Information Processing and Management 59 (2022) 103001

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

# TCACNet: Temporal and channel attention convolutional network for motor imagery classification of EEG-based BCI☆

Xiaolin Liu [a,1], Rongye Shi [b,1], Qianxin Hui [a], Susu Xu [c], Shuai Wang [a], Rui Na [a], Ying Sun [a], Wenbo Ding [d,e], Dezhi Zheng [a,*], Xinlei Chen [d,e,**]

[a] Beihang University, Xueyuan Road No. 37, Beijing 100191, China
[b] Beijing Institute of Technology, No. 5, South Street, Zhongguancun, Beijing 100081, China
[c] Stony Brook University, Nicolls Road No. 100, Stony Brook, NY 11794, USA
[d] Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China
[e] Peng Cheng Laboratory, Shenzhen 518055, China

ARTICLE INFO

*Keywords:*
Brain–computer interface
Electroencephalogram
Motor imagery classification
Deep learning
Attention mechanism

ABSTRACT

Brain–computer interface (BCI) is a promising intelligent healthcare technology to improve human living quality across the lifespan, which enables assistance of movement and communication, rehabilitation of exercise and nerves, monitoring sleep quality, fatigue and emotion. Most BCI systems are based on motor imagery electroencephalogram (MI-EEG) due to its advantages of sensory organs affection, operation at free will and etc. However, MI-EEG classification, a core problem in BCI systems, suffers from two critical challenges: the EEG signal's temporal non-stationarity and the nonuniform information distribution over different electrode channels. To address these two challenges, this paper proposes TCACNet, a temporal and channel attention convolutional network for MI-EEG classification. TCACNet leverages a novel attention mechanism module and a well-designed network architecture to process the EEG signals. The former enables the TCACNet to pay more attention to signals of task-related time slices and electrode channels, supporting the latter to make accurate classification decisions. We compare the proposed TCACNet with other state-of-the-art deep learning baselines on two open source EEG datasets. Experimental results show that TCACNet achieves 11.4% and 7.9% classification accuracy improvement on two datasets respectively. Additionally, TCACNet achieves the same accuracy as other baselines with about 50% less training data. In terms of classification accuracy and data efficiency, the superiority of the TCACNet over advanced baselines demonstrates its practical value for BCI systems.

## 1. Introduction

Over 200 million people in the world live with considerable disabilities in functioning, many of whom suffer from permanent and severe physical impairments (World Health Organization (WHO), 2011). People with severe physical impairments may experience

☆ An earlier version of this paper was presented at the conference UbiComp 2021, and was published in its proceedings (Liu et al., 2021). The conference version did not address the problem of nonuniform distribution of task-related information among multiple electrode channels. This manuscript significantly extends the earlier version by newly proposing the TCACNet model to address the issue via a novel channel attention mechanism. In addition, more thorough experiments and analysis are presented and enhanced results are achieved.

* Corresponding author at: Beihang University, Xueyuan Road No. 37, Beijing 100191, China.
** Corresponding author at: Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China.
*E-mail addresses:* liuxiaolin@buaa.edu.cn (X. Liu), rongyes@alumni.cmu.edu (R. Shi), zhengdezhi@buaa.edu.cn (D. Zheng), chen.xinlei@sz.tsinghua.edu.cn (X. Chen).

[1] X. Liu and R. Shi made equal contributions as co-first authors.

https://doi.org/10.1016/j.ipm.2022.103001
Received 31 January 2022; Received in revised form 17 June 2022; Accepted 18 June 2022
Available online 28 June 2022

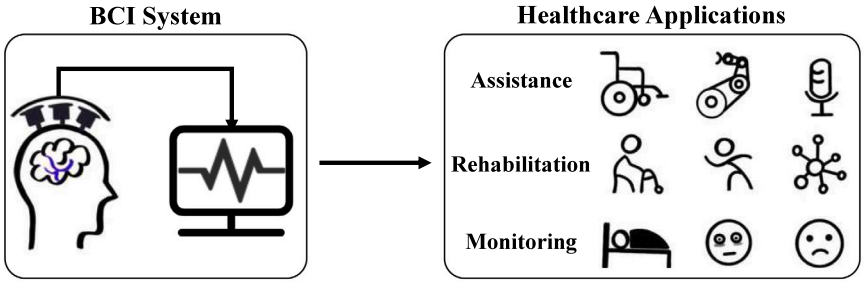


**Fig. 1.** Applications of BCI in healthcare scenarios. BCI can be applied to the healthcare of the disabled and patients, including the assistance of movement and communication, the rehabilitation of exercise and nerves, and the monitoring of sleep quality, fatigue and emotion.

difficulties in everyday life activity, such as walking, grabbing or maintaining body balance, and thus, long-term caring and assistance are in need. However, most of them do not receive the supports they need, because the cost of long-term and in-person care runs into the billions of dollars (Belkacem, Jamil, Palmer, Ouhbi, & Chen, 2020). Fortunately, new technologies, such as artificial intelligence, internet of things (IoT), mobile computing and wearable devices, have emerged to provide affordable assistive solutions to applications including smart health (Almathami, Win, & Vlahu-Gjorgievska, 2020; Kim & Xie, 2017; Orji & Moffatt, 2018; Tian et al., 2019), smart city (Chen et al., 2020, 2020; Shi, Mo and Di, 2021; Shi, Mo, Huang, Di and Du, 2021; Wang, Hu, Zheng, & Chen, 2021; Xu et al., 2019), smart living (Chen, Purohit, Dominguez, Carpin, & Zhang, 2015; Chen et al., 2017, 2020; Shi, Steenkiste, & Veloso, 2019; Shi, Steenkiste and Veloso, 2021) and etc. Among these intelligent technologies, brain–computer interface (BCI) is the one with great promise to considerably enhance the disabled people's quality of life (Belkacem et al., 2020).

As illustrated in Fig. 1, the BCI technology, which can monitor and translate brain signals, enables people with disabilities to control external assistive devices (e.g., wheelchairs, prosthetics, and robots) directly from brain activity, making it a useful healthcare tool to improve their personal autonomy and mobility across the lifespan (Mane, Chouhan, & Guan, 2020; Na et al., 2021; Swati, Kumar, & Namasudra, 2022). In BCI systems, electroencephalogram (EEG), a signal acquisition modality with the advantages of non-intrusiveness, high temporal resolution and relatively low cost, is commonly-used. As a typical example, motor imagery-EEG (MI-EEG) based BCI, a BCI paradigm that decodes human's mental imagination of an action, has been widely-investigated (Baig, Aslam, Shum, & Zhang, 2017; Cheng, Liu, & Zhang, 2018; Kumar, Sharma, Sharma, & Tsunoda, 2016; Oikonomou, Georgiadis, Liaros, Nikolopoulos, & Kompatsiaris, 2017). For MI-EEG based BCI systems, the decoding of users' thoughts of an imagined movement is a core technical component (Nijholt, 2016), which is a popular topic in literature. Classical machine learning approaches focus on manual feature extraction while data-driven deep learning approaches that can automatically extract features have recently been explored for MI-EEG classification. Due to the performance advantages, deep learning approaches, especially convolutional neural networks (CNNs) and their variants, have gained increasing interest in MI-EEG based BCI studies.

However, the improvements achieved by the deep learning approaches are still less than satisfactory. One challenge of developing high-performing MI-EEG classification algorithms is to tackle the EEG signal's non-stationarity issue, i.e., the signals of the same subject keep varying considerably and irregularly between trials or even within a single trial (Lotte et al., 2018). Most existing CNN models simply assume that the distribution of task-related information underlying the EEG signals is time-invariant, which could be problematic in the temporal non-stationary context (Szegedy et al., 2015). Another challenge for MI-EEG classification algorithms is to deal with the multi-channel signals, resulting from the fact that EEG signals are generally collected by a set of spatially distributed electrode channels. Therefore, the problem of nonuniform distribution of task-related information exists not only in the time dimension but also in the channel dimension, making the CNN models perform poorly in some multi-channel scenarios.

To address the above challenges, this paper proposes the temporal and channel attention convolutional network (TCACNet), a novel CNN-based classification framework for MI-EEG signals. On the one hand, to tackle the temporal non-stationarity issue of the signals, TCACNet is equipped with a temporal attention mechanism to adaptively adjust itself to focus on a small number of time slices that are task-related. Specifically, the TCACNet contains two sub-networks: global sub-network and local sub-network. The global sub-network is applied to the entire time horizon of the signal for a rough processing, while the local sub-network is applied to a small number of task-related time slices for a fine-grained processing. The task-related time slices are selected by the temporal attention mechanism based on the temporal salience level of each slice. On the other hand, to further resolve the problem of the nonuniform distribution of task-related information among electrode channels, we additionally introduce a novel channel attention mechanism in TCACNet. The channel attention mechanism employs the wavelet packet sub-band energy ratio (WPSER) to estimate the channel salience level of each electrode channel, based on which the weight coefficients of each channel in the network can be properly adjusted and learned. We verify the classification advantages of the proposed TCACNet over other state-of-the-art baselines on two open-source datasets collected from real experiments. Results show that TCACNet achieves 11.4% and 7.9% higher 4-class classification accuracy on the two datasets, respectively. Additionally, TCACNet presents advanced data efficiency, using 50% less training data to achieve the same accuracy as other baselines.

The main contributions of this paper are listed as follows:

- Propose a novel end-to-end CNN-based framework, called TCACNet, combining the global and local sub-networks to jointly process the global and local EEG information for better classification;
- Design a temporal attention mechanism to identify the time slices that are task-related, such that the network can focus on for an extra fine-grained processing, improving the ability of TCACNet to handle the EEG signal's temporal non-stationarity;
- Propose a channel attention mechanism based on WPSER, which adaptively adjusts the weight coefficients of each channel to make the network pay more attentions to task-related channels.

The rest of this paper is organized as follows: Section 2 defines the EEG classification problem; Section 3 introduces the TCACNet framework, as well as the temporal attention mechanism and channel attention mechanism in the proposed TCACNet; Section 4 presents and discusses the experimental results; Section 5 surveys the related work; and Section 6 concludes our work.

## 2. Research objectives

The goal of EEG classification considered in this paper is to correctly assign the inputted EEG of a trial performed by a participant to a category of motor imagery (MI) tasks of body moves. EEG signals are time series data collected by scalp surface electrodes. There are multiple subjects involved and each will perform several trials of mental tasks to generate EEG data. A trial corresponds to one of MI tasks, i.e., a class.

A given EEG dataset may contain the records from several subjects. For simplicity, we denote $D = \left\{\left(X^j, y^j\right) \mid j = 1, \ldots, N\right\}$ as the EEG data of $N$ trials for one subject. The input matrix $X^j \in \mathbb{R}^{C\times T}$ of trial $j$ contains the signals of $C$ electrode channels and $T$ discrete time points. The corresponding ground truth label of trial $j$ is denoted as $y^j$, which takes a value from a label set $Y$.

In general, $\left(X^j, y^j\right)$ can be viewed as a sample drawn i.i.d. from a joint probability distribution $P$ over $\mathbb{R}^{C\times T}$ and $Y$, associated to a subject. The ultimate goal of the classification algorithm is to find a decoder $f : \mathbb{R}^{C\times T} \to Y$ to minimize the risk of $f$, i.e.,

$$f^* = \arg\min_f R(f) = \arg\min_f \left[\mathop{\mathbb{E}}_{(X,y)\sim P}[L\left(f\left(X\right), y\right)]\right], \tag{1}$$

where $L(\cdot,\cdot)$ is a non-negative real-valued loss function, such as the cross-entropy loss. However, the risk cannot be computed because the distribution $P$ is unknown, and a proper strategy is to use the *empirical risk* as an approximation, i.e., to compute the expectation with respect to the dataset $D$. Under this setting, assuming that the decoder is parameterized by $\theta$, we would like to solve the empirical risk minimization problem:

$$\theta^* = \arg\min_\theta \frac{1}{N}\sum_{j=1}^{N} L\left(f\left(X^j;\theta\right), y^j\right). \tag{2}$$

The obtained decoder $f(\cdot;\theta^*)$ is a proper surrogate to $f^*$ because it approaches to $f^*$ as $N$ increases, and is evaluated by how well it can generalize to unseen trials. To improve the generalization performance of the decoder with limited data is a key challenge in classification, and there exit three strategies to steer the learning process on top of empirical risk minimization: to (1) train the decoder on a set of representative data points that are selected based on additional information about the subject characteristics, data distribution and other prior knowledge; (2) apply some inductive biases to design the internal structure of the decoder such that it can correctly predict the output that are not encountered; and (3) impose learning biases by adding regularization terms to the loss. As will be discussed in Section 3, this paper focuses on (2) and (3). Strategy (1) does not apply because no additional information about data and labels is available, and training data points are selected randomly.

A special notice is that, for further processing, we split a data point $X^j$ into $S$ equal time slices and each with $T'$ consecutive time points, such that $X^j = \left[X^j_k \in \mathbb{R}^{C\times T'} | k = 1, \ldots, S\right]$, where $T' \times S = T$. For the rest of the paper, without special notice, we make use of $X$ to denote $X^j$ by default for simplification, meaning that it is the signal of one trial performed by a subject. Accordingly, $X_k$ denotes the signal of the $k$th time slice.

To contribute to field of EEG classification for BCI systems, this paper focuses on addressing the following research problem: *how can we overcome the EEG signal's temporal non-stationarity and nonuniform information distribution over electrode channels to assign the MI-EEG input of a trial to its corresponding MI task category as accurately as possible?* As will be presented in the next section, we introduce the temporal and channel attention mechanisms to tackle the challenges, and design TCACNet, a novel CNN-based classification framework equipped with the two attention mechanisms, to fulfill the research goal.

## 3. Temporal and channel attention convolutional network

### 3.1. TCACNet framework overview

Fig. 2 shows our TCACNet framework, which is composed of an attention mechanism module to preprocess the raw EEG data and a follow-up network architecture to perform the classification. In the attention mechanism module, there are two types of attention mechanisms. As shown in yellow and green in Fig. 2, the two attention mechanisms are the *temporal attention mechanism* and *channel attention mechanism*, respectively. The temporal attention mechanism is a signal preprocessing method to identify the task-related time slices $X_{selected}$ from the raw EEG $X$. In this way, the follow-up network architecture is able to focus on the task-related information in $X_{selected}$ and reduce the interference from the task-unrelated slices, improving the ability of TCACNet to handle the temporal non-stationarity issue of EEG signals. Besides the temporal dimension, the channel dimension of multi-channel

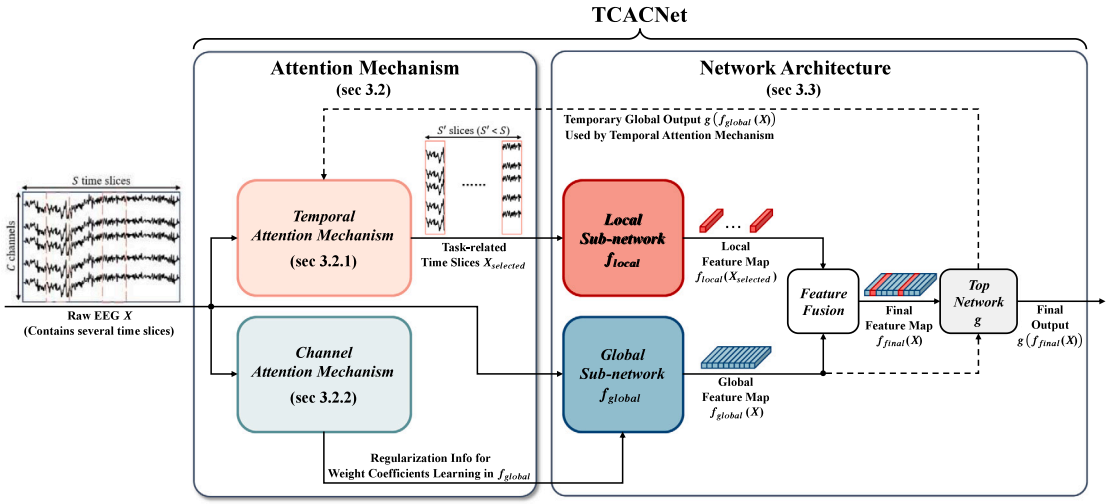


**Fig. 2.** TCACNet framework. The temporal attention mechanism is to identify the task-related time slices from the raw EEG $X$ such that the local sub-network $f_{local}$ can concentrate on, while the channel attention mechanism is to adjust the weight coefficients in the global sub-network $f_{global}$. The outputs of $f_{local}$ and $f_{global}$ are further fused and fed forward to the top network $g$ to obtain the classification result. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

EEG signals also contains varying information, leading to the spatial non-stationarity issue. To tackle this, the channel attention mechanism is employed to adaptively adjust the attention paid to different electrode channels. Specifically, the channel attention mechanism is a regularization method to adjust the weight coefficients in the network architecture, which are channel-related. We will discuss the details of attention mechanism module in Section 3.2.

The network architecture is the second main module in TCACNet. Both the whole-time data $X$ and selected time slices $X_{selected}$ serve as the input of the network architecture, and are processed by the *global sub-network* $f_{global}$ and *local sub-network* $f_{local}$, respectively, as shown in blue and red in Fig. 2. The global sub-network makes use of the regularization information from the channel attention mechanism for the learning of its weight coefficients, and is applied to the whole $X$ for a rough processing. In contrast, the local sub-network is applied only to a small number of time slices $X_{selected}$, that are task-related, for a fine-grained processing. The two sub-networks process and transform their corresponding inputs into a global feature map $f_{global}(X)$ and a local feature map $f_{local}(X_{selected})$, respectively. The two feature maps are further fused using a *feature fusion* operation to obtain the final feature map $f_{final}(X)$, which is expected to be an improved high-level abstract of the EEG signals compared to that from either $f_{global}$ or $f_{local}$ only. The final feature map is then fed forward to a fully-connected neural network, called *top network* $g$, to obtain the final output $g(f_{final}(X))$. The $g(f_{final}(X))$ is the classification distribution over class candidates, and based on which the classification decision can be made. We will discuss the details of network architecture in Section 3.3.

A special notice is that the temporal attention mechanism is coupled with the network architecture in terms of leveraging the temporary global output information $g(f_{global}(X))$ from the top network $g$. The corresponding coupling information flow is illustrated with dashed arrows in Fig. 2 and will be detailed in the next section.

### 3.2. Attention mechanism

The attention mechanism originated from the study of human vision in cognitive science, that is, humans will selectively focus on a part of all information while ignoring others due to the limitation of information processing ability. The attention mechanism has two main aspects: deciding which part of the input needs to be focused on, and allocating limited information processing resources to the important parts. Imitation of this information processing mechanism of humans, the attention mechanism has also become an important part of the neural network structure, and has achieved remarkable results in the analysis of images and natural language. Similarly, attention mechanisms can also be used in EEG signal analysis to assist models to focus on task-relevant information in the temporal or spatial (channel) dimension. Based on these motivations, we propose the following temporal attention mechanism and channel attention mechanism.

#### 3.2.1. Temporal attention mechanism

The goal of the temporal attention mechanism is to select a small number of task-related time slices (i.e., $X_{selected}$), such that the $f_{local}$ in the network architecture can concentrate on for a fine-grained processing. Specifically, this mechanism identifies and selects the salient vectors in the global feature map $f_{global}(X)$, which have the largest gradients of the loss objective. Then, the time slices associated with these salient vectors are selected and defined as task-related. The rest of this subsection details the content of the temporal attention mechanism, which is also illustrated in Fig. 3.

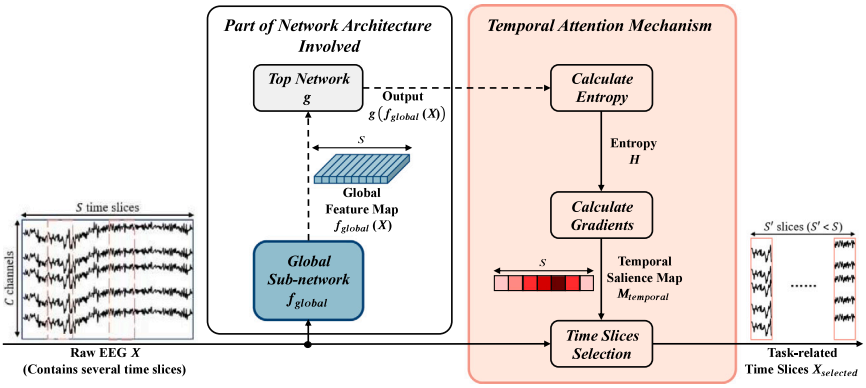


**Fig. 3.** Illustration of the temporal attention mechanism. The coupling information flow feedbacked from the network architecture is illustrated with dashed arrows.

Given the inputted raw signal $X$, the global sub-network $f_{global}$ (in the network architecture) is applied to the whole time horizon for a rough processing to compute the global feature map (consisting of $S$ vectors):

$$f_{global}(X) = \left[c_k \mid k = 1, \ldots, S\right], \tag{3}$$

where $c_k = f_{global}(X_k) \in \mathbb{R}^B$ is a global feature vector corresponding to $X_k$, and $S$ is the temporal dimension size of the global feature map. As can be seen from Eq. (3), the number of vectors is the same to the number of time slices in $X$, which establishes an one-to-one correspondence relationship between feature vectors in $f_{global}(X)$ and time slices in $X$. Then, the global feature map is directly fed forward to the top network $g$ to obtain the temporary global output $g(f_{global}(X))$, i.e., the preliminary classification distribution based only on the global sub-network $f_{global}$.

Next, in the temporal attention mechanism, a temporal salience map $M_{temporal}$ is calculated based on the $g(f_{global}(X))$, which echoes the coupling characteristic in Fig. 2, i.e., the information flow feedbacked from the network architecture to the temporal attention mechanism. Specifically, we choose the entropy of $g(f_{global}(X))$ as the preliminary loss objective:

$$H = -\sum_{l=1}^{L} o^l_{global} \log o^l_{global}, \tag{4}$$

where $o_{global} = g(f_{global}(X)) \in \mathbb{R}^L$ is the temporary global output (i.e., the distribution over class candidates). There are $L$ class candidates. We then make use of the entropy $H$'s gradient with respect to $c_k$ as the measure of the temporal salience level of the time slice $X_k$. The temporal salience level of time slice $X_k$ is given by:

$$\begin{aligned} M^k_{temporal} &= \left\| \nabla_{c_k} H \right\|_2 \\ &= \sqrt{\sum_{b=1}^{B} \left( -\frac{\partial}{\partial c^b_k} \sum_{l=1}^{L} o^l_{global} \log o^l_{global} \right)^2}, \end{aligned} \tag{5}$$

where $c^b_k \in \mathbb{R}$ is the $b$th element in the $k$th global feature vector $c_k$. The temporal salience map is then defined as $M_{temporal} = \left[M^1_{temporal}, \ldots, M^S_{temporal}\right]$.

Using the temporal salience map $M_{temporal}$, we select $S'$ time slices with the largest saliency as the task-related time slices in the operation of *time slices selection*. The list of indexes of the selected time slices is denoted as $I_{selected}$, such that $\left|I_{selected}\right| = S' < S$. The selected time slices is then denoted as $X_{selected} = \left[X_k \mid k \in I_{selected}\right]$, which is the output of the temporal attention mechanism.

Please note that it is not required to know the ground truth label for calculating the entropy $H$ of $g(f_{global}(X))$, and thus the $M_{temporal}$. Therefore, the formation of $X_{selected}$ and the corresponding fine-grained processing by $f_{local}$ do not require ground truth label information in the prediction phase.

### 3.2.2. Channel attention mechanism

The goal of the channel attention mechanism is to tackle the issue of nonuniform information distribution among electrode channels by adaptively adjusting the attention paid to different electrode channels. Considering that the signals from different electrode channels contain different levels of task-related information that affect the classification results, the channel attention mechanism explores the salience levels of different channels. Based on the channel salience information, the channel attention mechanism is able to impose a well-designed regularization term on the training loss to regularize the learning of the weight coefficients directly associated to each channel, such that the signals from different channels can be properly weighted, guiding the TCACNet to focus more on the task-related channels for improved classification accuracy. The rest of this subsection details the content of the channel attention mechanism, which is also illustrated in Fig. 4.

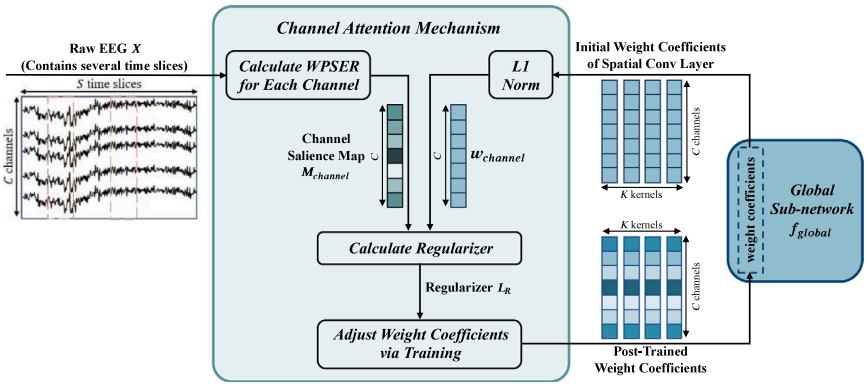


**Fig. 4.** Illustration of the channel attention mechanism. The $w_{channel}$ is learned and adjusted with the training of network parameters, while $M_{channel}$ is calculated before the training phase and then remains fixed.

From the perspective of the frequency domain, the task-related information contained in MI-EEG mainly concentrates in the range of 8–30 Hz, i.e., the range of $\mu$ rhythm and $\beta$ rhythm (Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002; Wolpaw, McFarland, Vaughan, & Schalk, 2003). The wavelet packet transform is a useful tool for time–frequency analysis in signal processing fields (Cody, 1994). We use the wavelet packet sub-band energy ratio (WPSER) to quantify the ratio of the energy within 8–30 Hz to the signal's total energy, and the WPSER is treated as the measure of the salience level of task-related information for the signal of each electrode channel.

The calculation procedure of WPSER approximates the raw signal of a channel to have $2^e$ signal points. Under this setting, the signal is decomposed by a $u$-layer ($u < e$) wavelet packet to obtain $2^u$ sub-node signals and each sub-node signal has the length of $2^{e-u}$. The sub-node signals are denoted as $[y^{u,v,n} \mid v = 0, \ldots, 2^u - 1; n = 0, \ldots, 2^{e-u} - 1]$, where $v$ is the index of sub-nodes, and $n$ is the index of points in sub-node $v$. For the signal to be decomposed, given a set of low-pass and high-pass conjugate orthogonal filter coefficients $\{p\}$ and $\{q\}$, the recurrence relationship between the sub-nodes of $u$-layer and those of $(u-1)$-layer is:

$$y^{u,v,n} = \begin{cases} \sum_{r=0}^{2^{e-u+1}-1} p^{r-2n} y^{u-1,\frac{v}{2},r}, & v = 0, 2, 4, \ldots, \quad (low-pass) \\ \sum_{r=0}^{2^{e-u+1}-1} q^{r-2n} y^{u-1,\frac{v-1}{2},r}, & v = 1, 3, 5, \ldots, \quad (high-pass) \end{cases} \tag{6}$$

Next, the energy proportion $P^{u,v}$ of each sub-node $v$ to the channel's total energy is obtained by:

$$P^{u,v} = \frac{E^{u,v}}{E^u} = \frac{\sum_n (y^{u,v,n})^2}{\sum_{v=0}^{2^u-1} E^{u,v}}, \tag{7}$$

where $E^{u,v} = \sum_n (y^{u,v,n})^2$ is the signal energy of each sub-node $v$ and $E^u = \sum_{v=0}^{2^u-1} E^{u,v}$ is the total energy of all the sub-nodes processed by the $u$-layer wavelet packet. Finally, we will have a set of sub-nodes belonging to the 8–30 Hz region, denoted as $V_{8--30\ \mathrm{Hz}}$, and the corresponding $P^{u,v}$ are summed up to obtain WPSER. We use the WPSER corresponding to 8–30 Hz as the measure of the salience level of task-related information in the signal. The salience level of a channel (e.g., channel $c$) is then defined as:

$$M^c_{channel} = \sum_{v \in V_{8--30\ \mathrm{Hz}}} P^{u,v}, \tag{8}$$

which has a positive value. We then define the channel salience map as $M_{channel} = \left[M^1_{channel}, \ldots, M^C_{channel}\right]$ for all the $C$ electrode channels, which is used to adjust the weight coefficients of in the global sub-network $f_{global}$. The implementation principle and processing process corresponding to Eqs. (6)–(8) are presented in Appendix A.

A special notice is that, $M_{channel}$ only adjusts certain part of the weight coefficients in $f_{global}$, i.e., the weights of the spatial convolutional layer in $f_{global}$. The spatial convolutional layer is the very first layer in $f_{global}$, and thus, its weight coefficients are directly channel-related. We skip the details of $f_{global}$ other than the spatial convolutional layer for now, which will be given later in Section 3.3. The spatial convolutional layer contains $K$ filters and each filter consists of $C$ weight coefficients corresponding to $C$ channels, which form a $C \times K$ weight coefficient matrix $W$:

$$W = \begin{bmatrix} w^{1,1} & \cdots & w^{1,K} \\ \vdots & \ddots & \vdots \\ w^{C,1} & \cdots & w^{C,K} \end{bmatrix}. \tag{9}$$

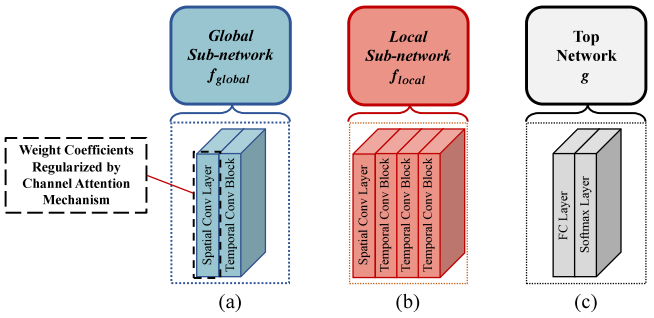


**Fig. 5.** Internal structures of global sub-network $f_{global}$, local sub-network $f_{local}$ and top network $g$.

As shown in Fig. 4, we compute the $L1$ norm of the weight coefficient matrix $W$ over the kernel dimension to obtain $w_{channel}$:

$$w_{channel} = \begin{bmatrix} w^1_{channel} \\ \vdots \\ w^C_{channel} \end{bmatrix} = \begin{bmatrix} \left|w^{1,1}\right| + \cdots + \left|w^{1,K}\right| \\ \vdots \\ \left|w^{C,1}\right| + \cdots + \left|w^{C,K}\right| \end{bmatrix}. \tag{10}$$

To regularize the training of weight coefficients based on $M_{channel}$, we design a regularizer $L_R$ to add to the loss function:

$$L_R = \frac{1}{Q}\begin{bmatrix} A^{\frac{1}{M^1_{channel}}} & \cdots & A^{\frac{1}{M^C_{channel}}} \end{bmatrix}\begin{bmatrix} w^1_{channel} \\ \vdots \\ w^C_{channel} \end{bmatrix} = \frac{1}{Q}\sum_{c=1}^{C}\left(A^{\frac{1}{M^c_{channel}}} \cdot w^c_{channel}\right), \tag{11}$$

where $A$ is a hyperparameter, which is a positive integer used as a base value, and $Q = \sum_{a=1}^{C} A^{\frac{1}{M^a_{channel}}}$ is for normalization. Using the regularizer $L_R$, the learning of each weight coefficient in the spatial convolutional layer is properly regularized in the training phase, and the regularization effect is guided by the channel salience map $M_{channel}$, which is calculated before the training phase and then remains fixed. Empirically, $A = 2$ is a proper choice for our scenarios.

By minimizing $L_R$, the channel with a larger $M^c_{channel}$ is expected to end up with a larger $w^c_{channel}$, while the channel with a smaller $M^c_{channel}$ is expected to end up with a smaller $w^c_{channel}$. In addition, we can further widen the gap among different channels through an exponential form. In this way, the channel containing more task-related information will enjoy a much larger weight coefficient in the spatial convolutional layer of $f_{global}$.

### 3.3. Network architecture

This section discusses the details of network architecture of TCACNet. In the network architecture, the global sub-network $f_{global}$ and the local sub-network $f_{local}$ are different in terms of internal structures and the scope of the inputted signal $X$ that the sub-networks will engage in and process.

As shown in Fig. 5(a), the global sub-network $f_{global}$ contains a spatial convolutional layer to learn the spatial patterns among electrode channels and a temporal convolutional block (a temporal convolutional layer followed by a mean pooling layer) to learn the temporal patterns. In the spatial convolutional layer, each filter covers all the electrode channels and processes the multi-channel information as a whole. Furthermore, the spatial convolutional layer is regularized by the channel attention mechanism as has been presented previously in Section 3.2.2. In the temporal convolutional layer, each filter performs a convolution over time and extracts temporal features. As shown in Fig. 5(b), the local sub-network $f_{local}$ has a similar structure but with several extra temporal convolutional blocks to process the temporal information in a fine-grained manner. Details of the structures of $f_{global}$ and $f_{local}$ can be found in Appendix B.

In the network architecture, the global sub-network $f_{global}$ processes the whole input $X$ and the global feature map $f_{global}(X)$ is obtained. The definition of the $f_{global}(X)$ has been given in Eq. (3). In contrast, the local sub-network $f_{local}$ only processes $X_{selected}$ and the local feature map is obtained:

$$f_{local}\left(X_{selected}\right) = \left[f_{local}\left(X_k\right) \mid k \in I_{selected}\right], \tag{12}$$

where $f_{local}(X_k) \in \mathbb{R}^B$ is a local feature vector. A local feature vector has the same dimension size as the global one.

To jointly process the global and local information contained in EEG signals, the global feature map $f_{global}(X)$ and the local feature map $f_{local}(X_{selected})$ will go through an operation of feature fusion to obtain the final feature map $f_{final}(X)$, denoted as $f_{final}(X)$:

$$f_{final}\left(X\right) = \left[f_{final}\left(X_k\right) \mid k = 1, \ldots, S\right], \tag{13}$$

where $f_{final}\left(X_k\right) = \begin{cases} f_{local}\left(X_k\right), k \in I_{selected} \\ f_{global}\left(X_k\right), k \notin I_{selected} \end{cases}$, which defines the operation of feature fusion in Fig. 2.

**Table 1**
Details of the EEG datasets for experiments.

| Dataset | Subject number | Electrode number | Trial number | Task time | Task category |
|---|---|---|---|---|---|
| BCIC IV 2a | 9 | 22 | 288 × 2 per subject | 4 s | 1: left hand 2: right hand 3: feet 4: tongue |
| HGD | 13 | 44 | About 1000 per subject | 4 s | 1: left hand 2: right hand 3: feet 4: rest |

To obtain the final output of TCACNet (i.e. the probability distribution over class candidates), the final feature map $f_{final}(X)$ will be fed forward to the top network $g$. As shown in Fig. 5(c), $g$ has a fully connected (FC) layer and a softmax output layer. TCACNet makes use of the final output $g(f_{final}(X))$ for decision-making. Details of the structure of the top network $g$ can be found in Appendix B.

### 3.4. Loss design and training

Based on the TCACNet framework, we design a training objective and then apply back-propagation algorithm to train the $f_{global}$, $f_{local}$ and $g$, jointly. Using $\theta$ to denote the learning parameters of $f_{global}$, $f_{local}$ and $g$ as a whole, we train the TCACNet by minimizing the negative log-likelihood on the ground truth labels:

$$L_J = -\sum_{j=1}^{m} \log p\left(y^j \mid X^j; \theta\right), \tag{14}$$

where $m$ is the number of trials in the training set and $p(y^j \mid X^j; \theta) = g(f_{final}(X^j))$ is the probability of label $y^j$ being assigned to the input $X^j$. Using gradient descent on the training objective, the parameters $\theta$ are updated and learned.

We additionally consider a hint-based loss term introduced by Romero et al. (2015) to reduce the disagreement between the global and local features. Specifically, the loss term is given by:

$$L_H = \sum_{X_k \in X_{selected}} \left\| f_{global}\left(X_k\right) - f_{local}\left(X_k\right) \right\|_2^2, \tag{15}$$

where $f_{global}\left(X_k\right)$ and $f_{local}\left(X_k\right)$ represent the global and local feature vectors, respectively.

As described in Section 3.2.2, we also introduce a regularizer $L_R$ to adjust the weight coefficients corresponding to each channel, which is the benefit brought by the proposed channel attention mechanism. Therefore, the complete form of the loss function is:

$$L_{complete} = L_J + \alpha L_H + \beta L_R, \tag{16}$$

where $\alpha$ and $\beta$ are hyperparameters used to balance the influence of the corresponding parts on the loss function. By training the TCACNet through $L_{complete}$, both the temporal and channel attention mechanisms take effects jointly.

## 4. Experiments and evaluation

### 4.1. Description of datasets and preprocessing

**Description of datasets:** Two popular public EEG datasets are used to test the TCACNet's classification performance: BCI competition IV dataset 2a (BCIC IV 2a) (Brunner, Leeb, Müller-Putz, Schlögl, & Pfurtscheller, 2008) and High Gamma Dataset (HGD) (Schirrmeister et al., 2017). The description of each dataset is as follows:

- BCIC IV 2a dataset: this EEG dataset is collected from 9 subjects, using an EEG acquisition device with 22 electrodes. It includes 288 × 2 trials of 4-second mental tasks for each subject. In each mental task, the subject was asked to image a movement of either the left hand, right hand, feet or tongue.
- HGD dataset: this EEG dataset is collected from 14 subjects, using an advanced EEG acquisition device with 128 electrodes. In each trial, the subject is also asked to perform a 4-second imagination of one of the 4 body movements, i.e., either the left hand, right hand, feet or rest. About 1000 trials are collected for each subject. For this dataset, we only use the 44 electrodes on the motor cortex, i.e., a brain area that controls voluntary movement. In addition, the data from subject 14 are excluded, because about half of the electrodes of this subject suffer from serious signal loss (see Table 1).

**Preprocessing:** We preprocess the raw EEG dataset to standardize and make it cleaner for classification. Specifically, the EEG signals of the HGD data are down-sampled to 250 Hz, which is the same to the BCIC IV 2a data, and in this way, we can use the same network structure in TCACNet to process both datasets. Moreover, an exponential moving operation is applied to standardize the data, which is introduced by Schirrmeister et al. (2017). We use the full-bandwidth data directly to maintain the data's fidelity as much as possible.

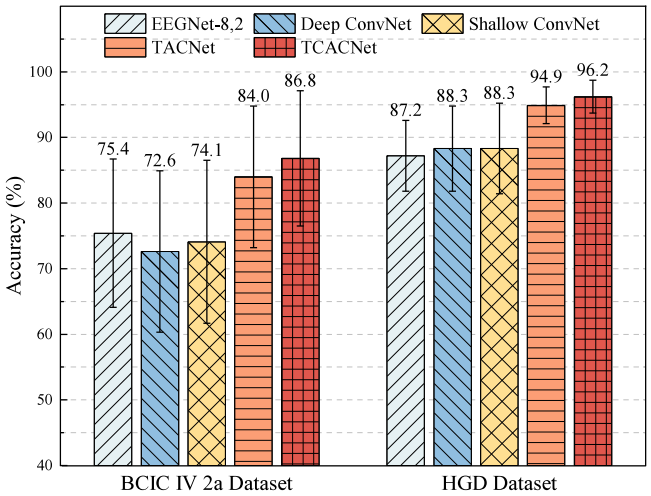


**Fig. 6.** Average subject-specific classification accuracy of TACNet, TCACNet and baseline models (mean±std).

### 4.2. Experimental settings

#### 4.2.1. Baselines

There are five baseline CNN-based models considered in the experiments to compare with the proposed TCACNet: *Shallow ConvNet* (Schirrmeister et al., 2017) and *Deep ConvNet* (Schirrmeister et al., 2017) are two state-of-the-art CNN models for processing MI-EEG, while *EEGNet-8,2* (Lawhern et al., 2018) is a CNN variant with a compact structure. We will discuss the details of Shallow ConvNet, Deep ConvNet and EEGNet-8,2 in Section 5. *Global model* is the basic component of TCACNet, in which the top network $g$ is applied directly to the global feature map for decision-making. Please note that this model only employs $L_J$ in the loss, and neither temporal nor channel attention mechanism is added. *TACNet* (Liu et al., 2021) is a downgraded version of TCACNet with no channel attention mechanism. Please note that this model employs $L_J$ and $L_H$ in the loss, and only the temporal attention mechanism takes effects. Also, this model uses the final feature map to obtain $L_J$, which is different from the global model.

#### 4.2.2. Experiments

A subject-specific analysis is conducted on both the BCIC IV 2a dataset and HGD dataset. Specifically, for each subject, we randomly divide the collected data into a training set and a test set with an approximate ratio of 8:2, resulting in 461 and 880 training trials as well as 115 and 160 test trails for the BCIC IV 2a dataset and the HGD dataset, respectively. All the baselines are built up on the optimal configurations as suggested by the corresponding literature. In the training phase, we make use of 20% of the training set as the validation set to prevent overfitting. We apply the Adam optimizer to train the classifiers. For more details of the experimental setup, we refer readers to our previous work in Liu et al. (2021).

### 4.3. Results

Fig. 6 shows the average subject-specific classification accuracy of TCACNet and baseline models on the two datasets (detailed subject-specific experimental results can be found in ). It can be observed that TACNet and TCACNet achieve the best average classification accuracy on both datasets with significant improvements over other baselines. TCACNet achieves 11.4% and 7.9% higher 4-class classification accuracy than the best baseline model on the two datasets, respectively (i.e. BCIC IV 2a: TCACNet-86.8% vs. EEGNet-75.4%, HGD: TCACNet-96.2% vs. Deep/Shallow ConvNet-88.3%). Furthermore, TACNet and TCACNet show less variance across subjects (i.e. smaller standard deviation, BCIC IV 2a: EEGNet-5.7% vs. Deep ConvNet-6.2% vs. Shallow ConvNet-6.2% vs. TACNet-5.4% vs. TCACNet-5.1%, HGD: EEGNet-2.7% vs. Deep ConvNet-3.3% vs. Shallow ConvNet-3.5% vs. TACNet-1.4% vs. TCACNet-1.2%), implying an improved performance in terms of reliability and stability. Fig. 6 exhibits an interesting phenomenon that, for all the models, the classification accuracy on the HGD dataset is generally greater than that on the BCIC IV 2a dataset. In addition, the corresponding standard deviation is smaller. This is because, compared with the MI signal, the signal-to-noise-ratio of the motor execution EEG is generally higher. To further verify the performance improvement of our model on multiple subjects from a statistical perspective, paired Wilcoxon signed-rank test is conducted on the experimental results to analyze the statistical significance of differences in classification accuracy achieved by different models. The results in Table 2 show that the improvement in accuracy achieved by TACNet and TCACNet over baseline models is statistically significant for both datasets. The statistical significance obtained on the HGD dataset is generally higher, which is also due to the quality difference between the two datasets. The performance improvement of TACNet results from the fact that the temporal attention mechanism makes the network pay more attention to specific time slices that contain more task-related information. In addition, the TCACNet takes advantages of the

**Table 2**
Performance comparison among different models.

| Dataset | Model | Accuracy | Significance of performance improvement | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | (mean ± std) | TACNet (without $L_H$) | | TACNet (with $L_H$) | | TCACNet | |
| BCIC IV 2a | EEGNet-8,2 | 75.4 ± 11.3 | | (p = 0.0547) | * | (p = 0.0117) | ** | (p = 0.0039) |
| | Deep ConvNet | 72.6 ± 12.3 | * | (p = 0.0117) | * | (p = 0.0117) | ** | (p = 0.0039) |
| | Shallow ConvNet | 74.1 ± 12.4 | ** | (p = 0.0039) | ** | (p = 0.0039) | ** | (p = 0.0039) |
| | Global Model | 79.5 ± 11.9 | ** | (p = 0.0039) | ** | (p = 0.0078) | ** | (p = 0.0039) |
| | TACNet (without $L_H$) | 83.3 ± 11.2 | – | – | | (p = 0.3105) | ** | (p = 0.0039) |
| | TACNet (with $L_H$) | 84.0 ± 10.8 | | (p = 0.3105) | – | – | ** | (p = 0.0039) |
| | TCACNet | 86.8 ± 10.3 | ** | (p = 0.0039) | ** | (p = 0.0039) | – | – |
| HGD | EEGNet-8,2 | 87.2 ± 5.4 | *** | (p = 0.0002) | *** | (p = 0.0002) | *** | (p = 0.0002) |
| | Deep ConvNet | 88.3 ± 6.5 | ** | (p = 0.0012) | *** | (p = 0.0002) | *** | (p = 0.0002) |
| | Shallow ConvNet | 88.3 ± 6.9 | *** | (p = 0.0007) | *** | (p = 0.0002) | *** | (p = 0.0002) |
| | Global Model | 89.1 ± 5.8 | *** | (p = 0.0002) | *** | (p = 0.0002) | *** | (p = 0.0002) |
| | TACNet (without $L_H$) | 91.9 ± 5.9 | – | – | ** | (p = 0.0081) | *** | (p = 0.0002) |
| | TACNet (with $L_H$) | 94.9 ± 2.8 | ** | (p = 0.0081) | – | – | * | (p = 0.0133) |
| | TCACNet | 96.2 ± 2.5 | *** | (p = 0.0002) | * | (p = 0.0133) | – | – |

The result of comparing the model with itself is meaningless, and - is used to represent the vacancy in the corresponding position.
*Indicate how much the classification performance of TACNet or TCACNet is significantly better than the given baseline models $p < 0.05$.
**Indicate how much the classification performance of TACNet or TCACNet is significantly better than the given baseline models $p < 0.01$.
***Indicate how much the classification performance of TACNet or TCACNet is significantly better than the given baseline models $p < 0.001$.

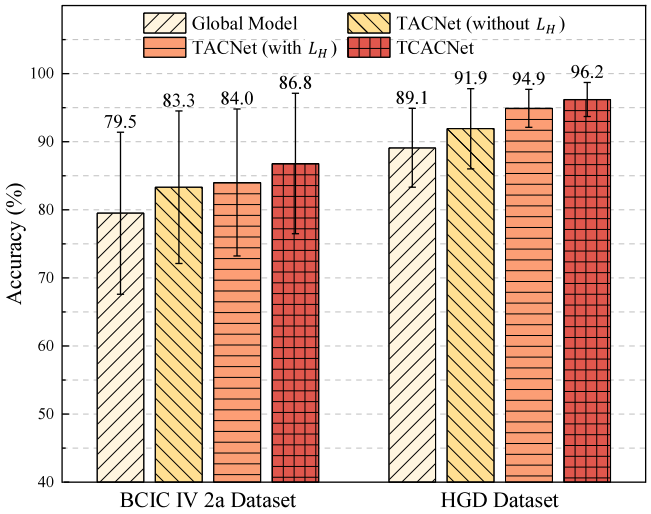


**Fig. 7.** Average subject-specific classification accuracy in ablation experiments (mean±std).

channel attention mechanism to further enhance the performance. In contrast, other baseline models treat each components of the input equally important and lack any attention mechanism to identify and focus on the task-related ones, limiting their classification ability.

To verify the effectiveness of each core component in TCACNet, ablation experiments are carried out. The experimental results are shown in Fig. 7 and more details can be found in . At first, as a part of TCACNet, global model is the most basic model. TACNet applies the proposed temporal attention mechanism to the global model, and introduces an additional term $L_H$ in the loss function for better model training. From Fig. 7, it can be observed that TACNet outperforms the global model, and this is because that the temporal attention mechanism in TACNet can help with the concentration on processing the task-related slices and reduce the interference from the task-unrelated slices. Furthermore, TACNet with $L_H$ presents a further improvement, demonstrating the advantages brought by the $L_H$, which reduces the disagreement between the global and local features. By comparing the accuracy achieved by TACNet and TCACNet, the latter stands out. This verifies the effectiveness of the proposed channel attention mechanism, in which WPSER is used as the measure of the channel salience level of task-related information to adjust the channel weight via the regularization term $L_R$ during the training phase. The pairwise comparison results in Table 2 show that the performance improvement brought by each core component in our model is overall statistically significant (i.e., TACNet without $L_H$ vs. global model on both datasets, TACNet with $L_H$ vs. TACNet without $L_H$ on HGD dataset and TCACNet vs. TACNet with $L_H$ on both datasets).

The temporal attention mechanism has a hyperparameter to optimize, i.e., the number of selected time slices $S'$. How this parameter affects the classification accuracy is studied and the results are presented in Fig. 8. For both datasets, the proposed

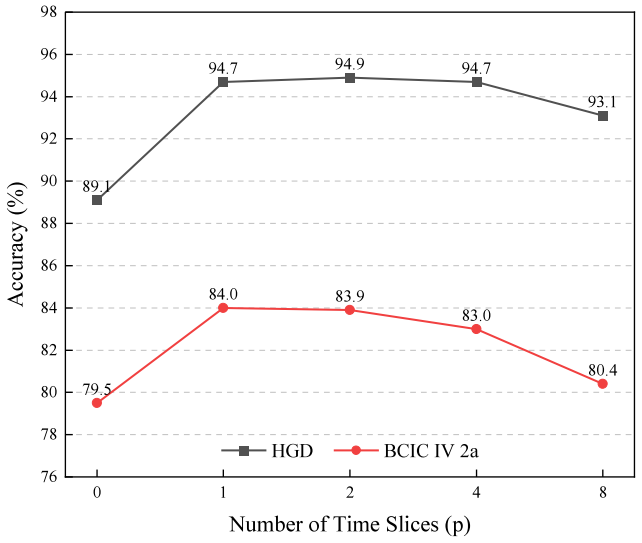


**Fig. 8.** Classification accuracy over selected time slice numbers.

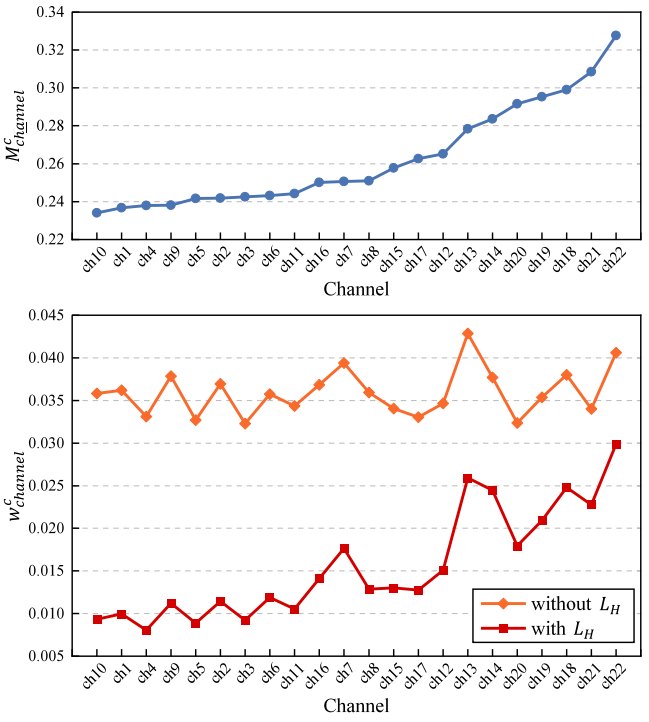


**Fig. 9.** Element-wise verification of the influence of channel attention mechanism on weight coefficients $w_{channel}$ (taking the parameters of one subject as an example). The "ch1" in $x$-axis denotes the 1-st channel, and correspondingly, the element value of $M^1_{channel}$ and $w^1_{channel}$ are plotted. Same for other channels.

TCACNet achieves the best accuracy at $S' = 1$ and $S' = 2$. Lower performance is observed when further increasing the $S'$. This property is beneficial because we can use just a small number of time slices to achieve optimal performance, implying a higher computational efficiency.

Fig. 9 shows the influence of channel attention mechanism on the learning of weight coefficients of the spatial convolutional layer in $f_{global}$. The proposed channel attention mechanism is designed to adjust $w_{channel}$ of the spatial convolutional layer through $M_{channel}$. The electrode channels in Fig. 9 are arranged in the increasing order with respect to $M_{channel}$, shown in the upper part of the figure. The lower part of the figure shows the post-trained $w_{channel}$ of each channel with or without the regularizer $L_R$. It can be observed that there is no correlation between the values of $M_{channel}$ and $w_{channel}$ of each channel in the model without

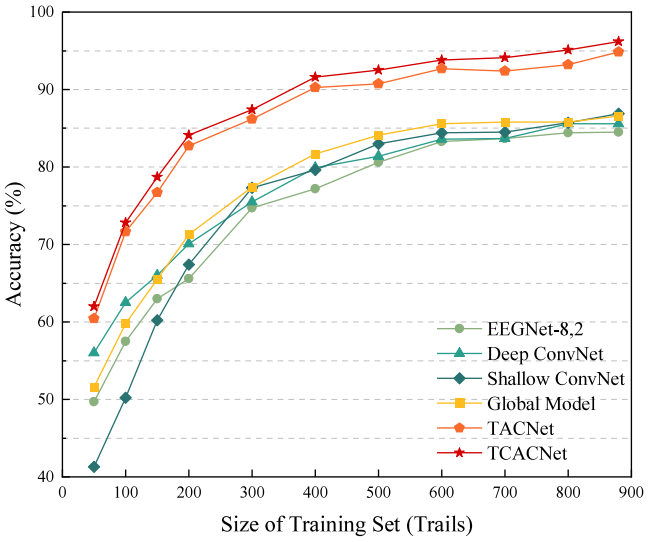


**Fig. 10.** Classification accuracy of models on training sets of different sizes.

$L_R$. In contrast, the two values are positively correlated when $L_R$ is imposed. Specifically, the overall trend of the post-trained $w_{channel}$ of each channel in the model with $L_R$ matches well with the overall trend of $M_{channel}$. One can also observe that the local characteristics of the curve of the trained $w_{channel}$ from TCACNet stays similar to those trained without the $L_R$. This verifies that the proposed channel attention mechanism does not negatively affect the local feature learning when enforcing the overall increasing trend of the weight coefficients. Another phenomenon observed in Fig. 9 is that the weight coefficient of each channel, e.g., $w^c_{channel}$ for the $c$th channel, decreases after channel attention mechanism. This is due to the fact that, according to Eqs. (10), (11), $L_R$ is an element-wise linear combination of $w_{channel}$, of which each element $w^c_{channel}$ is the $L1$ norm of a kernel, i.e., the row vector $\left[w^{c,1},\ldots,w^{c,K}\right]$ in $W$. As a result, by minimizing $L_R$, $w^c_{channel}$ will decrease accordingly after training the model.

To verify the data efficiency of models, classification accuracy of the proposed models and baselines on training sets of different sizes are evaluated using the HGD dataset. Each model is trained using various sizes of the training dataset ranging from 50 to approximately 880 trials, with 20% of which left out as the validation set. The post-trained models are then tested on the same test set and get the corresponding accuracy. The experimental results are presented in Fig. 10. From the perspective of classification accuracy, the TCACNet maintain the highest over all performance. Even when the training data is smaller than 100 trails, the performance of TCACNet is still better than any other baseline model. From another perspective of data efficiency, TCACNet achieves the same classification accuracy as other baseline models with about 50% less training data. For example, the classification accuracy of TCACNet on 200 trails (i.e., 84.1%) is slightly higher than that of any other baseline models on 400 trails (i.e., 83.7%, 81.9%, 81.6%, 79.2% for Global Model, Shallow ConvNet, Deep ConvNet, EEGNet, respectively). These results support that TCACNet can be applied to small-data scenarios, demonstrating its practical value for the BCI system development. Fig. 10 can also be interpreted in terms of generalizability: the proposed TCACNet achieves the best generalization performance, i.e., it is able to learn from small training data to generalize its classification decisions correctly to the test data and the future unseen trials. This generalization gain mainly comes from the temporal and channel attention mechanisms that make salient the task-related information behind the noisy EEG signals and properly regularize the training process to converge to a generalizable classifier.

The demo code for TCACNet and the corresponding experimental data are available at https://github.com/LiuXiaolin-lxl/TCACNet.git

## 5. Related work

In a BCI system, brain signals are extracted, decoded, and analyzed, mainly using non-invasive brain monitoring techniques like EEG (Li, Bao, Li, & Zhao, 2020), functional magnetic resonance imaging (fMRI) (Wang et al., 2022), and functional near-infrared spectroscopy (fNIRS) (Alzahab et al., 2021). Among these techniques, EEG is commonly used in BCI because it has advantages of high temporal resolution and relatively low cost (Na et al., 2021). There are different types of EEG paradigms to be used for BCI, such as MI, P300 evoked potentials, and steady-state visual evoked potentials (SSVEP) (Jin et al., 2021; Lotte et al., 2018). Different from the two evoked potentials, which requires external stimulus to induce strong brain signals, MI is a popular paradigm where users practice a spontaneous imagination to generate commands voluntarily and without any external stimulus, making it a more user-friendly choice. Therefore, we chose MI-EEG based BCI systems as our focus in this paper. For MI-EEG based BCI systems, features extraction and classification of users' MI states is a core technical component. MI-EEG classification research can be roughly divided into two categories: traditional machine learning approaches and deep learning approaches.

Traditional machine learning approaches typically involve manually extracting, which is usually done by human experts, neurophysiological features from EEG data to enhance signal-to-noise-ratio and reduce dimensionality in order to achieve acceptable accuracy and efficiency. It has been reported that MI can be decoded from the sensory-motor rhythms (SMR), an oscillatory idle rhythm of synchronized electric brain activity. Using the SMR features, common spatial patterns (CSP) based methods (Ang, Chin, Zhang, & Guan, 2008; Jin et al., 2019) and Riemannian geometry based methods (Barachant, Bonnet, Congedo, & Jutten, 2013; Xu, Grosse-Wentrup, & Jayaram, 2020) have achieved reasonably good performance. On top of the extracted features, many advanced machine learning techniques are applied to MI-EEG classification. Specifically, feature selection methods are used to screen some features related to the target mental state, which can also reduce the impact of possible over-training and thus improve performance, especially when the number of training samples is small (Jin et al., 2021, 2020). Adaptive classifiers are able to incrementally update parameters online to keep track of the EEG changes over time (Hsu, 2011). Transfer learning makes it possible to adapt a BCI classifier trained for a subject to be applied directly or indirectly to other subjects (Dai, Zheng, Liu, & Zhang, 2018). However, machine learning methods have a high dependency on handcrafted feature extraction, which is time-consuming and non-trivial.

Deep learning, which is an extensively data-driven approach, has the advantage of end-to-end training without manual feature extraction. Therefore, an increasing research interest has been there to leverage deep learning approaches, particularly those based on CNNs, for EEG classification. For example, Shallow ConvNet (Schirrmeister et al., 2017), Deep ConvNet (Schirrmeister et al., 2017), and EEGNet-8,2 (Lawhern et al., 2018) are three state-of-the-art models that have drawn increasing attention from the EEG signal processing community. Inspired by CSP based methods, Shallow ConvNet employs a 2-layer architecture, which is sophisticatedly tailored to extract the band-power features. Deep ConvNet is similar to Shallow ConvNet but has extra convolution layers and pooling layers added to constitute a 5-layer architecture. EEGNet is a more compact network with a small number of layers and parameters, and its high performance comes from the depthwise and separable convolutional operation that can extract EEG features efficiently. Based on the architectures of the three methods, more variant structures (Amin, Alsulaiman, Muhammad, Mekhtiche, & Hossain, 2019; Dai, Zhou, Huang, & Wang, 2020; Wu et al., 2019) were proposed to optimize the features extracted by the networks.

However, most existing deep learning approaches suffer from the non-stationarity issue of EEG signals, and studies to tackle this issue is lacking. To fill the gap, we have previously proposed the TACNet to tackle the temporal non-stationarity of EEG signals through a temporal attention mechanism (Liu et al., 2021). To further address the non-stationarity problem in channel dimensions, TCACNet with an additional channel attention mechanism has been introduced in this paper. The TCACNet combines both attention mechanism to help with the concentration on task-related time slices and channels of the signals, and thus, improved performance can be achieved.

## 6. Discussion on research implications and limitations

This section discusses the implications of our research and what effect it may have on the future work of EEG signal processing and BCI practice. Furthermore, we also discuss some limitations of our work.

**Theoretical implication:** As stated in Section 2, this paper aims to achieve high-performing MI-EEG classification by overcoming two EEG challenges: the EEG signal's temporal non-stationarity and the nonuniform information distribution over different electrode channels. To this end, we propose the temporal attention mechanism and the channel attention mechanism, which are in essence inspired by two underlying hypotheses (i.e., the *inductive biases* as they are usually called in the machine learning field): (1) the classification entropy gradient with respect to each time slice is positively correlated with the amount of temporal task-related information of that time slice, and (2) WPSER, the energy-oriented quantity of each channel, is positively correlated with the amount of task-related information of that channel. Following these hypotheses, we have successfully designed the TCACNet with the attention mechanism structure to enable itself to focus on the extracted time slices and give more weight to the important channels to process the intrinsic task-related information for improved classification performance. Although not rigorously proved, these hypotheses are empirically evidenced by the experimental results in terms of accuracy, reliability and stability. From the machine learning research perspective, these two hypotheses can be treated as a principle for designing deep learning frameworks to handle EEG signals and other types of data featured with nonuniform distribution of task-related information. From the brain science perspective, these two hypotheses, if supported by more evidence in the future, might inspire the potential establishment of new quantitative indexes to assist tracking the dynamics of task-related information flows in a brain. To decode how task-related information flows temporally and spatially among brain regions during imaginary, vision processing and other information processing types is an important topic in brain science (Ghuman et al., 2014; Rabinovich, Afraimovich, Bick, & Varona, 2012).

**Practical implication:** TCACNet is designed with engineering goals to support the development of a reliable, stable and accurate BCI system. The technical components involved, especially the introduced temporal and channel attention mechanisms, are aimed at pushing forward the upper limit of MI-EEG classification accuracy with good reliability and stability. From a software perspective, TCACNet serves as the core algorithm in the BCI system to process the EEG signals, i.e., to extract and integrate the information in EEG more effectively, assisting the system to make correct decisions and control the external assistive devices, appropriately. From a hardware perspective, the TCACNet can easily adapt to different patterns of electrode array on scalp surface or different skull shapes of users, because the attention mechanisms enable the classifier to learn and adjust itself to filter out noise and extract task-related information, making the system functionally insensitive to hardware variance in terms of electrode locations. In addition, as can be seen in Table 3, the post-trained TCACNet is a relatively small network, implying a lower requirement on the computing unit and a more energy-efficient system for smart healthcare across the lifespan.

**Table 3**
Details of layers in sub-networks.

| Sub-network | Layer name | Filter number/Node number | Filter size | Filter stride | Activation function |
|---|---|---|---|---|---|
| Global sub-network | Spatial convolutional layer | 40 | BCIC: 22 × 1<br>HGD: 44 × 1 | 1 × 1 | No |
| | Temporal convolutional layer | 40 | 1 × 50 | 1 × 1 | Squaring |
| | Mean pooling layer | | 1 × 150 | 1 × 90 | Logarithmic |
| Local sub-network | Spatial convolutional layer | 40 | BCIC: 22 × 1<br>HGD: 44 × 1 | 1 × 1 | No |
| | Temporal convolutional layer (1st) | 40 | 1 × 26 | 1 × 1 | Squaring |
| | Mean pooling layer (1st) | | 1 × 6 | 1 × 6 | Logarithmic |
| | Temporal convolutional layer (2nd) | 40 | 1 × 9 | 1 × 1 | Squaring |
| | Mean pooling layer (2nd) | | 1 × 3 | 1 × 3 | Logarithmic |
| | Temporal convolutional layer (3rd) | 40 | 1 × 5 | 1 × 1 | Squaring |
| | Mean pooling layer (3rd) | | 1 × 3 | 1 × 3 | Logarithmic |
| Top network | Fully connected layer | 400 | | | No |
| | Softmax output layer | 4 | | | Softmax |

**Table 4**
Summary of models' performance for each subject on BCIC IV 2a and HGD datasets.

| Dataset | Subject | EEGNet-8,2 | Deep ConvNet | Shallow ConvNet | Global model | TACNet (without $L_H$) | TACNet (with $L_H$) | TCACNet |
|---|---|---|---|---|---|---|---|---|
| BCIC IV 2a | 1 | 79.1 | 71.3 | 83.5 | 85.2 | 88.7 | 93.0 | 91.3 |
| | 2 | 61.7 | 51.3 | 51.3 | 55.7 | 60.0 | 62.6 | 65.2 |
| | 3 | 87.8 | 83.5 | 87.8 | 91.3 | 93.0 | 93.9 | 93.9 |
| | 4 | 62.6 | 61.7 | 66.1 | 76.5 | 80.9 | 77.4 | 84.3 |
| | 5 | 80.0 | 86.1 | 70.4 | 80.0 | 86.1 | 86.1 | 90.4 |
| | 6 | 60.9 | 60.0 | 63.5 | 67.8 | 72.2 | 73.0 | 75.7 |
| | 7 | 73.9 | 75.7 | 88.7 | 94.8 | 96.5 | 93.9 | 98.3 |
| | 8 | 82.6 | 82.6 | 78.3 | 84.3 | 87.8 | 87.8 | 90.4 |
| | 9 | 90.4 | 80.9 | 77.4 | 80.0 | 84.3 | 88.7 | 91.3 |
| | Accuracy (mean ± std) | 75.4 ± 11.3 | 72.6 ± 12.3 | 74.1 ± 12.4 | 79.5 ± 11.9 | 83.3 ± 11.2 | 84.0 ± 10.8 | **86.8 ± 10.3** |
| HGD | 1 | 87.5 | 87.5 | 91.5 | 89.4 | 91.9 | 95.0 | 95.0 |
| | 2 | 81.3 | 88.8 | 88.3 | 85.6 | 89.4 | 93.8 | 95.0 |
| | 3 | 91.3 | 96.3 | 95.6 | 96.3 | 96.9 | 97.5 | 99.4 |
| | 4 | 91.9 | 93.8 | 92.1 | 92.5 | 98.1 | 98.8 | 99.4 |
| | 5 | 93.1 | 93.1 | 91.1 | 93.1 | 96.3 | 96.9 | 98.1 |
| | 6 | 83.1 | 86.9 | 87.8 | 93.1 | 94.4 | 96.3 | 97.5 |
| | 7 | 86.2 | 88.1 | 87.3 | 86.8 | 91.8 | 88.7 | 92.5 |
| | 8 | 90.6 | 85.6 | 88.6 | 88.1 | 91.9 | 93.1 | 95.0 |
| | 9 | 90.6 | 94.4 | 93.7 | 92.5 | 93.1 | 96.9 | 97.5 |
| | 10 | 85.6 | 82.5 | 84.2 | 85.0 | 90.6 | 91.3 | 91.9 |
| | 11 | 73.8 | 71.3 | 68.4 | 73.1 | 74.4 | 95.6 | 93.8 |
| | 12 | 91.3 | 92.5 | 92.7 | 92.5 | 95.6 | 96.3 | 98.1 |
| | 13 | 87.4 | 86.8 | 86.6 | 89.9 | 90.6 | 93.1 | 96.9 |
| | Accuracy (mean ± std) | 87.2 ± 5.4 | 88.3 ± 6.5 | 88.3 ± 6.9 | 89.1 ± 5.8 | 91.9 ± 5.9 | 94.9 ± 2.8 | **96.2 ± 2.5** |

**Limitations:** There are two main limitations of our method. First, the current network architecture of TCACNet is designed specifically for the MI-EEG classification, which might not be applicable to other types of EEG signals and task types. How to apply this method to other types of EEG signals and tasks still needs further study. Second, the current training procedure of TCACNet is a subject-specific one, and the model needs to be trained on the EEG data of a specific subject before it can achieve excellent classification performance on that subject. If the model is to be used to another subject, new EEG data needs to be collected and the network needs to be retrained. It is expected that transfer learning techniques might be a solution to this limitation.

## 7. Conclusion and future work

This paper proposes the TCACNet, a novel CNN-based framework combined with the temporal and channel attention mechanisms for subject-specific MI-EEG classification. In TCACNet, we integrate the global sub-network and the local sub-network to process different scopes of the input EEG. With the help of the temporal attention mechanism, the local sub-network is able to concentrate on task-related time slices for a fine-grained processing. To further improve the performance, we propose a channel attention mechanism that can adjust the weight coefficients of each channel in the network according to the important level of the channel's task-related information. Experimental results on two public EEG datasets verify the advantage of the proposed method over baselines in terms of classification accuracy and data efficiency.

The proposed TCACNet framework is extendable: by adjusting the network structure, the attention mechanism in TCACNet can be transferred to other EEG signal analysis scenarios, such as SSVEP and P300 evoked potentials. In addition, TCACNet can be applied to other practical scenarios, such as grabbing and stop-and-go control of the wheelchair. In the future, we will focus on applying our model to more physiological signal analysis scenarios and practical scenarios, and leverage additional advanced deep learning frameworks, such as Transformers, to further improve model performance. These works are believed to positively promote the development of BCIs to assist people with severe physical impairments across the lifespan.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 61873021; No. 62088101), the Basic and Applied Basic Research Foundation of Guangdong Province (NO. 2020A1515110887), the Special Fund for Basic Scientific Research of Central Colleges, Universities-Youth Talent Support Program of Beihang University, the Key Laboratory of Precision Opto-mechatronics Technology (Beihang University), Ministry of Education, China and Beijing Advanced Innovation Center for Big Databased Precision Medicine, Beihang University.

## Appendix A. Procedure of WPSER calculation

In the proposed channel attention mechanism, we use WPSER to quantify the ratio of the energy within 8–30 Hz to the signal's total energy, and the WPSER is treated as the measure of the salience level of task-related information for the signal of each electrode channel. The calculation procedure of WPSER is given by Eqs. (6)–(8) in Section 3.2.2, and the pseudocode of WPSER calculation is presented in Algorithm 1. The input of the pseudocode is a one-dimensional time series signal to be decomposed (e.g., the $c$th channel of the input signal $X$ belonging to the space $\mathbb{R}^{1\times T}$), and a parameter to specify layers of decomposition (usually based on the desired frequency resolution). The output is the WPSER corresponding to 8–30 Hz, i.e. $M^{c}_{channel}$ in Eq. (6). The pseudocode of WPSER calculation is divided into three parts: (a) Line 1–5: wavelet packet decompose layer by layer starting from the input signal in Eq. (6) (which can be implemented using the "PyWavelets (pywt)" toolkit in Python and its details can be found on https://pywavelets.readthedocs.io/en/latest/ref/wavelet-packets.html); (b) Line 6–9: calculate the energy of each sub-node signal and the total energy of the final layer in Eq. (7); (c) calculate the energy proportion of each sub-node and the energy proportion corresponding to 8–30 Hz in Eqs. (7)(8).

**Algorithm 1: The pseudocode of WPSER calculation**

**Input:** $x$: one-dimensional time series, i.e., the raw EEG signal of a channel;
$layers$: layers of decomposition, set to 5 by default;
**Output:** $wpser$: WPSER corresponding to 8–30 Hz;
**Note:** $y(layer, node)$: decomposed signal of sub-node indexed by $layer$ and $node$;
$E(layer, node)$: energy of signal of sub-node indexed by $layer$ and $node$;
$E_{total}$: total energy of all the sub-nodes' signals in the final layer;
$P(layer, node)$: energy proportion of sub-node indexed by $layer$ and $node$;

// Initializing $y$ as a zero matrix and tut $x$ to $y(0,0)$ for decomposition.
1: $y \leftarrow [0]_{layers\times(2^{layers})}$; $y(0,0) \leftarrow x$

// Wavelet packet decomposition layer by layer (Eq. (6)).
2: **for** $layer$ **from** 1 **to** $layers$:
3: **for** $node$ **from** 0 **to** $2^{layer}-1$:
4: **if** $node$ is even: $y(layer, node) = \text{low-pass}(y(layer-1, node/2))$;
5: **if** $node$ is odd: $y(layer, node) = \text{high-pass}(y(layer-1, (node-1)/2))$;

// Calculate energy of each sub-node and total energy in the final layer (part of Eq. (7)).
6: $E_{total} \leftarrow 0$;
7: **for** $node$ **from** 0 **to** $2^{layers}-1$:
8: $E(layers, node) \leftarrow \text{energy}(y(layers, node))$;
9: $E_{total} \leftarrow E_{total} + E(layers, node)$;

// Calculate energy proportion of each sub-node and 8–30 Hz in $layers$ (Eqs. (7) (8)).
10: $wpser \leftarrow 0$;
11: **for** $node$ **from** 0 **to** $2^{layers}-1$:
12: $P(layers, node) \leftarrow E(layers, node)/E_{total}$;
13: **if** $node$ **belong to** 8–30 Hz: $wpser \leftarrow wpser + P(layers, node)$;

14: **return** $wpser$

## Appendix B. Details of network architecture

The global sub-network $f_{global}$ contains a spatial convolutional layer and a temporal convolutional block (a temporal convolutional layer followed by a mean pooling layer). The local sub-network $f_{local}$ contains a spatial convolutional layer and three temporal convolutional blocks (each has a temporal convolutional layer followed by a mean pooling layer). The top network $g$ has a fully connected (FC) layer and a softmax output layer. The details of these layers are as follows.

## Appendix C. Detailed subject-specific experimental results

The experimental results for each subject, in terms of classification accuracies of all models, are presented in Table 4. The average accuracies and standard deviations in Table 4 are consistent with those in Figs. 6 and 7.

## References


Almathami, H. K. Y., Win, K. T., & Vlahu-Gjorgievska, E. (2020). Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. *Journal of Medical Internet Research*, *22*(2), Article e16407.

Alzahab, N. A., Apollonio, L., Di Iorio, A., Alshalak, M., Iarlori, S., Ferracuti, F., Monteriù, A., & Porcaro, C. (2021). Hybrid deep learning (hDL)-based brain-computer interface (BCI) systems: a systematic review. *Brain Sciences*, *11*(1), 75.

Amin, S. U., Alsulaiman, M., Muhammad, G., Mekhtiche, M. A., & Hossain, M. S. (2019). Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Generation Computer Systems*, *101*, 542–554.

Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *2008 IEEE international joint conference on neural networks (IEEE World congress on computational intelligence)* (pp. 2390–2397). IEEE.

Baig, M. Z., Aslam, N., Shum, H. P., & Zhang, L. (2017). Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. *Expert Systems with Applications*, *90*, 184–195.

Barachant, A., Bonnet, S., Congedo, M., & Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, *112*, 172–178.

Belkacem, A. N., Jamil, N., Palmer, J. A., Ouhbi, S., & Chen, C. (2020). Brain computer interfaces for improving the quality of life of older adults and elderly patients. *Frontiers in Neuroscience*, *14*(692), 1–11.

Brunner, C., Leeb, R., Müller-Putz, G., Schlögl, A., & Pfurtscheller, G. (2008). *BCI Competition 2008–Graz data set A, Vol. 16* (pp. 1–6). Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology.

Chen, X., Purohit, A., Dominguez, C. R., Carpin, S., & Zhang, P. (2015). Drunkwalk: Collaborative and adaptive planning for navigation of micro-aerial sensor swarms. In *Proceedings of the 13th ACM conference on embedded networked sensor systems* (pp. 295–308).

Chen, X., Purohit, A., Pan, S., Ruiz, C., Han, J., Sun, Z., Mokaya, F., Tague, P., & Zhang, P. (2017). Design experiences in minimalistic flying sensor node platform through sensorfly. *ACM Transactions on Sensor Networks*, *13*(4), 1–37.

Chen, X., Ruiz, C., Zeng, S., Gao, L., Purohit, A., Carpin, S., & Zhang, P. (2020). H-DrunkWalk: Collaborative and adaptive navigation for heterogeneous MAV swarm. *ACM Transactions on Sensor Networks*, *16*(2), 1–27.

Chen, X., Xu, S., Han, J., Fu, H., Pi, X., Joe-Wong, C., Li, Y., Zhang, L., Noh, H. Y., & Zhang, P. (2020). Pas: Prediction-based actuation system for city-scale ridesharing vehicular mobile crowdsensing. *IEEE Internet of Things Journal*, *7*(5), 3719–3734.

Chen, X., Xu, S., Liu, X., Xu, X., Noh, H. Y., Zhang, L., & Zhang, P. (2020). Adaptive hybrid model-enabled sensing system (HMSS) for mobile fine-grained air pollution estimation. *IEEE Transactions on Mobile Computing*, *21*(6), 1927–1944.

Cheng, D., Liu, Y., & Zhang, L. (2018). Exploring motor imagery EEG patterns for stroke patients with deep neural networks. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2561–2565). IEEE.

Cody, M. A. (1994). The wavelet packet transform: extending the wavelet transform. *Dr Dobb's Journal*, *19*, 44–46.

Dai, M., Zheng, D., Liu, S., & Zhang, P. (2018). Transfer kernel common spatial patterns for motor imagery brain-computer interface classification. *Computational and Mathematical Methods in Medicine*, *2018*, Article 9871603.

Dai, G., Zhou, J., Huang, J., & Wang, N. (2020). HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification. *Journal of Neural Engineering*, *17*(1), Article 016025.

Ghuman, A. S., Brunet, N. M., Li, Y., Konecky, R. O., Pyles, J. A., Walls, S. A., Destefino, V., Wang, W., & Richardson, R. M. (2014). Dynamic encoding of face information in the human fusiform gyrus. *Nature Communications*, *5*(1), 5672.

Hsu, W. Y. (2011). EEG-Based motor imagery classification using enhanced active segment selection and adaptive classifier. *Computers in Biology and Medicine*, *41*(8), 633–639.

Jin, J., Fang, H., Daly, I., Xiao, R., Miao, Y., Wang, X., & Cichocki, A. (2021). Optimization of model training based on iterative minimum covariance determinant in motor-imagery BCI. *International Journal of Neural Systems*, *31*(07), Article 2150030.

Jin, J., Miao, Y., Daly, I., Zuo, C., Hu, D., & Cichocki, A. (2019). Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Networks*, *118*, 262–270.

Jin, J., Wang, Z., Xu, R., Liu, C., Wang, X., & Cichocki, A. (2021). Robust similarity measurement based on a novel time filter for SSVEPs detection. *IEEE Transactions on Neural Networks and Learning Systems*, 1–10.

Jin, J., Xiao, R., Daly, I., Miao, Y., Wang, X., & Cichocki, A. (2020). Internal feature selection method of CSP based on L1-norm and Dempster–Shafer theory. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4814–4825.

Kim, H., & Xie, B. (2017). Health literacy in the eHealth era: a systematic review of the literature. *Patient Education and Counseling*, *100*(6), 1073–1082.

Kumar, S., Sharma, R., Sharma, A., & Tsunoda, T. (2016). Decimation filter with common spatial pattern and fishers discriminant analysis for motor imagery classification. In *2016 international joint conference on neural networks (IJCNN)* (pp. 2090–2095). IEEE.

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. *Journal of Neural Engineering*, *15*(5), Article 056013.

Li, C., Bao, Z., Li, L., & Zhao, Z. (2020). Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. *Information Processing & Management*, *57*(3), Article 102185.

Liu, X., Hui, Q., Xu, S., Wang, S., Na, R., Sun, Y., Chen, X., & Zheng, D. (2021). TACNet: TAsk-aware electroencephalogram classification for brain-computer interface through a novel temporal attention convolutional network. In *Adjunct Proceedings of the 2021 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2021 ACM international symposium on wearable computers (UbiComp/ISWC 2021)* (pp. 660–665).

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, *15*(3), Article 031005.

Mane, R., Chouhan, T., & Guan, C. (2020). BCI For stroke rehabilitation: motor and beyond. *Journal of Neural Engineering*, *17*(4), Article 041001.

Na, R., Hu, C., Sun, Y., Wang, S., Zhang, S., Han, M., Yin, W., Zhang, J., Chen, X., & Zheng, D. (2021). An embedded lightweight SSVEP-BCI electric wheelchair with hybrid stimulator. *Digital Signal Processing*, *116*, Article 103101.

Na, R., Zheng, D., Sun, Y., Han, M., Wang, S., Zhang, S., Hui, Q., Chen, X., Zhang, J., & Hu, C. (2021). A wearable low-power collaborative sensing system for high-quality SSVEP-BCI signal acquisition. *IEEE Internet of Things Journal*, *9*(10), 7273–7285.

Nijholt, A. (2016). The future of brain-computer interfacing (keynote paper). In *2016 5th international conference on informatics, electronics and vision (ICIEV)* (pp. 156–161). IEEE.

Oikonomou, V. P., Georgiadis, K., Liaros, G., Nikolopoulos, S., & Kompatsiaris, I. (2017). A comparison study on EEG signal processing techniques using motor imagery EEG data. In *2017 IEEE 30th international symposium on computer-based medical systems (CBMS)* (pp. 781–786). IEEE.

Orji, R., & Moffatt, K. (2018). Persuasive technology for health and wellness: state-of-the-art and emerging trends. *Health Informatics Journal*, *24*(1), 66–91.

Rabinovich, M. I., Afraimovich, V. S., Bick, C., & Varona, P. (2012). Information flow dynamics in the brain. *Physics of Life Reviews*, *9*(1), 51–73.

Romero, A., Ballas, N., Kahou, S. E., Chassang, A., Gatta, C., & Bengio, Y. (2015). FitNets: HInts for thin deep nets. In *3rd international conference on learning representations (ICLR)* (pp. 1–13).

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. *Human Brain Mapping*, *38*(11), 5391–5420.

Shi, R., Mo, Z., & Di, X. (2021). Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(1), 540–547.

Shi, R., Mo, Z., Huang, K., Di, X., & Du, Q. (2021). A physics-informed deep learning paradigm for traffic state and fundamental diagram estimation. *IEEE Transactions on Intelligent Transportation Systems*, 1–11.

Shi, R., Steenkiste, P., & Veloso, M. M. (2019). SC-M*: A multi-agent path planning algorithm with soft-collision constraint on allocation of common resources. *Applied Sciences*, *9*(19), 4037.

Shi, R., Steenkiste, P., & Veloso, M. M. (2021). Improving the on-vehicle experience of passengers through SC-M*: A scalable multi-passenger multi-criteria mobility planner. *IEEE Transactions on Intelligent Transportation Systems*, *22*(2), 1026–1040.

Swati, S., Kumar, M., & Namasudra, S. (2022). Early prediction of cognitive impairments using physiological signal for enhanced socioeconomic status. *Information Processing & Management*, *59*(2), Article 102845.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–9).

Tian, S., Yang, W., Grange, J. M. L., Wang, P., Huang, W., & Ye, Z. (2019). Smart healthcare: Making medical care more intelligent. *Global Health Journal*, *3*(3), 62–65.

Wang, Z., Hu, C., Zheng, D., & Chen, X. (2021). Ultra-low-power sensing framework for internet of things: a smart gas meter as a case. *IEEE Internet of Things Journal*, *9*(10), 7533–7544.

Wang, L., Yuan, W., Zeng, L., Xu, J., Mo, Y., Zhao, X., & Peng, L. (2022). Dementia analysis from functional connectivity network with graph neural networks. *Information Processing & Management*, *59*(3), Article 102901.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, *113*(6), 767–791.

Wolpaw, J. R., McFarland, D. J., Vaughan, T. M., & Schalk, G. (2003). The Wadsworth Center brain-computer interface (BCI) research and development program. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(2), 1–4.

World Health Organization (WHO) (2011). Preface. In *World report on disability* (p. xi). WHO.

Wu, H., Niu, Y., Li, F., Li, Y., Fu, B., Shi, G., & Dong, M. (2019). A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification. *Frontiers in Neuroscience*, *13*(1275), 1–9.

Xu, S., Chen, X., Pi, X., Joe-Wong, C., Zhang, P., & Noh, H. Y. (2019). iLOCuS: INcentivizing vehicle mobility to optimize sensing distribution in crowd sensing. *IEEE Transactions on Mobile Computing*, *19*(8), 1831–1847.

Xu, J., Grosse-Wentrup, M., & Jayaram, V. (2020). Tangent space spatial filters for interpretable and efficient Riemannian classification. *Journal of Neural Engineering*, *17*(2), Article 026043.