Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 21, No. 1, 2026

An Enhanced CTNet for Motor Imagery EEG Classification

Kun Chen, Guimei Yin

College of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China

Abstract: To address the limited feature discriminability, substantial training fluctuations, and limited decision stability during inference in four-class motor imagery electroencephalogram (MI-EEG) classification, this paper proposes an enhanced CTNet method that integrates lightweight structural enhancement, stability-oriented training, and complementary fusion inference. Built upon CTNet, the proposed method introduces a lightweight token-channel gate to adaptively recalibrate high-level token features on top of the joint modeling of a convolutional front-end and a Transformer encoder, and further adopts a two-layer classification head to enhance final discriminative capability. During training, label smoothing, center loss, exponential moving average of parameters, Top-5 checkpoint averaging, composite-score-based model selection, and lightweight input augmentation are incorporated to reduce model fluctuations under small-sample conditions. During inference, the logits of the CTNet-Repro baseline model and the enhanced model are fused in a weighted manner to exploit their decision complementarity. Under the subject-dependent protocol on the BCI Competition IV 2a dataset, the proposed method achieves average Accuracy, Macro-F1, and Balanced Accuracy of 80.63%, 80.17%, and 80.41%, respectively, across 9 subjects, representing improvements of 4.52, 4.69, and 4.49 percentage points over CTNet-Repro. The results demonstrate that the proposed method can effectively improve discriminative performance, model-selection stability, and test-stage robustness for MI-EEG classification while largely preserving the original CTNet backbone. Highlights: (1) An enhanced CTNet is proposed for four-class MI-EEG classification. (2)A lightweight token-channel gate improves discriminative high-level token features. (3) Stabilized training combines label smoothing, center loss, EMA, and Top-5 averaging. (4) Composite-score-based model selection improves robustness under small-sample settings. (5) Logits-level post-fusion further boosts Accuracy, Macro-F1, and Balanced Accuracy.

Keywords: Brain-computer interface; Motor imagery; EEG; CTNet; Transformer; Stability-oriented training; Model fusion.

1. Introduction

Brain-computer interfaces (BCIs) enable direct interaction between humans and external devices by decoding brain signals and therefore have broad practical value in rehabilitation assistance, intelligent control, and human–computer interaction. Among different paradigms, motor imagery (MI) based on electroencephalogram (EEG) signals has become one of the most widely studied because of its noninvasiveness, relatively low cost, and high temporal resolution. However, MI-EEG decoding still faces several long-standing challenges. First, EEG signals usually suffer from a low signal-to-noise ratio, and clear differences exist across subjects and sessions. Second, in four-class MI tasks, the decision boundaries among some categories are not sufficiently clear, which easily leads to inter-class confusion. Third, under the limited sample size of public datasets, deep models may have stronger representational power, but they are also more likely to exhibit substantial validation fluctuations, unstable model selection, and notable variation in test performance. [23-26]

To address the above issues, MI-EEG classification methods have gradually evolved from traditional feature engineering to end-to-end deep learning frameworks. Recent studies show that temporal modeling, multi-scale representation, and attention mechanisms remain important directions for performance improvement. Meanwhile, with the rapid development of self-attention and Transformers in sequence modeling, increasing attention has been paid to hybrid architectures that combine local convolutional feature extraction with global dependency modeling. Representative methods, including EEG Conformer, hierarchical Transformer, MSVTNet, CTNet, compact convolutional Transformer, and two-stage Transformer, collectively indicate that the combination of convolution and Transformer can effectively balance local pattern extraction and global relation modeling, and has become an important technical route for current MI-EEG classification. [1-15]

Nevertheless, most existing improvements are concentrated on backbone design itself, whereas relatively limited attention has been paid to model-selection fluctuations under small validation sets, parameter instability in late-stage training, and the complementary use of different models. For EEG tasks characterized by small sample sizes and large inter-subject differences, improving the expressive power of the backbone alone does not necessarily translate into stable final test performance. Therefore, this study is built upon the CTNet framework and does not pursue complex backbone reconstruction; instead, it introduces coordinated improvements from three aspects: lightweight structural enhancement, stability-oriented training, and post-fusion inference.

The main contributions of this study are as follows:

(1) A lightweight token-channel gate is introduced into the CTNet framework to adaptively recalibrate high-level token features and strengthen the representation of key EEG responses.

(2) A stability-oriented training mechanism for small-sample MI-EEG classification is constructed by combining label smoothing, center loss, exponential moving average of parameters, Top-5 checkpoint averaging, composite-score-based model selection, and lightweight input augmentation, thereby improving training and testing stability.

(3) A logits-level post-fusion inference mechanism based

on model complementarity is designed to integrate the discriminative advantages of the CTNet-Repro baseline model and the enhanced model on different samples and to further improve final performance.

(4) Systematic experiments are conducted under a unified data protocol and reproducible experimental settings, and the results are analyzed from the perspectives of overall performance and module contribution.

2. Related Work

2.1. Motor Imagery EEG Classification Methods

Motor imagery EEG classification is a classic task in brain-computer interface research. Recent survey studies have systematically summarized the input construction, network architectures, evaluation protocols, and performance comparison issues in deep-learning-based MI-EEG decoding, and have pointed out that performance differences on public datasets are often related not only to the models themselves but also to preprocessing procedures, data partition strategies, and training details. At the method level, deep temporal networks, multi-scale channel-temporal attention, and wavelet-scattering-based feature fusion have improved motor imagery classification from the perspectives of temporal modeling, attention enhancement, and feature construction, respectively. [13,23,24,28,29]

2.2. Convolution- and Transformer-Based EEG Classification Methods

In recent years, Transformers have been gradually introduced into EEG classification tasks to enhance long-range temporal dependency modeling. To simultaneously exploit local feature extraction and global relation modeling, hybrid architectures that combine convolution and Transformer have become an important research direction. Methods such as EEG Conformer, hierarchical Transformer, local-global convolutional Transformer, MSVTNet, CTNet, subject-independent compact convolutional Transformer, two-stage Transformer, and EEG-VTTCNet have advanced this line of research from the perspectives of unified encoding, multi-scale feature extraction, lightweight backbones, and subject-independent generalization. Overall, the fusion of convolution and Transformer has become one of the mainstream technical routes in MI-EEG classification. [1-15]

2.3. Generalization and Transfer Learning Methods

In addition to classification models themselves, reducing distribution differences across subjects and sessions is also a key issue in EEG classification research. In recent years, methods such as dual-attention adversarial transfer, graph-structured domain adaptation, Wasserstein-distance-based alignment, correlation alignment, multi-stage transfer learning, and multi-source domain adaptation ensembles have improved cross-subject or cross-dataset performance from the perspectives of adversarial learning, graph modeling, statistical alignment, and transfer strategies. It should be noted that the present work mainly focuses on robust within-subject classification under a subject-dependent protocol rather than directly targeting cross-subject transfer. Therefore, this study is not intended to replace transfer-learning methods, but instead to improve model stability and system robustness under a unified training protocol. [16-22]

3. Method

3.1. Overall Framework

This study addresses the four-class motor imagery EEG classification task on the BCI Competition IV 2a dataset and proposes an enhanced CTNet method integrating lightweight structural enhancement, stability-oriented training, and complementary fusion inference. The overall workflow is as follows. First, EEG trials of 4 s are segmented from the original GDF files according to cue events, and artifact-contaminated trials are removed. Next, exponential moving standardization is applied to each trial to construct the model input. On the model side, a convolutional front-end is used to extract local temporal and spatial patterns, and the convolutional output is then mapped into a token sequence and fed into a Transformer encoder to model higher-level global dependencies. In the enhanced model, a lightweight token-channel gate is further introduced to adaptively recalibrate intermediate features, and a two-layer classification head is adopted for discrimination. During training, label smoothing, center loss, exponential moving average of parameters, Top-5 checkpoint averaging, composite-score-based model selection, and lightweight input augmentation are introduced to improve training stability, model-selection stability, and test-stage robustness. Finally, during inference, the logits of the CTNet-Repro baseline model and the enhanced model are fused in a weighted manner to obtain the final prediction.

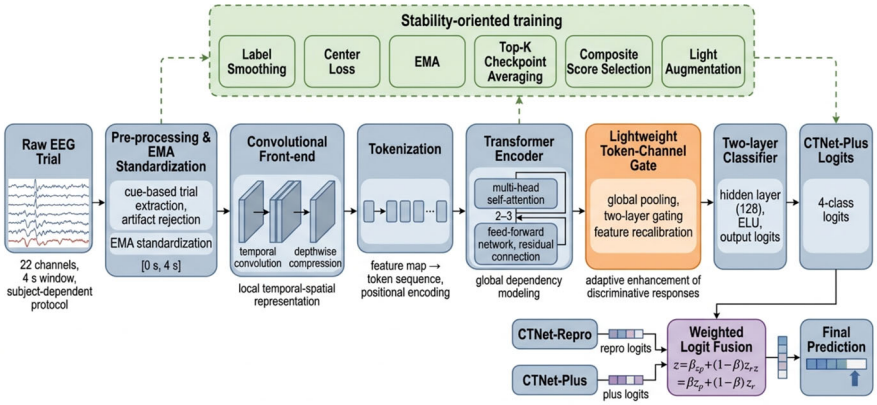

Figure 1. Overall framework of the proposed enhanced CTNet.

As shown in Figure 1, the proposed method does not require major reconstruction of the CTNet backbone. Instead,

it improves final performance through a coordinated three-part design—lightweight structural enhancement, stability-oriented training, and post-fusion inference—built on top of the original convolution–Transformer data flow. Specifically, the convolutional front-end, tokenization, and Transformer encoder establish local temporal-spatial representations and global dependency relations; the token-channel gate and the two-layer classification head enhance high-level discriminative representations; the stability-oriented training module runs throughout the training process to suppress model fluctuations under small-sample conditions; and at the test stage, final predictions are obtained by weighted fusion of the logits produced by CTNet-Repro and CTNet-Plus.

After cue-driven segmentation, artifact rejection, and EMA standardization, the original EEG trials are sequentially passed through the convolutional front-end, tokenization, and Transformer encoder. The lightweight token-channel gate and the two-layer classification head then produce the four-class logits of CTNet-Plus. During training, label smoothing, center loss, EMA, Top-5 checkpoint averaging, composite-score-based model selection, and lightweight input augmentation are jointly used to improve the stability of the model formation process. During testing, the logits of CTNet-Repro and CTNet-Plus are further fused in a weighted manner to output the final prediction.

Unlike common schemes that improve only the backbone, the proposed method consists of three jointly designed components: lightweight structural enhancement on the backbone, a stability-oriented training mechanism for small-sample EEG tasks, and lightweight inference fusion based on model complementarity. These three components together form the complete implementation framework of the proposed method.

3.2. CTNet Baseline Framework

$\mathbf{X}CTC = 22T = 1000$Let an input EEG trial be denoted by a two-dimensional temporal matrix, where represents the number of EEG channels and denotes the temporal length. In the current implementation, , the sampling rate is 250 Hz, and the time window length is 4 s; therefore, . The CTNet baseline framework consists of a convolutional feature extraction module, a Transformer encoding module, and a classification module. Its core idea is to use convolutional structures to extract local temporal and spatial correlation patterns while using the Transformer to model global dependencies among tokens within the same model:

$$\mathbf{X} \in \mathbb{R}^{C\times T}, \quad C = 22, \quad T = 1000.$$

$F_{\text{conv}}(\cdot)$ The convolutional front-end first performs lightweight convolutional encoding on the input. The first stage conducts temporal convolution to extract local temporal dynamics; the second stage applies depthwise separable cross-channel convolution to model spatial correlations; and the third stage continues temporal convolution and combines it with pooling to compress the temporal length, thereby obtaining a convolutional feature tensor. Denoting the convolutional encoding process by , the convolutional feature can be written as

$$\mathbf{H}_c = F_{\text{conv}}(\mathbf{X}).$$

NDThe model then rearranges the convolutional feature into a token sequence with length and token feature dimension . This process essentially reorganizes the convolutional output into a sequence form suitable for self-attention, enabling the model to capture relations among different local features from a sequence perspective:

$$\mathbf{H} = \text{Tokenize}(\mathbf{H}_c), \quad \mathbf{H} \in \mathbb{R}^{N\times D}.$$

$\mathbf{P}$After obtaining the token sequence, the model adds a learnable positional encoding and feeds the result into the Transformer encoder. Let the positional encoding be ; then the Transformer input can be written as

$$\mathbf{H}_0 = \mathbf{H} + \mathbf{P}.$$

The Transformer models global relations among tokens through multi-head self-attention, whose general form is

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d}}\right)\mathbf{V}.$$

$F_{\text{trans}}(\cdot)\mathbf{H}_t = F_{\text{trans}}(\mathbf{H}_0)$ Let the encoder mapping be ; then the encoded result is . In the current implementation, the final high-level representation used for classification adopts residual fusion, namely

$$\mathbf{H}_f = \mathbf{H}_t + \mathbf{H}.$$

The CTNet-Repro baseline model flattens the fused token feature and feeds it into a single-layer linear classifier to obtain four-class logits:

$$\mathbf{z}_r = \mathbf{W}_r \cdot \text{vec}(\mathbf{H}_f) + \mathbf{b}_r.$$

Overall, the CTNet baseline framework extracts local temporal and spatial patterns through lightweight convolutional modules and then models global token dependencies through the Transformer, thereby adapting well to the structural characteristics of motor imagery EEG signals. However, the baseline framework alone still has two limitations. First, different tokens or high-level feature dimensions do not contribute equally to classification, whereas the baseline model does not explicitly model such differences. Second, under small-sample and small-validation-set conditions, the training and model-selection processes are prone to fluctuation. Based on this, the present study further introduces structural enhancement and stability-oriented training on top of the baseline framework.

3.3. Lightweight Token-Channel Gate and Enhanced Classification Head

$ND\mathbf{h}$The enhanced model corresponds to CTNetPaperPlus in the code, and has two core modifications: enabling a lightweight token-channel gate and replacing the single-layer linear classification head with a two-layer head containing a hidden layer. Together, these two modifications constitute the main body of the enhanced CTNet. Suppose the token representation output by the convolution–Transformer backbone has length and feature dimension. The gating module first performs global aggregation along the token dimension to obtain a compact representation:

$$\mathbf{h} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{H}_{f,i}.$$

$\mathbf{g}$The gate weight vector corresponding to the feature dimension is then generated by two fully connected layers:

$$\mathbf{g} = \sigma(\mathbf{W}_2\,\delta(\mathbf{W}_1\mathbf{h} + \mathbf{b}_1) + \mathbf{b}_2),$$

$\delta(\cdot)$ Here, denotes the ELU activation function and represents the Sigmoid function. Finally, the gate coefficients are used to recalibrate the original token features:

$$\widehat{\mathbf{H}}_f = \mathbf{H}_f \odot \mathbf{g}.$$

It should be emphasized that although the module name contains the expression “channel gate,” its operating object is not the original EEG channels, but the high-level token feature dimensions that have already fused channel information. Therefore, we refer to it as a lightweight token-channel gate, which more accurately reflects its functional essence.

The second modification of the enhanced model lies in the classification head. The CTNet-Repro baseline model uses a single-layer linear classifier, whereas the enhanced model first passes the flattened feature through a hidden layer and an ELU activation and then performs final classification. Let the flattened feature be ; then the enhanced classification head can be expressed as

$$\mathbf{u} = \delta(\mathbf{W}_h \mathbf{x}_f + \mathbf{b}_h),$$
$$\mathbf{z}_p = \mathbf{W}_p \mathbf{u} + \mathbf{b}_p.$$

Therefore, the enhanced model does not redesign a new large-scale backbone, but improves feature representation quality through the lightweight gate and the enhanced classification head while keeping the CTNet data flow unchanged. This design controls the magnitude of structural changes and is also more suitable for the "moderate enhancement rather than excessive stacking" requirement in small-sample MI-EEG tasks.

3.4. Stability-Oriented Training Mechanism

In the current implementation, performance improvement does not rely solely on structural enhancement but also strongly depends on the stability-oriented design during training. Compared with many methods that emphasize only the backbone, the present study explicitly treats training stability as part of the method design.

3.4.1. Label Smoothing and Center Loss

The classification loss consists of cross-entropy with label smoothing and center loss. For the enhanced model, the label smoothing factor is set to 0.05 and the center-loss weight is set to 0.010; for the CTNet-Repro baseline model, the corresponding values are 0.02 and 0.005. The total loss is written as

$$\mathcal{L} = \mathcal{L}_{\text{ce}}^{\text{ls}} + \lambda \mathcal{L}_{\text{center}},$$

$\mathcal{L}_{\text{ce}}^{\text{ls}} \mathcal{L}_{\text{center}} \lambda$where denotes the cross-entropy loss with label smoothing, denotes the center loss, and is the center-loss weight. The feature constrained by the center loss comes from the global average representation after token fusion, which helps improve intra-class compactness and enlarge inter-class distances.

3.4.2. Exponential Moving Average of Parameters

During training, if exponential moving average (EMA) of parameters is enabled, the EMA parameters are updated using the current model parameters after each optimization step. The update formula is

$$\boldsymbol{\theta}_t^{\text{ema}} = \alpha \boldsymbol{\theta}_{t-1}^{\text{ema}} + (1-\alpha)\boldsymbol{\theta}_t,$$

$\alpha = 0.999$where is the smoothing coefficient, and in the current configuration. During validation, the EMA parameters are preferentially used for model evaluation, thereby reducing the effect of late-stage parameter oscillation on model selection.

3.4.3. Top-5 Checkpoint Averaging and Composite-Score-Based Model Selection

STo reduce the randomness introduced by selecting a single instantaneous best model under small validation sets, this study does not rely directly on validation accuracy alone for model selection. Instead, it uses a composite score jointly constructed from validation Accuracy, Macro-F1, and Balanced Accuracy to rank models. Let the composite score be ; then it can be written as

$$S = \omega_1 \,\text{Acc}_{\text{val}} + \omega_2 \,\text{F1}_{\text{macro,val}} + \omega_3 \,\text{BAcc}_{\text{val}}, \quad \omega_i \geq 0, \quad \sum_i \omega_i = 1.$$

During training, the Top-5 checkpoints with the highest scores are retained. After training, if parameter averaging is enabled, the parameters of these five checkpoints are averaged and the resulting average model is used as the final test model. Therefore, the final reported test results come from the Top-5 checkpoint-averaged model rather than from a single-epoch model.

3.4.4. Lightweight Input Augmentation

Training data augmentation mainly adopts online lightweight input augmentation, including weak Gaussian noise, small temporal shifts, and random channel dropout. For the enhanced model, the noise standard deviation, maximum temporal shift, and channel-dropout probability are set to 0.01, 12, and 0.10, respectively; for the CTNet-Repro baseline model, they are 0.008, 8, and 0.08. The augmentation is applied online for each sampled instance to improve the model's robustness to noise perturbation, temporal offset, and local channel absence.

In addition, the code retains an interface for segment recomposition augmentation, but under the main experimental configuration this interface is not a major source of gain. In other words, the core performance improvements reported in this paper mainly come from lightweight structural enhancement, stability-oriented training, and post-fusion inference rather than from large-scale data synthesis.

3.4.5. Early Stopping Strategy

Early stopping is enabled during training, with the patience value set to 20. Training is terminated in advance when the validation metric no longer improves for a number of consecutive epochs, so as to avoid performance fluctuation and overfitting caused by ineffective training epochs.

3.5. Complementary Fusion Inference

Final fusion is performed at the test stage. Let the logits output by the enhanced model be and those output by the CTNet-Repro baseline model be ; then the fusion result is written as

$$\mathbf{z} = \beta \mathbf{z}_p + (1-\beta)\mathbf{z}_r,$$

where is the fusion weight? In the current implementation, , meaning that the enhanced model is assigned a higher weight. The fused logits are converted to the final prediction category through. It should be emphasized that this fusion is not jointly optimized during training; rather, it is a post-fusion inference strategy executed at the test stage. The design is simple to implement and incurs little additional cost, while effectively exploiting the decision complementarity between the CTNet-Repro baseline model and the enhanced model on different samples.

As shown in Figure 2, the three key components of the proposed method can be further decomposed into independent yet coordinated mechanism layers. Panel (a) illustrates the computation process of the lightweight token-channel gate: high-level token features are first globally aggregated, then two fully connected layers and a Sigmoid function are used to generate gating coefficients, and finally the original high-level features are recalibrated dimension by dimension. Panel (b) presents the components of the stability-oriented training mechanism, including label smoothing, center loss, EMA, Top-5 checkpoint averaging, composite-score-based model selection, and lightweight input augmentation; these modules jointly act on the training and model-selection processes. Panel (c) shows the complementary fusion inference procedure at the test stage,

where the logits of CTNet-Repro and CTNet-Plus are fused in a weighted manner to obtain a more robust final decision. Together with Figure 1, this figure forms a pair of "overall framework + key mechanism decomposition," which helps explain the proposed method from both the system level and the module level.

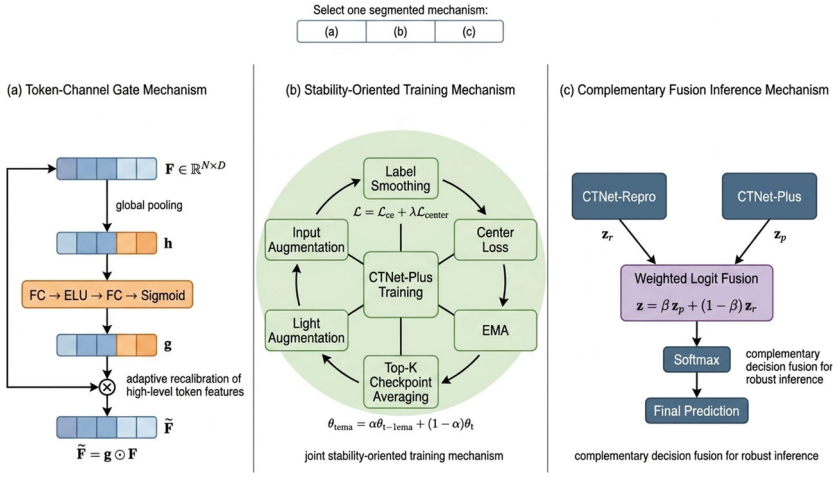

Figure 2. Decomposition of the three key mechanisms in the proposed method

(a) The lightweight token-channel gate enhances high-level token representations through "global aggregation + two-layer gating mapping + feature recalibration." (b) The stability-oriented training mechanism consists of label smoothing, center loss, EMA, Top-5 checkpoint averaging, composite-score-based model selection, and lightweight input augmentation, and is used to reduce parameter oscillation and model-selection fluctuation in small-sample EEG training. (c) The complementary fusion inference strategy performs weighted post-fusion on the logits of CTNet-Repro and CTNet-Plus at the test stage, thereby improving the stability and robustness of final predictions by exploiting the decision complementarity of the two models.

4. Experiments and Results Analysis

4.1. Dataset and Experimental Settings

The proposed method is validated on the BCI Competition IV 2a dataset. This dataset contains 9 subjects, 22 EEG channels, and 4 motor imagery classes. Data reading and trial construction are carried out by extracting the four MI cue events according to event codes 769/770/771/772 and segmenting EEG trials of 4 s starting from the cue onset, namely cue_offset_sec = 0.0 and window_sec = 4.0. For both training and test trials, those marked by artifact events are removed.

Table 1. Training and implementation parameters.

Item	CTNet-Repro	CTNet-Plus
Input and protocol	raw EEG; subject-dependent; AxxT training / AxxE testing; val_ratio = 0.2	same as left
Standardization	ema_standardize: factor_new = 0.001, init_block_size = 250, eps = 1e-4	same as left
Backbone parameters	d_model = 16, temporal_filters = 8, depth_multiplier = 2, n_heads = 2, n_layers = 3, dropout = 0.4	same as left
Optimizer	AdamW, lr = 0.0003, weight_decay = 0.0001	same as left
Epochs / batch size	120 / 32	120 / 32
Scheduler	warmup_cosine, warmup_epochs = 10, min_lr_ratio = 0.05	same as left
Label smoothing	0.02	0.05
Center-loss weight	0.005	0.010
EMA	use_ema = true, ema_decay = 0.999	same as left
Top-5 and model selection	Top-5 checkpoint averaging; composite-score-based model selection	same as left
Lightweight input augmentation	noise_std = 0.008; max_shift = 8; channel_dropout_p = 0.08	noise_std = 0.01; max_shift = 12; channel_dropout_p = 0.10
Post-fusion	Not applicable	weighted post-fusion with CTNet-Repro, β = 0.65

The main experiment adopts a subject-dependent setting. For each subject, samples from session T are used as the

training source and samples from session E are used as the test set; meanwhile, 20% of the training-session samples are randomly selected in a stratified manner as the validation set. In other words, one model is trained for each subject, where the training set is drawn from that subject's training session, the test set is drawn from the corresponding test session, and the validation set is a stratified subset within the training session. This protocol is consistent with the objective of this study, which focuses on within-subject classification performance and model stability under a unified data protocol.

The input feature mode is the raw time-domain EEG signal. In preprocessing, exponential moving standardization (ema_standardize) is applied to each trial with factor_new = 0.001, init_block_size = 250, and eps = 1e-4. Because this standardization is performed independently for each trial, it does not introduce additional information based on full-dataset statistics.

In terms of model settings, the CTNet-Repro baseline model and the enhanced model share the same backbone: d_model = 16, temporal_filters = 8, depth_multiplier = 2, temporal_kernel = 64, spatial_kernel = 16, pool1 = 8, pool2 = 6, n_heads = 2, n_layers = 3, ff_mult = 4, and dropout = 0.4. Compared with the baseline model, the main difference of the enhanced model is that it enables the token-channel gate and sets the hidden dimension of the classification head to 128.

In terms of training settings, both models use the AdamW optimizer with a learning rate of 0.0003, a weight decay of 0.0001, a batch size of 32, and a maximum of 120 training epochs. The learning-rate scheduler is warmup_cosine, where the warmup epochs are set to 10 and the minimum learning-rate ratio is 0.05. Both models enable EMA, Top-5 parameter averaging, composite-score-based model selection, and early stopping. Finally, a complete system output is formed at the test stage by weighted fusion of the logits from the CTNet-Repro baseline model and the enhanced model.

4.2. Evaluation Metrics

This study uses Accuracy and Macro-F1 as the primary evaluation metrics and further reports Balanced Accuracy as a supplementary metric. Accuracy measures the overall classification correctness; Macro-F1 computes the equally weighted average of the F1 scores of all classes and can better reflect balanced recognition ability in multi-class tasks; Balanced Accuracy is used to characterize the overall balance at the class-recall level. In Table 2, Std (%) denotes the standard deviation of subject-wise Accuracy across 9 subjects.

4.3. Comparison Experiments

To verify the effectiveness of the proposed method, traditional EEG classification methods, classical deep-learning models, and representative convolution–Transformer-based methods in recent years are selected as comparison baselines. All deep models are trained and tested under the same data partition, preprocessing procedure, and evaluation protocol. For traditional methods such as FBCSP+SVM, their standard training pipelines are implemented under the same data partition.

Table 2. Classification results of different methods on the BCI Competition IV 2a dataset.

Method	Acc (%)	Macro-F1 (%)	BAcc (%)	Std (%)
FBCSP + SVM	66.24	65.81	65.53	10.44
ShallowConvNet	69.53	68.97	69.12	9.63
DeepConvNet	71.08	70.46	70.83	9.18
EEGNet	73.42	72.88	73.01	8.74
EEG Conformer	75.36	74.95	75.10	8.37
CTNet-Repro	76.11	75.48	75.92	8.21
CTNet-Plus	79.24	78.83	79.01	7.54
Ours (Plus + Ensemble)	80.63	80.17	80.41	7.12

As shown in Table 2, traditional feature-engineering methods achieve relatively limited performance on this four-class motor imagery task, indicating the high intrinsic difficulty of the task. Compared with traditional methods, deep-learning-based models perform better overall, suggesting that end-to-end feature learning can more effectively extract discriminative patterns from EEG signals. Among them, models combining convolution and Transformer further outperform purely convolutional networks, indicating that the combination of local feature extraction and global dependency modeling is beneficial for motor imagery EEG classification.

As shown in Table 2, CTNet-Repro already achieves a strong baseline result. Further, the proposed CTNet-Plus outperforms CTNet-Repro in both Accuracy and Macro-F1 by 3.13 and 3.35 percentage points, respectively, indicating that the lightweight token-channel gate and the enhanced classification head can improve high-level representation quality, while the stability-oriented training mechanism helps reduce fluctuations during model selection. With the addition of post-fusion inference, the best result is obtained, with an average Accuracy of 80.63% and a Macro-F1 of 80.17%, showing that the decision patterns of the CTNet-Repro baseline model and the enhanced model do not completely overlap across samples and that logits-level fusion can further integrate their complementary advantages.

4.4. Ablation Experiments

To further analyze the contribution of each key design to model performance, ablation experiments are conducted around the lightweight token-channel gate, center loss, exponential moving average of parameters, Top-5 checkpoint averaging, composite-score-based model selection, lightweight input augmentation, and post-fusion inference. It should be noted that, to control training fluctuations and ensure the interpretability of the ablation comparison, CTNet-Repro in Table 3 denotes the CTNet backbone reproduced under the same data protocol and the same basic stability-oriented training framework as in the main experiments, meaning that it still retains the shared optimization control items. In other words, Table 3 is intended to evaluate the marginal gains of the key enhancement modules under a shared stabilized training control, rather than to compare against the raw original CTNet implementation. Label smoothing is treated as a basic setting for all enhanced training configurations and is therefore not listed as a separate ablation item.

As shown in Table 3, removing the token-channel gate leads to decreases in both Accuracy and Macro-F1, indicating that this module can effectively strengthen the model's focus on key high-level responses and positively contributes to the improvement of feature representation quality. Removing the center loss causes a more pronounced drop in Macro-F1, suggesting that the center constraint plays an important role in improving intra-class compactness and inter-class separability.

Table 3. Ablation results of the proposed method.

Variant	Gate	Ctr	EMA	TopK	Score	Aug	Ens	Acc / Macro-F1 (%)
CTNet-Repro	×	×	√	√	√	√	×	76.11 / 75.48
w/o Gate	×	√	√	√	√	√	×	77.02 / 76.34
w/o Ctr	√	×	√	√	√	√	×	78.31 / 77.58
w/o EMA	√	√	×	√	√	√	×	78.05 / 77.21
w/o TopK	√	√	√	×	√	√	×	78.46 / 77.73
w/o Score	√	√	√	√	×	√	×	78.58 / 77.92
w/o Aug	√	√	√	√	√	×	×	78.74 / 78.08
CTNet-Plus	√	√	√	√	√	√	×	79.24 / 78.83
Full (Ensemble)	√	√	√	√	√	√	√	80.63 / 80.17

Furthermore, removing EMA or Top-5 checkpoint averaging also results in varying degrees of performance degradation, indicating that parameter smoothing and multi-checkpoint averaging can effectively reduce late-stage parameter oscillation and instability in model selection in MI-EEG scenarios with limited sample size and small validation sets. Removing composite-score-based model selection results in a slight decrease in overall Accuracy but a more evident decrease in Macro-F1, suggesting that relying solely on a single validation metric cannot sufficiently guarantee balanced recognition performance in multi-class settings.

In addition, removing lightweight input augmentation also reduces performance, indicating that conservative augmentation methods such as weak Gaussian noise, small temporal shifts, and random channel dropout can improve the model's adaptability to noise perturbation, local distortion, and temporal offset without destroying the original EEG patterns. Finally, when post-fusion inference is further added on top of the complete enhanced model, Accuracy and Macro-F1 increase by another 1.39 and 1.34 percentage points, respectively, demonstrating that effective decision complementarity indeed exists between the CTNet-Repro baseline model and the enhanced model.

4.5. Discussion of Results

From the overall results, models combining convolution and Transformer generally outperform traditional feature-engineering methods and purely convolutional networks, indicating that jointly modeling local temporal patterns and global dependency relations is important for motor imagery EEG classification. The convolutional front-end effectively extracts local temporal and spatial patterns from EEG, whereas the Transformer encoder further integrates long-range relations among different local features. Their combination can better adapt to the complex structural characteristics of MI-EEG signals.

The results further show that the performance gain does not come solely from a single structural module. Unlike strategies that emphasize only backbone modification, the present study jointly designs structural enhancement, stability-oriented training, and post-fusion inference as one system. First, the lightweight gate and the enhanced classification head improve the discriminability of high-level representations. Second, EMA, Top-5 parameter averaging, and composite-score-based model selection jointly improve the stability of the final model formation process. Third, post-fusion inference further exploits the decision complementarity of different models. Therefore, the improvement of the proposed method should be understood as the integrated gain of "lightweight enhancement + stable training + lightweight system fusion" rather than as an isolated gain brought by a single module.

It should be noted that the experiments in this paper are conducted only under the subject-dependent protocol of BCI Competition IV 2a, and have not yet been extended to cross-subject transfer, cross-dataset generalization, or online deployment scenarios. Therefore, the conclusions of this paper should be limited to the current protocol: the proposed method can effectively improve the overall performance and model stability of within-subject four-class MI-EEG classification.

5. Conclusion

To address the limited feature discriminability, substantial training fluctuations, and limited decision stability during inference in motor imagery EEG classification, this paper proposes an enhanced CTNet method integrating lightweight gating, stability-oriented training, and complementary fusion. While largely preserving the original convolution–Transformer backbone of CTNet, the proposed method introduces a token-channel gate and a two-layer classification head to improve high-level representation capability; meanwhile, label smoothing, center loss, EMA, Top-5 parameter averaging, composite-score-based model selection, and lightweight input augmentation are used to improve the stability of training and model selection; finally, logits-level post-fusion of the CTNet-Repro baseline model and the enhanced model is used to further exploit their decision complementarity.

Under the subject-dependent protocol on the BCI Competition IV 2a dataset, the proposed method achieves average Accuracy, Macro-F1, and Balanced Accuracy of 80.63%, 80.17%, and 80.41%, respectively, across 9 subjects, improving CTNet-Repro by 4.52, 4.69, and 4.49 percentage points. The experimental results demonstrate that the proposed method can effectively improve discriminative performance, model-selection stability, and test-stage robustness for MI-EEG classification under a unified data protocol.

Future work may further explore finer-grained spatiotemporal joint modeling, more robust cross-subject alignment, and lightweight deployment and online

application, so as to enhance the practical value of the proposed method in real brain-computer interface systems.

References

[1] Song Y, Zheng Q, Liu B, et al. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 710-719. DOI: 10.1109/TNSRE.2022.3230250.

[2] Zhang D, Li H, Xie J, et al. MI-CAT: A Transformer-Based Domain Adaptation Network for Motor Imagery Classification[J]. Neural Networks, 2023, 165: 451-462. DOI: 10.1016/j.neunet.2023.06.005.

[3] Liu K, Yang T, Yu Z, et al. MSVTNet: Multi-Scale Vision Transformer Neural Network for EEG-Based Motor Imagery Decoding[J]. IEEE Journal of Biomedical and Health Informatics, 2024, 28(12): 7126-7137. DOI: 10.1109/JBHI.2024.3450753.

[4] Zhao W, Jiang X, Zhang B, et al. CTNet: A Convolutional Transformer Network for EEG-Based Motor Imagery Classification[J]. Scientific Reports, 2024, 14: 20237. DOI: 10.1038/s41598-024-71118-7.

[5] Keutayeva A, Fakhrutdinov N, Abibullaev B. Compact Convolutional Transformer for Subject-Independent Motor Imagery EEG-Based BCIs[J]. Scientific Reports, 2024, 14: 25775. DOI: 10.1038/s41598-024-73755-4.

[6] Zhang J, Li K, Yang B, et al. Local and Global Convolutional Transformer-Based Motor Imagery EEG Classification[J]. Frontiers in Neuroscience, 2023, 17: 1219988. DOI: 10.3389/fnins.2023.1219988.

[7] Liu M, Liu Y, Shi W, et al. EMPT: A Sparsity Transformer for EEG-Based Motor Imagery Recognition[J]. Frontiers in Neuroscience, 2024, 18: 1366294. DOI: 10.3389/fnins.2024.1366294.

[8] Shi X, Li B, Wang W, et al. EEG-VTTCNet: A Loss Joint Training Model Based on the Vision Transformer and the Temporal Convolution Network for EEG-Based Motor Imagery Classification[J]. Neuroscience, 2024, 556: 42-51. DOI: 10.1016/j.neuroscience.2024.07.051.

[9] Altaheri H, Karray F, Karimi A H. Temporal Convolutional Transformer for EEG Based Motor Imagery Decoding[J]. Scientific Reports, 2025, 15: 32959. DOI: 10.1038/s41598-025-16219-7.

[10] Zhao W, Jiang X, Zhang B, et al. Multi-Scale Convolutional Transformer Network for Motor Imagery Brain-Computer Interface[J]. Scientific Reports, 2025, 15: 96611. DOI: 10.1038/s41598-025-96611-5.

[11] Chaudhary P, et al. A Two-Stage Transformer Based Network for Motor Imagery Classification[J]. Medical Engineering & Physics, 2024, 128: 104154. DOI: 10.1016/j.medengphy.2024.104154.

[12] Deny P, Cheon S, Son H, et al. Hierarchical Transformer for Motor Imagery-Based Brain Computer Interface[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(11): 5459-5470. DOI: 10.1109/JBHI.2023.3304646.

[13] Sharma N, Upadhyay A, Sharma M, et al. Deep Temporal Networks for EEG-Based Motor Imagery Recognition[J]. Scientific Reports, 2023, 13: 18813. DOI: 10.1038/s41598-023-41653-w.

[14] Luo J, Wang Y, Xia S, et al. A Shallow Mirror Transformer for Subject-Independent Motor Imagery BCI[J]. Computers in Biology and Medicine, 2023, 164: 107254. DOI: 10.1016/j.compbiomed.2023.107254.

[15] Sartipi S, Cetin M. Subject-Independent Deep Architecture for EEG-Based Motor Imagery Classification[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024, 32: 718-727. DOI: 10.1109/TNSRE.2024.3360194.

[16] Li H, Zhang D, Xie J, et al. MI-DABAN: A Dual-Attention-Based Adversarial Network for Motor Imagery Classification[J]. Computers in Biology and Medicine, 2023, 152: 106420. DOI: 10.1016/j.compbiomed.2022.106420.

[17] Zhang D, Li H, Xie J, et al. MI-DAGSC: A Domain Adaptation Approach Incorporating Comprehensive Information of MI-EEG Signals[J]. Neural Networks, 2023, 167: 183-198. DOI: 10.1016/j.neunet.2023.08.008.

[18] She Q, Chen T, Fang F, et al. Improved Domain Adaptation Network Based on Wasserstein Distance for Motor Imagery EEG Classification[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 1137-1148. DOI: 10.1109/TNSRE.2023.3241846.

[19] Zhong X C, Wang Q, Liu D, et al. A Deep Domain Adaptation Framework with Correlation Alignment for EEG-Based Motor Imagery Classification[J]. Computers in Biology and Medicine, 2023, 163: 107235. DOI: 10.1016/j.compbiomed.2023.107235.

[20] Li J, Shi J, Yu P, et al. Feature-Aware Domain Invariant Representation Learning for EEG Motor Imagery Decoding[J]. Scientific Reports, 2025, 15: 10664. DOI: 10.1038/s41598-025-95178-5.

[21] Li J, She Q, Meng M, et al. Three-Stage Transfer Learning for Motor Imagery EEG Recognition[J]. Medical & Biological Engineering & Computing, 2024, 62(6): 1689-1701. DOI: 10.1007/s11517-024-03036-9.

[22] Miao M, Yang Z, Sheng Z, et al. Multi-Source Deep Domain Adaptation Ensemble Framework for Cross-Dataset Motor Imagery EEG Transfer Learning[J]. Physiological Measurement, 2024, 45(5). DOI: 10.1088/1361-6579/ad4e95.

[23] Wang X, Liesaputra V, Liu Z, et al. An In-Depth Survey on Deep Learning-Based Motor Imagery Electroencephalogram (EEG) Classification[J]. Artificial Intelligence in Medicine, 2024, 147: 102738. DOI: 10.1016/j.artmed.2023.102738.

[24] Hameed I, Khan D M, Ahmed S M, et al. Enhancing Motor Imagery EEG Signal Decoding Through Machine Learning: A Systematic Review of Recent Progress[J]. Computers in Biology and Medicine, 2025, 185: 109534. DOI: 10.1016/j.compbiomed.2024.109534.

[25] Vafaei E, Hosseini M. Transformers in EEG Analysis: A Review of Architectures and Applications in Motor Imagery, Seizure, and Emotion Classification[J]. Sensors, 2025, 25(5): 1293. DOI: 10.3390/s25051293.

[26] Pfeffer M A, Wong J K W, Ling S H. Trends and Limitations in Transformer-Based BCI Research[J]. Applied Sciences, 2025, 15(20): 11150. DOI: 10.3390/app152011150.

[27] Song Y, Zheng Q, Wang Q, et al. Global Adaptive Transformer for Cross-Subject Enhanced EEG Classification[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 2767-2777. DOI: 10.1109/TNSRE.2023.3285309.

[28] Wu R, Jin J, Daly I, et al. Classification of Motor Imagery Based on Multi-Scale Feature Extraction and the Channel-Temporal Attention Module[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 3075-3085. DOI: 10.1109/TNSRE.2023.3294815.

[29] Pham T D. Classification of Motor-Imagery Tasks Using a Large EEG Dataset by Fusing Classifiers Learning on Wavelet-Scattering Features[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 1097-1107. DOI: 10.1109/TNSRE.2023.3241241.

[30] Gómez-Morales Ó W, Collazos-Huertas D F, Álvarez-Meza A M, et al. EEG Signal Prediction for Motor Imagery Classification in Brain-Computer Interfaces[J]. Sensors, 2025, 25(7): 2259. DOI: 10.3390/s25072259.