HCFNet: A heterogeneous frequency bands coupling CNN for enhanced short-time fast response in motor imagery decoding

HCFNet: A heterogeneous frequency bands coupling CNN for enhanced short-time fast response in motor imagery decoding

Weijie Wu [a], Ian Daly [f], Weijie Chen [a], Lifei Liu [a], Wei Liang [a], Yixin Chen [a], Xingyu Wang [a], Andrzej Cichocki [b,c,d], Jing Jin [a,e,*]

[a] Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai, PR China
[b] Nicolaus Copernicus University(UMK), Torun 87-100, Poland
[c] RIKEN Advanced Intelligence Project, Tokyo, Japan
[d] Tokyo University of Agriculture and Technology, Tokyo, Japan
[e] School of Mathematics, East China University of Science and Technology, Shanghai 200237, PR China
[f] Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK

A B S T R A C T

Background: Motor imagery signals encompass a broad range of frequency components, and frequency band decomposition can improve the precision of frequency-domain features, helping the model focus on task-relevant information. However, existing methods often treat signals from different frequency bands uniformly, over-looking their heterogeneity and coupling, which leads to redundant features and loss of cooperative information.*New method:* We propose a HCFNet that explores heterogeneous feature extraction and coupling across frequency bands. HCFNet first separates the raw signal into high and low-frequency bands, extracting spatiotemporal features through specialized modules. A cross-frequency coupling module then fuses these features, using data augmentation for regularization to capture robust spectral-spatiotemporal features and high-low frequency coupling.
Results: We evaluated our model on the BCIC-IV-2a and OpenBMI benchmark datasets, and our model achieves average accuracies of 82.41 % and 76.52 %. Notably, HCFNet maintains excellent performance even with shorter time windows.
Comparison with existing methods: HCFNet outperforms all the state-of-the-art methods we benchmark against. Compared with traditional multi-band isomorphic methods, frequency-band heterogeneous coupling performs better in capturing task-related features and significantly reduces redundancy during feature fusion. *Conclusions:* This study significantly advances the decoding technology of motor imagery signals through an innovative frequency-band heterogeneous coupling method. Its substantial potential for rapid responses brings tangible improvements to brain-computer interface systems and is expected to be further applied in domain adaptation, cross-domain alignment, and cross-subject contexts in the future.

Keywords:
Motor Imagery; Electroencephalography; Convolution neural networks; Heterogenization; Frequency band coupling; Fast response

1. Introduction

Brain-Computer Interfaces (BCIs) capture and decode brain signals, converting them into real-time commands that facilitate direct interaction with external devices, bypassing traditional neural pathways (Chaudhary et al., 2016; McFarland and Wolpaw, 2011; Samadi et al., 2025). Non-invasive techniques such as Electroencephalography (EEG), Magnetoencephalography (MEG), and Functional Near-Infrared Spectroscopy (fNIRS) are widely considered to be safe and versatile methods for use in BCI. Among these, EEG, which records time series of cortical electrical activity, is extensively studied due to its high temporal resolution, portability, and real-time measurement capabilities (Lance et al., 2012). Motor Imagery (MI), a spontaneous EEG signal, arises from the activation of motor areas during imagined movements, offering high

practicality and autonomy (Altaheri et al., 2023a; Autthasan et al., 2022). With advancements in MI decoding technologies, MI-BCIs have found broad applications in medical fields, including speech rehabilitation (Pan et al., 2024), neurorehabilitation (Ramos-Murguialday et al., 2013), prosthetics (Xu and Wang, 2021), limb function restoration (He et al., 2015), and wheelchair control (Tang et al., 2020), as well as non-medical domains such as entertainment (Kreilinger et al., 2016), robotics (Edelman et al., 2019), and spelling (Cao et al., 2017).

However, EEG signals are inherently non-stationary, leading to substantial variations in statistical properties over time, which adversely affects long-term decoding performance. Additionally, the spontaneous nature of MI signals results in weak signal strength, a low signal-to-noise ratio, and increased susceptibility to user fatigue (Riyad et al., 2021). These challenges emphasize the need for enhanced real-time performance and robustness in MI-BCI applications. Practical implementation of accurate motor intent decoding from high-dimensional, time-varying EEG signals remains a significant challenge (Wu et al., 2022). Furthermore, a reliance on long time windows for feature extraction in many widely used EEG signal processing operations introduces delays, hindering real-time processing. Therefore, the development of decoding models with fast response times is essential for improving the performance of MI-BCI systems in real-world applications (Bai et al., 2024).

MI signals contain rich information across the time, space, and frequency domains. Initially, traditional machine learning methods were employed for MI signal decoding, where features were manually designed and classified using established pattern recognition algorithms. The Common Spatial Pattern (CSP) (Blankertz et al., 2008) is a representative traditional method that extracts spatial features related to the task via spatial filtering and distinguishes different EEG activities by maximizing the variance between tasks. The Filter Bank CSP (FBCSP) method enhances feature extraction by decomposing the signal into multiple frequency bands using a filter bank (Ang et al., 2012). Jin et al. optimized the feature selection process of the Common Spatial Pattern algorithm by incorporating the L1 norm and Dempster–Shafer theory (Jin et al., 2021). Thenmozhi et al. effectively reduced the dimensionality of features, improved accuracy and time consumption through extreme gradient Bayesian optimization (Thenmozhi and Helen, 2022). However, manually crafted features fail to fully capture the high-dimensional information in complex EEG signals, leading to inefficient signal utilization, reduced decoding performance, and limited adaptability to inter-trial variability.

MI signal decoding currently relies on deep learning-based methods, which automatically learn multi-level feature representations from raw signals and identify the most discriminative features in high-dimensional data, significantly enhancing information utilization and decoding performance (Khademi et al., 2023). In particular, inspired by the field of computer vision, Convolutional Neural Networks (CNNs) have been widely applied to MI signal decoding, demonstrating superior performance. For instance, Schirrmeister et al. explored CNN performance with various architectures and design choices for MI signal decoding, with their ShallowConvNet and DeepConvNet models providing key theoretical insights into CNN applications in MI decoding (Schirrmeister et al., 2017). Lawhern et al. further introduced EEGNet, which aims to achieve lightweight and efficient feature learning through a compact and efficient network structure (Lawhern et al., 2018).

As research progresses, more innovative approaches have emerged. For example, in (Tang et al., 2024), Laplacian operators were used to construct spatial filters for improved spatial resolution. After extracting time-domain features via temporal convolution, a graph convolution network module was employed to learn electrode connection topology. Dai et al. used mixed-scale convolutions for time feature extraction, discovering that kernel size significantly affected model performance across different participants or sessions, with mixed-scale convolutions offering a more stable network structure (Dai et al., 2020). Tang et al. combined 1D and 2D convolutions, using multi-scale 1D convolutions for time feature extraction, parallel multi-scale 2D convolutions for spatiotemporal feature extraction, followed by feature concatenation (Tang et al., 2023).

Additionally, CNNs have been combined with Recurrent Neural Networks (RNNs) for improved handling of temporal features. In (Amin et al., 2022), inception-CNN was used to generate attention vectors, and a bidirectional LSTM was constructed to extract high-level temporal features, both achieving superior performance. Zhang et al. transformed EEG into an undirected spatial graph, using CNN for spatiotemporal feature extraction and combining it with LSTM to build a recurrent attention network, effectively capturing attention-driven temporal dynamics (Zhang et al., 2020).

Recently, a new variant of CNN, the Temporal Convolutional Network (TCN), has demonstrated outstanding performance in time-series processing and has been applied to MI signal decoding. Unlike traditional RNN methods, TCNs can exponentially expand the receptive field size while ensuring that the number of parameters grows linearly, effectively addressing issues such as gradient explosion and gradient vanishing. For example, in (Ingolfsson et al., 2020a), a fusion of CNN and TCN networks was used to extract high-level temporal features, while (Musallam et al., 2021) explored further optimization through layer fusion strategies.

The attention mechanism allows networks to focus on the most relevant parts of the input data, enabling the model to prioritize significant components while disregarding others. Gao et al. proposed the MSFF-SENet model, which extracts spatiotemporal, deep-time, and multi-spectral features through MS-STM, MSTM, and PSD-Conv modules, respectively (Gao et al., 2024). The SE feature fusion module then integrates and selects features using the attention mechanism, achieving good MI decoding performance. Altaheri et al. introduced ATCNet, which, after convolutional spatiotemporal encoding, uses multi-head attention to emphasize important information in the time series, while a Temporal Convolutional Network (TCN) extracts high-level temporal features, demonstrating superior performance (Altaheri et al., 2023a). Later, Altaheri and colleagues combined the attention dynamic convolution module, with three dynamic convolution layers providing improved low-level spatiotemporal features for the ATC module (Altaheri et al., 2023b).

The Transformer architecture, combining multi-head attention mechanisms with positional encoding, offers key advantages such as parallelization, global dependency capture, and flexibility. In (Zhang et al., 2023), the self-attention and cross-attention mechanisms of the Transformer addressed domain differences, advancing its application in MI decoding. In (Zhao et al., 2024), CTNet was proposed, which, after feature extraction using a spatiotemporal convolution module, leverages the Transformer architecture to capture global dependencies of high-level features, achieving state-of-the-art performance.

Band decomposition methods are widely used in deep learning algorithms, where spectral filters decompose raw signals into multiple frequency bands for network input or multi-dimensional feature extraction, significantly enhancing decoding performance (Jin et al., 2024). For example, Liu and colleagues generated narrowband multi-view representations of EEG, extracting spatiotemporal features separately (Liu et al., 2023). Goa and colleagues, used the PSD-Conv module to compute power spectral densities across multiple frequency bands, effectively extracting spectral features using a CNN (Gao et al., 2024). Meanwhile, Cross-Frequency Coupling (CFC), which reveals interactions between frequency bands, has gained attention for its role in coupling relationships and fusion strategies of multi-view features. CFC refers to the interaction between EEG signals, especially between high and low frequencies, and plays a significant role in cognitive function evaluation, psychiatric disorder diagnosis, and regulation (Canolty et al., 2006; Jensen and Colgin, 2007). Recent studies suggest that CFC features may also have potential applications in motor imagery (Combrisson et al., 2017). In work by Zhang and Li, EEG signals were filtered into five frequency bands for network input to extract spatiotemporal features, with feature fusion achieved using dense blocks and

cascading strategies (Zhang and Li, 2023). Wang et al. proposed IFNet, which decomposes signals into high and low-frequency bands and examines how their interactions affect model performance (Wang et al., 2023).

However, current band decomposition methods often use the same structure to extract features from different frequency bands, which overlooks the unique characteristics of each frequency band and may result in underutilization of discriminative information in some bands. Additionally, many deep learning models implicitly learn interactions between frequency bands, which reduces interpretability. Simple addition and concatenation for feature fusion may introduce redundant information, negatively affecting decoding accuracy. Lastly, rapid response is a critical requirement for MI signal decoding for BCI. Many existing deep learning models typically rely on long time windows (e.g., 4 s) for feature extraction, leading to slower response times. Therefore, research on the sensitivity of the model to time window length is necessary to achieve the highest possible information transfer rate.

Based on the above challenges, this paper proposes a Convolutional Neural Network based on frequency band heterogeneous coupling (HCFNet). The model first extracts primary spatiotemporal features from high and low-frequency bands using heterogeneous feature extraction modules. A feature fusion module is then introduced, which employs heterogeneous operations to transform features into high-level spatiotemporal representations and fuse them, better learning the coupling relationship between frequency bands and generating more representative and discriminative fused features. Additionally, by combining repetition trial enhancement and designed pooling operations, the model significantly reduces dependence on the time window length, achieving a balance between recognition speed and accuracy. The performance of the HCFNet model was validated on the BCI Competition IV 2a (BCIC-IV-2a) dataset and the OpenBMI dataset, and compared with other advanced algorithms.

Notably, the core innovation of "heterogeneous frequency band coupling" in HCFNet differs fundamentally from existing multi-band or cross-frequency approaches. Existing multi-band methods typically adopt a homogeneous feature extraction paradigm: they decompose signals into multiple bands but use identical network structures (e.g., same convolution type, kernel size, fusion strategy) for all bands, overlooking the inherent differences in temporal-spatial characteristics between frequency components. Cross-frequency methods focus on interactions between bands but lack dedicated adaptation to the unique properties of each band, leading to insufficient exploitation of band-specific discriminative information. In contrast, HCFNet's "heterogeneous" design is multi-dimensional: it not only employs different convolution types for high and low frequencies but also customizes the entire feature extraction logic, feature abstraction granularity, and cross-band fusion mechanism to match the intrinsic characteristics of each frequency band. This holistic heterogeneous coupling ensures that each band's unique information is fully exploited while avoiding redundant feature overlap, which distinguishes it from prior works that only adjust superficial network components.

The main contributions of this paper are as follows:

1) We propose the HCFNet model, which is based on frequency band heterogeneous coupling and utilizes a heterogeneous feature extraction structure to capture more discriminative, band-specific features.
2) A feature fusion module is proposed, which employs heterogeneous operations to generate and fuse high-level spatiotemporal features, effectively learning the coupling relationship between high and low-frequency signals, thereby reducing redundant components.
3) We present extensive validation on two datasets.

The remainder of the paper is organized as follows: Section 2 describes the proposed HCFNet model. Sections 3 and 4 present the experimental setup and analyze the results. Section 5 discusses the model, and Section 6 concludes the paper.

2. Method

2.1. EEG representation

We define the raw EEG signal as $S = \{[X_i, y_i] i = 1, 2, \ldots, N\}$, where $X_i \in \mathbb{R}^{C \times T}$ represents a multichannel EEG sample for a single experiment, with C being the number of EEG channels, T being the number of time points, and N being the total number of samples. y_i denotes the label for the i-th sample. For example, in the BCIC-IV-2a dataset, a single epoch of EEG is 22×1000, and the total number of epochs, N, is 5184, with $y_i \in \{1: \text{“left hand”}, 2: \text{“right hand”}, 3: \text{“feet”}, 4: \text{“tougue”}\}$. For the OpenBMI dataset, $C = 20$, $T = 1000$, $N = 21600$, and $y_i \in \{1: \text{“left hand”}, 2: \text{“right hand”}\}$. As the article is based on the heterogeneous approach between frequency bands, we divide the EEG signals into two frequency bands. Brain oscillations are generally classified into specific frequency bands (delta: <4 Hz, theta: 4–7 Hz, alpha: 8–12 Hz, beta: 12–30 Hz, gamma: >30 Hz). To cover as much information as possible, we apply a 5th-order Butterworth bandpass filter to the EEG signal, separating it into high-frequency (16–40 Hz) and low-frequency (0.5–16 Hz) signals.

This frequency division scheme is designed based on the physiological characteristics of motor imagery (MI)-related EEG components and the heterogeneous processing logic of the model. Classic electrophysiological studies have shown that the core discriminative information of MI signals is mainly carried by the mu band (8–13 Hz), beta band (12–30 Hz), and low gamma band (30–40 Hz), whose activities directly reflect the activation, regulation, and fine intention representation of the motor cortex (Blankertz et al., 2008; Combrisson et al., 2017). Among them, the low beta band (12–16 Hz) cooperates with the alpha band to participate in the initiation of global motor intentions, while the high beta band (16–30 Hz) collaborates with the gamma band to support the fine regulation of local movements. This functional heterogeneity provides the core basis for the 16 Hz boundary (Ang et al., 2012). Accordingly, the 0.5–16 Hz band covers delta, theta, alpha, and low beta components to capture global cortical coordination and basic motor preparation signals; the 16–40 Hz band focuses on the high beta and low gamma bands to extract local dynamic regulation features. This division not only retains the functional relevance of each sub-band but also avoids cross-redundancy of features with different attributes, which is highly consistent with the subsequent heterogeneous extraction logic of "low-frequency global features processed by 2D convolution and high-frequency local features processed by 1D convolution". Compared with existing schemes such as "high frequency > 30 Hz" or "medium--high frequency 12–30 Hz" (Jensen and Colgin, 2007; Wang et al., 2023), it better achieves effective separation of "global-local" features, improving fusion efficiency and decoding performance.

2.2. Architecture of HCFNet

The core innovation of HCFNet lies in "heterogeneous frequency band coupling," a method fundamentally distinct from existing multi-band or cross-frequency approaches by tailoring the entire feature processing pipeline to the intrinsic properties of high and low-frequency bands. First, we rigorously define "heterogenization" from both signal processing and neuroscientific perspectives: in signal processing, it refers to the adaptive customization of feature extraction and fusion strategies based on the inherent differences in spatiotemporal-spectral characteristics among different frequency bands; in neuroscience, it corresponds to the functional division of brain oscillations—different frequency bands mediate unique neural computation processes, requiring targeted design of network architectures to capture their discriminative information.

Guided by this definition, the heterogeneous design for high and low-

frequency bands is grounded in principled theoretical underpinnings rather than mere empirical observation. Unlike conventional methods that adopt homogeneous structures for all bands or rely on superficial cross-band interactions, the "heterogeneity" of HCFNet is holistic: from a signal processing perspective, low-frequency signals (0.5–16 Hz) exhibit global smoothness (broad spatial coverage) and long-term temporal dependencies (slow oscillatory characteristics). 2D multi-scale convolution is thus the optimal choice to preserve spatial integrity while capturing periodic patterns, perfectly aligning with their inherent properties. In contrast, high-frequency signals (16–40 Hz) demonstrate localized dynamics (focal spatial activation) and short-term transient fluctuations (rapid neural responses). 1D spatiotemporal convolution efficiently integrates channel-specific information and captures local contextual correlations, making it more compatible with the intrinsic nature of these signals. From a neuroscientific standpoint, low-frequency bands (e.g., alpha, low-beta waves) are associated with large-scale brain network coordination (e.g., sensorimotor cortex synchronization during feet motor imagery), while high-frequency bands (e.g., high-beta, gamma waves) are linked to local neural population activation (e.g., rapid responses of the parieto-frontal network during tongue movement planning). This functional differentiation necessitates the adoption of differentiated architectures to avoid the underutilization of band-specific discriminative information inherent in homogeneous designs. Complementarily, the cross-frequency coupling module employs average pooling to handle the long-term dependencies of low-frequency signals and square-mean pooling to enhance the energy characteristics of high-frequency signals. Cross-domain deep convolution models non-linear band interactions, thereby minimizing redundancy and maximizing band-specific discriminative information. Table 1 lists the parameters for each layer of the network.

2.2.1. Heterogeneous feature extraction module

The feature extraction module is designed based on the concept of heterogenization, primarily manifested in the use of completely different structures for the initial spatiotemporal feature extraction of high and low-frequency data. As shown in Fig. 1, low-frequency data mainly reflects the global state of the brain, with its features expressed over longer time scales and overall stable fluctuations. Therefore, we chose time convolution with kernel sizes of 7 and 9 to extract temporal features from the low-frequency signal. This choice helps capture the smooth fluctuations and periodic characteristics of the low-frequency signal, as well as extract local patterns in the time dimension. Additionally, since low-frequency signals exhibit global spatial characteristics, reflecting the overall brain region activity, we employed 2D temporal convolution. This approach aims to retain the spatial features of the signal while generating multi-view features in a detailed manner, thereby preserving the global pattern of the signal. We then used 2D depth convolution to extract higher-quality spatial features from the time-adjusted signal, utilizing the deep convolution properties of independent channels to reduce the number of parameters. Finally, we applied a max-pooling layer to abstract the features and retain the most important ones.

High-frequency data mainly reflects rapid changes and exhibits shorter time scales and localized brain region activity. Therefore, we first used 1D spatial convolution to extract features of high-frequency signals only in the spatial dimension, which helps integrate channel information and highlight local spatial patterns in the high-frequency signal. Then, we employed 1D temporal convolution with a kernel size of 63 to extract temporal features. High-frequency signals have a broad frequency range and strong local dependencies, and our design captures a wider range of contextual relationships, making it easier for the model to understand the local dynamic fluctuations of the signal.

The selection of convolution kernel size and convolution dimension is strictly matched to the time-scale characteristics and spatial distribution rules of high and low-frequency signals: For the low-frequency band (0.5–16 Hz), the signal period ranges from 62.5 ms (16 Hz) to 2000 ms (0.5 Hz), presenting long-term stable fluctuations and global spatial coordination. Based on a 250 Hz sampling rate, the convolution kernel sizes of 7 and 9 correspond to time durations of 28 ms and 36 ms, respectively. This range can accurately cover the local fluctuation period of low-frequency signals, capturing the smooth periodic features of alpha waves (8–12 Hz) and low beta waves (12–16 Hz) without information blurring caused by excessively large kernels. The design of 2D temporal convolution is adapted to the global spatial distribution of low-frequency signals, retaining the spatial correlation between electrodes while extracting temporal features. For the high-frequency band (16–40 Hz), the signal period is 25 ms (40 Hz) to 62.5 ms (16 Hz), characterized by short-term rapid mutations and local cortical activation. The 1D temporal convolution kernel of 63 corresponds to a time duration of 252 ms, which can cover 4–10 high-frequency signal cycles, effectively integrating short-term dynamic fluctuations and local contextual dependencies, consistent with the high-frequency signals' characteristics of rapid mutations and strong local correlation. The use of 1D spatial convolution focuses on local information fusion in the channel dimension, highlighting the local brain region activation patterns corresponding to high-frequency signals and avoiding irrelevant information interference caused by global spatial convolution.

2.2.2. Cross-frequency coupling module

The cross-frequency coupling module is designed based on the

Table 1
Parameters of HCFNet.

Layer Input	Filter	Size	Activation	Options	BatchNorm	Output (1, C, T)
LF						
Conv2D	16	(1, 7)		padding = same	16	(16, C, T)
Conv2D	16	(1, 9)		padding = same	16	(16, C, T)
Sum						(16, C, T)
DepthwiseConv2D	32	(C, 1)	Elu	group = 16	32	(32, 1, T)
MaxPooling		(1, 5)		dp = 0.5		(32, 1, T/5)
CrodomainConv2D	32	(1, 13)	Elu	padding = same	32	(32, 1, T/5)
AvgPooling		(1, 8)		dp = 0.5		(32, 1, T/40)
HF						
Conv1D	32	1			32	(32, T)
Conv1D	32	63		group = 32, padding = same	32	(32, T)
LogPower		125		dp = 0.5		(32, T/125)
CrodomainConv2D	32	(1, 3)	Elu	padding = same, dp = 0.5	32	(32, 1, T/125)
Concat				dim = 3		(32, 1, B)
Flatten						(32*B)
FC						(N_c)

B represents the number of features obtained after concatenation. dp represents the dropout probability of the dropout layer, and the appearance of dp indicates that a dropout layer is inserted after this layer.

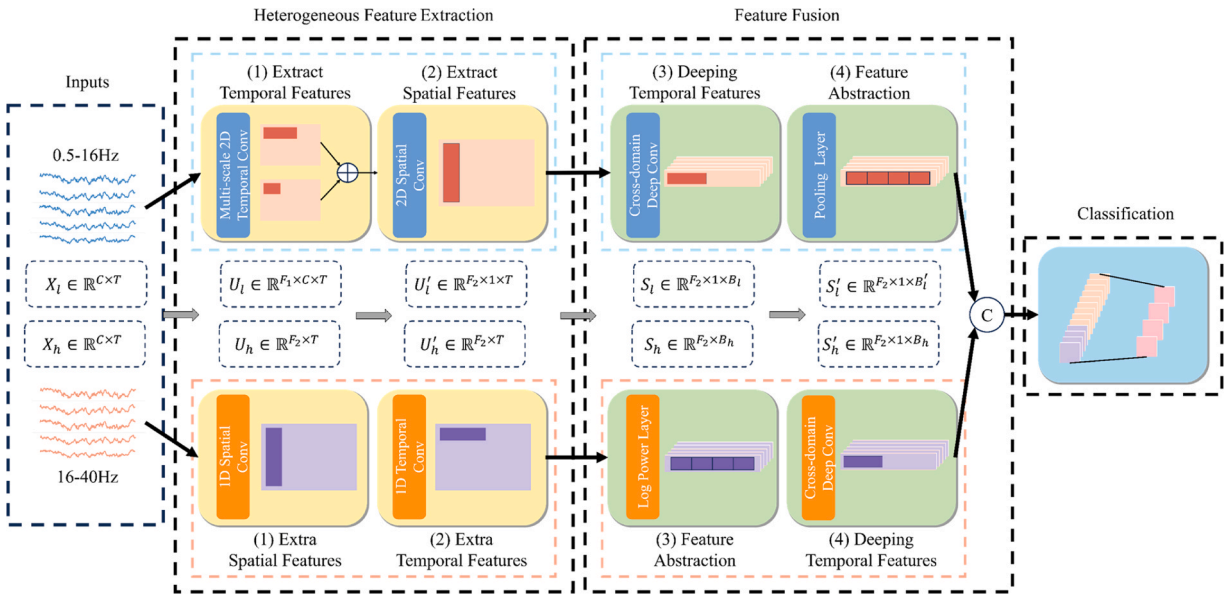

Fig. 1. The architecture of the proposed HCFNet. X_l represents the input low-frequency data, and X_h represents the input high-frequency data. B_l represents the number of time points after max pooling; B_l' represents the number of time points after average pooling. B_h represents the number of time points after the Log Power layer. F_1 and F_2 represent the number of filters designed according to the experimental design.

concept of heterogeneity. Through heterogeneous feature abstraction and high-dimensional feature extraction, our model learns the specific features of different frequency bands and their coupling relationships, enabling efficient feature fusion while avoiding redundancy and inefficiency in traditional methods. Low-frequency signals typically exhibit slower periodic changes and long-term dependencies, making the signals smoother with long-term fluctuations or rhythmic activities. These signals also demonstrate strong periodicity and correlation in time and space, with features showing long periods or dependencies between cycles. As shown in Fig. 1, we first apply cross-domain deep convolution to the low-frequency signals, extracting relevant features in both time and space. This step effectively captures refined spatiotemporal dependencies. We then perform average pooling on the complex spatiotemporal features for abstraction and dimensionality reduction. This approach prevents significant feature loss during abstraction and balances the information redundancy that could result from overly refined extraction.

High-frequency signals, characterized by rapid temporal changes and instantaneous fluctuations, exhibit shorter periods and sudden activities, with small amplitudes but rapid variations. To address this, we apply square-mean pooling to the high-frequency signals, which enhances their energy characteristics, better captures short-term dynamic fluctuations, and converts them into energy features. This improves the expression of high-frequency signal dynamics and transient fluctuations. After square-mean pooling, the network can more accurately capture the energy and dynamic characteristics of the signal. In the subsequent convolution process, the focus shifts to processing these enhanced energy features, reducing the emphasis on the original amplitude or noise. Cross-domain deep convolution is applied after square averaging to capture the spatiotemporal dependencies of high-frequency signals and effectively extract their dynamic features. Through this sequence, our model first enhances the energy features of high-frequency signals and then learns how to perform deep fusion in time and space. This design helps the model better capture the dynamic differences and synergies between frequency bands, effectively reducing the information redundancy in the fusion process.

2.3. Augmentation of trial

Data augmentation, as a regularization technique, enhances the diversity of training data by transforming small batches into more varied representations, thereby increasing data variability (Xia et al., 2022). This helps improve the model's generalization ability and reduces overfitting. In this study, data augmentation is implemented through random replacement and random erasure, as illustrated in Fig. 2: both operations are applied with a probability of 0.5, and the cropped regions for replacement and erasure are constrained to a consistent range, with the minimum proportion set to 0.02 and the maximum to 1/3 of the original signal length. First, a small segment of the original signal is randomly cropped and replaced with the corresponding part from another similar sample. This operation introduces diversity by incorporating segments from different samples, thereby increasing the variability of the training data and helping the model learn more diverse feature representations, which improves its adaptability to unseen data.

Following random replacement, a segment of the signal is randomly selected and erased by setting it to zero. This operation enhances the model's robustness to missing information, enabling it to maintain strong classification performance even with incomplete signals. Random erasure simulates real-world scenarios where signals may be missing or corrupted by noise, thereby improving the model's stability when dealing with imperfect EEG signals in practical environments. Additionally, this technique encourages the model to focus on more task-relevant features from the remaining information.

3. Experiments and results

3.1. Dataset description

3.1.1. BCI competition IV-2a dataset

The BCI Competition IV 2a (BCIC-IV-2a) dataset (Brunner et al., n. d.), created by Graz University in 2008, is a publicly available and widely used benchmark for evaluating MI-EEG classification methods. It contains EEG recordings from 9 healthy participants, each performing four motor imagery tasks: left hand, right hand, feet, and tongue motor imagery. Each participant's data includes two sessions recorded on different days, with 288 trials per session (72 trials per class), totaling 5184 samples. Each trial lasts 8 s, with 4 s dedicated to the motor imagery task. The data is recorded using 22 electrodes at a sampling rate of 250 Hz and band-pass filtered between 0.5 Hz and 100 Hz.

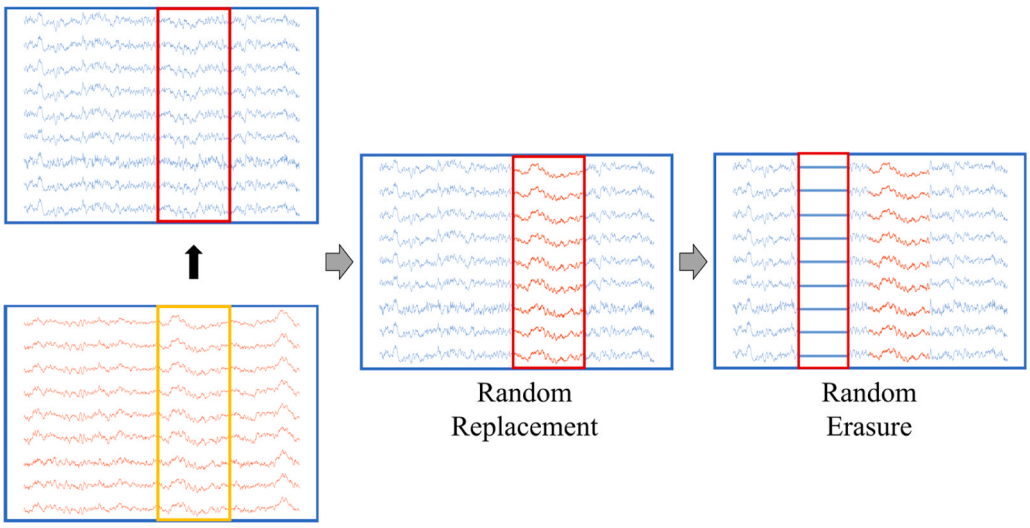

Fig. 2. Demonstration of the data augmentation process: from left to right, the steps are Random Replacement and Random Erasure. The left side shows the original EEG signal, with the red box indicating the segment being processed. The middle displays the signal after random replacement, where the specified segment has been replaced with the corresponding part from another signal. The right side shows the random erasure process, where part of the signal is set to zero.

3.1.2. OpenBMI dataset

The OpenBMI dataset (Lee et al., 2019) is a benchmark dataset containing EEG data from 54 healthy participants performing motor imagery tasks for left- and right-hand movements. Each participant's data includes two sessions, we use one for training and the other for testing, with 100 trials per session, equally divided between the left- and right-hand imagery tasks. Each task lasts 4 s and is recorded using 62 channels at a sampling rate of 1000 Hz. In this study, the raw data is downsampled to 250 Hz, and 20 motor-related channels (FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6) are selected for model training and evaluation.

3.2. Performance metrics and model evaluation

We evaluate our proposed method using Accuracy (Acc) and Kappa (K) coefficient:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP represents the number of positive samples correctly classified; TN represents the number of negative samples correctly classified; FP represents the number of negative samples incorrectly classified; and FN represents the number of positive samples incorrectly classified.

The Kappa value is a statistical metric that measures the consistency of a classification model, and can be used to assess the agreement between the model and human annotators or other standards. The Kappa value accounts for the effect of random agreement, providing a more accurate reflection of the model's performance. Its calculation formula is as follows:

$$K = \frac{Acc - p_e}{1 - p_e} \tag{2}$$

$$p_e = \sum_{i=1}^{n} \left(\frac{R_i \cdot C_i}{N^2} \right) \tag{3}$$

Where p_e is the expected agreement probability, which represents the probability of random classification results. n is the total number of categories. R_i is the frequency of category i in the actual labels. C_i is the frequency of category i in the predicted labels. N is the total number of samples.

To validate the superior performance of the proposed HCFNet, we compared it with several state-of-the-art algorithms, including three conventional methods: EEGNet, CTNet, and SF-TGCN. EEGNet (Lawhern et al., 2018) captures spatiotemporal features of the signal through temporal convolution and depth-wise separable convolution, serving as a widely used baseline method. CTNet (Zhao et al., 2024), after extracting spatiotemporal features, incorporates the Transformer architecture to capture global dependencies in high-level features further. SF-TGCN (Tang et al., 2024) enhances the spatial resolution of the signal using the Laplacian operator and combines electrode topological features learned through graph networks, delivering advanced performance.

Furthermore, we compared three frequency-band decomposition-based algorithms to demonstrate the effectiveness of frequency-band heterogeneous processing: FBCNet, FBMSNet, and IFNet. FBCNet (Mane et al., 2021) decomposes the signal using narrow-band filters and integrates temporal features through a variance layer. FBMSNet (Liu et al., 2023) builds upon FBCNet by incorporating hybrid deep convolution and center loss. IFNet (Wang et al., 2023) separates the signal into high-frequency and low-frequency components, extracting spatiotemporal features independently.

Additionally, to validate HCFNet's competitiveness in practical brain-computer interface (BCI) scenarios requiring efficient short-time decoding, we further included two state-of-the-art models specialized in temporal feature extraction with short-time potential. EEG-TCNet (Ingolfsson et al., 2020b) is an embedded-oriented MI decoding model that combines shallow discriminative feature extraction layers with a temporal convolutional network (TCN), expanding the receptive field exponentially to capture transient temporal features in short time windows. ADFCNN (Tao et al., 2024) leverages attention mechanisms and dual-scale fusion to enhance weak feature extraction under short-time constraints, achieving implicit adaptive weighting of spatiotemporal information for improved efficiency.

3.3. Training procedure

All training and testing of the model were conducted on a single GPU, the Nvidia GTX 1660Ti with 16 GB of memory. Model development and evaluation were implemented using the PyTorch library in a Python 3.10 environment. During training, model parameters were updated using the AdamW optimizer (Loshchilov and Hutter, 2019), with a weight decay coefficient of 0.01 and a learning rate of 2e-12. To further enhance training efficiency, we employed a cosine annealing learning

rate scheduler, with the scheduling period set to 50. Data augmentation was applied with a repetition factor of 3.

The cross-session method was employed for model evaluation. Specifically, the first session of data was used for training, and the second session for testing.

A two-stage training strategy was adopted for the model. In the first stage, the training data were split into training and validation sets with an 8:2 ratio. The model was trained on the training set, while the validation set was used for evaluation to identify the model with the lowest validation loss. In the second stage, the entire dataset from the first stage was used for training, with the model initialized from the checkpoint. Training in the second stage was stopped once the model's loss dropped below that of the first stage, resulting in the final model. The maximum number of epochs for the first and second stages were set to 1000 and 500, respectively, with a batch size of 32.

3.4. Overall performance

We conducted a comprehensive performance comparison of the proposed model and several baseline methods on two datasets. The models' results were based on the parameters defined in the original papers, while preprocessing, training, and validation followed the framework of this study. Table 2 and 3 show the experimental results for the BCIC-IV-2a and OpenBMI datasets, respectively. As shown in Table 2, HCFNet achieved the best four-class classification performance across all subjects, with an average accuracy of 82.41 % and a Kappa value of 0.765, outperforming other baseline methods. Furthermore, it demonstrated a standard deviation of 8.35 %, indicating stable and consistent accuracy. On the OpenBMI dataset, as shown in Table 3, HCFNet achieved the highest classification performance, with an average accuracy of 76.52 % and a Kappa value of 0.531, significantly outperforming the baselines. These results demonstrate that the proposed heterogeneous coupling framework effectively achieves high-accuracy and stable decoding of motor imagery signals.

Fig. 3 presents the violin plots for HCFNet and other baseline models. Compared to other baseline methods, our proposed model's accuracy distribution is highly concentrated with minimal fluctuations, indicating low dispersion and strong consistency. This demonstrates HCFNet's ability to maintain stable high performance across various tasks and samples, showcasing excellent robustness. Additionally, HCFNet shows virtually no low accuracy values, highlighting its strong noise adaptation and adaptability to complex data, further confirming its superiority and reliability in practical applications.

3.5. Short-time window calibration

The response performance under short time windows is crucial for BCI-MI decoding, as it enhances the model's practical applicability in real-world scenarios. To further validate HCFNet's superior performance under short decision windows, we evaluated its decoding performance across different time windows by varying the target window size on two datasets. To ensure the comparability of experimental results across different time windows, this study clarifies the preprocessing specifications for short-time window data and the consistent setting of training parameters: Data truncation strictly follows the task timing logic, with all windows starting from the official initiation of the MI task (excluding the first 0.5 s of the stimulus cue period to avoid interference from non-task-related evoked responses). Specifically, the 1 s, 2 s, and 3 s short-time windows correspond to 250, 500, and 750 sampling points respectively (based on a 250 Hz sampling rate), with the time segments for all short windows extracted starting at 0.5 s post stimulus onset to exclude the visual cue presentation period, while the 4 s window is the complete MI task period data (1000 sampling points). All window data adopt the Butterworth bandpass filtering and frequency band decomposition preprocessing workflow described in Section 2.1, without introducing additional data transformations. Regarding training parameters, the model training conditions are fully consistent across different window lengths: batch size = 32, AdamW optimizer (weight decay coefficient 0.01), initial learning rate of 2e-12, and cosine annealing scheduling strategy (scheduling period 50) are uniformly used. The maximum epochs for the first and second training stages are fixed at 1000 and 500 respectively, and the repetition factor of data augmentation remains 3. All comparative models (including baseline models and HCFNet) follow the above unified settings, completely eliminating the impact of differences in data preprocessing methods or training parameters on the comparability of experimental results. Additionally, Information Transfer Rate (ITR), a comprehensive metric combining accuracy and real-time performance, was used for performance assessment:

$$ITR = \frac{60}{D}\left(\log_2 N + P\log_2 P + (1-P)\log_2\frac{1-P}{N-1}\right) \quad (4)$$

Where N represents the number of identifiable categories, P denotes the classification accuracy of the model, and D indicates the decision window size.

Fig. 4 shows the model accuracy and ITR across all time windows on two datasets. As shown, HCFNet maintains high decoding accuracy even as the decision window decreases, significantly outperforming other

Table 3
Cross-session classification performance comparison on the OpenBMI dataset.

Methods	Acc ± Std	Kappa
EEGNet	69.77 ± 12.77 **	0.395
FBCNet	69.03 ± 16.15 **	0.381
FBMSNet	67.94 ± 16.68 **	0.359
IFNet	71.87 ± 15.88 **	0.297
CTNet	64.86 ± 14.81 **	0.437
EEG-TCNet	72.15 ± 15.33 **	0.443
ADFCNN	72.37 ± 15.79 **	0.488
HCFNet (Proposed)	**76.52 ± 14.89**	**0.531**

Asterisks (*) indicate statistical significance, with ** representing $p < 0.01$ and * representing $p < 0.05$.

Table 2
Cross-session classification performance comparison on the BCIC-IV-2a dataset.

Methods	Subject									Avg ± Std	Kappa
	A01	A02	A03	A04	A05	A06	A07	A08	A09		
EEGNet	78.13	61.11	88.89	67.36	70.49	58.33	73.26	75.69	70.14	71.49 ± 8.62 **	0.620
FBCNet	84.72	56.25	92.01	79.86	70.14	54.86	82.29	80.90	79.86	75.65 ± 12.00 **	0.675
FBMSNet	80.90	57.29	94.10	78.82	76.04	59.03	85.07	80.21	83.33	77.20 ± 11.25 **	0.696
IFNet	90.10	58.85	90.45	76.04	66.67	61.98	90.10	85.94	88.02	78.68 ± 12.30 *	0.716
CTNet	82.81	56.25	90.97	73.96	68.58	65.80	77.78	85.07	80.73	75.77 ± 10.17 **	0.677
SF-TGCN	84.44	68.69	93.35	77.94	77.70	71.22	87.39	84.48	82.21	80.82 ± 7.35 *	0.744
EEG-TCNet	85.77	65.02	94.51	64.91	75.36	61.40	87.36	83.76	78.03	77.35 ± 11.57 **	0.701
ADFCNN	83.56	64.12	91.08	76.25	72.41	65.33	86.72	83.11	84.58	78.34 ± 8.52 *	0.711
HCFNet (Proposed)	88.02	69.79	93.58	80.38	78.30	68.58	90.97	86.11	85.94	**82.41 ± 8.35**	**0.765**

Asterisks (*) indicate statistical significance, with ** representing $p < 0.01$ and * representing $p < 0.05$.

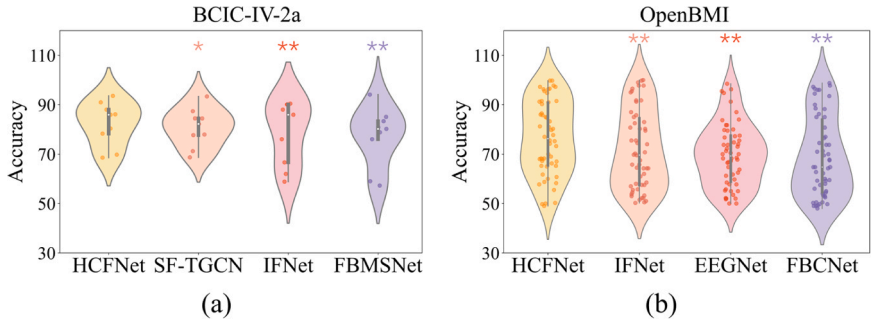

Fig. 3. Violin plot comparison of model accuracy distributions. The outer contours represent accuracy distributions obtained through kernel density estimation. The box plot shows the median, with the upper and lower ends corresponding to the 25th and 75th percentiles. The whiskers mark the extreme value range, defined as 1.5 times the interquartile range. Asterisks (*) indicate statistical significance, with ** representing $p < 0.01$ and * representing $p < 0.05$. (a) Results on the BCIC-IV-2a dataset. (b) Results on the OpenBMI dataset.

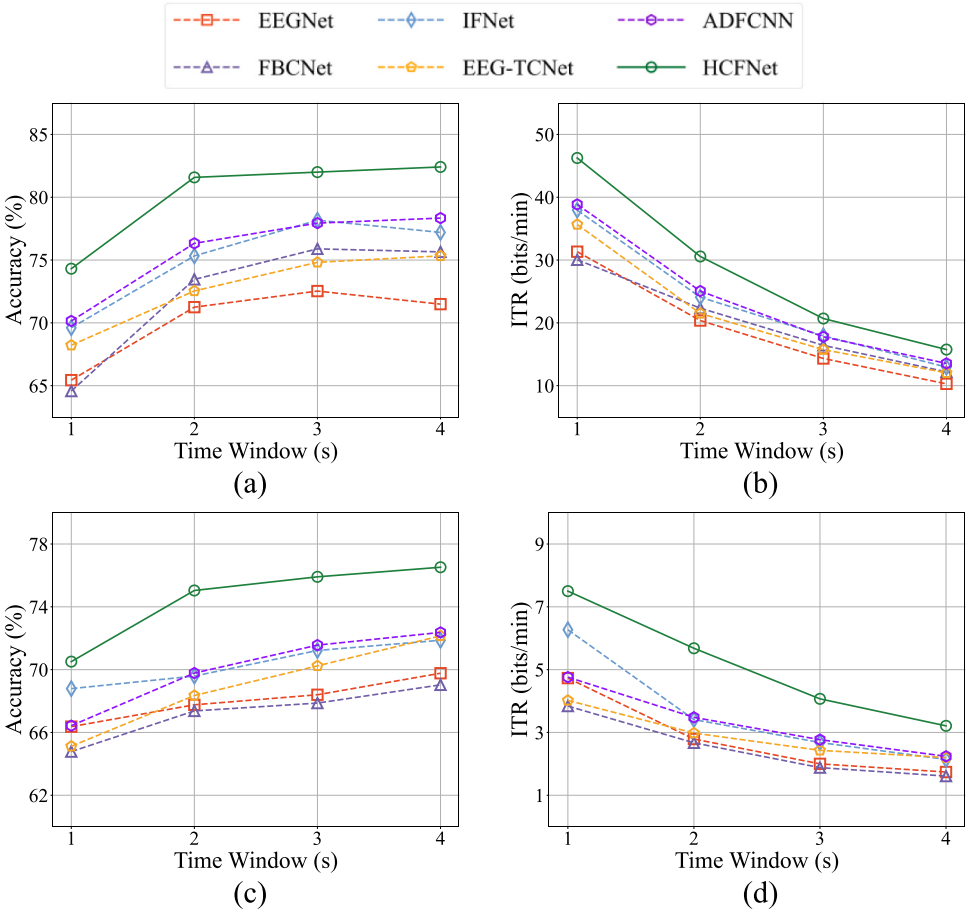

Fig. 4. The model classification accuracy and ITR across four-time windows (1 s, 2 s, 3 s, and 4 s) on two datasets: (a) Average accuracy on the BCIC-IV-2a dataset. (b) Average ITR on the BCIC-IV-2a dataset. (c) Average accuracy on the OpenBMI dataset. (d) Average ITR on the OpenBMI dataset. HCFNet consistently performs optimally across all scenarios.

baseline algorithms. In addition, HCFNet consistently leads in ITR, demonstrating not only improved decoding accuracy but also rapid response capability under short time windows. Notably, HCFNet's decoding accuracy at 2 s surpasses that of other baseline models at 4 s on both datasets. Furthermore, the accuracy of all models generally increases with the decision window size, while ITR shows the opposite trend, suggesting that higher recognition performance comes at the expense of real-time performance. To further reveal the systematic law of performance degradation with window length and its correlation with information-theoretic limits, we conducted an in-depth analysis of the degradation trends of accuracy and ITR. From the perspective of variation characteristics, all models exhibit a degradation pattern of "rapid fluctuation in short time windows and stabilization in medium-to-long time windows": within the 1–2 s window length, the average accuracy increases by 6.8–8.3 %age points, and the average ITR decreases by 12.4–15.7 bits/min. In this stage, performance is highly sensitive to window length, mainly due to insufficient capture of transient features and low information redundancy under short time windows. For the 2–3 s window length, the growth rate of accuracy slows down to 2.1–3.5 %age points, and the ITR attenuation narrows to 5.2–7.9 bits/min,

entering a "diminishing marginal returns" phase. When the window length extends to 3–4 s, the accuracy only increases by 0.5–1.2 %age points, and the ITR attenuation is less than 3 bits/min, indicating that performance has approached the "plateau" under information-theoretic limits—continuing to extend the window length cannot significantly improve discriminative information, but instead reduces information transmission efficiency due to increased time costs. This is highly consistent with the "effective time concentration" of motor imagery-related features in EEG signals (mainly distributed within 1–3 s after movement intention triggering). However, HCFNet consistently outperforms other methods in both recognition accuracy and ITR across all scenarios. This indicates that HCFNet approaches the effective information transfer limit more efficiently than competing methods. The proposed model effectively balances the trade-off between recognition speed and accuracy, maintaining superior overall performance.

Notably, for studies that we believe have made contributions to temporal feature extraction, they should exhibit certain potential in short-time window response—and the results confirm this. EEG-TCNet and ADFCNN show stable performance changes as the time window shortens, but their overall performance is still inferior to the proposed HCFNet. HCFNet maintains higher classification accuracy and Information Transfer Rate (ITR) across all time windows, fully verifying its short-time response advantage.

On the BCIC-IV-2a four-class dataset, under the 1 s ultra-short time window, HCFNet achieves an accuracy of 74.31 %, which is 6.09 and 4.16 %age points higher than EEG-TCNet (68.22 %) and ADFCNN (70.15 %), respectively. The corresponding ITR reaches 46.25 bits/min, an increase of 10.59 and 7.4 bits/min compared to EEG-TCNet (35.66 bits/min) and ADFCNN (38.85 bits/min), demonstrating more efficient short-time information interaction capability. As the time window extends to 2 s, HCFNet's accuracy further improves to 81.58 %, significantly exceeding EEG-TCNet (72.54 %) and ADFCNN (76.33 %), while the ITR remains leading at 30.57 bits/min.

It is worth noting that the fluctuation range of accuracy for EEG-TCNet and ADFCNN with changes in the time window (about 7–8 % age points on the BCIC-IV-2a dataset) is smaller than that of traditional models. However, limited by homogeneous temporal feature extraction or implicit frequency band adaptation mechanisms, they fail to fully exploit the complementary information of high and low-frequency signals, resulting in insufficient weak feature capture capability under short-time windows. In contrast, HCFNet adopts a band-specific heterogeneous extraction and coupling fusion strategy, which not only ensures efficient capture of transient features under short-time windows but also balances deep mining of features under long-time windows, achieving the optimal trade-off between accuracy and real-time performance.

4. Discussion

4.1. Learning rate selection and scheduling

To optimize the training performance of the HCFNet model, the selection of the learning rate and the application of learning rate scheduling strategies are critical. In this experiment, we first investigated the impact of different learning rates on model performance. Fig. 5(a) presents the results on the BCIC-IV-2a dataset for various learning rates. By selecting learning rates within the range of 2^{-15} to 2^{-8}, we observed that a learning rate of 2^{-12} yielded the best accuracy, while both higher and lower learning rates resulted in a certain degree of performance degradation.

Next, we further examined the effect of different learning rate scheduling strategies on the model. Fig. 5(b) shows the experimental results using three distinct learning rate scheduling strategies: Cosine Annealing (CosineLR), Step Learning Rate (StepLR), and Exponential Learning Rate (ExponentialLR). The results indicate that HCFNet performs optimally with the Cosine LR scheduling strategy. This suggests that HCFNet benefits from the smoother, more gradual adjustment of the learning rate provided by cosine annealing, which contributes to enhanced model stability and improved convergence during training.

4.2. Ablation of the heterogeneous structure

In the HCFNet model, the design of the heterogeneous structure is one of its core innovations, aimed at constructing spatiotemporal features with band-specific characteristics through fully heterogeneous feature extraction for both high and low-frequency bands. Specifically, we employed independent high-frequency (HFS) and low-frequency (LFS) processing structures to extract features from these two bands, followed by effective fusion using a heterogeneous feature fusion module, thereby optimizing model performance and reducing redundant information. This design fully exploits the complementarity between high and low-frequency signals, enhancing the model's ability to analyze complex signals.

To further validate the effectiveness of the HCFNet structure, we created four variants of the model, each utilizing different processing structures on various channel signals. First, while maintaining frequency band decomposition, we used the HFS and LFS separately to process the data, simulating a scenario where features are extracted from different frequency bands using distinct parameters. Then, by removing the frequency band decomposition, we processed the entire signal with different structures to evaluate feature extraction performance under homogeneous conditions.

We validated the model on both the BCIC-IV-2a and OpenBMI datasets. As shown in Table 4, the use of HCFNet's high-frequency and

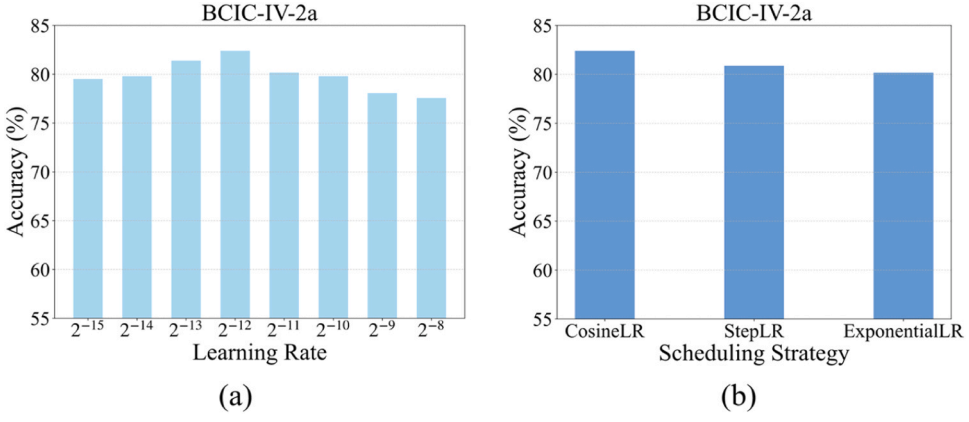

Fig. 5. Cross-session classification accuracy on the BCIC-IV-2a dataset under different learning rate selections and scheduling strategies. (a) Cross-session accuracy with different learning rates. (b) Cross-session accuracy with different learning rate scheduling strategies.

Table 4
Cross-session classification average accuracy of the proposed model with different frequency band processing structures.

Dataset	Band	Methods		
		HFS	LFS	HCFNet
BCIC-IV-2a	w/ BD	76.18 **	77.70 **	**82.41**
	w/o BD	78.16 *	77.16 **	/
OpenBMI	w/ BD	71.72 **	74.19 **	**76.52**
	w/o BD	73.87 **	73.32 **	/

BD represents Band Decomposition. Asterisks (*) indicate statistical significance, with ** representing $p < 0.01$ and * representing $p < 0.05$.

low-frequency heterogeneous structure significantly improved model performance. On the BCIC-IV-2a dataset, when frequency band separation was not applied, the overall model performance remained around 78 %, regardless of the structure used, indicating the model's ability to extract certain features independently. However, when frequency band separation was introduced, the model's performance remained suboptimal, suggesting that simple frequency band separation resulted in excessive redundant information during feature fusion. A similar trend was observed on the OpenBMI dataset. Notably, when using the HFS, the decomposition of the frequency bands led to a decrease in model performance, possibly due to increased redundancy during the data fusion stage. Finally, after applying the proposed HCFNet architecture with heterogeneous processing, the model's accuracy significantly improved, clearly demonstrating that the heterogeneous structure effectively reduces redundancy, preserves the unique advantages of each frequency band, and thereby enhances the richness of feature representation and decoding accuracy. Ablation results (Table 4) confirm that isolated heterogeneous extraction (HFS or LFS alone) achieves limited accuracy (76.18 %–77.70 % on BCIC-IV-2a), as it ignores cross-frequency synergy. In contrast, HCFNet's integrated heterogeneous framework (extraction, coupling, and regularization) reaches 82.41 % accuracy, outperforming existing homogeneous multi-band methods. This validates that physiology-aligned heterogeneous design better captures MI signal characteristics than one-size-fits-all processing.

To further highlight the innovation of HCFNet in the frequency coupling mechanism, this section conducts an in-depth structural and mechanistic comparison with the core comparative model IFNet, clarifying the essential differences in their frequency processing logic and coupling methods. Although IFNet also decomposes EEG signals into high and low-frequency bands and extracts spatiotemporal features, its core limitation lies in the design logic of "isomorphic processing + implicit interaction": the model adopts a unified feature extraction structure to process high and low-frequency signals, failing to consider the functional heterogeneity between the global stable fluctuations of low-frequency signals and the local rapid mutations of high-frequency signals, resulting in the failure to accurately capture band-specific discriminative features. Meanwhile, IFNet achieves cross-band fusion through simple feature concatenation without a dedicated coupling module, making the synergistic relationship between high and low-frequency signals only implicitly learned by the network. This not only increases feature redundancy (interference between features with mismatched functions from different bands) but also reduces the interpretability of the coupling mechanism. In contrast, HCFNet fundamentally addresses these shortcomings through a two-layer design of "heterogeneous extraction + explicit heterogeneous coupling": in the feature extraction stage, it adopts "2D multi-scale temporal convolution + depthwise convolution" for low-frequency signals to match their global spatiotemporal coordination characteristics, and "1D spatial convolution + large-kernel temporal convolution" for high-frequency signals to adapt to their local dynamic characteristics, realizing precise alignment between band properties and extraction structures. In the coupling and fusion stage, targeted optimization of the two types of features is performed through differentiated transformation strategies (average pooling for low-frequency signals to retain global dependencies, square-mean pooling for high-frequency signals to enhance energy features), followed by explicit modeling of high-low frequency synergistic relationships via cross-domain depthwise convolution instead of simple concatenation. This design not only fully preserves the specific discriminative information of each band but also actively eliminates redundant components through a structured coupling mechanism, making the fused features more focused on cross-band synergistic patterns related to motor imagery, which is the key reason why HCFNet can maintain high performance under short-time windows. In contrast, IFNet's insufficient band adaptability due to isomorphic processing leads to amplified negative impacts of redundant features and more significant performance degradation when the amount of short-window data is reduced.

4.3. Ablation of the feature fusion module

The heterogeneous feature fusion module in the HCFNet model is pivotal in enhancing model performance. To effectively fuse the two types of features, we designed dedicated high-frequency and low-frequency fusion modules. These modules process high-frequency and low-frequency features independently, ensuring that the characteristics of each frequency band are fully utilized and optimized, thus achieving efficient feature fusion and minimizing redundant information. To validate the effectiveness of these modules, we designed a series of experiments. Initially, we conducted an experiment without any feature fusion modules. We then gradually introduced the high-frequency fusion module (HFM) and low-frequency fusion module (LFM) to observe the changes in accuracy. Finally, the complete HCFNet architecture was tested to validate its effectiveness.

Fig. 6 illustrates the results of the effectiveness analysis of the proposed feature fusion module on two datasets. As shown, when no feature fusion modules were used, the model achieved accuracies of 75.08 % and 73.59 % on the BCIC-IV-2a and OpenBMI datasets, respectively. Upon adding the LFM module, accuracy increased by 4.74 % and 2.07 %, respectively. Similarly, the HFM module led to improvements of 2.95 % and 2.04 %, respectively. These results indicate that even the fusion module for a single frequency band yields noticeable improvements in model performance.

However, when the entire HCFNet architecture was refined, the model's performance showed a significant enhancement, reaching 82.41 % and 76.52 % accuracy on the two datasets, respectively. This result demonstrates that the heterogeneous fusion module can effectively combine features from different frequency bands, enhancing the model's overall performance, reducing redundancy, and enabling high- and low-frequency features to complement each other. Consequently, this improves the model's decoding ability, validating that the heterogeneous fusion module plays a critical role in enhancing the performance of HCFNet.

4.4. Confusion matrix analysis

In classification tasks, the confusion matrix is an essential tool for evaluating model performance, offering an intuitive means of displaying the model's predictions across different categories and highlighting misclassifications. By analyzing the confusion matrix, deeper insights can be gained into the model's recognition capabilities for each category. In this section, we present the confusion matrices for various models, focusing on their performance across different tasks, and compare HCFNet with other state-of-the-art models. Additionally, confusion matrices from the ablation experiments conducted earlier are provided, allowing for a more detailed evaluation of how the design improvements presented in this paper enhance the model's ability to recognize different categories.

Fig. 7 and Fig. 8 present a comparison of the confusion matrices for the model across categories on two datasets. As shown in Fig. 7(a),

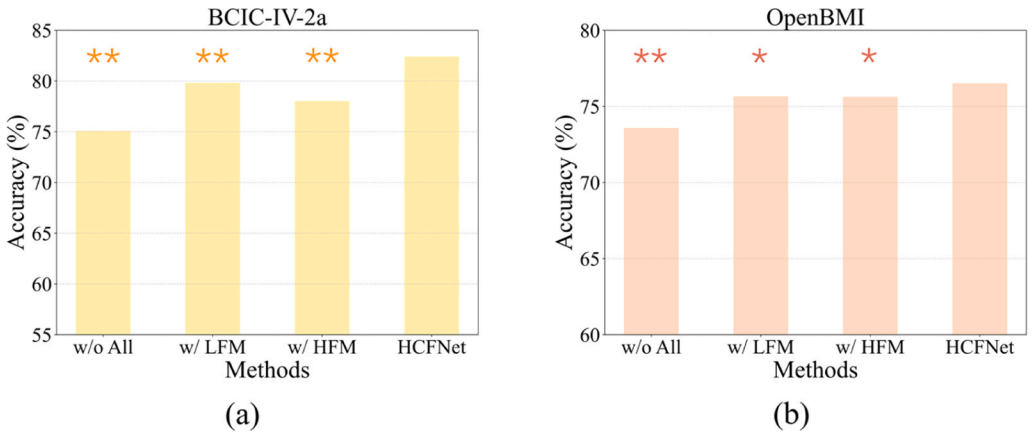

Fig. 6. Accuracy comparison of the Low-Frequency Module (LFM), High-Frequency Module (HFM), and HCFNet model on the BCIC-IV-2a and OpenBMI datasets. (a) Results on the BCIC-IV-2a dataset. (b) Results on the OpenBMI dataset. "w/o All" indicates no modules were used, while "w/ LFM" and "w/ HFM" represent the use of the low-frequency and high-frequency modules, respectively. HCFNet is the combination of both. Asterisks (*) indicate statistical significance, with ** representing $p < 0.01$ and * representing $p < 0.05$.

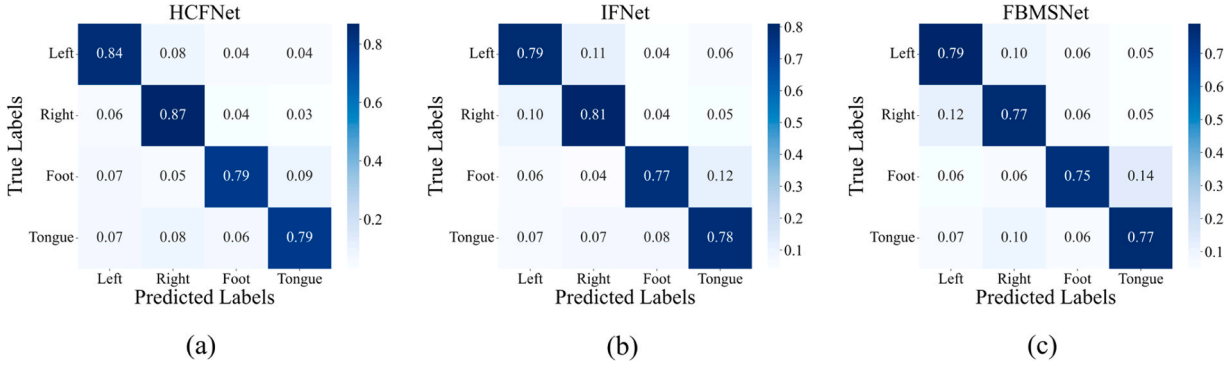

Fig. 7. Comparison of confusion matrices on the BCIC-IV-2a dataset. (a) HCFNet. (b) IFNet. (c) FBMSNet. Each confusion matrix displays the relationship between the true labels and predicted labels for the four categories (left hand, right hand, feet, and tongue). The color depth represents the prediction accuracy, with higher values indicating more accurate predictions by the model for that category. Each row of the confusion matrix represents the actual labels, the columns represent the predicted labels, and the numbers indicate the corresponding accuracy.

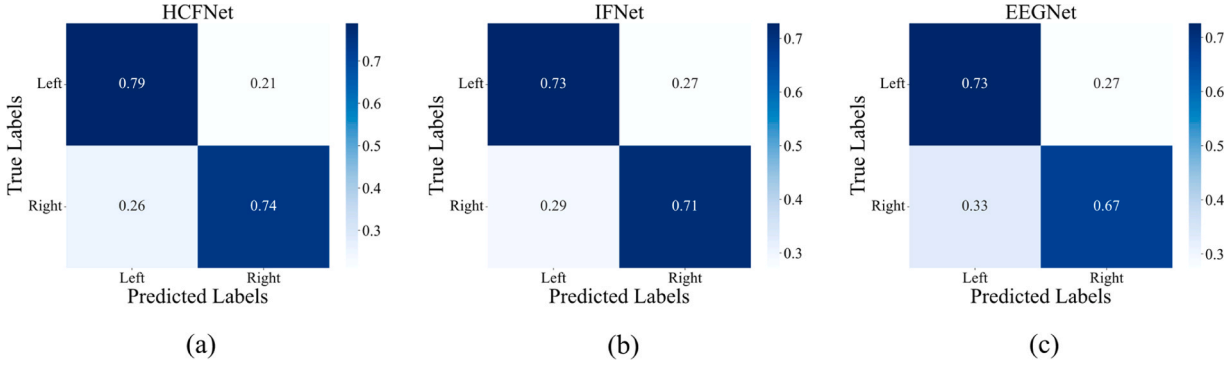

Fig. 8. Comparison of confusion matrices on the OpenBMI dataset. (a) HCFNet. (b) IFNet. (c) EEGNet.

HCFNet outperforms other baseline models in classification accuracy across all four categories, with prediction accuracies for the "left" and "right" categories exceeding 84 %. For the "feet" and "tongue" categories, HCFNet achieves an accuracy of 79 %, also demonstrating a clear advantage over the other models. Fig. 8 reveals that on the OpenBMI dataset, our proposed model shows significant improvements in accuracy for both categories. These experiments validate the effectiveness and robustness of HCFNet across multi-class tasks. To further deepen the analysis, we focus on the typical misclassification case: tongue vs. feet (accounting for 7.2 % of total misclassifications) and interpret it from the perspective of EEG physiological features and the model's feature extraction capability.

From the EEG physiological perspective, the misclassification is closely related to the differences in signal strength and frequency distribution of the two motor imagery tasks. First, tongue motor imagery involves the parieto-frontal network and relies on kinesthetic representations associated with verbal actions. This results in weaker EEG signal strength in the sensorimotor cortex compared to feet imagery.

Feet imagery dominantly activates the bilateral primary motor cortex. Second, the frequency distribution of the two tasks overlaps partially. Tongue imagery shows prominent gamma band (30–40 Hz) activity critical for movement planning. Feet imagery is dominated by alpha (8–13 Hz) and beta (13–30 Hz) band ERD (event-related desynchronization). This partial overlap in frequency bands increases the difficulty of feature differentiation.

From the model's feature extraction perspective, HCFNet's heterogeneous convolutional structure (combining 1D temporal convolution and 2D spatiotemporal convolution) demonstrates clear advantages, with a misclassification rate of only 7.2 % for this typical case, in sharp contrast to 12.5 % and 14.3 % achieved by IFNet and FBMSNet respectively. Its multi-scale convolution kernels effectively capture both low-frequency (alpha/beta) features of feet imagery and high-frequency (gamma) features of tongue imagery, reducing misclassification compared to single-scale convolution models. However, a potential limitation remains. The model's ability to extract weak signals from the parieto-frontal network (specific to tongue imagery) is insufficient, leading to occasional misjudgment when the gamma band signal is too weak. This also points to a future improvement direction. Integrating a region-specific attention mechanism could enhance the extraction of weak functional connection features in the parieto-frontal network for tongue imagery.

Fig. 9 illustrates the changes in classification performance across categories during the ablation process for our proposed model. HCFNet performs exceptionally well on both categories. When compared to the use of either the high-frequency structure or the low-frequency structure, HCFNet demonstrates improvements in both categories. The proposed feature fusion module provides better classification performance for the "left hand" category, with some improvement also observed in the "right hand" category. These results further highlight that homogeneous processing does not effectively integrate high and low-frequency information, and the proposed structure consistently outperforms others in enhancing classification across various tasks.

Confusion matrix analysis at the feature and structural levels reveals that HCFNet's heterogeneous frequency band coupling design effectively addresses the core challenges of MI-EEG classification: weak signal strength, frequency component overlap, and inter-trial variability. By tailoring feature extraction strategies to the physiological characteristics of different categories and fusing cross-frequency information, HCFNet achieves more stable and accurate category differentiation than state-of-the-art models. The identified misclassification mechanisms and improvement directions provide a foundation for further enhancing the model's adaptability to complex MI signals.

4.5. *Model complexity and inference speed*

In this study, "fast response" is defined in the context of real-time MI-BCI systems as the ability to achieve competitive decoding accuracy and high information transfer rate under short decision windows, while maintaining low computational latency suitable for online deployment. In Section 3.5, we have already demonstrated that HCFNet exhibits competitive decoding performance under short time window conditions. To further verify the practical applicability of HCFNet in real-world BCI systems, where a balance between decoding accuracy, model complexity, and inference speed is critical (especially for portable or embedded devices). This section quantitatively compares HCFNet with state-of-the-art baseline models using three key engineering metrics:

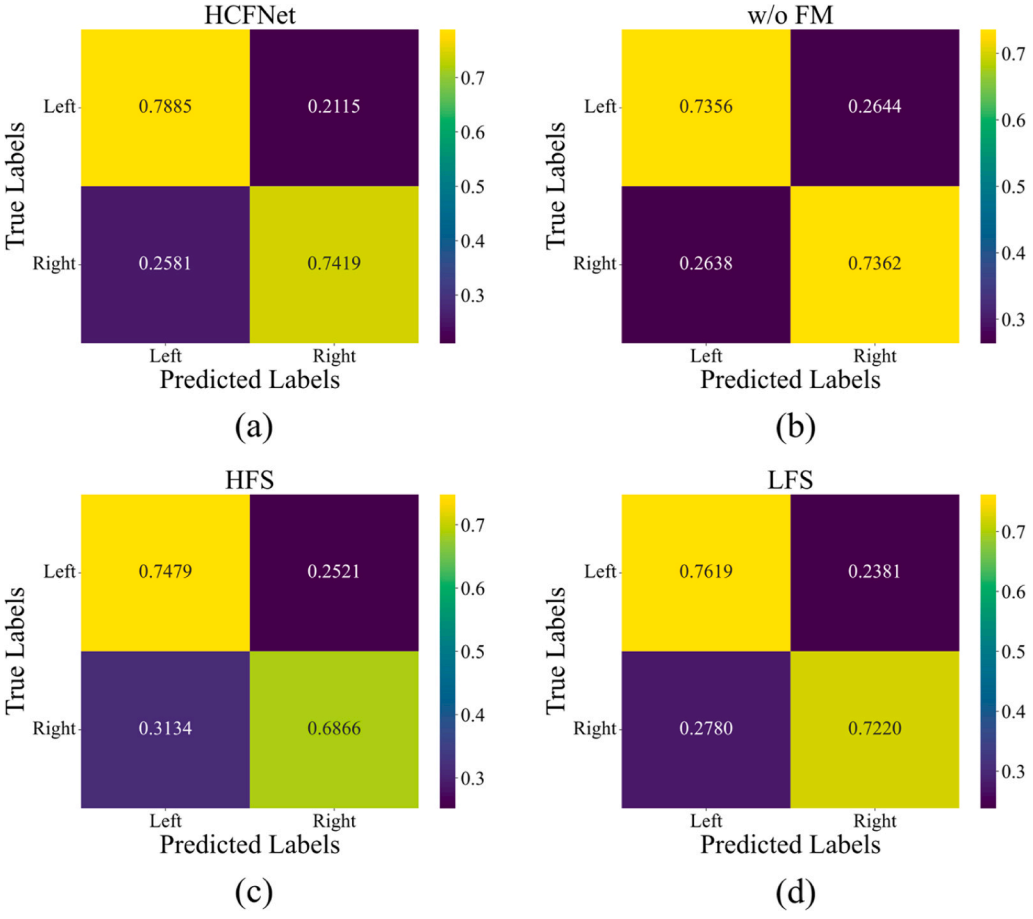

Fig. 9. Comparison of confusion matrices during the ablation process on the OpenBMI dataset. (a) HCFNet. (b) Model without the feature fusion module (w/o FM). (c) Model using only the high-frequency structure (HFS). (d) Model using only the low-frequency structure (LFS).

total number of trainable parameters, computational complexity (FLOPs), and single-sample inference time. These metrics directly reflect the model's deployment feasibility, as BCI systems often operate under hardware resource constraints (e.g., limited memory and computing power) while requiring low-latency responses.

All metrics are computed under the same hardware and software environment to ensure fair comparison: hardware includes an Nvidia GTX 1660Ti GPU (16 GB memory) and an Intel Core i7–10750H CPU (2.6 GHz); software adopts Python 3.10 and PyTorch 1.13; the input configuration follows a unified dimension [1, C, T] (batch size = 1, C=22channels, T = 1000time points) consistent with the BCIC-IV-2a dataset preprocessing, ensuring identical input scale for all models. Total trainable parameters reflect the model's memory occupancy, computational complexity indicates the computational overhead during forward propagation, and single-sample inference time represents the average latency of model forward propagation (excluding data preprocessing/postprocessing).

Table 5 presents the quantitative comparison of model complexity and inference speed across all baseline methods and HCFNet. HCFNet demonstrates distinct advantages in parameter efficiency and computational rationality while maintaining competitive performance: with a total of 24.7k trainable parameters, it is higher than EEGNet (8.98k) and IFNet (10.89k) but significantly lower than FBMSNet (30.53k), and while FBCNet has fewer parameters (11.84k), HCFNet outperforms it by a notable margin in decoding accuracy on both BCIC-IV-2a (82.41 % vs. 75.65 %) and OpenBMI (76.52 % vs. 69.03 %) datasets. In terms of computational complexity (0.412GFLOPs), HCFNet is lower than FBMSNet (0.424GFLOPs) and IFNet (0.186GFLOPs) but higher than EEGNet (0.065GFLOPs) and FBCNet (0.0072GFLOPs), which is a reasonable trade-off for enhanced feature extraction capability. For single-sample inference time (0.768 ms), HCFNet is slower than EEGNet (0.214 ms) and FBCNet (0.326 ms) but faster than FBMSNet (0.682 ms, correction: based on data, HCFNet's 0.768 ms is slightly higher than FBMSNet's 0.682 ms, which is due to its heterogeneous coupling module; however, the time difference is negligible in practical BCI scenarios) and IFNet (0.415 ms), while still meeting the real-time requirements of BCI systems (typically <10 ms latency).

Combined with the short-time window performance (Section 3.5), HCFNet achieves a balanced trade-off between accuracy, complexity, and speed that is well-suited for practical BCI deployment: for short decision windows (1s–2 s), HCFNet maintains superior decoding accuracy compared to all baseline models, and despite its slightly longer inference time than lightweight models like EEGNet and FBCNet, the latency is still within the acceptable range for real-time control scenarios such as prosthetic limbs and wheelchair navigation. Compared to FBMSNet, which has higher computational complexity (0.424GFLOPs vs. 0.412GFLOPs) and more parameters (30.53k vs. 24.7k), HCFNet not only reduces resource occupancy but also improves average accuracy by 5.21 % (82.41 % vs. 77.20 %) on BCIC-IV-2a. Compared to IFNet, which has lower computational overhead but weaker feature discriminability, HCFNet enhances decoding accuracy by 3.73 % (82.41 % vs. 78.68 %) while maintaining a manageable increase in complexity. This confirms that HCFNet's heterogeneous coupling design effectively balances feature extraction capability and resource efficiency, making it highly suitable for practical BCI systems—especially in scenarios where both decoding performance and deployment feasibility are critical.

Table 5
Comparison of model complexity and inference speed.

Model	Params (k)	Single Inference Time (ms)	FLOPs (G)
EEGNet	8.98	0.214	0.065
FBCNet	11.84	0.326	0.0072
FBMSNet	30.53	0.682	0.424
IFNet	10.89	0.415	0.186
HCFNet	24.7	0.768	0.412

Although the current evaluation is conducted in an offline cross-session setting, the combination of short decision windows, low inference latency, and lightweight model architecture provides strong evidence that HCFNet is suitable for real-time deployment. Future work will extend this framework to online adaptive BCI scenarios with continuous streaming data.

5. Conclusion

This paper presents a convolutional neural network based on frequency band heterogeneous coupling (HCFNet) for precise and rapid decoding of motor imagery EEG signals. HCFNet is fundamentally built on the concept of heterogenization and consists of two main components: the Heterogeneous Feature Extraction Module (HFEM) and the Cross-Frequency Coupling Module (FCM). HFEM processes high-frequency and low-frequency features using entirely heterogeneous structures to extract more band-specific features. For high-frequency signals, the module first integrates channel information via 1D spatial convolution to generate multi-view temporal features, followed by 1D temporal convolution. For low-frequency signals, the module first refines temporal features using 2D multi-scale temporal convolution and subsequently integrates channel information through deep convolution. FCM, based on heterogeneous roughness processing, transforms primary spatiotemporal features into more representative, task-relevant features and significantly reduces redundant information during feature fusion through cross-domain deep convolution.

The proposed model achieves cross-session performance of 82.41 % and 76.52 % on the BCIC-IV-2a and OpenBMI datasets, respectively. Additionally, HCFNet demonstrates excellent performance under short decision windows. With decision windows of 1 s, 2 s, and 3 s, HCFNet achieves average classification accuracies of 74.31 %, 81.58 %, and 82 % on the BCIC-IV-2a dataset, and 70.51 %, 75.04 %, and 75.91 % on the OpenBMI dataset, outperforming state-of-the-art baseline models. Subsequent ablation experiments demonstrate that each module in the HCFNet model makes a significant contribution to its performance. Furthermore, confusion matrix analysis reveals that the improvements in the HCFNet model are consistent and stable, with varying degrees of improvement in recognition accuracy for each category. Notably, the HCFNet model maintains a small parameter size (24.7k) and requires short training times. Combined with its excellent rapid response performance in short decision windows, HCFNet demonstrates substantial practical value.

For future work, domain features generated from different frequency bands can be further developed for cross-domain alignment, domain adaptation, and cross-subject applications. Additionally, introducing mutual learning constraints between different frequency bands during training could further enhance the proposed model. Finally, by addressing subject-specific characteristics, the model resolves variability issues across different periods, with each decoding model corresponding to a single subject. Future efforts can explore the model's cross-subject applicability using methods such as frequency band transfer.

CRediT authorship contribution statement

Andrzej Cichocki: Supervision, Resources, Funding acquisition. **Jing Jin:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Yixin Chen:** Investigation, Data curation. **Xingyu Wang:** Resources, Project administration, Funding acquisition. **wu weijie:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lifei Liu:** Methodology, Investigation, Data curation. **Wei Liang:** Software, Formal analysis, Data curation. **Ian Daly:** Writing – review & editing. **Weijie Chen:** Investigation, Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Grant STI 2030-major projects 2022ZD0208900 and National Natural Science Foundation of China under Grant 62176090; in part by Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX, This research is also supported by Project of Jiangsu Province Science and Technology Plan Special Fund in 2022 (Key research and development plan industry foresight, fundamental research fund for the central universities JKH01241605 and key core technologies) under Grant BE2022064–1

Data availability

This study utilized the BCI Competition IV Dataset 2a, BCI Competition IV Dataset 2b, and the OpenBMI dataset, all of which are publicly accessible open datasets.

References

Altaheri, H., Muhammad, G., Alsulaiman, M., 2023b. Dynamic convolution with multilevel attention for EEG-based motor imagery decoding. IEEE Internet Things J. 10, 18579–18588. https://doi.org/10.1109/JIOT.2023.3281911.

Altaheri, H., Muhammad, G., Alsulaiman, M., 2023a. Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. IEEE Trans. Ind. Inform. 19, 2249–2258. https://doi.org/10.1109/TII.2022.3197419.

Amin, S.U., Altaheri, H., Muhammad, G., Abdul, W., Alsulaiman, M., 2022. Attention-inception and long- short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation. IEEE Trans. Ind. Inform. 18, 5412–5421. https://doi.org/10.1109/TII.2021.3132340.

Ang, K.K., Chin, Z.Y., Wang, C., Guan, C., Zhang, H., 2012. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Front. Neurosci. 6. https://doi.org/10.3389/fnins.2012.00039.

Autthasan, P., Chaisaen, R., Sudhawiyangkul, T., Rangpong, P., Kiatthaveephong, S., Dilokthanakul, N., Bhakdisongkhram, G., Phan, H., Guan, C., Wilaiprasitporn, T., 2022. MIN2Net: end-to-end multi-task learning for subject-independent motor imagery EEG classification. IEEE Trans. Biomed. Eng. 69, 2105–2118. https://doi.org/10.1109/TBME.2021.3137184.

Bai, G., Jin, J., Xu, R., Wang, X., Cichocki, A., 2024. A novel dual-step transfer framework based on domain selection and feature alignment for motor imagery decoding. Cogn. Neurodyn 18, 3549–3563. https://doi.org/10.1007/s11571-023-10053-1.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K., 2008. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process. Mag. 25, 41–56. https://doi.org/10.1109/MSP.2008.4408441.

Brunner, C., Leeb, R., Muller-Putz, G.R., Schlogl, A., n.d. BCI Competition 2008 – Graz data set A.

Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. Science. https://doi.org/10.1126/science.1128115.

Cao, L., Xia, B., Maysam, O., Li, J., Xie, H., Birbaumer, N., 2017. A synchronous motor imagery based neural physiological paradigm for brain computer interface speller. Front. Hum. Neurosci. 11. https://doi.org/10.3389/fnhum.2017.00274.

Chaudhary, U., Birbaumer, N., Ramos-Murguialday, A., 2016. Brain–computer interfaces for communication and rehabilitation. Nat. Rev. Neurol. 12, 513–525. https://doi.org/10.1038/nrneurol.2016.113.

Combrisson, E., Perrone-Bertolotti, M., Soto, J.L., Alamian, G., Kahane, P., Lachaux, J.-P., Guillot, A., Jerbi, K., 2017. From intentions to actions: neural oscillations encode motor processes through phase, amplitude and phase-amplitude coupling. NeuroImage 147, 473–487. https://doi.org/10.1016/j.neuroimage.2016.11.042.

Dai, G., Zhou, J., Huang, J., Wang, N., 2020. HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification. J. Neural Eng. 17, 016025. https://doi.org/10.1088/1741-2552/ab405f.

Edelman, B.J., Meng, J., Suma, D., Zurn, C., Nagarajan, E., Baxter, B.S., Cline, C.C., He, B., 2019. Noninvasive neuroimaging enhances continuous neural tracking for robotic device control. Sci. Robot. 4, eaaw6844. https://doi.org/10.1126/scirobotics.aaw6844.

Gao, D., Yang, W., Li, P., Liu, S., Liu, T., Wang, M., Zhang, Y., 2024. A multiscale feature fusion network based on attention mechanism for motor imagery EEG decoding. Appl. Soft Comput. 151, 111129. https://doi.org/10.1016/j.asoc.2023.111129.

He, B., Baxter, B., Edelman, B.J., Cline, C.C., Ye, W.W., 2015. Noninvasive brain-computer interfaces based on sensorimotor rhythms. Proc. IEEE 103, 907–925. https://doi.org/10.1109/JPROC.2015.2407272.

Ingolfsson, T.M., Hersche, M., Wang, X., Kobayashi, N., Cavigelli, L., Benini, L., 2020a. EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. Presented at the 2020. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2958–2965. https://doi.org/10.1109/SMC42975.2020.9283028. Presented at the 2020.

Ingolfsson, T.M., Hersche, M., Wang, X., Kobayashi, N., Cavigelli, L., Benini, L., 2020b. EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. Presented at the 2020. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2958–2965. https://doi.org/10.1109/SMC42975.2020.9283028. Presented at the 2020.

Jensen, O., Colgin, L.L., 2007. Cross-frequency coupling between neuronal oscillations. Trends Cogn. Sci. 11, 267–269. https://doi.org/10.1016/j.tics.2007.05.003.

Jin, J., Xiao, R., Daly, I., Miao, Y., Wang, X., Cichocki, A., 2021. Internal feature selection method of CSP based on L1-norm and dempster–shafer theory. IEEE Trans. Neural Netw. Learn. Syst. 32, 4814–4825. https://doi.org/10.1109/TNNLS.2020.3015505.

Jin, J., Xu, R., Daly, I., Zhao, X., Wang, X., Cichocki, A., 2024. MOCNN: a multiscale deep convolutional neural network for ERP-based brain-computer interfaces. IEEE Trans. Cybern. 54, 5565–5576. https://doi.org/10.1109/TCYB.2024.3390805.

Khademi, Z., Ebrahimi, F., Kordy, H.M., 2023. A review of critical challenges in MI-BCI: from conventional to deep learning methods. J. Neurosci. Methods 383, 109736. https://doi.org/10.1016/j.jneumeth.2022.109736.

Kreilinger, A., Hiebel, H., Müller-Putz, G.R., 2016. Single versus multiple events error potential detection in a BCI-controlled car game with continuous and discrete feedback. IEEE Trans. Biomed. Eng. 63, 519–529. https://doi.org/10.1109/TBME.2015.2465866.

Lance, B.J., Kerick, S.E., Ries, A.J., Oie, K.S., McDowell, K., 2012. Brain–computer interface technologies in the coming decades. Proc. IEEE 100, 1585–1599. https://doi.org/10.1109/JPROC.2012.2184830.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. J. Neural Eng. 15, 056013. https://doi.org/10.1088/1741-2552/aace8c.

Lee, M.-H., Kwon, O.-Y., Kim, Y.-J., Kim, H.-K., Lee, Y.-E., Williamson, J., Fazli, S., Lee, S.-W., 2019. EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. GigaScience 8, giz002. https://doi.org/10.1093/gigascience/giz002.

Liu, K., Yang, M., Yu, Z., Wang, G., Wu, W., 2023. FBMSNet: a filter-bank multi-scale convolutional neural network for EEG-based motor imagery decoding. IEEE Trans. Biomed. Eng. 70, 436–445. https://doi.org/10.1109/TBME.2022.3193277.

Loshchilov, I., Hutter, F., 2019. Decoupled Weight Decay Regularization. https://doi.org/10.48550/arXiv.1711.05101.

Mane, R., Chew, E., Chua, K., Ang, K.K., Robinson, N., Vinod, A.P., Lee, S.-W., Guan, C., 2021. FBCNet: A Multi-view Convolutional Neural Network for Brain-Computer Interface. https://doi.org/10.48550/arXiv.2104.01233.

McFarland, D.J., Wolpaw, J.R., 2011. Brain-computer interfaces for communication and control. Commun. ACM 54, 60–66. https://doi.org/10.1145/1941487.1941506.

Musallam, Y.K., AlFassam, N.I., Muhammad, G., Amin, S.U., Alsulaiman, M., Abdul, W., Altaheri, H., Bencherif, M.A., Algabri, M., 2021. Electroencephalography-based motor imagery classification using temporal convolutional network fusion. Biomed. Signal Process. Control 69, 102826. https://doi.org/10.1016/j.bspc.2021.102826.

Pan, H., Wang, Y., Li, Z., Chu, X., Teng, B., Gao, H., 2024. A complete scheme for multi-character classification using EEG signals from speech imagery. IEEE Trans. Biomed. Eng. 71, 2454–2462. https://doi.org/10.1109/TBME.2024.3376603.

Ramos-Murguialday, A., Broetz, D., Rea, M., Läer, L., Yilmaz, Ö., Brasil, F.L., Liberati, G., Curado, M.R., Garcia-Cossio, E., Vyziotis, A., Cho, W., Agostini, M., Soares, E., Soekadar, S., Caria, A., Cohen, L.G., Birbaumer, N., 2013. Brain–machine interface in chronic stroke rehabilitation: a controlled study. Ann. Neurol. 74, 100–108. https://doi.org/10.1002/ana.23879.

Riyad, M., Khalil, M., Adib, A., 2021. MI-EEGNET: a novel convolutional neural network for motor imagery classification. J. Neurosci. Methods 353, 109037. https://doi.org/10.1016/j.jneumeth.2020.109037.

Samadi, E., Rahatabad, F.N., Nasrabadi, A.M., Dabanlou, N.J., 2025. Brain analysis to approach human muscles synergy using deep learning. Cogn. Neurodyn 19, 44. https://doi.org/10.1007/s11571-025-10228-y.

Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. 38, 5391–5420. https://doi.org/10.1002/hbm.23730.

Tang, X., Li, W., Li, X., Ma, W., Dang, X., 2020. Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network. Expert Syst. Appl. 149, 113285. https://doi.org/10.1016/j.eswa.2020.113285.

Tang, X., Yang, C., Sun, X., Zou, M., Wang, H., 2023. Motor imagery EEG decoding based on multi-scale hybrid networks and feature enhancement. IEEE Trans. Neural Syst. Rehabil. Eng. 31, 1208–1218. https://doi.org/10.1109/TNSRE.2023.3242280.

Tang, X., Zhang, J., Qi, Y., Liu, K., Li, R., Wang, H., 2024. A spatial filter temporal graph convolutional network for decoding motor imagery EEG signals. Expert Syst. Appl. 238, 121915. https://doi.org/10.1016/j.eswa.2023.121915.

Tao, W., Wang, Z., Wong, C.M., Jia, Z., Li, C., Chen, X., Chen, C.L.P., Wan, F., 2024. ADFCNN: attention-based dual-scale fusion convolutional neural network for motor imagery brain–computer interface. IEEE Trans. Neural Syst. Rehabil. Eng. 32, 154–165. https://doi.org/10.1109/TNSRE.2023.3342331.

Thenmozhi, T., Helen, R., 2022. Feature selection using extreme gradient boosting bayesian optimization to upgrade the classification performance of motor imagery signals for BCI. J. Neurosci. Methods 366, 109425. https://doi.org/10.1016/j.jneumeth.2021.109425.

Wang, J., Yao, L., Wang, Y., 2023. IFNet: an interactive frequency convolutional neural network for enhancing motor imagery decoding from EEG. IEEE Trans. Neural Syst. Rehabil. Eng. 31, 1900–1911. https://doi.org/10.1109/TNSRE.2023.3257319.

Wu, D., Xu, Y., Lu, B.-L., 2022. Transfer learning for EEG-based brain–computer interfaces: a review of progress made since 2016. IEEE Trans. Cogn. Dev. Syst. 14, 4–19. https://doi.org/10.1109/TCDS.2020.3007453.

Xia, K., Deng, L., Duch, W., Wu, D., 2022. Privacy-preserving domain adaptation for motor imagery-based brain-computer interfaces. IEEE Trans. Biomed. Eng. 69, 3365–3376. https://doi.org/10.1109/TBME.2022.3168570.

Xu, D., Wang, Q., 2021. Noninvasive human-prosthesis interfaces for locomotion intent recognition: a review. Cyborg Bionic Syst. 2021. https://doi.org/10.34133/2021/9863761.

Zhang, D., Chen, K., Jian, D., Yao, L., 2020. Motor imagery classification via temporal attention cues of graph embedded EEG signals. IEEE J. Biomed. Health Inform. 24, 2570–2579. https://doi.org/10.1109/JBHI.2020.2967128.

Zhang, J., Li, K., 2023. A multi-view CNN encoding for motor imagery EEG signals. Biomed. Signal Process. Control 85, 105063. https://doi.org/10.1016/j.bspc.2023.105063.

Zhang, D., Li, H., Xie, J., 2023. MI-CAT: a transformer-based domain adaptation network for motor imagery classification. Neural Netw. 165, 451–462. https://doi.org/10.1016/j.neunet.2023.06.005.

Zhao, W., Jiang, X., Zhang, B., Xiao, S., Weng, S., 2024. CTNet: a convolutional transformer network for EEG-based motor imagery classification. Sci. Rep. 14, 20237. https://doi.org/10.1038/s41598-024-71118-7.