settings for the PhysioNet and BCI IV-2a experiments, including preprocessing parameters, federated training configurations, and VQCS hyperparameters, are summarized in Table 3.

Table 3. Implementation settings used in the experiments.

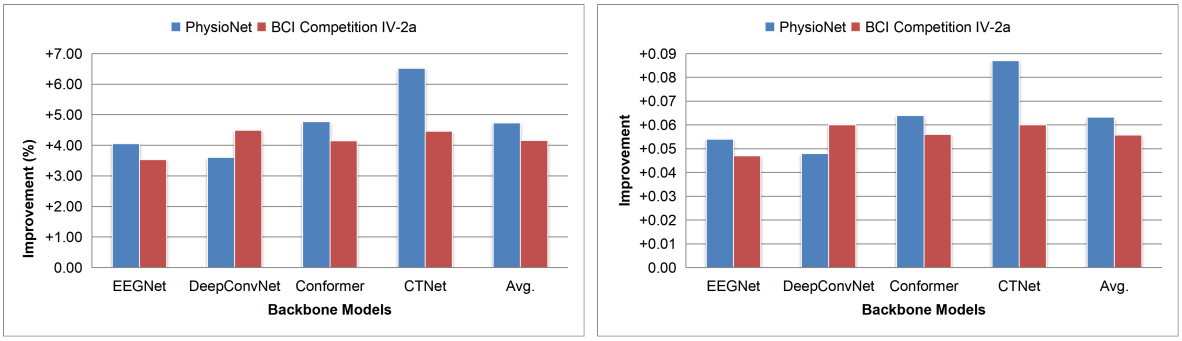
Setting	PhysioNet	BCI IV-2a
Band-pass range	4–40 Hz	4–40 Hz
Trial window	0–4 s from event	2–6 s after cue
Client data split	70/10/20%	A0xT/A0xE split
Client participation	20%/round	100%/round
Local epochs	5	5
Communication rounds	120	150
Batch size	32	32
Optimizer	AdamW	AdamW
Initial learning rate	10^{-3}	10^{-3}
Codebook size K_c	128	128
Embedding dimension d	64	64
$\beta/\lambda_c/\lambda_{\mathrm{VQ}}/\lambda_{\mathrm{adv}}$	0.25 / 0.5 / 1.0 / 0.1	
Random seeds	2022–2026 (five runs)	

4.1.5. Implementation Details

All methods use AdamW with an initial learning rate of 1×10^{-3}, weight decay 1×10^{-4}, and cosine decay to 1×10^{-5}. The local mini-batch size is 32. Unless varied in the sensitivity study, the VQCS codebook contains $K_c = 128$ prototypes with embedding dimension $d = 64$; the commitment coefficient is $\beta = 0.25$, and the loss weights are $\lambda_c = 0.5$, $\lambda_{\mathrm{VQ}} = 1.0$, and $\lambda_{\mathrm{adv}} = 0.1$. The auxiliary backbone-level regularizer in Equation (9) is disabled in the reported experiments. All reported experiments are repeated with five independent seeds, {2022, 2023, 2024, 2025, 2026}. We report mean $\pm$ standard deviation over seeds for accuracy and Cohen's κ. Paired two-sided t-tests are conducted between matched-seed runs, followed by Holm correction within each dataset; $p < 0.05$ is considered statistically significant. The implementation and configuration files will be released in an anonymized repository and made publicly available upon publication.

4.2. Main Results

We first evaluate whether the proposed VQCS module consistently improves federated EEG decoding across different backbone architectures. Table 4 and Figure 3 report the final-round performance on PhysioNet and BCI Competition IV-2a. Each backbone-specific entry in Table 4 is reported as the mean $\pm$ standard deviation over five matched random seeds. For each backbone, the original federated model and its VQCS-enhanced counterpart use identical data partitions, client samples, initialization seeds, and optimization schedules, enabling paired statistical comparison of the incremental contribution of VQCS.



(a) Absolute accuracy gain (ΔAcc.). **(b)** Absolute Cohen's kappa gain ($\Delta\kappa$).

Figure 3. Absolute performance improvement yielded by the integration of the proposed VQCS module across various backbone architectures on the PhysioNet and BCI Competition IV-2a datasets.

Table 4. Main performance across backbones on PhysioNet and BCI Competition IV-2a. Results for individual backbones are reported as mean $\pm$ standard deviation over five independent seeds. The "Avg." row reports the arithmetic mean of the four backbone-specific mean values for each metric, with no standard deviation reported for this descriptive summary. "+VQCS" denotes insertion of the proposed module under the same federated protocol. ΔAcc. is computed from the mean accuracies. Paired baseline-to-VQCS improvements are significant after Holm correction ($p < 0.05$).

Dataset	Backbone	Acc. (%)	Acc. (+VQCS) (%)	ΔAcc. (%)	κ	κ (+VQCS)
PhysioNet	EEGNet	54.98 ± 0.61	59.03 ± 0.54	+4.05	0.400 ± 0.008	0.454 ± 0.007
	DeepConvNet	55.85 ± 0.72	59.46 ± 0.63	+3.61	0.411 ± 0.010	0.459 ± 0.009
	Conformer	59.46 ± 0.57	64.23 ± 0.48	+4.77	0.459 ± 0.008	0.523 ± 0.007
	CTNet	60.35 ± 0.53	66.87 ± 0.45	+6.52	0.471 ± 0.007	0.558 ± 0.006
	Avg.	57.66	62.40	+4.74	0.435	0.499
BCI IV-2a	EEGNet	65.84 ± 0.83	69.37 ± 0.71	+3.53	0.545 ± 0.011	0.592 ± 0.010
	DeepConvNet	66.92 ± 0.76	71.41 ± 0.65	+4.49	0.559 ± 0.010	0.619 ± 0.009
	Conformer	70.11 ± 0.64	74.26 ± 0.56	+4.15	0.601 ± 0.009	0.657 ± 0.008
	CTNet	72.36 ± 0.58	76.82 ± 0.49	+4.46	0.631 ± 0.008	0.691 ± 0.007
	Avg.	68.81	72.97	+4.16	0.584	0.640

As shown in Table 4 and Figure 3, VQCS improves the mean performance of all four backbones on both datasets. On PhysioNet, the average final accuracy increases from 57.66% to 62.40%, an absolute gain of 4.74

percentage points. The largest improvement is obtained with CTNet, from $60.35 \pm 0.53\%$ to $66.87 \pm 0.45\%$. On BCI IV-2a, the average final accuracy increases from 68.81% to 72.97%, with gains observed for convolutional and convolution-transformer backbones alike. All paired baseline-to-VQCS accuracy comparisons remain significant after Holm correction, indicating that the improvement is not driven by a single favorable initialization.

Cohen's κ follows the same trend. Across both datasets, every VQCS-enhanced backbone exhibits a higher mean κ than its matched baseline, and the paired differences are statistically significant at $p < 0.05$ after correction. Together with the accuracy results, these repeated runs show that VQCS improves task decoding consistently across architectures while keeping the underlying federated protocol unchanged.

4.3. Comparison with Representative Federated Learning Methods

To assess whether the gains extend beyond a backbone-only comparison, we compare Fed-VQCS with representative methods that address client heterogeneity through personalization, normalization, or representation disentanglement. CTNet is used as the common EEG backbone to control for feature-extractor capacity. Table 5 reports mean $\pm$ standard deviation over the same five random seeds.

Table 5. Comparison with representative federated learning strategies using CTNet as the common backbone. Results are mean $\pm$ standard deviation over five seeds. Fed-VQCS is significantly better than the strongest non-VQCS competitor under paired testing ($p < 0.05$ after Holm correction).

Method	PhysioNet Acc. (%)	BCI IV-2a Acc. (%)
FedAvg (CTNet)	60.35 ± 0.53	72.36 ± 0.58
FIXBN [39]	61.82 ± 0.60	73.10 ± 0.63
HBN [40]	62.74 ± 0.55	74.08 ± 0.59
FLEEG [35]	63.11 ± 0.58	74.21 ± 0.61
DFL [24]	64.03 ± 0.50	75.06 ± 0.52
Fed-VQCS	66.87 ± 0.45	76.82 ± 0.49

FedAvg with CTNet provides the direct backbone baseline. Normalization-based methods improve performance by reducing client-dependent statistic mismatch, while FLEEG and DFL obtain larger gains by introducing personalized or disentangled learning mechanisms. Nevertheless, Fed-VQCS achieves the highest mean accuracy on both datasets, exceeding the strongest non-VQCS competitor by 2.84 percentage points on PhysioNet and 1.76 points on BCI IV-2a. The paired differences remain significant after correction. These results complement the backbone study and demonstrate the effectiveness of the proposed representation-level content bottleneck across representative federated heterogeneity-handling strategies.

4.4. Low-Data Adaptation Results

In practical BCI applications, labeled target-domain data for calibration or adaptation can be limited. We therefore evaluate three transfer/adaptation settings: 0-shot, 5-shot, and 10-shot, all starting from the trained global model. The target trials used for this evaluation are not included in the original federated training data. In the 0-shot setting, no labeled target-domain trial is provided for adaptation and the trained global model is evaluated directly on the target trials. In the 5-shot setting, five labeled target trials from each class of each target client are sampled uniformly without replacement and used for local adaptation; the 10-shot setting analogously uses ten labeled trials per class. Evaluation is then performed only on the remaining target trials, with no overlap between adaptation and evaluation samples. The target adaptation samples are re-sampled independently for each of the five random seeds. This protocol tests both zero-shot transfer and the ability of the shared content representation to support adaptation when only limited client-specific supervision is available.

Table 6 shows that VQCS improves the mean accuracy of every backbone under 0-shot, 5-shot, and 10-shot calibration on both datasets. On PhysioNet, the average gains across the three calibration budgets range from 3.57 to 4.53 percentage points; on BCI IV-2a, they range from 3.78 to 4.16 points. Because calibration samples are re-drawn across seeds, the reported standard deviations reflect both optimization randomness and low-shot sampling variability.

The gains persist after client-specific adaptation and remain consistent across all three calibration budgets. This behavior is consistent with the intended role of the discrete content space: global training provides a stable task-semantic starting point, while a small calibration set supports efficient client-specific adaptation. The results therefore show that VQCS remains effective when target-domain supervision is limited.

Table 6. Low-data transfer/adaptation performance under 0-shot, 5-shot, and 10-shot settings. The 0-shot setting uses no labeled target trial for adaptation; 5-shot and 10-shot use five and ten labeled target trials per class per target client, respectively. Results are mean ± standard deviation over five seeds with adaptation samples re-drawn for each seed. "+VQCS" denotes insertion of VQCS into the corresponding backbone. "Avg. Δ (pp)" represents the arithmetic mean of the absolute accuracy improvements of VQCS over the corresponding baseline across the 0-shot, 5-shot, and 10-shot settings.

Dataset	Backbone	0-Shot (%)	+VQCS (%)	5-Shot (%)	+VQCS (%)	10-Shot (%)	+VQCS (%)	Avg. Δ (pp)
PhysioNet	EEGNet	54.60 ± 0.69	58.10 ± 0.58	54.80 ± 0.76	59.40 ± 0.63	56.50 ± 0.72	61.10 ± 0.61	+4.23
	DeepConvNet	55.50 ± 0.73	59.80 ± 0.60	58.40 ± 0.70	63.00 ± 0.57	58.20 ± 0.68	62.60 ± 0.55	+4.43
	Conformer	51.50 ± 0.81	55.00 ± 0.72	55.60 ± 0.75	59.00 ± 0.64	59.30 ± 0.66	63.10 ± 0.58	+3.57
	CTNet	57.10 ± 0.70	61.50 ± 0.60	59.70 ± 0.64	64.40 ± 0.52	61.00 ± 0.61	65.50 ± 0.49	+4.53
BCI IV-2a	EEGNet	64.80 ± 0.92	68.77 ± 0.78	66.10 ± 0.85	69.63 ± 0.74	67.40 ± 0.80	71.39 ± 0.69	+3.83
	DeepConvNet	66.20 ± 0.88	69.67 ± 0.76	68.40 ± 0.80	72.86 ± 0.68	69.10 ± 0.76	73.65 ± 0.64	+4.16
	Conformer	62.70 ± 0.95	66.97 ± 0.82	66.80 ± 0.84	70.41 ± 0.72	69.00 ± 0.78	73.19 ± 0.65	+4.02
	CTNet	67.50 ± 0.86	70.74 ± 0.73	70.10 ± 0.78	73.88 ± 0.66	71.60 ± 0.71	75.91 ± 0.59	+3.78

4.5. Ablation Study

To further investigate the contribution of each component in VQCS, we conduct an ablation study using CTNet as the backbone because it achieves the strongest overall performance in the main comparison. All ablation variants are repeated with the same five seeds and federated protocol as the full model. We remove the VQ-layer, replace the Bi-Mamba semantic extractor with a lightweight feed-forward projection, remove the complementary residual pathway, remove the adversarial residual constraint, use separate classifiers instead of the shared classifier, remove the continuous semantic supervision on $\tilde{Z}_c$, or remove the discrete content supervision on Z_c. The corresponding ablation results and the configurations of the content pathway and residual constraint are summarized in Table 7.

Table 7. Ablation study with CTNet as the backbone. Results are mean ± standard deviation over five seeds. ΔAcc. denotes the difference between the full VQCS mean and the corresponding ablated mean. In "Content Path", a checkmark denotes retention of the standard pathway; "continuous" denotes removal of the VQ-layer, while "quantized" denotes retention of the quantized pathway with Bi-Mamba replaced by a lightweight feed-forward projection; "unstable" and "weak" indicate removal of continuous and discrete semantic supervision, respectively. In "Residual Constraint", a checkmark denotes retention of the adversarial constraint; a dash denotes removal of the residual pathway; "weak" denotes removal of the adversarial style loss; and "separate" denotes the use of separate classifiers instead of the shared classifier.

Dataset	Variant	Final Acc.	Final κ	Δ Acc.	Content Path	Residual Constraint
PhysioNet	Full VQCS	66.87 ± 0.45	0.558 ± 0.006	–	✓	✓
	w/o VQ-layer	63.14 ± 0.58	0.508 ± 0.008	+3.73	continuous	✓
	w/o Bi-Mamba	62.76 ± 0.62	0.503 ± 0.009	+4.11	quantized	✓
	w/o residual Z_s	63.52 ± 0.56	0.514 ± 0.008	+3.35	✓	–
	w/o $\mathcal{L}_{\text{style}}$	64.08 ± 0.54	0.521 ± 0.008	+2.79	✓	weak
	w/o shared classifier	64.41 ± 0.51	0.525 ± 0.007	+2.46	✓	separate
	w/o $\ell(\tilde{Z}_c, y)$	64.95 ± 0.49	0.533 ± 0.007	+1.92	unstable	✓
	w/o $\ell(Z_c, y)$	63.86 ± 0.55	0.518 ± 0.008	+3.01	weak	✓
BCI IV-2a	Full VQCS	76.82 ± 0.49	0.691 ± 0.007	–	✓	✓
	w/o VQ-layer	73.62 ± 0.63	0.648 ± 0.009	+3.20	continuous	✓
	w/o Bi-Mamba	72.94 ± 0.68	0.639 ± 0.010	+3.88	quantized	✓
	w/o residual Z_s	73.71 ± 0.61	0.649 ± 0.009	+3.11	✓	–
	w/o $\mathcal{L}_{\text{style}}$	74.29 ± 0.59	0.657 ± 0.008	+2.53	✓	weak
	w/o shared classifier	74.61 ± 0.57	0.661 ± 0.008	+2.21	✓	separate
	w/o $\ell(\tilde{Z}_c, y)$	75.07 ± 0.55	0.668 ± 0.008	+1.75	unstable	✓
	w/o $\ell(Z_c, y)$	74.02 ± 0.60	0.654 ± 0.009	+2.80	weak	✓

The full VQCS model achieves the highest mean accuracy on both datasets. Removing the VQ-layer reduces accuracy by 3.73 points on PhysioNet and 3.20 points on BCI IV-2a, supporting the role of the discrete bottleneck

in constraining the shared content space. Replacing Bi-Mamba with a feed-forward projection produces the largest degradation (4.11 and 3.88 points), indicating that temporally contextualized semantic modeling is important before quantization.

Removing the complementary residual pathway reduces accuracy by 3.35 points on PhysioNet and 3.11 points on BCI IV-2a, while removing $\mathcal{L}_{\text{style}}$ also produces a clear drop. Together, these ablations show that allocating complementary variation to a residual pathway and suppressing task predictability in that pathway improves the shared predictive representation. The shared-classifier and dual semantic-supervision ablations further show that the decomposition benefits from the joint optimization of content supervision, residual regularization, and a unified decision space.

4.6. Sensitivity Analysis

We further evaluate the sensitivity of Fed-VQCS to the codebook size K_c, embedding dimension d, and the two loss weights most directly associated with the proposed decomposition, λ_{adv} and λ_{VQ}. CTNet is used as the backbone, one factor is varied at a time, and all other settings follow Table 3. Table 8 reports five-seed mean accuracy.

Table 8. Sensitivity analysis of key VQCS hyperparameters with CTNet. Results are mean accuracy (%) ± standard deviation over five seeds.

Parameter	Value	PhysioNet	BCI IV-2a
K_c	32	64.51 ± 0.61	74.33 ± 0.66
	64	65.92 ± 0.52	75.74 ± 0.58
	128	$\mathbf{66.87 \pm 0.45}$	$\mathbf{76.82 \pm 0.49}$
	256	66.31 ± 0.49	76.45 ± 0.53
d	32	65.40 ± 0.57	75.31 ± 0.61
	64	$\mathbf{66.87 \pm 0.45}$	$\mathbf{76.82 \pm 0.49}$
	128	66.42 ± 0.51	76.39 ± 0.55
λ_{adv}	0	64.08 ± 0.54	74.29 ± 0.59
	0.05	65.84 ± 0.50	75.66 ± 0.56
	0.10	$\mathbf{66.87 \pm 0.45}$	$\mathbf{76.82 \pm 0.49}$
	0.20	66.10 ± 0.53	76.09 ± 0.57
λ_{VQ}	0.5	66.12 ± 0.50	76.08 ± 0.55
	1.0	$\mathbf{66.87 \pm 0.45}$	$\mathbf{76.82 \pm 0.49}$
	2.0	66.21 ± 0.48	76.25 ± 0.52

Boldface marks the default hyperparameter values and their corresponding results used in the main experiments: $K_c = 128$, $d = 64$, $\lambda_{\text{adv}} = 0.10$, and $\lambda_{\text{VQ}} = 1.0$.

The method remains stable over a practical range of codebook capacities and embedding dimensions. Very small codebooks reduce performance, consistent with insufficient semantic capacity, whereas increasing K_c beyond the default yields no additional gain. A moderate adversarial weight performs best: setting $\lambda_{\text{adv}} = 0$ recovers the no-style-constraint ablation, while an overly large adversarial weight slightly degrades task performance. The VQ loss shows a similarly broad optimum around the default value. These results indicate that the improvements are not tied to a single fragile hyperparameter choice.

4.7. Disentanglement Analysis

The classification results demonstrate the performance advantage of VQCS. We next examine the information allocation induced by the learned decomposition. The analysis focuses on whether task information is concentrated in the shared content pathway and whether the complementary residual exhibits lower task predictability together with stronger client association.

After training, the backbone feature H is extracted from the CTNet baseline model, whereas the continuous semantic representation $\tilde{Z}_c$, quantized content representation Z_c, and complementary residual representation Z_s are extracted from the corresponding CTNet+VQCS model using held-out trials. Task Acc. and Task κ are calculated using the task classifier associated with each model: the original CTNet classifier for H and the shared classifier g for the VQCS-derived representations. Subject Acc. is evaluated separately using a lightweight linear probe

trained on each frozen representation to predict client identity. The subject-ID probing experiments are repeated over the same five random seeds. A desirable shared content representation should retain high task-classification performance while reducing subject-ID predictability; the residual pathway should exhibit substantially lower task-classification performance than the content pathway.

As shown in Table 9, H from the CTNet baseline model retains task-discriminative information while exhibiting considerable subject-ID predictability. Compared with H, the VQCS-derived representations $\tilde{Z}_c$ and Z_c show higher task-classification performance and lower subject-ID predictability, with Z_c achieving the highest task-classification performance and lowest subject-ID predictability. In contrast, Z_s has substantially lower task-probe accuracy than Z_c while retaining higher subject-ID predictability. This pattern is consistent across both datasets and supports the intended allocation of task semantics to the shared content pathway and subject-associated complementary variation to the residual pathway.

The lower subject-ID predictability of $\tilde{Z}_c$ and Z_c compared with H, together with their strong task-classification performance, is consistent with the intended role of the VQCS semantic and quantization components. Combined with the low task predictability and higher subject-ID predictability of Z_s, these results provide complementary evidence for the content–style separation encouraged by the training objective.

Table 9. Representation analysis on held-out trials using CTNet and CTNet+VQCS. H is extracted from the CTNet baseline, while $\tilde{Z}_c$, Z_c, and Z_s are extracted from CTNet+VQCS. Task Acc. and Task κ are obtained using the corresponding model's task classifier (the original CTNet classifier for H and the shared classifier g for VQCS-derived representations). Subject Acc. is obtained using independently trained lightweight linear probes on frozen representations. Results are reported as mean $\pm$ standard deviation over five seeds.

Dataset	Representation	Task Acc.	Subject Acc.	Task κ
PhysioNet	H	60.35 ± 0.53	47.82 ± 1.10	0.471 ± 0.007
	$\tilde{Z}_c$	64.55 ± 0.49	27.36 ± 0.96	0.527 ± 0.007
	Z_c	66.87 ± 0.45	13.92 ± 0.74	0.558 ± 0.006
	Z_s	29.84 ± 0.88	43.71 ± 1.04	0.064 ± 0.012
BCI IV-2a	H	72.36 ± 0.58	73.18 ± 1.21	0.631 ± 0.008
	$\tilde{Z}_c$	75.24 ± 0.52	45.67 ± 1.08	0.670 ± 0.007
	Z_c	76.82 ± 0.49	24.13 ± 0.91	0.691 ± 0.007
	Z_s	31.67 ± 0.94	68.41 ± 1.16	0.089 ± 0.013

5. Conclusions

In this paper, we proposed Fed-VQCS, a federated EEG decoding framework for heterogeneous motor imagery EEG learning. Fed-VQCS learns a task-semantic pathway with Bi-Mamba, maps the semantic representation to vector-quantized shared content prototypes, and uses a complementary residual pathway with adversarial suppression of task-discriminative information. The residual pathway is evaluated locally during training, while the shared backbone, semantic extractor, codebook, and classifier are optimized through federated aggregation. Repeated experiments on PhysioNet and BCI Competition IV-2a show consistent improvements across four EEG backbones, and representative federated-method comparisons, low-data adaptation, sensitivity analysis, ablation studies, and task/subject probing provide complementary evidence for the proposed design.

Several limitations remain. First, the "style" residual is an operational complementary representation and may contain a mixture of individual physiology, recording/session effects, device or electrode differences, noise, and other variation outside the semantic pathway. A more factor-specific decomposition is an important direction for future work. Second, the current subject-ID analysis measures linear probe predictability; formal privacy guarantees would require dedicated privacy mechanisms and evaluation. Third, the evaluation focuses on motor imagery EEG and does not yet cover other tasks such as emotion recognition, sleep staging, seizure detection, or event-related potential analysis. Finally, real-world federated BCI systems may exhibit dynamic participation, cross-session drift, device changes, and resource constraints beyond the present protocol. Future work will investigate factor-aware residual modeling, privacy-preserving extensions, adaptive or hierarchical codebooks, and more realistic longitudinal federated evaluation.

Author Contributions: J.L.: conceptualization, methodology, software, investigation, writing—original draft preparation; J.C.: software, data curation, formal analysis; R.Y.: methodology, supervision, funding acquisition; H.C.: validation, visualization, writing—reviewing and editing; J.L.: resources, writing—reviewing and editing.

Funding: This research was funded by the Jiangsu Provincial Scientific Research Center of Applied Mathematics (BK20233002) and the State Key Laboratory of Novel Software Technology (KFKT2026B79).

Data Availability Statement: The original EEG recordings analyzed in this study are publicly available from the PhysioNet EEG Motor Movement/Imagery Dataset (version 1.0.0; https://doi.org/10.13026/C28G6P) and the BCI Competition IV-2a dataset (https://www.bbci.de/competition/iv/download/index.html). The processed data and derived materials generated in this study are available from the corresponding author upon reasonable request. The implementation and configuration files supporting this study will be made publicly available in an anonymized repository upon publication.

Conflicts of Interest: The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies:During the preparation of this work, the authors used ChatGPT (OpenAI) solely to assist with language proofreading, including wording and grammar correction, and manuscript formatting. After using this tool the authors reviewed and edited the AI-assisted content as needed and take full responsibility for the content of the published article.

References

1. Abiri, R.; Borhani, S.; Sellers, E.W.; et al. A Comprehensive Review of EEG-Based Brain–Computer Interface Paradigms. *J. Neural Eng.* **2019**, *16*, 011001. https://doi.org/10.1088/1741-2552/aaf12e.
2. Altaheri, H.; Muhammad, G.; Alsulaiman, M.; et al. Deep Learning Techniques for Classification of Electroencephalogram (EEG) Motor Imagery (MI) Signals: A Review. *Neural Comput. Appl.* **2023**, *35*, 14681–14722. https://doi.org/10.1007/s00521-021-06352-5.
3. Värbu, K.; Muhammad, N.; Muhammad, Y. Past, Present, and Future of EEG-Based BCI Applications. *Sensors* **2022**, *22*, 3331. https://doi.org/10.3390/s22093331.
4. Yadav, H.; Maini, S. Electroencephalogram Based Brain-Computer Interface: Applications, Challenges, and Opportunities. *Multimed. Tools Appl.* **2023**, *82*, 47003–47047. https://doi.org/10.1007/s11042-023-15653-x.
5. Hossain, K.M.; Islam, M.A.; Hossain, S.; et al. Status of Deep Learning for EEG-Based Brain–Computer Interface Applications. *Front. Comput. Neurosci.* **2023**, *16*, 1006763. https://doi.org/10.3389/fncom.2022.1006763.
6. Luo, Y.; Wang, H.; Chen, Z.; et al. Integrating Channel Priors and Shared Pattern Learning for Enhanced Motor Imagery EEG Classification. *Neurocomputing* **2026**, *669*, 132532. https://doi.org/10.1016/j.neucom.2025.132532.
7. Qin, C.; Yang, R.; Zhu, L.; et al. EEG-Infinity: A Mathematical Modeling-Inspired Architecture for Addressing Cross-Device Challenges in Motor Imagery. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2025**, *33*, 4669–4686. https://doi.org/10.1109/tnsre.2025.3635018.
8. Liu, B.; Lv, N.; Guo, Y.; et al. Recent Advances on Federated Learning: A Systematic Survey. *Neurocomputing* **2024**, *597*, 128019. https://doi.org/10.1016/j.neucom.2024.128019.
9. Lyu, L.; Yu, H.; Ma, X.; et al. Privacy and Robustness in Federated Learning: Attacks and Defenses. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 8726–8746. https://doi.org/10.1109/tnnls.2022.3216981.
10. Chen, K.; Liu, M.; Zhu, Z.; et al. Distributed Observers for Linear Systems with Communication Noises over Markovian Switching Topologies. *Int. J. Syst. Sci.* **2025**, *56*, 1973–1985. https://doi.org/10.1080/00207721.2024.2436639.
11. Liu, S.; Wang, L.; Zhang, Y.; et al. Recursive Filtering of Networked Systems with Communication Protocol Scheduling: A Survey. *Int. J. Syst. Sci.* **2025**, *56*, 2499–2516. https://doi.org/10.1080/00207721.2024.2448775.
12. Chen, W.; Hu, J.; Wu, Z.; et al. A Survey on Fault Detection for Networked Systems under Communication Constraints. *Syst. Sci. Control. Eng.* **2025**, *13*, 2460434. https://doi.org/10.1080/21642583.2025.2460434.
13. Yang, Q.; Liu, Y.; Chen, T.; et al. Federated Machine Learning. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. https://doi.org/10.1145/3298981.
14. Zhan, S.; Huang, L.; Luo, G.; et al. A Review on Federated Learning Architectures for Privacy-Preserving AI: Lightweight and Secure Cloud–Edge–End Collaboration. *Electronics* **2025**, *14*, 2512. https://doi.org/10.3390/electronics14132512.
15. Yan, X.; Wang, C.; Jin, Y. Federated Bimodal Graph Neural Networks for Text-Image Retrieval. *Int. J. Netw. Dyn. Intell.* **2025**, *4*, 100009. https://doi.org/10.53941/ijndi.2025.100009.
16. Jia, T.; Meng, L.; Li, S.; et al. Federated Motor Imagery Classification for Privacy-Preserving Brain-Computer Interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2024**, *32*, 3442–3451. https://doi.org/10.1109/tnsre.2024.3457504.
17. Zhao, P.; Chen, Y.; Ding, D.; et al. Privacy-Preserving Distributed State Estimation for Microgrids Based on Encrypted Measurements under Bit-Rate Constraints. *Int. J. Syst. Sci.* **2025**, *56*, 2481–2498. https://doi.org/10.1080/00207721.2024.2448773.
18. Gao, P.; Jia, C.; Zhou, A. Encryption–Decryption-Based State Estimation for Nonlinear Complex Networks Subject to Coupled Perturbation. *Syst. Sci. Control. Eng.* **2024**, *12*, 2357796. https://doi.org/10.1080/21642583.2024.2357796.
19. Zou, L.; Song, B.; Suo, J.; et al. A Survey on Outlier-Resistant State Estimation and Its Applications. *Syst. Sci. Control. Eng.* **2025**, *13*, 2474471. https://doi.org/10.1080/21642583.2025.2474471.

20. Song, B.; Zhao, S.; Dang, L.; et al. A Survey on Learning from Data with Label Noise via Deep Neural Networks. *Syst. Sci. Control. Eng.* **2025**, *13*, 2488120. https://doi.org/10.1080/21642583.2025.2488120.
21. Cai, Y.; Yang, X.; Yang, Y.; et al. Leader-Following Privacy-Preserving Consensus Control of Nonlinear Multi-Agent Systems: A State Decomposition Approach. *Int. J. Syst. Sci.* **2025**, *56*, 2284–2295. https://doi.org/10.1080/00207721.2024.2445726.
22. Wan, Z.; Yang, R.; Huang, M.; et al. A Review on Transfer Learning in EEG Signal Analysis. *Neurocomputing* **2021**, *421*, 1–14. https://doi.org/10.1016/j.neucom.2020.09.017.
23. Zhan, Y.; Yang, R.; You, J.; et al. A Systematic Literature Review on Incomplete Multimodal Learning: Techniques and Challenges. *Syst. Sci. Control. Eng.* **2025**, *13*, 2467083. https://doi.org/10.1080/21642583.2025.2467083.
24. Luo, Z.; Wang, Y.; Wang, Z.; et al. Disentangled Federated Learning for Tackling Attributes Skew via Invariant Aggregation and Diversity Transferring. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; Volume 162, pp. 14527–14541.
25. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; et al. EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain–Computer Interfaces. *J. Neural Eng.* **2018**, *15*, 056013. https://doi.org/10.1088/1741-2552/aace8c.
26. Riyad, M.; Khalil, M.; Adib, A. MI-EEGNET: A Novel Convolutional Neural Network for Motor Imagery Classification. *J. Neurosci. Methods* **2021**, *353*, 109037. https://doi.org/10.1016/j.jneumeth.2020.109037.
27. Ding, Y.; Robinson, N.; Zhang, S.; et al. TSception: Capturing Temporal Dynamics and Spatial Asymmetry from EEG for Emotion Recognition. *IEEE Trans. Affect. Comput.* **2023**, *14*, 2238–2250. https://doi.org/10.1109/taffc.2022.3169001.
28. Yue, J.; Xiao, X.; Wang, K.; et al. Augmenting Electroencephalogram Transformer for Steady-State Visually Evoked Potential-Based Brain–Computer Interfaces. *Cyborg Bionic Syst.* **2025**, *6*, 0379. https://doi.org/10.34133/cbsystems.0379.
29. Gao, M.; Gao, C.; Yao, Z.; et al. FA-STTM: A Hybrid Transformer-Mamba Network for Motor Imagery EEG Classification. *Neurocomputing* **2026**, *669*, 132481. https://doi.org/10.1016/j.neucom.2025.132481.
30. Xu, C.; Liu, H.; Qi, W. EEG Emotion Recognition Based on Federated Learning Framework. *Electronics* **2022**, *11*, 3316. https://doi.org/10.3390/electronics11203316.
31. Yin, K.; Ding, Z.; Yang, X.; et al. Sleep Staging Method for Imbalanced EEG Data Based on Differential Privacy Federated Learning. *Int. J. Artif. Intell. Tools* **2022**, *31*, 2240018. https://doi.org/10.1142/s0218213022400188.
32. Ma, B.; Duan, L.; Huang, Z.; et al. A Federated Semi-Supervised Automatic Sleep Staging Method Based on Relationship Knowledge Sharing. *Expert Syst. Appl.* **2024**, *237*, 121427. https://doi.org/10.1016/j.eswa.2023.121427.
33. Moon, S.; Kim, T.S.; Ryu, J.; et al. Federated Learning for Sleep Stage Classification on Edge Devices via a Model-Agnostic Meta-Learning-Based Pre-Trained Model. In Proceedings of the 2023 IEEE 13th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), Berlin, Germany, 3–5 September 2023; pp. 188–192. https://doi.org/10.1109/icce-berlin58801.2023.10375664.
34. López Bernal, S.; Hidalgo Rogel, J.M.; Martínez Beltrán, E.T.; et al. Privacy-Preserving Hierarchical Federated Learning with Biosignals to Detect Drowsiness while Driving. *Neural Comput. Appl.* **2024**, *36*, 20425–20437. https://doi.org/10.1007/s00521-024-10282-3.
35. Liu, R.; Chen, Y.; Li, A.; et al. Aggregating Intrinsic Information to Enhance BCI Performance through Federated Learning. *Neural Netw.* **2024**, *172*, 106100. https://doi.org/10.1016/j.neunet.2024.106100.
36. Liang, Y.; Chen, Y.; Pan, D.; et al. Privacy-Preserving Distributed Entropy Filtering for Microgrids with Innovation Decomposition. *Int. J. Netw. Dyn. Intell.* **2025**, *4*, 100004. https://doi.org/10.53941/ijndi.2025.100004.
37. Aziz, A.; Chhaglani, B.; Radmehr, A.; et al. Hardware-Assisted Privacy-Preserving Multi-Channel EEG Computational Headwear. In Proceedings of the 2024 IEEE 20th International Conference on Body Sensor Networks (BSN), Chicago, IL, USA, 15–17 October 2024; pp. 1–4. https://doi.org/10.1109/bsn63547.2024.10780473.
38. Zhang, W.; Wang, Z.; Wu, D. Multi-Source Decentralized Transfer for Privacy-Preserving BCIs. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2710–2720. https://doi.org/10.1109/tnsre.2022.3207494.
39. Zhong, J.; Chen, H.Y.; Chao, W.L. Making Batch Normalization Great in Federated Deep Learning. *arXiv* **2023**, arXiv:2303.06530v4.
40. Chen, H.; Xu, T.; Wu, X.; et al. Hybrid Batch Normalisation: Resolving the Dilemma of Batch Normalisation in Federated Learning. In Proceedings of the 42nd International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 9500–9515. https://proceedings.mlr.press/v267/chen25by.html.
41. Bao, W.; Wei, T.; Wang, H.; et al. Adaptive Test-Time Personalization for Federated Learning. In Proceedings of the Advances in Neural Information Processing Systems 36, New Orleans, LA, USA, 10–16 December 2023; Volume 36, pp. 77882–77914. https://doi.org/10.52202/075280-3403.
42. Perslev, M.; Darkner, S.; Kempfner, L.; et al. U-Sleep: Resilient High-Frequency Sleep Staging. *npj Digit. Med.* **2021**, *4*, 72. https://doi.org/10.1038/s41746-021-00440-5.
43. Qin, C.; Yang, R.; You, W.; et al. EEGUnity: Open-Source Tool in Facilitating Unified EEG Datasets toward Large-Scale EEG Model. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2025**, *33*, 1653–1663. https://doi.org/10.1109/tnsre.2025.3565158.
44. Hu, Y.; Zhang, C.; Liu, S. Privacy-Preserving Distributed Recursive Filtering for State-Saturated Systems with Quantization Effects. *Int. J. Netw. Dyn. Intell.* **2025**, *4*, 100012. https://doi.org/10.53941/ijndi.2025.100012.
45. Fan, H.; Liu, S.; Cao, X.; et al. A Prototype-Guided Federated Learning Based Fault Diagnosis Method of Me-

chanical Transmission System under Label Distribution Skew. *Neurocomputing* **2025**, *656*, 131532. https://doi.org/10.1016/j.neucom.2025.131532.

46. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **2000**, *101*, e215–e220. https://doi.org/10.1161/01.cir.101.23.e215.
47. Brunner, C.; Leeb, R.; Müller-Putz, G.; et al. BCI Competition 2008–Graz data set A. Available online: https://www.bbci.de/competition/iv/desc_2a.pdf (accessed on 7 November 2026).
48. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; et al. Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420. https://doi.org/10.1002/hbm.23730.
49. Song, Y.; Zheng, Q.; Liu, B.; et al. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 710–719. https://doi.org/10.1109/tnsre.2022.3230250.
50. Zhao, W.; Jiang, X.; Zhang, B.; et al. CTNet: A Convolutional Transformer Network for EEG-Based Motor Imagery Classification. *Sci. Rep.* **2024**, *14*, 20237. https://doi.org/10.1038/s41598-024-71118-7.