IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 April 2024, accepted 20 May 2024, date of publication 23 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404634

RESEARCH ARTICLE

Dual-Branch Convolution Network With Efficient Channel Attention for EEG-Based Motor Imagery Classification

KAI ZHOU, AIERKEN HAIMUDULA, AND WANYING TANG
College of Intelligent Manufacturing Modern Industry, Xinjiang University, Ürümqi 830017, China
Corresponding authors: Aierken Haimudula (arkin@xju.edu.cn) and Kai Zhou (zhoukai_yc@163.com)

ABSTRACT Brain-Computer Interface (BCI) is a revolutionary technique that employs wearable electroencephalography (EEG) sensors and artificial intelligence (AI) to monitor and decode brain activity. EEG-based motor imagery (MI) brain signal is widely utilized in various BCI fields including intelligent healthcare, robot control, and smart homes. Yet, the limited capability of decoding brain signals remains a significant obstacle to BCI techniques expansion. In this study, we describe an architecture known as the dual-branch attention temporal convolutional network (DB-ATCNet) for EEG-based MI classification. DB-ATCNet improves MI classification performance with relatively fewer parameters by utilizing a dual-branch convolutional network and channel attention. The DB-ATCNet model consists of two primary modules: attention dual-branch convolution (ADBC) and attention temporal fusion convolution (ATFC). The ADBC module utilizes a dual-branch convolutional network to extract low-level MI-EEG features and incorporates channel attention to improve spatial feature extraction. ATFC employs sliding windows with self-attention to obtain the high-level temporal features, and utilizes feature fusion strategies to minimize information loss. The DB-ATCNet achieved subject-independent accuracies of 87.33% and 69.58% in two-class and four-class classification tasks, respectively, on the PhysioNet dataset. On the BCI Competition IV-2a dataset, it achieved an accuracy of 71.34% and 87.54% for subject-independent and subject-dependent evaluations, respectively, surpassing existing methods. The code is available at https://github.com/zk-xju/DB-ATCNet.

INDEX TERMS Intelligent healthcare, MI-EEG classification, channel attention, dual-branch convolutional network, multi-head self-attention.

I. INTRODUCTION

Edge computing, communication technologies, cloud computing, and artificial intelligence (AI) are propelling us into an era of unprecedented technological convergence. These advances are bringing about significant changes in various fields, including the domain of Brain-Computer Interface (BCI). BCI is an interdisciplinary field involving AI, neuroscience, and other disciplines. BCI utilizes AI algorithms and wearable sensors for monitoring and decoding brain activity, translating it into control commands for interacting with external devices. As a revolutionary and cutting-edge technology, BCI has a variety of applications, such as prosthetic control, rehabilitation training, virtual reality experiences, and gaming entertainment.

Electroencephalography (EEG) is a key technique used in obtaining brain signals in the BCI system. In this method, the brain's electrical activity is recorded by measuring subtle changes in electrical potential on the surface of the scalp. EEG is low-cost, non-invasive, low risk, and excellent temporal resolution, hence it is widely implemented in the BCI field.

Motor imagery (MI) is the mental simulation of motion without any physical execution. When a person imagines doing a specific action, their brain produces neural activity comparable to that during the actual execution of the action. Medical and non-medical applications of EEG-based MI (MI-EEG) activities are extensive. Medical facets include

The associate editor coordinating the review of this manuscript and approving it for publication was Roberta Palmeri.

thought-to-text translation, stroke recovery, and the control of a variety of assistive devices including electrical stimulation devices, prosthetics, screen pointers, and wheelchairs [1]. Non-medical examples involve controlling environments in smart homes, games, enhancing human abilities with exoskeletons or robotic arms, and even authentication and security identification [1].

Although BCI is advancing through innovation, its practical application is currently constrained by the decoding capability of brain signals. MI-EEG signals are particularly difficult to decode for several reasons. Firstly, MI-EEG signals are sensitive to various interferences involving muscle movement, eye blinking, and environmental noise, leading to a decline in signal quality and increased decoding difficulty. Secondly, individual differences in brain signals impose high demands on the generalization capability of decoding models. Additionally, EEG signals exhibit channel correlation, high dimensionality, and nonstationary, which further complicates the detection and decoding of MI-EEG signals.

To meet MI-EEG signals decoding difficulties, researchers have proposed some traditional machine learning (ML) and deep learning (DL) techniques. The traditional ML process for analyzing EEG signals typically has three steps: preprocessing, feature extraction, and classification. Preprocessing technology can effectively remove artifacts and interference in EEG signals while retaining the original and true EEG information. Preprocessing methods for EEG signals typically include channel selection, signal filtering, signal normalization, and artifact removal [1]. The most commonly used artifact removal method is independent component analysis (ICA) [2], [3]. Feature extraction methods for MI-EEG signals include signal processing algorithms such as power spectral density (PSD), short-time Fourier transform (STFT), common spatial pattern (CSP) and filter bank CSP (FBCSP), which can extract frequency, spatial frequency, or time-frequency features from MI-EEG signals. Commonly used feature classification methods include random forest (RF), support vector machine (SVM), extreme learning machine (ELM) and k-nearest neighbor (KNN), which are used to classify extracted features. However, traditional ML is labor-intensive and requires extensive expertise, which limits classification performance. By contrast, DL can obtain richer features from raw EEG data without requiring preprocessing or manual feature extraction, and also has excellent feature classification performance, which usually improves recognition accuracy. Furthermore, DL integrates feature extraction and classification into a unified framework that enables end-to-end decoding of MI-EEG signals. DL has been applied in many fields, including video recognition and speech processing, and has produced excellent results [4], [5].

In recent years, research utilizing DL to identify MI tasks has developed rapidly, drawing from the successful applications of DL in other domains [1]. Statistical analysis indicates that among these DL models, convolutional neural network (CNN) has emerged as the predominant method for decoding MI-EEG signals [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. For example, Schirrmeister et al. [6] proposed Deep ConvNet architecture and Shallow ConvNet architecture for MI task classification, and found that their performance is highly susceptible to interference from CNN depth. Lawhenn et al. [7] integrated separable convolution operations into CNN and proposed the EEG-Net framework, which has been successfully applied to various task classifications. In addition, some studies have exploited the unique characteristics of EEG signals in developing CNN architectures for MI task classification. Such as, Mane et al. [8] proposed the FBCNet architecture by using filter-bank CSP to divide the EEG signal into multiple frequency bands and using CNN to extract features and classify MI tasks. Wang et al. [9] proposed the IFNet architecture to improve the representation of MI features by further exploring cross-frequency interactions. Additionally, several other CNN varieties have been proposed, such as attention-based CNN [10], [11], [12], [13], [14], residual-based CNN [15], inception-based CNN [10], [11], multi-branch CNN [14], [15], multi-scale CNN [13], and multi-layer CNN [16].

Among these CNN varieties, the multi-branch CNN structure has attracted attention due to its unique advantages. Multi-branch structures not only facilitate feature extraction across multiple dimensions but also effectively alleviate overfitting issues in MI classification, enhancing model trainability [17]. Some researchers have found that multi-branch CNN outperforms single-branch CNN for multi-class MI tasks. For instance, Liu et al. [15] designed a three-branch densely connected CNN, where its combination of multi-branches exhibited superior classification accuracy compared to a single branch, and an increase in dense block branches improved performance. Amin et al. [16] used four branch CNN with varying numbers of convolutional blocks to acquire multi-level information from MI-EEG signal and fused these features utilizing multi-layer perceptron (MLP) and auto-encoder (AE) for classification. Zhao et al. [18] designed a three-dimensional CNN with multiple branches based on three different receptive fields to extract features. Inspired by these studies, we propose a dual-branch CNN module as the feature extraction module of the MI-EEG classification algorithm.

Besides CNN, the methods used for MI classification include deep belief network (DBN) [19], auto-encoder (AE) [20], recurrent neural network (RNN) [21], [22], and hybrid DL models [12], [16]. Hassanpour et al. [20] employed stacked AE (SAE) to decode MI tasks utilizing spectral features. Xu et al. [19] designed a DBN that utilized restricted Boltzmann machines to decode four types of MI tasks. In some studies, RNN was utilized to obtain temporal features of MI-EEG signals and demonstrated excellent results. For instance, Kumar et al. [21] utilized FBCSP and long short-term memory (LSTM) network for feature extraction, with SVM as classifiers. Luo and Chao [22] utilized the FBCSP algorithm to obtain spatial-frequency features before input the extracted features to a gated recurrent unit (GRU) model. This research found that GRU surpassed LSTM in EEG signal decoding. Overall, CNN models perform better in MI task identification than other DL methods [1]. Furthermore, many researchers have explored the integration of other DL models with CNN, for instance, LSTM in [12] and SAE in [16], and hybrid DL networks show excellent results.

In recent times, a temporal convolutional network (TCN) has achieved outstanding results in modeling and classifying temporal data [23]. Unlike traditional CNN, TCN can increase parameter numbers linearly while expanding receptive fields exponentially. This allows them to have a broader receptive field with fewer parameters. Furthermore, TCN does not encounter issues such as gradient explosion or vanishing gradients when handling long input sequences, as may occur in other time-series classification networks, such as RNN [24], [25]. Compared with other RNN models including GRU and LSTM, TCN has achieved superior results in various sequence-related tasks [23]. Consequently, recent research has applied TCN architectures to decode MI-EEG signals with satisfactory performance. Ingolfsson et al. [24] designed the EEG-TCNET architecture, which integrates the TCN with the EEGNet architecture. Musallam et al. [25] improved upon this with the TCNet-Fusion model, enhancing the EEG-TCN model using TCN structure and feature fusion techniques. Altaheri et al. [26] utilized attention integrated TCN and CNN for decoding MI-EEG signals, achieving superior performance.

Over the past few years, researchers have discovered the unexpected benefits of integrating attention mechanisms into DL models. Attention mechanisms simulate the selective focus process in human information processing, allowing the model to selectively attend to important elements while disregarding irrelevant content. Integrating attention mechanisms with DL models automatically highlights the key information in the input data. Luong et al. [27] and Bahdanau et al. [28] proposed the initial algorithms based on attention, called multiplicative and additive attentions. Vaswani et al. [29] proposed the Transformer model with multilayer perceptron (MLP) and multi-head self-attention (MSA). Initially designed for natural language processing, these models based on attention have also been applied to other domains like speech recognition and computer vision. Recently, some attention methods have been designed, particularly in the domain of computer vision, including squeeze-and-excitation (SE) [30], efficient channel attention (ECA) [31], and convolutional block attention module (CBAM) [32].

Recent research explores the potential of DL methods based on attention for classification of MI [14], [26], [33]. For example, Altuwaijri et al. [14] proposed decoding raw EEG signals with a triple-branch CNN model based on SE attention blocks. Jia et al. [33] introduced a multi-branch CNN model combining ECA and LightGBM, where the ECA module assigns weights to features, obtaining more discriminative features and improving feature recognition. Altaheri et al. [26] introduced an MI-EEG signal decoding architecture based on MSA, achieving excellent performance.

Feature fusion is a process of combining features from various branches or layers to improve model performance by leveraging their complementary nature. In computer vision, feature fusion has been shown to be essential technique for boosting performance [34], [35]. In addition, feature fusion methods have been widely used in the medical field, such as cardiovascular disease assessment [36], [37]. Recently, researchers have integrated feature fusion techniques with DL models to decode MI-EEG. For example, the TCNet-Fusion model [25] enhances the EEG-TCN model's performance by introducing fusion layers that combine features from different layers to construct rich feature mappings. Li et al. [13] proposed a multi-scale fusion CNN combining an attention mechanism to decode MI-EEG signals. The network extracts multiscale spatio-temporal features from the signal and uses a multiple feature fusion method to maintain maximum information flow.

In this paper, we design a dual-branch attentional temporal convolutional network (DB-ATCNet) for decoding MI-EEG brain signals. MI-EEG signal is processed by the proposed model in three steps. Firstly, a dual-branch CNN combined with channel attention mechanisms is used to encode MI-EEG signals into high-level temporal representations. Secondly, attention layers are utilized to emphasize the most valuable content in the time series. Finally, a temporal convolutional layer with multi-scale information fusion is employed to obtain advanced temporal features from the highlighted content. The model utilizes sliding window to enhance MI classification performance. Contributions of this study include:

1) We designed an excellent DB-ATCNet framework, which incorporates dual-branch CNN, channel attention mechanism, MSA mechanism, TCN with multi-scale feature fusion, and sliding window.
2) Dual-branch CNN can acquire and integrate features from EEG signals at multiple scales, enriching feature information and effectively improving accuracy.
3) Channel attention mechanisms can improve a model's feature extraction across the spatial dimensions of EEG signals.
4) Adding multi-level residual connections in the TCN enhances feature reuse, thereby improving the model's representational capacity and generalization ability.
5) The DB-ATCNet architecture has achieved excellent results with the BCI Competition IV-2a dataset [38] and Physionet MI-EEG dataset [39].

The paper is structured as follows: the proposed DB-ATCNet model is introduced in Section II, results are presented and discussed in Section III, and the conclusion is presented in Section IV.

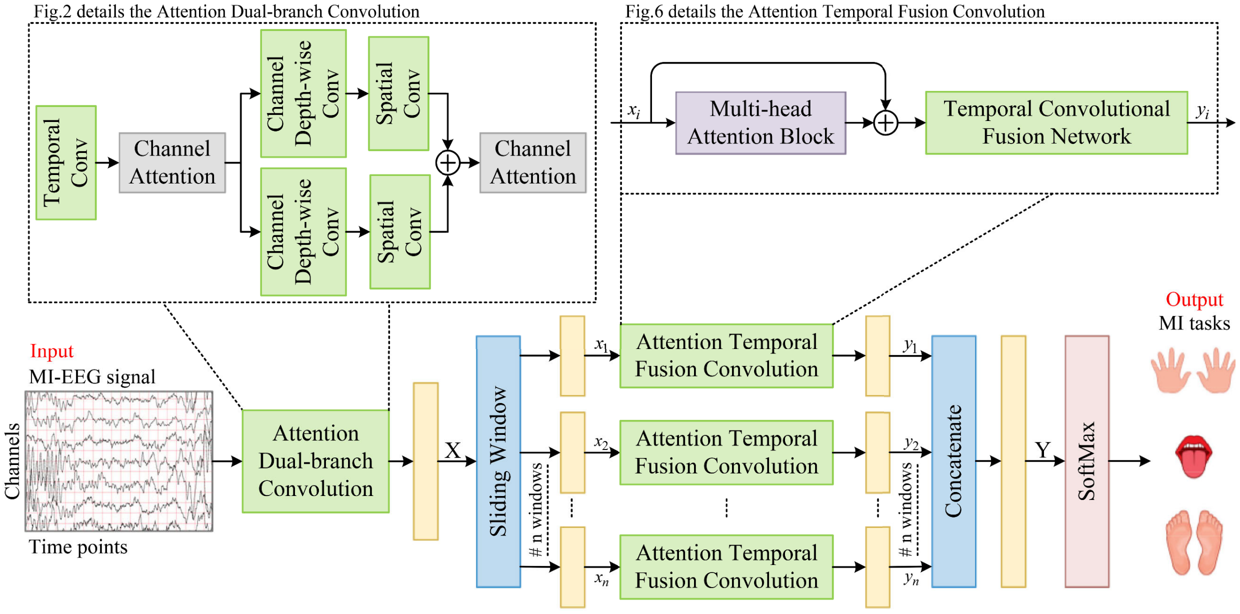
Fig.2 details the Attention Dual-branch Convolution
Temporal Conv
Channel Attention
Channel Depth-wise Conv
Spatial Conv
Channel Depth-wise Conv
Spatial Conv
Channel Attention
Fig.6 details the Attention Temporal Fusion Convolution
x_i
Multi-head Attention Block
Temporal Convolutional Fusion Network
y_i
Input
MI-EEG signal
Channels
Time points
Attention Dual-branch Convolution
X
Sliding Window
# n windows
x_1
x_2
x_n
Attention Temporal Fusion Convolution
Attention Temporal Fusion Convolution
Attention Temporal Fusion Convolution
y_1
y_2
y_n
# n windows
Concatenate
Y
SoftMax
Output
MI tasks

FIGURE 1. Components of the DB-ATCNet architecture.

II. METHOD

The paper proposes the DB-ATCNet model, which consists of two main modules: attention dual-branch convolution (ADBC) module and attention temporal fusion convolution (ATFC) module. The ADBC module includes a dual-branch convolutional network (DBCN) and two channel attention modules. The DBCN extracts low-level spatio-temporal features of MI-EEG signals, while the channel attention modules enhance EEG signals' spatial information selection ability. Ultimately, the ADBC module outputs a temporally sequenced series of high-level representations. Subsequently, a sliding window (SW) is employed to segment the series into multiple windows. These windows are then input in parallel into independent ATFC modules, as shown in Fig. 1. The ATFC module comprises a MSA and a subsequent temporal convolution fusion network (TCFN). The MSA automatically concentrate on the most critical information in every window and inputs the features into TCFN to further extract high-level temporal features from time series. Finally, a fully connected (FC) layer with a SoftMax activation function integrates all features from all windows. Probability predictions are generated using SoftMax for the performed MI tasks. Following are detailed descriptions of the DB-ATCNet architecture.

A. INPUT REPRESENTATION AND PREPROCESSING

This study employs the same input representation as ATCNet [26]. The DB-ATCNet model takes a MI trial $X_i \in \mathbb{R}^{C\times T}$ as input, consisting of C channels (EEG electrodes) and T time points. The objective of the DB-ATCNet model is to map the input MI trial X_i to its corresponding class y_i. The set of m labeled MI trials $S = \{X_i, y_i\}_{i=1}^{m}$ is given, where $y_i \in \{1, \ldots, n\}$ is the corresponding class label for trial X_i and n is the total number of classes defined in set S. For the BCI-2a [38] dataset, T = 1125 time points, C = 22 EEG channels, n = 4 MI classes, and m = 5184 MI trials. For the Physionet dataset [39], T = 640 time points, C = 64 EEG channels, n = 4 MI classes, and m = 9241 MI trials.

Recent studies related to deep learning have shown that utilizing raw EEG signals from public datasets without preprocessing (except for the preprocessing of the dataset itself (for example, before the release of the BCI-2a dataset, the data was subjected to 0.5-100 HZ bandpass filtering and 50 Hz notch filtering), no other preprocessing methods are used) as model input can lead to more competitive results [1]. This approach has been adopted by numerous studies, such as G-CRAM [12], MBEEGSE [14], ATCNet [26]. In this study, we followed this approach by using raw, unprocessed EEG data that spans the entire frequency range (0.5 - 100 Hz in the BCI-2a dataset and 0 - 80 Hz in the Physionet dataset) and encompasses all channels (22 in the BCI-2a dataset and 64 in the Physionet dataset), without any artifact removal. Before input into the DB-ATCNet model, MI-EEG signals are standardized, as follows:

$$x_i' = \frac{x_i - mean\,(x_i)}{std\,(x_i)}, i = 1, 2, 3, \ldots, C \tag{1}$$

where C represents the number of EEG channels.

B. ATTENTION DOUBLE-BRANCH CONVOLUTION BLOCK

The ADBC module draws inspiration from ATCNet [26] and MBEEGSE [14]. It consists of the dual-branch convolutional network (DBCN) and two channel attention modules, as depicted in Fig. 2. The DBCN comprises two components. The first component is a temporal convolutional layer that acquires spectral features of X_i across different frequency

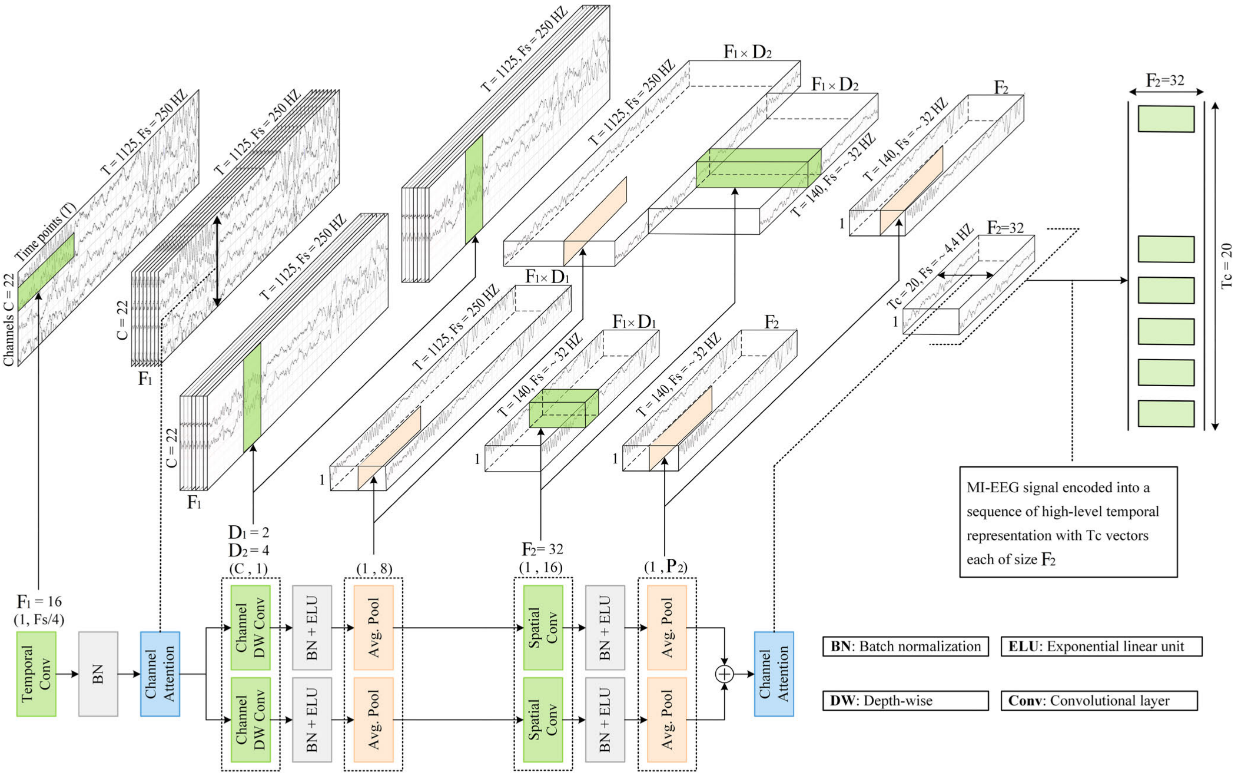

FIGURE 2. Attention double-branch convolution (ADBC) block.

bands. The second part is the dual-branch network, where each branch includes a spatial (depth-wise) convolutional layer which obtains spatial features from the feature maps and a spatio-temporal convolutional layer that learns how to properly fuse the spatial-temporal features. Each branch outputs a spatio-temporal sequence, which is then fused into a single advanced spatio-temporal representation sequence through summation. The first and second parts of DBCN are both followed by channel attention modules used to improve the learning ability of EEG signal spatial (channel) information. Details of the DBCN and channel attention are discussed in the subsequent sections.

1) DUAL-BRANCH CONVOLUTIONAL NETWORK

The DBCN module and the convolutional blocks explained in ATCNet [26] share similar kernel parameters and both utilize batch normalization (BN) [40], exponential linear units (ELU), and average pooling after spatial (depth-wise) convolutional layers and spatio-temporal convolutional layers to respectively enhance trainability, introduce non-linearity, and reduce dimensionality. However, they differ significantly in network structure. The DBCN utilizes a dual-branch network structure to enhance feature information and reduce network overfitting, resulting in improved model accuracy and generalization performance.

The first part of the DBCN module consists of a temporal convolutional layer that employs F_1 filters of size $(1, K_c)$, where K_c represents the length of the filter along the time axis. Enabling the filter to capture frequency information of 4 Hz and above, K_c is configured to be one-fourth of the sampling rate (64 in the BCI-2a dataset). As a result of the temporal convolutional layer, we obtain the F_1 temporal feature map. The DBCN module's second part comprises a dual-branch network, with both convolutional branches sharing the same structure. The initial layer in both branches is a depth-wise convolutional layer that extracts spatial features from each feature map (related to EEG channels). Each depth kernel learns spatial features from specific frequency bands. The output of the first part, processed through spatial depth convolution, results in a time series $S_{i,2} \in \mathbb{R}^{T\times D\times F_1}$, where each feature map is connected to D spatial kernels. In the experiments, the values of D for the two branches were set as $D_1 = 2$ and $D_2 = 4$, as explained in Section III-D. After the depth-wise convolutional layer, both branches were down-sampled using an average pooling layer with a size of (1, 8) for decreasing dimensionality. This decreased the time data and sampling frequency to 1/8 of the input, which means the signal's sampling rate is approximately 32 Hz. Both branches' second convolutional layers consist of spatio-temporal convolutional layers with filters of size $(1,K_{c2})$ for F_2. To ensure consistent output formats, we set the value of K_{c2} to be the same for both branches. Moreover, for decoding MI activity in 500 ms (sampled data at 32 Hz), we set K_{c2} to 16 for both branches. The layer of spatio-temporal convolution learns to optimally integrate spatio-temporal features, resulting in a high-level spatio-temporal representation sequence $S_{i,3} \in \mathbb{R}^{T_1 \times F_2}$. Then, an average pooling layer of size $(1, P_2)$ decreases the sampling rate to $\sim 32/P_2 Hz$. Based on the experiences of EEGNet [7] and ATCNet [26], we set P_2 to 7. Finally, the spatio-temporal sequences from both branches are fused through addition. Both the channel depth convolutional layer and the spatio-temporal convolutional layer use BN [40] to accelerate network training. Then, the ELU is activated to introduce nonlinearity.

The DBCN block produces a time series $z_i \in \mathbb{R}^{T_c \times d}$, comprising $\mathrm{T_c}$ time vectors, each of length d. Based on experience, we set $d = DF_1 = F_2 = 32$.

2) CHANNEL ATTENTION

Regarding neural network design, the feature extraction components of ATCNet [26], EEGNet [7], and the proposed DBCN module follow a similar pattern. They all start with temporal convolution, followed by depth-wise convolution, and finally spatio-temporal convolution. However, depth-wise convolutions and temporal convolutions have limited capacity to capture spatial information from EEG signals, which may result in the loss of spatial features.

To overcome this limitation, LMDA [41] successfully applied channel attention mechanisms to select channels in MI-EEG, which yielded promising results. Motivated by this success, we introduce channel attention mechanisms to recalibrate spatial and spectral information within the proposed DBCN module. We have integrated a channel attention module after the temporal convolution layer of the DBCN module to capture inter-channel interactions in MI-EEG data, enhancing its ability to learn spatial (channel) information. Furthermore, we have added another channel attention module after the dual-branch convolutional network of the DBCN module to further focus on key information in the high-level spatio-temporal representations produced by the network.

Commonly used channel attention mechanisms in EEG data processing include SE [30], CBAM [32], and ECA [31]. ECA module is preferred over SE and CBAM due to its ability to capture cross-channel interactions in an extremely lightweight manner while avoiding channel reduction. Additionally, ECA has fewer parameters than CBAM [31]. The experiments in Section III-E. demonstrate that both channel attention modules (ECA1 and ECA2) using the ECA module outperform other methods. Therefore, this study introduces channel attention modules (ECA1 and ECA2) that both adopt the ECA module. Fig. 3 depicts their specific structures.

In the ECA module, acquired spectral-temporal features are assigned attention weights using a neural network, and the network parameters are adaptively optimized based on the importance of each feature. The implementation process is detailed as follows:

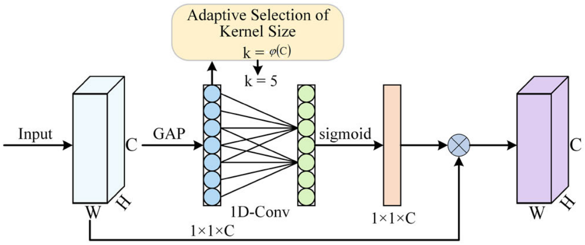

FIGURE 3. The architecture of ECA module. Dimensions of feature maps are represented by parameters C, H, and W. The k represents the 1D convolution size, and GAP refers to the global average pooling layer.

1) Fig. 3 illustrates that the ECA module input dimension is (N, H, W, C), where C is the number of feature maps. The GAP layer performs global average pooling, compressing the input data without reducing the dimension of the feature maps and aggregating features for each channel. Following the GAP layer, the input dimension becomes $(N, 1, 1, C)$.
2) Feature maps are automatically learned using a one-dimensional convolutional layer. Feature channels in the attention mechanism are influenced by the kernel size of one-dimensional convolutional layers. To calculate the value of k, which is proportional to the number of feature channels, an adaptive algorithm is proposed. The mathematical expression is as follows:

$$k = \psi(C) = \left| \frac{log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{2}$$

where b represents the offset of the linear mapping, and $\psi(C)$ represents the linear mapping between k and the number of channels C.

1) Following the calculation of cross-channel interaction ranges using 1D convolution in step (b), feature redistribution across non-aggregated channels is performed using σ. σ represents the relationship between the effect obtained after channel interactions and the assigned weights. In this study, σ employs the sigmoid function to map weights, as follows:

$$\omega = \sigma(C1D_k(y)) \tag{3}$$

where $C1D_k$ represents a one-dimensional convolution of size k.

In summary, the ECA module utilizes GAP to aggregate convolutional features, determines the kernel size k adaptively, performs one-dimensional convolution, and applies the sigmoid function to learn attention for local neighboring channels. This approach ensures the range of channel interactions, addresses dimensionality reduction, and enhances performance and speed by reducing computational and parameter overhead.

C. ATTENTIONAL TEMPORAL FUSION CONVOLUTION BLOCK

The Attention Temporal Fusion Convolution (ATFC) module contains a MSA block and a temporal convolution fusion network (TCFN). The MSA extracts key temporal features from the time series Z_i output by the AMDC block. Subsequently, the TCFN capture advanced temporal features from the sequence output by the MSA. To enhance the performance

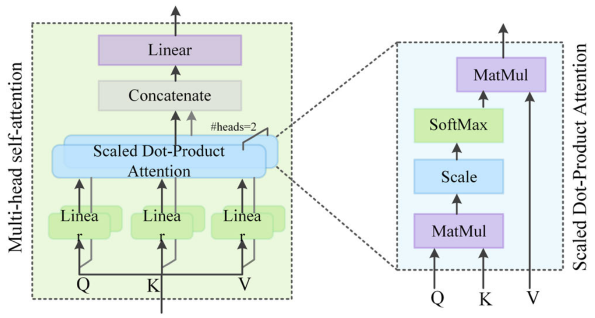
Multi-head self-attention
Linear
Concatenate
#heads=2
Scaled Dot-Product Attention
Linea r
Linea r
Linea r
Q
K
V
MatMul
SoftMax
Scale
MatMul
Q
K
V
Scaled Dot-Product Attention

FIGURE 4. Multi-head self-attention.

of DB-ATCNet, the input sequence Z_i is segmented into local windows by utilizing a sliding window (SW) approach. These local windows are used as inputs to parallel ATFC blocks, which will demonstrate improved performance as shown in the subsequent Experimental section. The SW, MSA, and TCFN will be explained in the subsequent paragraphs.

1) SLIDING WINDOW

The time series Z_i is divided into multiple local sequences $Z_i^w \in \mathbb{R}^{T_w \times d}$ using a sliding window (SW). This facilitates the extraction of individual local features. A SW of size $T_w = T_c - 5$ with one element step was employed, segmenting Z_i into 5 local windows. For further settings and a detailed discussion about sliding windows, please reference the research in [26].

2) MULTI-HEAD SELF-ATTENTION

The attention mechanism is an excellent method for capturing interactions in sequential data or images. The MSA employed in this study has demonstrated promising results in ATCNet [26]. Multiple self-attention heads make up the MSA layer. Each head executes scaled dot-product attention [29]. Each attention head contains three important parts: query(Q), key(K), and value(V). The interaction of keys and queries generates attention scores which emphasize the valuable aspects of the values, as shown in Fig. 4. The following is a detailed description of this interaction.

Initially, compute the query/key/value vector for every local window z_i^w through a linear projection, as follows:

$$q_t^h = W_Q^h LN\left(z_{i,t}^w\right) \in \mathbb{R}^{d_H}, W_Q^h \in \mathbb{R}^{d \times d_H} \quad (4)$$

$$k_t^h = W_K^h LN\left(z_{i,t}^w\right) \in \mathbb{R}^{d_H}, W_K^h \in \mathbb{R}^{d \times d_H} \quad (5)$$

$$v_t^h = W_V^h LN\left(z_{i,t}^w\right) \in \mathbb{R}^{d_H}, W_V^h \in \mathbb{R}^{d \times d_H} \quad (6)$$

where the variable $h = 1, \ldots, H$ denotes the head index, where H represents how many attention heads there are. $t = 1, \ldots, T_w$ represents the index of elements in the local window z_i^w, where T_w refers to the window's length (the whole number of vectors within the window). The head dimension is experimentally set to $d_H = d/2H$. LN represents layer normalization [42].

Next, the vector of context for every head can be calculated by multiplying the value V with the attention score. Assume a mini-batch with m key-value pairs ($K \in \mathbb{R}^{m \times d_H}$, $V \in \mathbb{R}^{m \times v}$) and n queries $Q \in \mathbb{R}^{m \times d_H}$, each head's context vector C^h is computed as follows:

$$C^h = softmax\left(\frac{Q^h \left(K^h\right)^T}{\sqrt{d_H}}\right) V^h \in \mathbb{R}^{n=T_w \times v=d_H} \quad (7)$$

where$Q^h \in \mathbb{R}^{n \times d_H}$, $V^h \in \mathbb{R}^{m \times v}$, and $K^h \in \mathbb{R}^{m \times d_H}$

In this study, we configure $v = d_H = 8$ and $n = m = T_w$.

Following this, by combining vector of context from all heads then linearly projecting the resulting vector, they are added to the input sequence z_i^w, thereby achieving multi-head self-attention, as follows:

$$z_i^w = W_0\left[C^1, \ldots, C^H\right] + z_i^w \in \mathbb{R}^{T_w \times d}, W_0 \in \mathbb{R}^{d_H \times d} \quad (8)$$

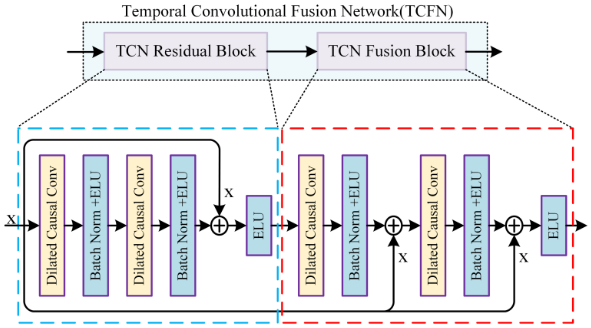

FIGURE 5. The temporal convolution fusion network.

3) TEMPORAL CONVOLUTION FUSION NETWORK

The design of TCFN model resembles the TCN network proposed in [26], as they share the same set of hyperparameters. The key difference lies in the replacement of the second TCN residual block in the TCN network described in ATCNet [26] with a TCN fusion block in the TCFN module. This TCN fusion block modifies the residual connections in the TCN residual block, replacing them with multi-level residual connections, as depicted in Fig. 5. These multi-level residual connections in the TCN fusion block enable multi-level feature fusion, enriching the feature information while mitigating model overfitting.

TCFN comprises a series of residual blocks, where every block connects two causal dilated convolutional layers to an exponential linear unit (ELU) and a BN [40], as shown in Fig. 5. For further information on the TCFN structure, please refer to [26].

Fig. 6 shows 16 temporal components ($T_w = 16$) that enter the TCFN module. Every component is a vector of size F_2 (equivalent to the kernel number in the final transformation layer of the ADBC block). The last component in the TCFN module output sequence is a vector of length F_T. For our research, we set $F_T = F_2 = 32$.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. DATASET

We evaluated the DB-ATCNet model using the BCI-IV2a dataset [38] and Physionet EEG motor movement/imagery dataset [39].

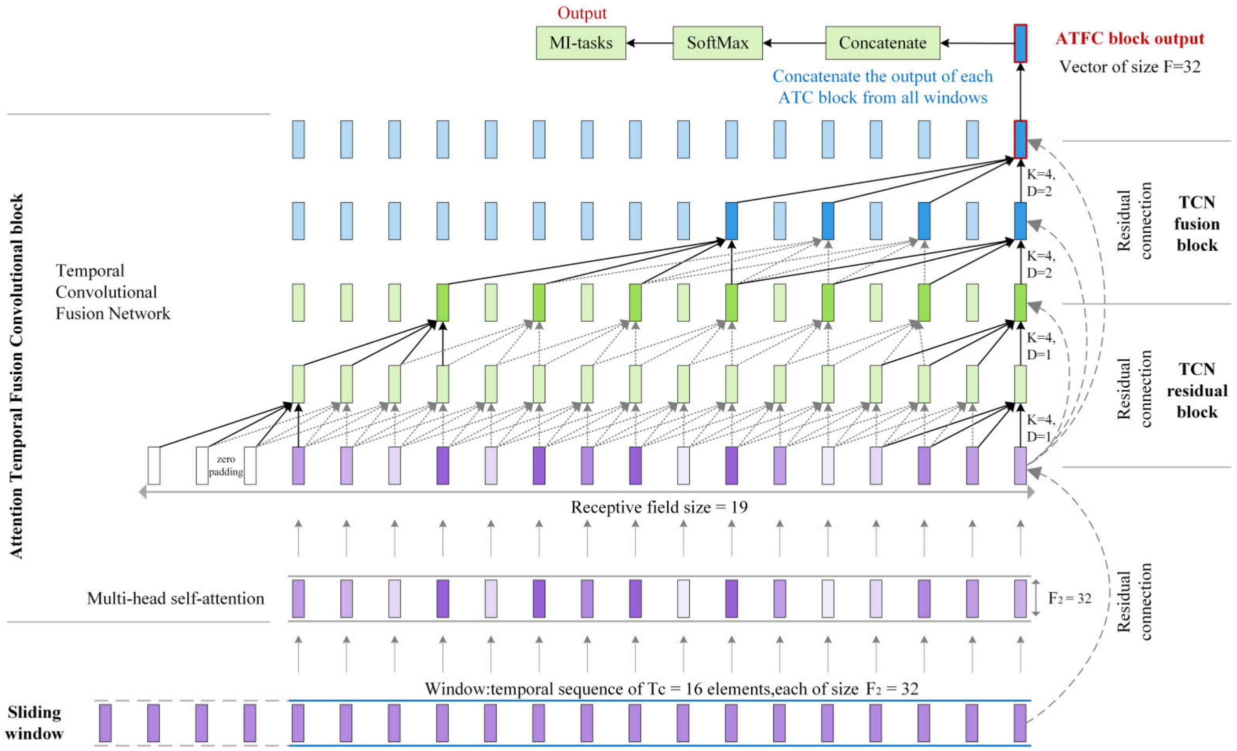
Output
MI-tasks
SoftMax
Concatenate
ATFC block output
Vector of size F=32
Concatenate the output of each ATC block from all windows
Attention Temporal Fusion Convolutional block
Temporal Convolutional Fusion Network
K=4, D=2
K=4, D=2
K=4, D=1
K=4, D=1
Residual connection
TCN fusion block
Residual connection
TCN residual block
zero padding
Receptive field size = 19
Multi-head self-attention
F2 = 32
Residual connection
Window:temporal sequence of Tc = 16 elements,each of size F2 = 32
Sliding window

FIGURE 6. Visualization of feature maps in the attention temporal fusion convolution network (ATFC) block using a window of 16 elements ($T_W = 16$).

BCI-IV2a dataset: This dataset contains 5184 MI trials across 4 classes of MI tasks. Every trial lasting 8 seconds with the MI activity occurring within the middle 4 seconds. The BCI-IV2a dataset consists of two sessions, both recorded by 9 subjects with 22 EEG sensors. DB-ATCNet model was trained on one session and evaluated for its performance on another session.

Physionet EEG motor movement/imagery dataset: The dataset comprises EEG recordings from 109 participants who performed 4 tasks across 14 experimental runs. Participants utilized the BCI 2000 system to record 64-channel EEG signals at a sampling rate of 160 Hz while performing tasks. The 14 experiments consisted of 2 baseline runs, 6 actual movement runs, and 6 motor imagery (MI) runs, which included four types of MI tasks: left fist (L), right fist (R), both fists (LR), and both feet (F). Each type of MI task consisted of 21 trials, each lasting 4 seconds and containing 640 time points. Data related to participants 38, 88, 89, 92, 100, and 104 were excluded from the sample due to annotation errors.

B. EVALUATION METHOD AND PERFORMANCE METRICS

On the BCI-IV 2a dataset, the proposed model was evaluated through subject-independent and subject-dependent methods. Subject-dependent evaluation was conducted using original competition testing and training data. The model underwent training on 9 × 288 trials in Phase 1 and was then evaluated in Phase 2 using another set of 9 × 288 trials. The 'Leave-One-Subject-Out' (LOSO) evaluation method was used for subject-independent evaluation [26].

On the Physionet dataset, the proposed model was evaluated through subject-independent methods. Specifically, we employ 10-fold cross-validation to evaluate the model performance. In each validation, 10% of the data is randomly selected for testing, while the remaining 90% is used for training. This process is repeated ten times, and the average of the ten accuracies obtained is taken as the classification result.

In this research, accuracy and Kappa scores are utilized to evaluate the proposed models, which are described as follows:

$$ACC = \frac{\sum_{i=1}^{n} TP_i/I_i}{n} \tag{9}$$

where n indicates the number of classes, I_i is the number of samples in class i, and TP_i is the true positive, i.e., the number of correctly predicted samples in class i.

$$k_{score} = \frac{1}{n}\sum_{a=1}^{n}\frac{P_a - P_e}{1 - P_e} \tag{10}$$

where P_e is the expected percentage chance of agreement, P_a is the actual percentage of agreement, and n is the number of classes.

TensorFlow framework and Python 3.8 were utilized to train and test the model on a single Nvidia GTX 3080 10GB GPU. Initialize the weights using the Glorot uniform initializer. The Adam optimizer was utilized with a learning rate of 0.0009, and a categorical cross-entropy loss function. The BCI-IV 2a dataset was trained for 1000 epochs with a

TABLE 1. Contribution of each block in the DB-ATCNet model to the performance of MI Classification using the BCI-2a dataset.

Removed block	Accuracy %	k-score
None (DB-ATCNet)	87.54	0.834
ECA 1	87.31	0.831
ECA 2	87.11	0.828
ECA 1+ECA 2	87.23	0.830
SW	85.49	0.807
MSA	85.69	0.809
SW+MSA	84.95	0.799
TCFN	79.55	0.727
SW+TCFN	79.90	0.732
MSA+TCFN	83.22	0.776
SW+MSA+TCFN (ADBC)	82.52	0.767
SW+MSA+ECA 1+TCFN	81.75	0.757
SW+MSA+ECA 2+TCFN	81.02	0.747
SW+MSA+ECA 1+ECA 2+TCFN (DBCN)	81.37	0.752

ECA 1: efficient channel attention 1, ECA 2: efficient channel attention 2, SW: sliding window, MSA: multi-head self-attention, TCFN: temporal convolution fusion network.

patience of 300 and a batch size of 64, while the Physionet dataset was trained for 500 epochs with a patience of 100 and a batch size of 32.

C. THE CONTRIBUTIONS OF DB-ATCNet BLOCKS

This section evaluates the performance of each module in the DB-ATCNet model through ablation analysis. Table 1 shows the effect of deleting single or multiple blocks in the DB-ATCNet model on MI classification performance applying the BCI-2a dataset. The results show that the MSA and SW modules improve the overall accuracy by 1.85% and 2.05%, respectively. Compared to the absence of the ECA 1 and ECA 2 modules, adding only the ECA 1 module decreases the overall accuracy by 0.12%, adding only the ECA 2 module increases it by 0.08%, and including both the ECA 1 and ECA 2 modules together increases it by 0.31%.

In addition, when considering the DBCN module alone, adding the ECA 1 module decreases accuracy by 0.35%, adding the ECA 2 module increases it by 0.39%, and adding both the ECA 1 and ECA 2 modules together increases it by 1.16%. This indicates that the ECA 1 and ECA 2 modules affect the overall accuracy of the model by working together with the DBCN module, adding the ECA 1 and ECA 2 modules simultaneously can achieve the best results of the model. Additionally, the introduction of the TCFN module results in a 2.43% increase in accuracy compared to the ATBC module alone.

Overall, with the exception of the MSA and ECA 1 modules, the independent addition of the other modules improves the overall accuracy. When the TCFN module is added following the MSA module, it improves accuracy; conversely, when the MSA module is removed before the TCFN module, accuracy decreases and may fall below that achieved with the DBCN module alone. This suggests that placing the MSA module directly in the end may degrade performance, but adding an extra classification layer could potentially improve model performance. In conclusion, our ablation experiments show that the proposed modules have a beneficial effect on the results of MI-EEG classification tasks.

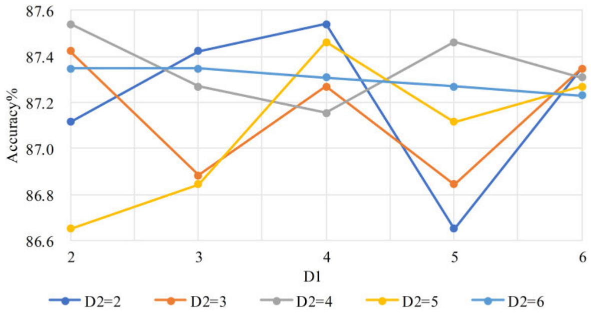

FIGURE 7. Accuracy of the BCI-2a as a function of the number of output channels in the deep-wise convolutional layers.

D. VARYING THE NUMBER OF OUTPUT CHANNELS IN THE DEEP-WISE CONVOLUTIONAL LAYERS OF DBCN

This section examines the effect of varying the output channels of the two deep convolutional layers of the DBCN module on the accuracy of DB-ATCNet on the BCI-IV 2a dataset. The output channels of deep-wise convolutional layers are determined by $F \times D$, where F is the number of input channels and D is the number of filters connected to every feature map in the previous layer. In our study, both deep convolutional layers had F set to 16, so the output channels were uniquely determined by the parameters D_1 and D_2. We used a controlled experimental design to analyze the influence of D_1 and D_2 on the performance of DB-ATCNet. Based on experience, we set the range of D_1 and D_2 to integers between 2 and 6. Fig. 7 illustrates the change in DB-ATCNet accuracy when we fixed the value of D_2 and then increased D_1 from 2 to 6. The graph shows that when D_2 is fixed, the accuracy is lower when D_1 is the same as D_2 than when they are different. This indicates that when both branches have the same number of channels, the learned features are similar, limiting the advantage of the dual-branch structure in enriching features. In addition, DB-ATCNet achieves optimal performance when D_1 and D_2 are both set to 2 and 4.

E. COMPARING DIFFERENT CHANNEL ATTENTION SCHEMES

In this study, we increased the performance of the DB-ATCNet by integrating channel attention modules before and after the dual-branch convolutional network (DBCN) module. For the sake of clarity in the following experimental descriptions, we will refer to these two channel attention modules as CA 1 and CA 2.

In this section, we used three different attention mechanisms (ECA [31], CBAM [32], and SE [30]) to validate the effectiveness of the insertion positions of CA 1 and CA 2, and to explore the optimal choices for CA 1 and CA 2. Fig. 8 illustrates our experimental results on the BCI-IV 2a dataset, where “None” indicates the absence of an attention module.

When either CA 1 or CA 2 was added individually, the use of ECA for CA 2 resulted in improved accuracy. When both CA 1 and CA 2 were added simultaneously, selecting CBAM for either CA 1 or CA 2 decreased accuracy, whereas

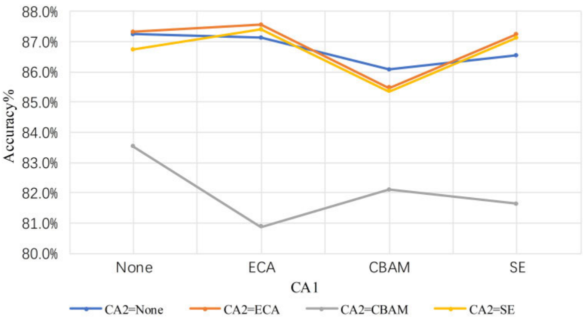

FIGURE 8. Performance of DB-ATCNet model with different channel attention schemes: ECA, CBAM, and SE.

selecting ECA or SE for both CA 1 and CA 2 increased accuracy. Notably, when ECA was used for both CA 1 and CA 2, accuracy improved by 0.31% and the highest accuracy was achieved. These results suggest that strategically inserting channel attention at specific locations can improve model performance. Furthermore, CBAM was found to be inappropriate for this model, while SE and ECA were found to be appropriate for the EEG representation in this model, with ECA being the optimal choice for channel attention in this context.

F. COMPARISON TO RECENT STUDIES ON THE BCI-IV 2a DATASET

In this section, we use the BCI-IV2a dataset to evaluate the performance of DB-ATCNet and compare it with other reproduced models, including EEGNet [7], EEG-TCNet [24], TCNet_Fusion [25], and ATCNet [26]. These models were preprocessed, trained, and evaluated according to the methods specified in this study, while their results depend on the parameters specified within the original papers. Table 2 presents the average and best performance of every model according to 10 random runs. The results show that DB-ATCNet achieves an accuracy of 87.5% and a κ-score of 0.83 and outperforms the other models in all subjects. DB-ATCNet also has better average performance than other models. This shows that the designed model has excellent learning capabilities as well as is able to achieve the same stable performance over multiple runs. Additionally, the proposed model showed minimal standard deviation, meaning better stability across subjects than the other models.

Fig. 9 presents the confusion matrix of reproduced networks and DB-ATCNet. In comparison, DB-ATCNet presented gains in classification performance for all MI classes.

T-distributed stochastic neighbor embedding (t-SNE) [43] is a widely employed statistical technique for reducing dimensionality and visualizing features. t-SNE maps high-dimensional data points to a low-dimensional space while retaining the local structure of the data. Fig. 10 illustrates the feature distributions of the BCI-IV2a dataset after separate training using reproduced networks and DB-ATCNet. The visualization results indicate that the features extracted by the proposed DB-ATCNet model have clearer boundaries and more distinct clustering. This further demonstrates the model's ability to effectively extract feature information and enhance the separability of different categories.

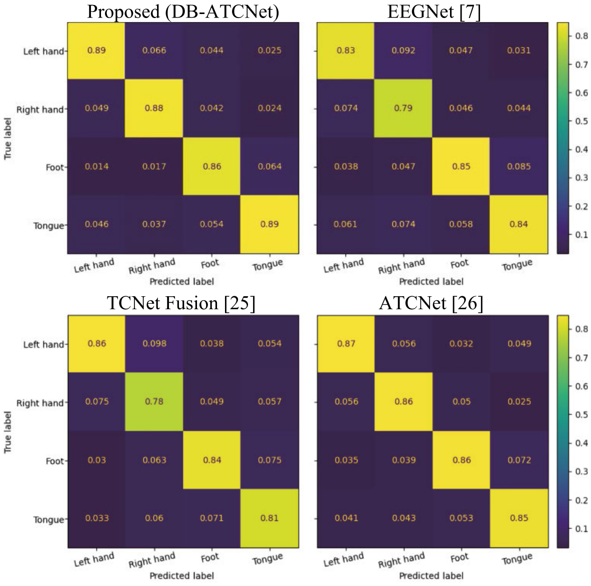

FIGURE 9. Confusion matrices of the reproduced models and DB-ATCNet.

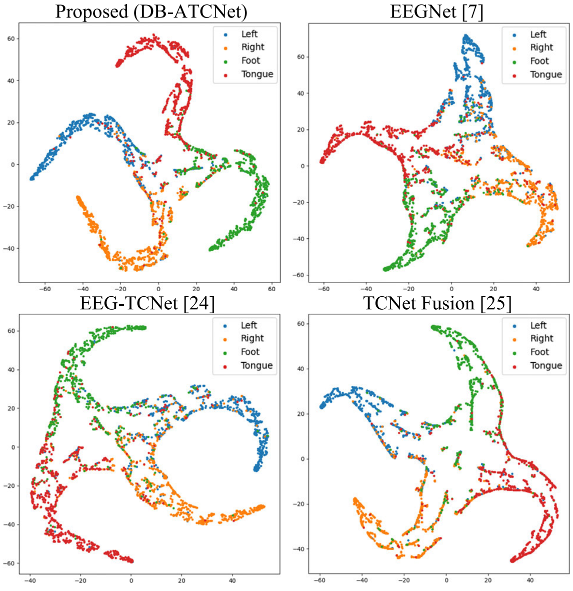

FIGURE 10. The results of t-SNE visualization of the extracted features using the reproduced networks and DB-ATCNet for the subject-specific classification task on the BCI-2a dataset.

Table 3 presents a comparison of recent research in MI decoding, considering preprocessing techniques, input formulations, model architectures, and model performance in subject-independent and subject-dependent methods. For all researches, leave-one-subject-out (LOSO) cross-validation was utilized for subject-independent evaluation and the original BCI-IV2a competition division (Session 1 for training and Session 2 for testing) was utilized for subject-dependent evaluation. Our proposed DB-ATCNet outperforms previous studies utilizing raw MI-EEG signals without preprocessing.

TABLE 2. Performance (accuracy (%) and-score (k)) comparison of subject-specific classification using BCI-2a dataset for the proposed model with other reproduced models.

Sub.	Proposed (DB-ATCNet)				ATCNet [26]				EEGNet [7]				EEG-TCNet [24]				TCNet-Fusion[25]			
	best		average		best		average		best		average		best		average		best		average	
	%	k	%	k	%	k	%	k	%	k	%	k	%	k	%	k	%	k	%	k
1	91.3	0.88	89.4	0.86	89.9	0.87	87.7	0.84	89.6	0.86	86.6	0.82	86.1	0.81	82.3	0.76	87.2	0.83	84.5	0.79
2	74.3	0.66	70.8	0.61	72.6	0.63	69.3	0.59	66.3	0.55	61.3	0.48	63.5	0.51	60.2	0.47	66.0	0.55	64.4	0.53
3	97.9	0.97	96.6	0.96	96.9	0.96	95.2	0.94	96.9	0.96	91.9	0.89	95.1	0.94	93.6	0.91	93.8	0.92	92.1	0.89
4	85.1	0.80	82.6	0.77	82.4	0.77	78.3	0.71	70.8	0.61	66.8	0.56	72.2	0.63	68.9	0.59	74.3	0.66	71.3	0.62
5	84.0	0.79	81.3	0.75	82.3	0.76	79.6	0.73	79.9	0.73	71.6	0.62	77.8	0.70	73.9	0.65	79.9	0.73	77.1	0.69
6	78.5	0.71	76.5	0.69	75.4	0.67	73.5	0.65	65.3	0.54	60.3	0.47	63.2	0.51	60.6	0.47	66.3	0.55	64.7	0.53
7	96.2	0.95	94.3	0.92	93.8	0.92	91.5	0.89	90.3	0.87	88.6	0.85	91.0	0.88	87.5	0.83	92.7	0.91	90.4	0.87
8	89.2	0.86	87.7	0.84	89.6	0.86	88.4	0.84	86.8	0.82	84.3	0.79	85.4	0.81	83.1	0.77	88.5	0.85	86.0	0.81
9	91.3	0.88	90.1	0.87	89.6	0.86	88.2	0.84	88.2	0.84	82.8	0.77	85.8	0.81	80.0	0.73	88.5	0.85	83.6	0.78
Mean	**87.5**	**0.83**	**85.5**	**0.81**	85.8	0.81	83.5	0.78	81.6	0.75	77.1	0.69	80.0	0.73	76.7	0.69	81.9	0.76	79.3	0.72
St. D.	**7.8**	**0.10**	**8.4**	**0.11**	8.2	0.11	8.7	0.12	11.5	0.15	12.2	0.16	11.6	0.16	11.7	0.15	10.8	0.15	10.5	0.14

TABLE 3. Subject-dependent and subject-independent performance comparison between DB-ATCNet and recent studies using the BCI-IV2a dataset. The average score (k) and accuracy (%) of all subjects is presented.

Study (* Reproduced)	Description			Performance			
	Pre-processing [1]	Input data [2]	DL approach	Subject-dependent		Subject-independent	
				%	k	%	k
Lawhern et al.2018, [7] *	FB:8-35	Rs	CNN (EEGNet)	80.1	0.73	68.8	0.58
Hassanpour et al.2019, [20]	FB:8-35 Hz, AR: SWT	SF	DBN-AE	71.0	-	-	-
Amin et al.2019, [16]	FB:0.5-40Hz	Rs	Multi-layer-CNN, MLP	75.0	-	55.3	-
Ingolfsson et al.2020, [24] *	AR: manual	Rs	CNN+TCN (EEG-TCNet)	80.0	0.73	69.5	0.59
Zhang et al.2020, [12]	no preprocessing	TM	Attention, graph CNN, LSTM	-	-	60.1	-
Musallam et al.2021, [25] *	AR: manual	Rs	Multi-layer CNN, TCN (TCNet_Fusion)	80.9	0.75	70.6	0.61
Amin et al.2022, [10]	FB:8-35	Rs	Attention, inception CNN, LSTM	82.8	-	-	-
Altuwaijri et al.2022, [14] *	no preprocessing	Rs	Attention, multi-branch CNN	82.2	0.76	68.7	0.58
Altaheri et al.2023 [26] *	no preprocessing	Rs	Attention, CNN, TCN (ATCNet)	85.7	0.81	70.9	0.61
Proposed method	no preprocessing	Rs	Dual-branch. CNN, Attention, TCN (DB-ATCNet)	**87.5**	**0.83**	**71.3**	**0.62**

[1] Preprocessing: signal filtering (**SF**), channel selection (**CS**), and artifact removal (**AR**). **SWT**: Synchrosqueezed wavelet transforms.

[2] Input formulation: Topological maps (**TM**), Spectral features (**SF**), Raw signal (**RS**).

In particular, proposed model demonstrates superior performance in subject-independent and subject-dependent evaluations, demonstrating its robust generalization capability to new subjects.

G. COMPARISON TO RECENT STUDIES ON THE PHYSIONET MI-EEG DATASET

In this section, we conducted two-class (L and R) and four-class (L, R, LR, and F) classification tasks using the Physionet MI-EEG dataset to test the generalization performance of the DB-ATCNet model. We compared the performance of DB-ATCNet with other state-of-the-art models on the same dataset.

Table 4 shows that the proposed model achieved accuracies of 87.33% and 69.58% in the two-class and four-class classification tasks, respectively. Compared to other state-of-the-art models, the proposed model demonstrated an accuracy improvement of 0.63% and 1.04% in the two-class and four-class tasks, respectively.

Fig. 11 illustrates the confusion matrices of DB-ATCNet in the two-class and four-class classification tasks. The results indicate that DB-ATCNet performed well in classifying the

TABLE 4. Comparison of DB-ATCNet and state-of-the-art methods on the PhysioNet dataset.

Models	Subjects	Class Type	Max. ACC	Avg. ACC
SUT-CCSP + SVM (2013) [44]	56	MI 2-class	90%	72.37%
Phase information (2014) [45]	103	MI 2-class	71.55%	-
SUT-CCSP random forest (2016) [46]	24	MI 2-class	-	80.05%
IMOCS (2016) [47]	85	MI 2-class	-	63%
	35	MI 2-class	-	79.9%
CNNs (2018) [48]	105	MI 2-class	-	80.38%
G-CRAM (2020) [12]	105	MI 2-class	74.71%	-
EEGNet Fusion (2020) [49]	103	MI 2-class	-	83.80%
BENDR (2021) [50]	105	MI 2-class	-	86.70%
s-CTrans (2022) [51]	109	MI 2-class	-	83.31%
CNNs (2018) [48]	105	MI 4-class	-	58.58%
EEGNet (2020) [52]	105	MI 4-class	-	65.07%
ConTraNet CNN-Transformer (2022) [53]	105	MI 4-class	-	65.44%
t-CTrans (2022) [51]	109	MI 4-class	-	68.54%
Proposed method	103	**MI 2-class**	**88.58%**	**87.33%**
	103	**MI 4-class**	**70.38%**	**69.58%**

Subjects: number of subjects used. Class Type: number of MI classes used. Max. ACC: maximum classification accuracy. Avg. ACC: average classification accuracy. In our study, Max. ACC and Avg. ACC are the best accuracy and average accuracy of 10 experiments, respectively.

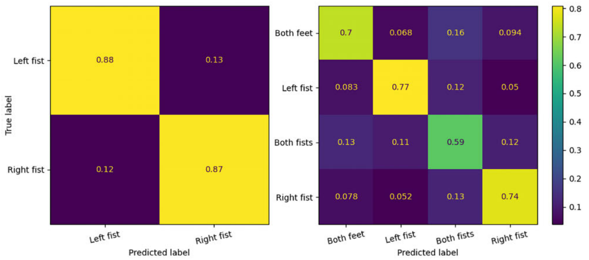

FIGURE 11. Confusion matrix of DB-ATCNet for two-class and four-class classification tasks on the Physionet dataset.

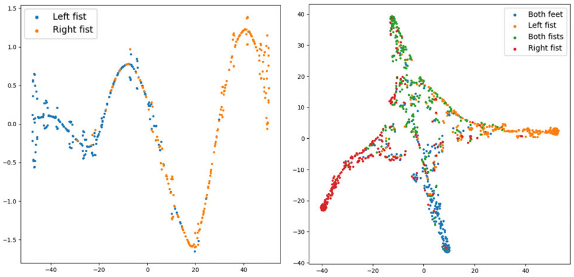

FIGURE 12. The results of t-SNE visualization of the extracted features using the DB-ATCNet for the two-class and four-class classification tasks on the Physionet dataset.

left fist, right fist, and feet categories, but exhibited poorer performance in identifying the both fists category. This suggests that the proposed model has a good ability to recognize neural activity in bilateral brain regions during single-hand movements, but lacks the ability to effectively recognize neural activity during both-hand movements. Additionally, this model demonstrates better capacity in processing neural activity information from the unilateral motor cortex.

We visualized the features extracted by the DB-ATCNet model during two-class and four-class classification tasks on the Physionet dataset using the t-SNE method, and the results are shown in Fig. 12. In both tasks, most samples from the left fist and right fist categories exhibit high feature distinguishability, indicating that the DB-ATCNet model can effectively extract features for these two types of motor imagery tasks. However, in the four-class classification tasks, the features extracted for both feet motor imagery demonstrate high distinguishability, while those for both fists motor imagery overlap significantly and have low distinguishability, which hinders accurate classification.

IV. CONCLUSION

This paper proposes a dual-branch attentional temporal convolutional network (DB-ATCNet) model to recognize the MI activities based on EEG signals. The proposed model consists of two modules: attention dual-branch convolution (ADBC) module and attention temporal fusion convolution (ATFC) module. The ADBC module employs a dual-branch convolutional network (DBCN) to extract low-level spatio-temporal features and enhances spatial feature extraction using the ECA module. The ATFC module utilizes sliding windows, self-attention mechanisms, and the temporal convolutional fusion network (TCFN) module to extract high-level temporal features. Ablation experiments indicate the contributions of each module to the overall performance of the DB-ATCNet model. When evaluated on the challenging BCI-IV2a dataset with little preprocessing and without removing artifacts, the model achieves a subject-independent accuracy of 71.34% and a subject-dependent accuracy of 87.54%, outperforming current DL architectures. This indicates the robust capability of the DB-ATCNet model to recognize MI activities from raw brain signals. The DB-ATCNet model improves EEG decoding performance for all subjects in the BCI-IV2a dataset, showing its capability to extract generalized MI features from different categories and subjects. Furthermore, we obtained accuracies of 87.33% and 69.58% in the two-class and four-class classification tasks, respectively, on the PhysioNet dataset, which further confirms the strong generalization and feature recognition abilities of DB-ATCNet. With high performance and relatively fewer parameters (150k), the DB-ATCNet model is well-suited for resource-constrained Internet of Things (IoT) and edge devices.

Although our research has yielded promising results, there are still some limitations that impede its practical application in engineering. One such limitation is the length of the time window, which is currently too long. Future work will focus on reducing the length of time windows (4.5s for the BCI 2a dataset and 4s for the Physionet dataset) used by DB-ATCNet and other CNN-based BCI methods, as these windows result in excessive latency for online BCI applications. We will also address the challenge of handling continuous incoming data streams from EEG recordings, which makes aligning the onset of imagined events with the data window fed to the CNN difficult. Despite these challenges, continuing this work will contribute to the future development of affordable, portable EEG-based BCI systems.

REFERENCES

[1] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, vol. 35, no. 20, pp. 14681–14722, Jul. 2023.

[2] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *NeuroImage*, vol. 34, no. 4, pp. 1443–1449, Feb. 2007.

[3] A. Jafarifarmand and M. A. Badamchizadeh, "EEG artifacts handling in a real practical brain–computer interface controlled vehicle," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1200–1208, Jun. 2019.

[4] H. Altaheri, M. Alsulaiman, and G. Muhammad, "Date fruit classification for robotic harvesting in a natural environment using deep learning," *IEEE Access*, vol. 7, pp. 117115–117133, 2019.

[5] M. A. Qamhan, H. Altaheri, A. H. Meftah, G. Muhammad, and Y. A. Alotaibi, "Digital audio forensics: Microphone and environment classification using deep learning," *IEEE Access*, vol. 9, pp. 62719–62733, 2021.

[6] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[8] R. Mane, N. Robinson, A. P. Vinod, S.-W. Lee, and C. Guan, "A multi-view CNN with novel variance layer for motor imagery brain computer interface," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 2950–2953.

[9] J. Wang, L. Yao, and Y. Wang, "IFNet: An interactive frequency convolutional neural network for enhancing motor imagery decoding from EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1900–1911, 2023.

[10] S. U. Amin, H. Altaheri, G. Muhammad, W. Abdul, and M. Alsulaiman, "Attention-inception and long- short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5412–5421, Aug. 2022.

[11] S. U. Amin, H. Altaheri, G. Muhammad, M. Alsulaiman, and W. Abdul, "Attention based inception model for robust EEG motor imagery classification," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2021, pp. 1–6.

[12] D. Zhang, K. Chen, D. Jian, and L. Yao, "Motor imagery classification via temporal attention cues of graph embedded EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2570–2579, Sep. 2020.

[13] D. Li, J. Xu, J. Wang, X. Fang, and Y. Ji, "A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2615–2626, Dec. 2020.

[14] G. A. Altuwaijri, G. Muhammad, H. Altaheri, and M. Alsulaiman, "A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for EEG-based motor imagery signals classification," *Diagnostics*, vol. 12, no. 4, p. 995, Apr. 2022.

[15] T. Liu and D. Yang, "A densely connected multi-branch 3D convolutional neural network for motor imagery EEG decoding," *Brain Sci.*, vol. 11, no. 2, p. 197, Feb. 2021.

[16] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Gener. Comput. Syst.*, vol. 101, pp. 542–554, Dec. 2019.

[17] X. Liu, S. Xiong, X. Wang, T. Liang, H. Wang, and X. Liu, "A compact multi-branch 1D convolutional neural network for EEG-based motor imagery classification," *Biomed. Signal Process. Control*, vol. 81, Aug. 2023, Art. no. 104456.

[18] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164–2177, Oct. 2019.

[19] J. Xu, H. Zheng, J. Wang, D. Li, and X. Fang, "Recognition of EEG signal motor imagery intention based on deep multi-view feature learning," *Sensors*, vol. 20, no. 12, p. 3496, Jun. 2020.

[20] A. Hassanpour, M. Moradikia, H. Adeli, S. R. Khayami, and P. Shamsinejadbabaki, "A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals," *Expert Syst.*, vol. 36, no. 6, Dec. 2019, Art. no. e12494.

[21] S. Kumar, R. Sharma, and A. Sharma, "OPTICAL+: A frequency-based deep learning scheme for recognizing brain wave signals," *PeerJ Comput. Sci.*, vol. 7, p. e375, Feb. 2021.

[22] T.-J. Luo, C.-L. Zhou, and F. Chao, "Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network," *BMC Bioinf.*, vol. 19, no. 1, p. 344, Sep. 2018.

[23] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[24] T. Mar Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces," 2020, *arXiv:2006.00622*.

[25] Y. K. Musallam, N. I. AlFassam, G. Muhammad, S. U. Amin, M. Alsulaiman, W. Abdul, H. Altaheri, M. A. Bencherif, and M. Algabri, "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," *Biomed. Signal Process. Control*, vol. 69, Aug. 2021, Art. no. 102826.

[26] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-informed attention temporal convolutional network for EEG-based motor imagery classification," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 2249–2258, Feb. 2023.

[27] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1412–1421.

[28] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaise, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[30] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[31] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.

[32] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[33] H. Jia, S. Yu, S. Yin, L. Liu, C. Yi, K. Xue, F. Li, D. Yao, P. Xu, and T. Zhang, "A model combining multi branch spectral–temporal CNN, efficient channel attention, and LightGBM for MI-BCI classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1311–1320, 2023.

[34] S. Huang, Z. Lu, R. Cheng, and C. He, "FaPN: Feature-aligned pyramid network for dense image prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 844–853.

[35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944.

[36] S. Amal, L. Safarnejad, J. A. Omiye, I. Ghanzouri, J. H. Cabot, and E. G. Ross, "Use of multi-modal data and machine learning to improve cardiovascular disease care," *Frontiers Cardiovascular Med.*, vol. 9, Apr. 2022, Art. no. 840262.

[37] M. Milosevic, Q. Jin, A. Singh, and S. Amal, "Applications of AI in multi-modal imaging for cardiovascular disease," *Frontiers Radiol.*, vol. 3, Jan. 2024, Art. no. 1294068.

[38] C. Brunner, R. Leeb, G. M´uller-Putz, A. Schlogl, and G. Pfurtscheller, "BCI competition 2008-graz data set a," *Inst. Knowl. Discov. Graz Univ.Technol.*, vol. 16, pp. 1–6, 2008.

[39] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, p. E215, Jun. 2000.

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[41] Z. Miao, M. Zhao, X. Zhang, and D. Ming, "LMDA-Net: A lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability," *NeuroImage*, vol. 276, Aug. 2023, Art. no. 120209.

[42] J. Lei Ba, J. Ryan Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.

[43] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, Nov. 2008.

[44] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 1–10, Jan. 2014.

[45] A. Loboda, A. Margineanu, G. Rotariu, and A. Mihaela, "Discrimination of EEG-based motor imagery tasks by means of a simple phase information method," *Int. J. Adv. Res. Artif. Intell.*, vol. 3, no. 10, p. 10, 2014.

[46] Y. Kim, J. Ryu, K. K. Kim, C. C. Took, D. P. Mandic, and C. Park, "Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–13, Sep. 2016.

[47] V. S. Handiru and V. A. Prasad, "Optimized bi-objective EEG channel selection and cross-subject generalization with brain–computer interfaces," *IEEE Trans. Hum.-Mach. Syst.*, vol. 46, no. 6, pp. 777–786, Dec. 2016.

[48] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Syst. Appl.*, vol. 114, pp. 532–542, Dec. 2018.

[49] K. Roots, Y. Muhammad, and N. Muhammad, "Fusion convolutional neural network for cross-subject EEG motor imagery classification," *Computers*, vol. 9, no. 3, p. 72, Sep. 2020.

[50] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, "BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data," *Frontiers Human Neurosci.*, vol. 15, Jun. 2021, Art. no. 653659.

[51] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, and Y. Zhan, "A transformer-based approach combining deep learning network and spatial–temporal information for raw EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2126–2136, 2022.

[52] X. Wang, M. Hersche, B. Tömekce, B. Kaya, M. Magno, and L. Benini, "An accurate EEGNet-based motor-imagery brain–computer interface for low-power edge computing," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, Jun. 2020, pp. 1–6.

[53] O. Ali, M. Saif-ur-Rehman, T. Glasmachers, I. Iossifidis, and C. Klaes, "ConTraNet: A single end-to-end hybrid network for EEG-based and EMG-based human machine interfaces," 2022, *arXiv:2206.10677*.

KAI ZHOU received the B.S. degree in mechanical engineering from Henan Polytechnic University of Science and Technology. He is currently pursuing the M.S. degree. His research interests include brain–computer interfaces, signal analysis, and artificial intelligence (machine learning and deep learning).

AIERKEN HAIMUDULA received the B.S. degree from Xinjiang University, China, and the M.S. degree from Huazhong University of Science and Technology. From April 2004 to June 2005, he was a Visiting Scholar with Tokyo University of Science, Japan. He is currently an Associate Professor with the College of Intelligent Manufacturing and Modern Industry, Xinjiang University. His research interests include brain–computer interfaces, intelligent robotics, and artificial intelligence (machine learning and deep learning).

WANYING TANG received the B.S. degree in industrial engineering from Shenyang Aerospace University. She is currently pursuing the M.S. degree. Her research interests include rehabilitation robotics and artificial intelligence (machine learning and deep learning).