



An adaptive session-incremental broad learning system for continuous motor imagery EEG classification

Yufei Yang¹ · Mingai Li^{1,2,3} · Linlin Wang¹

Received: 17 June 2024 / Accepted: 8 November 2024 / Published online: 29 November 2024
© International Federation for Medical and Biological Engineering 2024

Abstract

Motor imagery electroencephalography (MI-EEG) is usually used as a driving signal in neuro-rehabilitation systems, and its feature space varies with the recovery progress. It is required to endow the recognition model with continuous learning and self-updating capability. Broad learning system (BLS) can be remodeled in an efficient incremental learning way. However, its architecture is intractable to change automatically to adapt to new incoming MI-EEG with time-varying and complex temporal–spatial characteristics. In this paper, an adaptive session-incremental BLS (ASiBLS) is proposed based on mutual information theory and BLS. For the initial session data, a compact temporal–spatial feature extractor (CTS) is designed to acquire the temporal–spatial features, which are input to a baseline BLS (bBLS). Furthermore, for new session data, a mutual information maximization constraint (MIMC) is introduced into the loss function of CTS to make the features' probability distribution sufficiently similar to that of the previous session, a new incremental BLS sequence (iBLS) is obtained by adding a small number of nodes to the previous model, and so on. Experiments are conducted based on the BCI Competition IV-2a dataset with two sessions and IV-2b dataset with five sessions, ASiBLS achieves average decoding accuracies of 79.89% and 87.04%, respectively. The kappa coefficient and forgetting rate are also used to evaluate the model performance. The results show that ASiBLS can adaptively generate an optimized and reduced model for each session successively, which has better plasticity in learning new knowledge and stability in retaining old knowledge as well.

Keywords Motor imagery EEG · Incremental learning · Broad learning system · Mutual information theory · Temporal–spatial features

1 Introduction

Motor imagery electroencephalography (MI-EEG)-based brain–computer interface (BCI) allows the human brain to control devices without activating muscles [1]. BCI technology can recognize and translate motor intentions to control rehabilitation assistive devices [2]. However, during the rehabilitation process, MI-EEG signals exhibit high-dimensional and time-varying characteristics, leading to significant differences in temporal and spatial feature distributions [3].

Most existing recognition models, which process all data simultaneously, work in batch mode, the changes in feature distribution are ignored. It is essential to develop more efficient incremental learning methods to improve the accuracy of MI task recognition across different sessions. Such advancements are critical for enhancing the clinical utility and adaptability of MI-BCI systems.

Incremental learning (IL) strategies strive to find a trade-off between plasticity and stability in neural networks. Plasticity refers to a model's ability to adapt and learn new knowledge, while stability ensures that previously learned knowledge is preserved even as the model is updated with new data, i.e., the ability to mitigate catastrophic forgetting [4, 5].

At present, IL strategies are generally categorized into three types: replaying old data, isolating parameters (or structural expansions), and adding regularization terms [6]. Replay-based methods combine old and new data during training to maintain model generalization and prevent

✉ Mingai Li
limingai@bjut.edu.cn

¹ School of Information Science and Technology, Beijing University of Technology, Beijing 100124, China

² Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China

³ Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China

overfitting [7, 8]. However, these methods may become impractical when data privacy or storage limitations are a concern. Parameter isolation methods freeze key parameters for previous tasks while expanding new network branches to accommodate new data [9]. If these frozen parameters are not properly aligned with the new branches, feature representations for earlier tasks may degrade, leading to catastrophic forgetting [10]. Regularization-based methods introduce constraints to limit the update direction of parameters, maintaining consistency with previous tasks [4, 11]. However, relying solely on regularization may limit the model's adaptability to new data. At the current stage, numerous researchers are focusing on hybrid methods, which integrate multiple strategies to mitigate individual limitations [12–14]. Despite these efforts, developing effective IL methods for incremental MI-EEG decoding remains a significant challenge.

Deep convolutional neural network (DCNN) provides a powerful framework for continuous EEG decoding, with their strong feature representation and high recognition performance resulting from numerous hyperparameters and layers [15]. Thus, the training and deployment of DCNN-based models is costly and time consuming, especially for complex and incremental tasks. Consequently, compact DCNN architectures are increasingly advocated for IL scenarios. Lawhern et al. [16] first proposed EEGNet, a compact and robust CNN, to extract interpretable temporal and spatial features from EEG-based BCI. Later, EEGNet was deployed as a feature extractor for widespread applications in emotion recognition [17], fatigue driving detection [18], epilepsy prediction [19], and so on. In MI-EEG classification, the classification accuracies can reach 77% [20] and 74.93% [21] on the binary and four-category MI tasks, respectively. The performance of the traditional EEGNet needs to be further improved. Moreover, when updated with new data that differ from previous data distributions, EEGNet is vulnerable to catastrophic forgetting.

In addition, to explore alternative models for DCNN, the broad learning system (BLS) was proposed in [22]. BLS not only shows high recognition performance but also offers an expandable architecture suitable for incremental learning. Its convergence can be improved by adding nodes in the broad direction, while previously added nodes are frozen [23]. Notably, the weights and bias of output layer are optimized by recursive least squares rather than iterative updating [24]. This “one-step” update rule balances high performance with low computational complexity. Furthermore, compared with DCNN, BLS has the higher interpretability [25].

In recent years, more efficient derivative algorithms based on BLS have been developed. Fu et al. [26] introduced a novel transformer–BLS network aimed at balancing training speed and accuracy. A stacking multi-view BLS with residual structures was proposed to enhance classification performance on complex datasets [27]. However, in EEG-based classification, the reported researches show that BLS-based methods can

hardly surpass DCNN-based methods [28–30]. This is primarily due to the limitations of BLS's flat model structure, which makes it challenging to capture the implicit information within high-dimensional, time-varying EEG signals.

Inspired by BLS and the three aforementioned incremental learning strategies, an adaptive recognition method, called adaptive session-incremental BLS (ASiBLS), is developed for incremental MI-EEG from various sessions. ASiBLS integrates mutual information theory with BLS-based dynamic remodeling mechanisms to improve adaptability and continuous learning capability. First, raw MI-EEG is preprocessed through filtering and data augmentation. A compact temporal–spatial convolutional block (CTS) is designed to capture the temporal and spatial features of MI-EEG on the initial session. Then, an optimized baseline BLS (bBLS), enhanced by sparse autoencoder (SAE) and grid search (GS), is built to map and enhance the global features and then classify them. For the incremental sessions, CTS is updated to an incremental CTS (iCST) by using mutual information to reduce distributional discrepancies between the current and adjacent sessions. Finally, an expandable incremental BLS sequence (iBLS) is constructed by adding minimal nodes to the existing model, ensuring efficient adaptation to new session while preserving knowledge from previous session.

The main contributions of this work are summarized as follows:

A session-incremental method, ASiBLS, is proposed to automatically classify the time-varying and complex MI-EEG across incremental sessions.

A novel CTS is constrained by the additional MIMC in the loss function, making the feature distributions of two adjacent sessions more similar.

For each new session MI-EEG, iBLS with optimized structure is generated based on the previous session by adding a small number of neural nodes.

The incremental learning performance of ASiBLS is evaluated on two public BCI competition datasets.

The rest of the paper is organized as follows: Section 2 summarizes the research progress of related work. Section 3 presents details of the ASiBLS. Large numbers of related experiments and results are presented in Section 4. Section 5 discusses the present work. Finally, Section 6 concludes this paper.

2 Related works

2.1 Incremental learning

At present, the replay-based methods have achieved significant progress. On the one hand, real old data can be directly replayed to train with new data. Rebuffi et al. [31]

gave an incremental classifier and representation learning; Tian et al. [7] proposed a novel pre-trained model knowledge distillation method instead. Other studies, such as the prototype similarity replay and similarity-adjusted regularization (PSSR) [32] and the adaptive-experience replay and distribution character-guided projection replay network (AdaER) [33], involved playing back the prototypes or representative features from each category of old tasks. Furthermore, some works have deployed generative models to generate pseudo-old data. Kang et al. [34] proposed a recall-oriented continual learning framework to address the stability–plasticity dilemma. He et al. [35] introduced an introspective generative adversarial network to perform incremental generation and classification. Yang et al. [36] proposed a dynamic regeneration with task-adaptive distillation for class incremental learning.

Parameter-isolated incremental learning methods perform well in alleviating catastrophic forgetting, especially for old and new tasks with high variability. The classic methods include progressive neural networks (PNN) [9], dynamically expandable networks (Der) [10], piggyback networks [37], and so on. In recent years, as researchers have increased their computational resources, some more advanced modeling improvements have been proposed, such as an efficient perturbation inference and expandable network (EPIE-Net) that achieved superior performance with fewer parameters on class incremental learning benchmarks [38]; an adaptive parameter multiplexing model was proposed for dynamic parameter adjustments in [39]. More researchers are working to reduce its resource consumption while improving its applicability in more complex scenarios.

Regularization-based strategies are most popularly used in the current phase of incremental learning research. There are three main mechanisms: regularization of model weights, such as classic elastic weight consolidation [4] and gradient reweighting [11]; regularization based on knowledge distillation that includes the learning without memorizing [40], the dual distillation structure for few-shot class incremental learning [41]; and task correlation regularization, which involves the proposed task-based hard attention mechanisms [42] and so on.

Currently, of course, more research is being done on the integration of two or more incremental learning strategies, which is superior to methods using only one strategy [12–14].

2.2 Incremental learning in EEG classification

To advance the real-world application of EEG-based BCI technology, several DCNNs incorporating IL strategies have been developed. In EEG-based emotion recognition, incremental learning preserving the learned feature space method (IL2FS) was established by adding a bias correction in the

loss function and replaying a small number of old samples [43]. An incremental multimodal autoencoder (AE) was proposed to diagnose chronic diseases; it provided an extensible modeling structure for new tasks [44]. Incremental vocabulary learning in imagined speech was facilitated by a class-incremental neural network, which employed a novel loss function designed to minimize the center-of-mass distance between old and new tasks [45]. In EEG-based seizure prediction, Deng et al. [46] proposed an experience replay (ER) method combined with centroid matching (ER-CM) to build a unified epilepsy prediction incremental learning system. Sun et al. [47] designed a meta self-attention prototype incrementer (MAPIC) to enhance few-shot class-incremental learning for medical time series, achieving significant performance improvements. A memory-based approach for online sequential EEG decoding is proposed, utilizing a dynamic memory buffer and kernel-based subject shift detection to address catastrophic forgetting and maintain performance across sequential subjects [48].

2.3 Motor imagery EEG classification

Motor imagery decoding and classification technology has made significant progress, especially as derived artificial intelligence algorithms, represented by CNN, have been widely used to extract temporal and spatial features automatically. Based on some classical, EEGNet [16], Shallow-ConvNet [15], the long short-term memory (LSTM) [49], and so on, many state-of-the-art models have been developed in recent years. Fu et al. proposed an iterative weighted sparse-group lasso (iWSGL) to optimize high-dimensional features by adaptively grouping features and calculating weights based on conditional entropy [50]. A novel network framework (IFBCLNet), which integrates CNN and LSTM, is presented to extract spatio-temporal-frequency features of MI-EEG [51]. A manifold attention-based multi-domain CNN (MAMCNet) was introduced to decode brain intention [52]. To boost the robustness of the decoding MI model, the parallel multi-scale temporal convolutional network (TCN) layers were introduced to capture different detailed features [53]. To address low signal-to-noise ratio and inter-subject variability, a novel end-to-end multi-branch 1D convolutional neural network (EEG-FMCNN) was presented in [21]. Moreover, a novel Riemannian geometry and deep domain adaptation network (RGDDANet) was proposed to classify MI, where two 1-D convolutions were exploited to obtain temporal and spatial features [54]. Although CNN-based deep learning exhibits high recognition performance, the multilayer structure inevitably increases the computational costs.

Broad learning system has been successfully applied in MI-EEG classification, owing to its advantages in efficiency, interpretability, and easy implementation. She et al. [28] gave

a graph-based semi-supervised broad learning system, which utilizes graph label propagation to generate pseudo-labels for unlabeled data and trains the classifier with both labeled and pseudo-labeled samples. Based on GSS-BLS, a novel semi-supervised broad learning system with transfer learning (TSS-BLS) was further proposed in [29]. Moreover, a session-incremental broad learning system (SIBLS) was proposed to recognize incremental MI-EEG by incorporating the advantages of the principal component analysis network (PCANet) and SAE [30]. However, SIBLS covers multiple layers of structure, which makes model training more time consuming.

3 Materials and methodology

This section first presents two public EEG datasets, which are used to evaluate our method, and describe the preprocessing techniques applied to the raw MI-EEG. The implementation details of ASiBLS on the initial and incremental sessions are separately given in the subsequent subsections. An illustration of ASiBLS is shown in Fig. 1.

3.1 EEG datasets

The public datasets BCI Competition IV-2a (IV-2a) [55] and BCI Competition IV-2b (IV-2b) [56], which were provided by the Brain-Computer Interface Laboratory of Graz University of Technology's Knowledge Discovery Institute, are used to evaluate the proposed method. Dataset details are summarized in Table 1. Each session MI-EEG was recorded

on different days; the sampling time schemes of one session are shown in Fig. 2. IV-2a can be downloaded from <http://www.bbc.de/competition/iv>, and IV-2b can be downloaded from <http://www.bbc.de/competition/iv/#dataset2b>.

3.2 Preprocessing and input definition

MI-EEG records motor-related perceptual rhythmic changes in the cerebral cortex during MI, mostly covering both Mu and Beta rhythms. In most reported studies, Mu and Beta frequency bands are considered to 8–13 Hz and 14–32 Hz, respectively [15]. To filter raw MI-EEG, a finite impulse response (FIR) bandpass filter with a bandwidth of 8–32 Hz is applied. In addition, due to the limited data available in independent session studies, models are prone to overfitting, thus data augmentation plays an important role. The sliding window method is introduced by using an overlapping window to slice filtered signals. Given a time sequence of length T , the number of augmented sub-sequences $\mathcal{N}$ can be calculated by the following equation:

$$\mathcal{N} = \left(\frac{T - l_w}{l_s} \right) + 1 \quad (1)$$

where l_w is the window length and l_s is the step size of the window movement.

Suppose that a session stream of labeled MI-EEG chunks is represented as $\{\mathbb{D}^{(1)}, \dots, \mathbb{D}^{(n)}, \dots, \mathbb{D}^{(N)}\}$, where N is the number of sessions, for the n th session $\mathbb{D}^{(n)} = \{X^{(n)}, Y^{(n)}\}$,

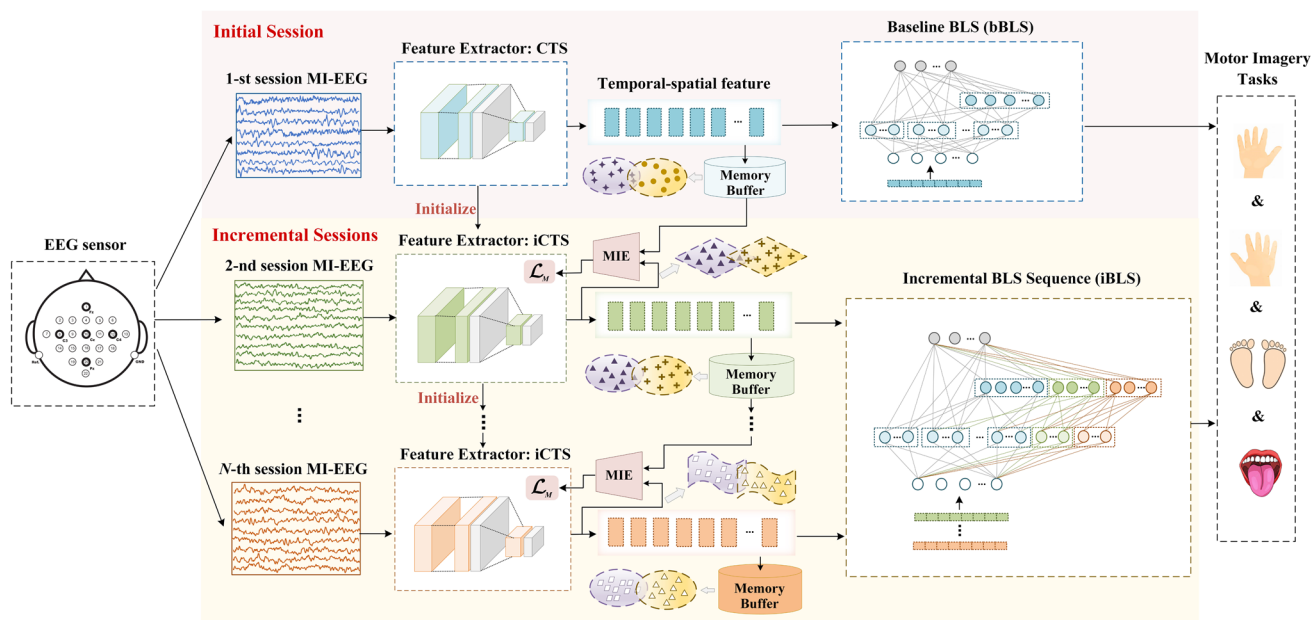
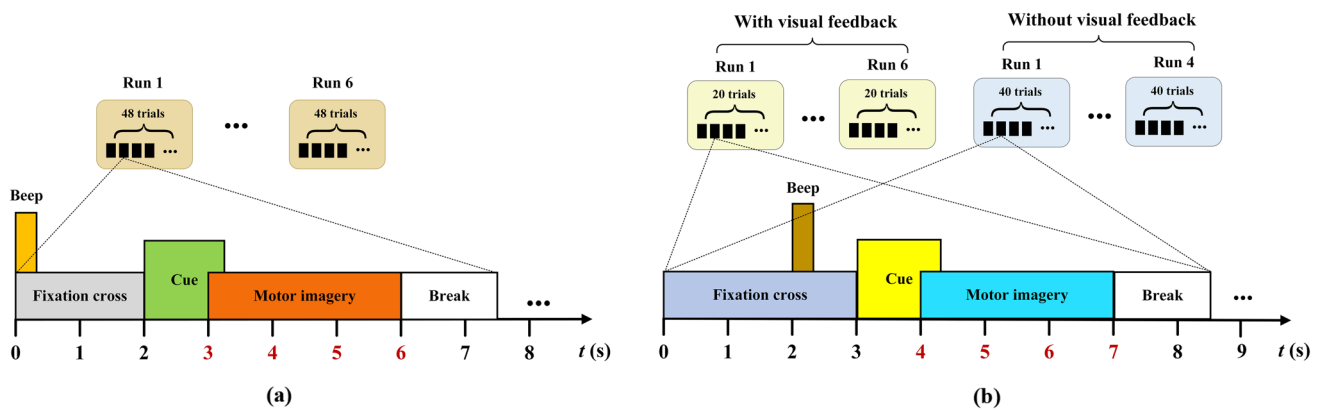


Fig. 1 Illustration of ASiBLS, where CTS is a compact temporal–spatial convolutional block, iCTS is an incremental CTS, MIE is the mutual information estimator, and $\mathcal{L}_M$ is the objective function of mutual information maximization constraint term

Table 1 Details of EEG datasets

Parameters	IV-2a	IV-2b	Explanation
—	9	9	No. of subjects
N	2	5	No. of sessions
S	288	120 (with feedback)/160 (without feedback)	No. of sampling trials for each session
f_s	250 Hz	250 Hz	Sampling frequency
—	3–6 s	4–7 s	Duration of MI
T	750	750	No. of MI sampling points
C	22	3	No. of EEG channels
L	Left hand, right hand, foot, tongue	Left hand, right hand	Categories of MI tasks


Fig. 2 The sampling time scheme for one session. **a** IV-2a dataset, **(b)** IV-2b dataset

the raw sampling signal $X^{(n)} \in \mathbb{R}^{S \times C \times T}$, and the label matrix $Y^{(n)} \in \mathbb{R}^{S \times L}$. On each session of IV-2a, the number of sampling trials S , EEG channels C , sampling points T , and MI categories L are 288, 22, 250, and 4, respectively. Additionally, on each session of IV-2b, C , T , and L are 3, 250, and 2, respectively; S is 120 on the 1st and 2nd sessions; and S is 160 on the 3rd, 4th, and 5th sessions. Then, $X^{(n)}$ is pre-processed. According to Eq. (1), when l_w and l_s are set to 1 s (250 sampling points) and 0.2 s (50 sampling points), respectively, $\mathcal{N}$ is calculated as 11. More specifically, after data augmentation, the number of sampling trials S' can be obtained by $S \times \mathcal{N}$ for each session of IV-2a, i.e., S' is equal to 288×11 . Similarly, for each session of IV-2b, S' is equal to 120×11 on the 1st and 2nd sessions and 160×11 on the 3rd, 4th, and 5th sessions, respectively.

3.3 Modeling on the initial session

3.3.1 Compact temporal–spatial feature extractor

Inspired by EEGNet [16], a compact temporal–spatial feature extractor (CTS) is structured to extract low-dimensional temporal–spatial features, which record the local distribution

characteristics of MI-EEG. CTS mainly includes a temporal convolution block (T-Conv) and a spatial convolution block (S-Conv). Unlike EEGNet constructed by a two-dimensional convolution (Conv2D) and Depthwise and Separable convolutions, a Conv2D in S-Conv replaces the Separable convolution in EEGNet. The Conv2D, Depthwise Conv2, and another Conv2D can capture sequentially temporal features along the time dimension, spatial features along the EEG channels dimension, and the feature map channel dimension. The implementation details are listed in Table 2 and the illustration of CTS is depicted in Fig. 3.

The first convolutional block, T-Conv, is used to extract features in the temporal dimension. T-Conv is composed of a Conv2D layer with D_1 temporal filters of kernel size $(1, k_1)$, where k_1 is the filter length, and a batch normalization (BN) layer. In this work, k_1 is set to a quarter of the sampling frequency f_s . The output of T-Conv is D_1 temporal feature maps.

Sequentially, the second convolutional block, S-Conv, is designed to independently capture the spatial features of the temporal feature maps, which correspond to the EEG channels. It then integrates all feature maps simultaneously, ensuring that both temporal and spatial information are effectively combined for more comprehensive feature

Table 2 Implementation details of CTS, where T-Conv is the temporal convolution block; S-Conv is the spatial convolution block; “C,” “T,” and “S” are the number of MI-EEG channels, sampling points,

and sampling trials, respectively; and “D₁,” “D₂,” and “D₃” are the number of temporal feature filters, spatial feature filters in the depthwise Conv2D layer and the Conv2D layer, respectively

Blocks	Layers	No. of filters	Kernel sizes	No. of parameters	Output	Activation
Input	Input	—	—	—	(C, T)	—
	Reshape	—	—	—	(1, C, T)	—
T-Conv	Conv2D	D ₁	(1, f _s /4)	D ₁ × f _s /4	(D ₁ , C, T)	Linear
	BN1	—	—	D ₁ × 2	(D ₁ , C, T)	—
S-Conv	Depthwise Conv2D	D ₂ × D ₁	(C, 1)	D ₁ × D ₂ × C	(D ₁ × D ₂ , 1, T)	Linear
	BN2	—	—	D ₁ × D ₂ × 2	(D ₁ × D ₂ , 1, T)	ELU
	Avg. pool	—	(1, 4)	—	(D ₁ × D ₂ , 1, T // 4)	—
	Conv2D	D ₃	(D ₂ × D ₁ , f _s /16)	f _s /16 × D ₂ × D ₁ + D ₃ × (D ₂ × D ₁)	(D ₃ , 1, T // 4)	Linear
	BN3	—	—	2 × D ₃	(D ₃ , 1, T // 4)	ELU
	Avg. pool	—	(1, 8)	—	(D ₃ , 1, T // 32)	—
Output	Flatten	—	—	—	(D ₃ × (T // 32), 1)	—

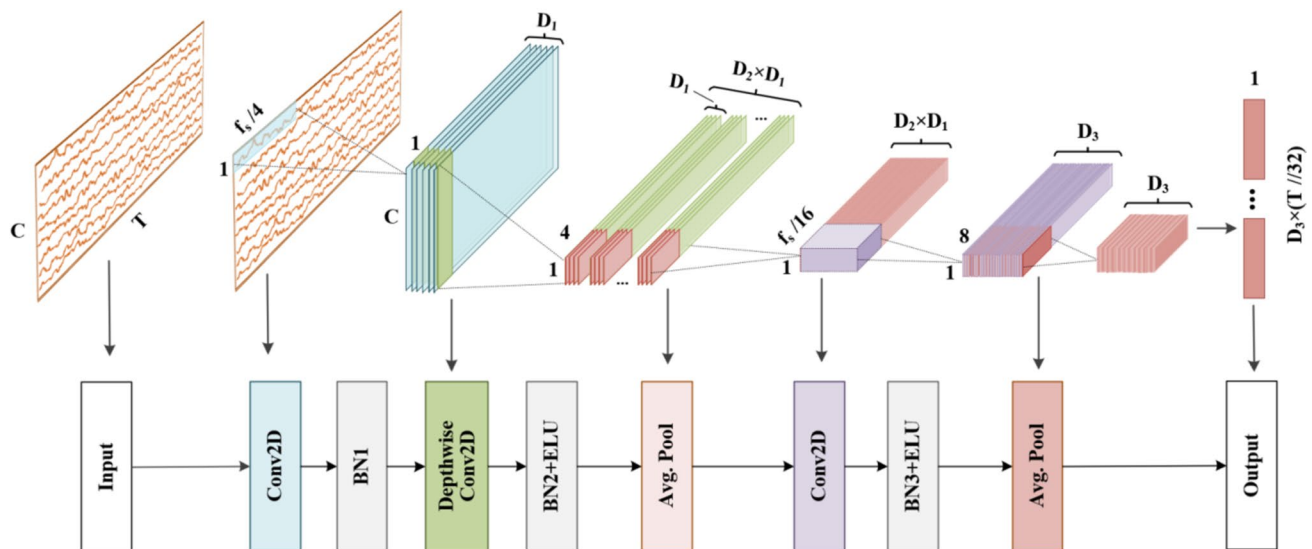


Fig. 3 Illustration of CTS, where the input of CTS is preprocessed MI-EEG and the outputs are temporal–spatial features

extraction. Two convolutional layers, Depthwise Conv2D and Conv2D, are incorporated in S-Conv. In the Depthwise Conv2D layer, the number of spatial filters is D_2 and kernel size is $(C, 1)$; the output dimension is controlled via $D_1 \times D_2$. The exponential linear unit (ELU) is used for nonlinear activation. An average pool layer of size $(1, 4)$ is attached to Depthwise Conv2D for feature compression. Next, in the Conv2D layer, the number of filters is D_3 and kernel size is $(D_1 \times D_2, k_2)$, where the length of temporal filter k_2 is set to one-sixteenth of the frequency. ELU is also used as an activation function. Another average pool layer of size $(1, 8)$ is introduced to reduce the sampling rate. All convolutional layers are performed batch normalization to accelerate convergence.

Finally, a flatten layer is exploited to flat temporal–spatial features $F^{(n)} \in \mathbb{R}^{S \times (8 \times D_3)}$. Moreover, an auxiliary dense layer with softmax activation is appended to CTS for supervised feature representation. CTS is first trained on the initial session by minimizing the softmax cross-entropy loss $\mathcal{L}_{CE}$, i.e.,

$$\text{argmin: } \mathcal{L}_{CE} = - \sum_{s=1}^S p_s(X_s^{(1)}; \mathcal{W}^{(1)}) \log \hat{p}_s(X_s^{(1)}; \mathcal{W}^{(1)}) \quad (2)$$

where $p_s(X_s^{(1)}; \mathcal{W}^{(1)})$ and $\hat{p}_s(X_s^{(1)}; \mathcal{W}^{(1)})$ are the actual and predicted probabilities for the s – th MI sampling trial. Finally, the temporal–spatial features $F^{(n)}$ of the initial session are outputted from CTS.

CTS offers significant advantages in EEG feature extraction, including automatic feature extraction, efficiency, and

enhanced performance. Different from traditional methods that rely on expert knowledge and experience, CTS simplifies the feature extraction process and effectively captures the complex characteristics of EEG signals through deep learning. This makes it more practical for BCI applications and clinical diagnosis, overcoming the limitations associated with conventional approaches.

3.3.2 Baseline BLS

In contrast to deep learning, BLS greatly reduces computational costs by adopting forward ridge regression approximation instead of backward iterative updating and randomly generating weights and biases [22]. In this paper, an optimized baseline BLS (bBLS) by SAE and GS is developed to further capture features from a global perspective and classify MI tasks on the initial session. SAE, as a powerful feature optimization tool, can effectively extract high-quality and sparse features to reduce redundancy and enhance feature differentiation, thereby improving model performance and reducing computational complexity [25]. Depending on the feature distribution of the different sessions, the width (scale) of bBLS can be adaptively determined by GS. The illustration of bBLS is shown in Fig. 4.

Given the feature set of the initial session is $\{F^{(1)}, Y^{(1)}\}$, in the feature mapping layer, $F^{(1)}$ first is mapped to K groups of feature nodes, which are computed by

$$Z_{\kappa}^{(1)} = \ell \left(F^{(1)} W_{Z_{\kappa}}^{(1)} + \beta_{Z_{\kappa}}^{(1)} \right), \quad \kappa = 1, \dots, K \quad (3)$$

where $\ell(\cdot)$ is the linear activation function, the randomly generated connection weight and bias $W_{Z_{\kappa}}^{(1)}$ and $\beta_{Z_{\kappa}}^{(1)}$ [22]. To overcome the unpredictability of random $W_{Z_{\kappa}}^{(1)}$ and $\beta_{Z_{\kappa}}^{(1)}$, SAE is used to fine-tune randomly generated $Z_{\kappa}^{(1)}$ into a set of parse and compact features, which can better characterize the data distributions. The sparse features can be achieved by

$$\underset{\tilde{W}_{Z_{\kappa}}}{\operatorname{argmin}} \| Z_{\kappa}^{(1)} \tilde{W}_{Z_{\kappa}} - F^{(1)} \|_2^2 + \gamma \| \tilde{W}_{Z_{\kappa}} \| \quad (4)$$

where $\tilde{W}_{Z_{\kappa}}$ is the connection weight of SAE and γ is the regularization coefficient. The mapped features of all groups are concatenated and represented as $Z^{(1)} = [Z_1^{(1)}, Z_2^{(1)}, \dots, Z_K^{(1)}]$. Then, $Z^{(1)}$ is applied to generate M groups enhancement nodes in the feature enhancement layer, and the calculation equation is given as

$$H_m^{(1)} = \xi \left([Z_1^{(1)}, Z_2^{(1)}, \dots, Z_K^{(1)}] W_{H_m}^{(1)} + \beta_{H_m}^{(1)} \right), \quad m = 1, \dots, M \quad (5)$$

where $W_{H_m}^{(1)}$ is randomly generated as an orthogonal matrix and $\xi(\cdot)$ is the nonlinear activation function. Same as $Z^{(1)}$, $H^{(1)}$ is also concatenated as $H^{(1)} = [H_1^{(1)}, H_2^{(1)}, \dots, H_M^{(1)}]$. The mapping and enhancement features are combined to

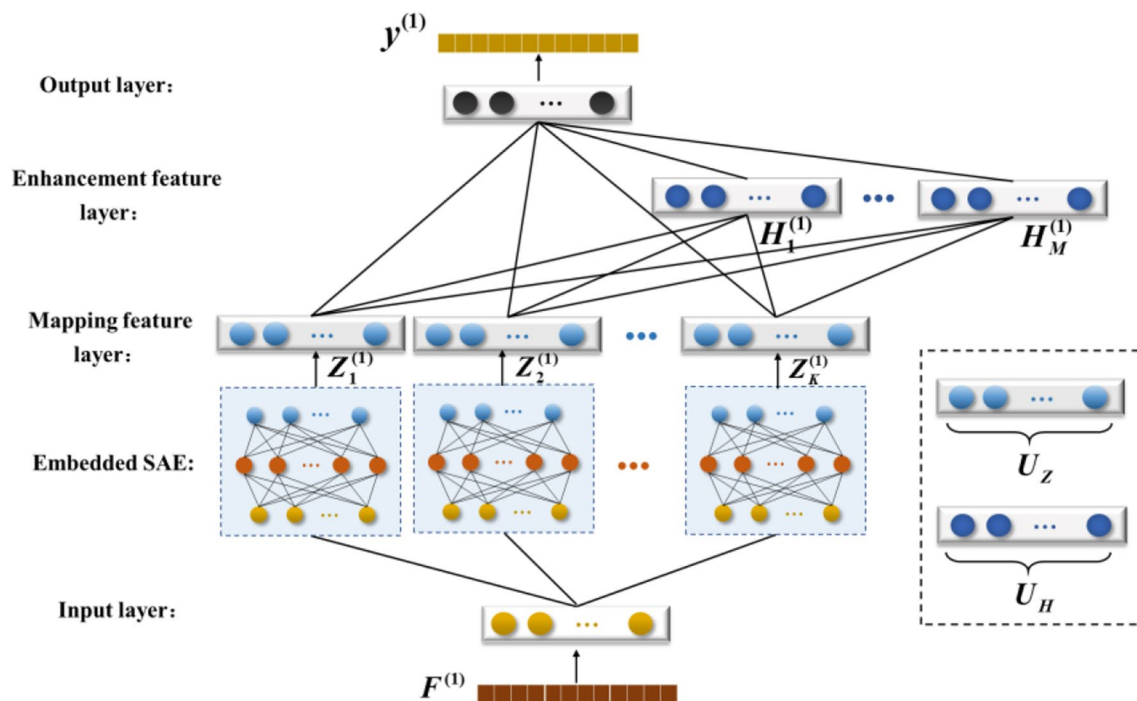


Fig. 4 Illustration of bBLS, where $F^{(1)}$ is the temporal-spatial features from the initial session, $Z_1^{(1)}, \dots, Z_K^{(1)}$ are the mapping features, $H_1^{(1)}, \dots, H_M^{(1)}$ are the enhancement features, $y^{(1)}$ is the predictions,

and U_Z and U_H are the number of mapping and enhancement feature nodes in a group

constitute the features matrix $A^{(1)} \triangleq [Z_1^{(1)}, \dots, Z_K^{(1)} | H_1^{(1)}, \dots, H_M^{(1)}]$. Finally, the outputs of bBLS are solved by $y^{(1)} = A^{(1)}W_O^{(1)}$, where $W_O^{(1)}$ is the output layer connection weights. To avoid model overfitting and boost the generalization ability, $W_O^{(1)}$ are derived and optimized by the ridge regression, which is the least squares estimate of linear equations, and the objective function is as follows:

$$\operatorname{argmin}_{W_O} : \|A^{(1)}W_O^{(1)} - Y\|^2 + \mu \|W_O^{(1)}\|^2 \quad (6)$$

then

$$W_O^{(1)} = (A^{(1)})^+ Y^{(1)} = ((A^{(1)})^T A^{(1)} + \mu I)^{-1} (A^{(1)})^T Y^{(1)} \quad (7)$$

where the pseudo-inverse matrix $(A^{(1)})^+$ can be approximated as $\lim_{\mu \rightarrow 0} ((A^{(1)})^T A^{(1)} + \mu I)^{-1} (A^{(1)})^T$, and μ is the regularization coefficient, which will be selected by repetitive experiments. The selectable hyper-parameters K and M will be determined via GS for the specific session.

The modeling process on the initial session is shown in Algorithm 1.

Algorithm 1 Modeling on the initial session

Input: Training dataset $\mathbb{D}^{(1)} = \{X^{(1)}, Y^{(1)}\}$, learning rate η , epochmax, initial K , M , U_z , U_h and search steps.

Output: $\mathcal{W}^{(1)}, \mathcal{W}_z^{(1)}, \mathcal{W}_h^{(1)}, \mathcal{W}_o^{(1)}$ and $F^{(1)}$

CTS:

- 1 **for** epoch = 1, 2, ..., epochmax **do**
- 2 Set random $\mathcal{W}^{(1)}$;
- 3 Calculate $\mathcal{L}_c$ by Eq.(2);
- 4 Update $\mathcal{W}^{(1)}$ by $\mathcal{W}^{(1)} = \mathcal{W}^{(1)} - \eta \nabla \mathcal{L}_c(\mathcal{W}^{(1)})$;
- 5 epoch = epoch + 1;
- 6 **end**
- 7 Save $F^{(1)}$ in memory buffer;

bBLS:

- 8 Search the optimal K , M , U_z , U_h by GS
- 9 Randomly generate $\mathcal{W}_z^{(1)}$, $\beta_z^{(1)}$;
- 10 Calculate $Z^{(1)} = [Z_1^{(1)}, Z_2^{(1)}, \dots, Z_K^{(1)}]$ by Eqs.(3) and (4);
- 11 Randomly generate $\mathcal{W}_h^{(1)}$, $\beta_h^{(1)}$;
- 12 Calculate $H^{(1)} = [H_1^{(1)}, H_2^{(1)}, \dots, H_M^{(1)}]$ by Eq.(5);
- 13 Set $\mathcal{A}^{(1)} \triangleq [Z^{(1)} | H^{(1)}]$
- 14 calculate $\mathcal{W}_o^{(1)}$ by Eqs.(6) and (7).

3.4 Modeling on the incremental sessions

3.4.1 Incremental CTS

Time-varying MI-EEG exhibits significant feature distribution variations on different sessions. The trained CTS by the

old sessions is hard to converge rapidly on the new session. However, directly retraining a new subnetwork may cause the modified model to deviate from the optimal solutions for the old sessions. Motivated by the mutual information theory and the regularization-based incremental learning, an extra regularization term, called mutual information maximization constraint (MIMC), is imposed on the loss function of GST, yielding an incremental CTS (iCTS) to improve the features distribution similarity between new and previous sessions.

Mutual information theory based on Shannon entropy-based statistics [57] is usually used to measure the correlation of two nonlinear random variables by calculating probability [58]. When the labeled MI-EEG from the n th session $\{X^{(n)}, Y^{(n)}\}$ is added to ASiBLS, the mutual information between temporal-spatial features $F^{(n)}$ and $F^{(n-1)}$ can be calculated by

$$\mathcal{I}(F^{(n)}, F^{(n-1)}) = \mathbb{H}(F^{(n)}) + \mathbb{H}(F^{(n-1)}) - \mathbb{H}(F^{(n)}, F^{(n-1)}) \quad (8)$$

where $\mathbb{H}(\cdot)$ denotes the Shannon entropy, which is utilized to quantify the uncertainty or randomness of variates. However, the direct calculation of Shannon entropy requires a known prior probability distribution, which is a significant challenge for complex high-dimensional variables or sparse data. Thus, a mutually informed neural estimation (MIME) is introduced to estimate $\mathcal{I}(F^{(n)}, F^{(n-1)})$. MIME reformulates mutual information into Donsker–Varadhan representation by using a mutual information estimator (MIE), which is a parameterized neural network, and the reformulated equation is given as follows:

$$\mathcal{I}(F^{(n)}, F^{(n-1)}) \geq \mathcal{D}(P_J \| P_Q) \quad (9)$$

$$\mathcal{D}(P_J \| P_Q) = \sup \mathbb{E}_{P_J}(\mathcal{M}(F^{(n)}, F^{(n-1)}))$$

$$-\log\left(\mathbb{E}_{P_Q}\left(e^{\mathcal{M}(F^{(n)}, F^{(n-1)})}\right)\right) \quad (10)$$

where $\sup \mathbb{E}(\cdot)$ indicates the supremum of expectation, P_J and P_Q are the probabilities of joint distribution and product of marginal distributions of $F^{(n)}$ and $F^{(n-1)}$, respectively, and $\mathcal{M}$ stands for the parameterized MIE. Furthermore, $\mathcal{I}(F^{(n)}, F^{(n-1)})$ is approximated as a probabilistic form:

$$\begin{aligned} \mathcal{I}(F^{(n)}, F^{(n-1)}) &\approx -\frac{1}{S'} \sum_{s'=1}^{S'} \mathcal{M}(F_{s'}^{(n)}, F_{s'}^{(n-1)}) \\ &+ \log\left(\frac{1}{S'} \sum_{s'=1}^{S'} \mathcal{M}(F_{s'}^{(n)}, F_{s'}^{(n-1)})\right) \end{aligned} \quad (11)$$

where S' is the number of sampling trials after data augmentation. Then, the objective function of MIMC is given as

$$\operatorname{argmax}:\mathcal{L}_{MIMC} = \mathcal{I}(F^{(n)}, F^{(n-1)}) \quad (12)$$

The overall objective function of iCTS on continuous sessions is defined as follows:

$$\operatorname{argmin}:\mathcal{L} = \mathcal{L}_{CE} - \alpha \mathcal{L}_{MIMC} \quad (13)$$

where α is the trade-off coefficient.

3.4.2 Incremental BLS sequence

To further improve the continuous learning and self-updating ability of our method for session-incremental MI-EEG, an adaptive incremental BLS sequence (iBLS) is built by adding a few extra nodes to bBLS. The illustration of iBLS is given in Fig. 5.

When the temporal-spatial features of the n -th session $F^{(n)}$ are sequentially incorporated, according to Eqs. (2) and (4), the mapped features $\hat{Z}_{k*}^{(n)}$ and enhancement features $\hat{H}_{m*}^{(n)}$ on the trained iBLS by the $(n-1)$ -th session are calculated by $\hat{Z}_{k*}^{(n)} = \ell(F^{(n)} \hat{W}_{Z_{k*}}^{(n-1)} + \hat{\beta}_{Z_{k*}}^{(n-1)})$, $k^* = K+1, \dots, K^*$ and $\hat{H}_{m*}^{(n)} = \xi([\hat{Z}_1^{(n)}, \hat{Z}_2^{(n)}, \dots, \hat{Z}_{K^*}^{(n)}] \hat{W}_{H_{m*}}^{(n-1)} + \hat{\beta}_{H_{m*}}^{(n-1)})$, $m^* = M+1, \dots, M^*$, and the output can be obtained by

$$\hat{y}^{(n)} = \hat{A}^{(n)} \hat{W}_O^{(n-1)} = \hat{A}^{(n)} (\hat{A}^{(n-1)})^T \hat{A}^{(n)} + \mu I)^{-1} (\hat{A}^{(n-1)})^T Y^{(n)} \quad (14)$$

If the model fails to converge completely, new mapping and enhancement nodes are added. For the K^*+1 group feature mapping nodes, the mapped feature and the corresponding enhancement nodes are randomly generated as

$$Z_{K^*+1}^{(n)} = \ell(F^{(n)} \hat{W}_{Z_{K^*+1}}^{(n)} + \hat{\beta}_{Z_{K^*+1}}^{(n)}) \quad (15)$$

$$H_{M^*+G}^{(n)} = [\xi(Z_{K^*+1}^{(n)} \hat{W}_{H_{M^*+G}}^{(n)} + \hat{\beta}_{H_{M^*+G}}^{(n)}), \dots, \xi(Z_{K^*+1}^{(n)} \hat{W}_{H_{M^*+G}}^{(n)} + \hat{\beta}_{H_{M^*+G}}^{(n)})] \quad (16)$$

Then, denote $A^{(n)} \triangleq [\hat{A}^{(n)} | Z_{K^*+1}^{(n)} | H_{M^*+G}^{(n)}]$, the pseudoinverse matrix $(A^{(n)})^+$ is updated as

$$(A^{(n)})^+ = \begin{bmatrix} \hat{A}^{(n)} - G \mathcal{V}^T \\ \mathcal{V}^T \end{bmatrix} \quad (17)$$

where $G = (\hat{A}^{(n)})^+ [Z_{K^*+1}^{(n)} | H_{M^*+G}^{(n)}]$,

$$\mathcal{V}^T = \begin{cases} ([Z_{K^*+1}^{(n)} | H_{M^*+G}^{(n)}] - \hat{A}^{(n)} G)^+ & \text{if } [Z_{K^*+1}^{(n)} | H_{M^*+G}^{(n)}] - \hat{A}^{(n)} G \neq 0 \\ (1 + G^T G)^{-1} G^T (\hat{A}^{(n)})^+ & \text{if } [Z_{K^*+1}^{(n)} | H_{M^*+G}^{(n)}] - \hat{A}^{(n)} G = 0 \end{cases} \quad (18)$$

thus, the output weight $W_O^{(n-1)}$ is updated as $W_O^{(n)}$:

$$\hat{W}_O^{(n)} = \begin{bmatrix} \hat{W}_O^{(n-1)} - G \mathcal{V}^T Y^{(n)} \\ \mathcal{V}^T Y^{(n)} \end{bmatrix} \quad (19)$$

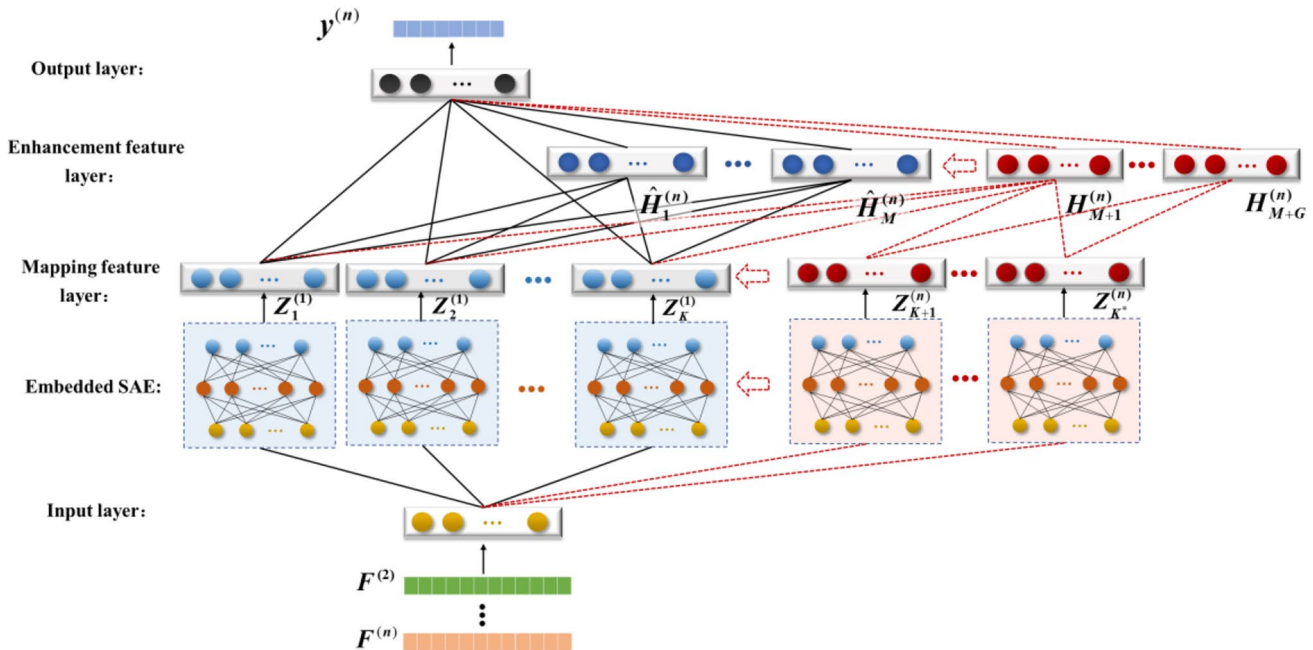


Fig. 5 Illustration of iBLS, where the temporal-spatial features $F^{(2)}, \dots, F^{(n)}$ are from the 2nd to n th sessions, $Z_{K+1}^{(n)}, \dots, Z_{K^*}^{(n)}$ are the additional mapping features generated by incremental feature map-

ping nodes, and $H_{M+1}^{(n)}, \dots, H_{M+G}^{(n)}$ are the enhancement features from incremental feature enhancement nodes

If the model is still unconverging, repeat Eq. (14) to Eq. (18) until the convergence condition is satisfied. In particular, only the pseudo-inverse of the newly mapped and enhanced features needs to be computed, the trained iBLS without retraining. The number of new nodes is adaptively determined by GS as well.

The modeling process on the incremental sessions is shown in Algorithm 2.

Algorithm 2 Modeling on the incremental sessions

Input: Incremental session dataset $\{X^{(n)}, Y^{(n)}\}_{n=2}^N$, temporal-spatial feature $F^{(1)}$, initial the group number of incremental nodes K^* , M^* , G and search steps.

Output: $W_z^{(N)}, W_h^{(N)}, W_o^{(N)}$ and $F^{(N)}$

```

1  for  $n=2, \dots, N$  do
    iCTS:
2    Calculate  $\mathcal{L}_{MMC}$  by Eqs.(11),(12);
3    Update  $W^{(n)}$  by Eq.(13);
4    Get temporal-spatial feature  $F^{(n)}$ ;
    iBLS:
5    Search incremental  $K^*$ ,  $M^*$  by GS;
6    Randomly generate new  $\hat{W}_z^{(n)}$ ,  $\hat{\beta}_z^{(n)}$ ;
7    Calculate  $Z^{(n)} = [Z_{K+1}^{(n)}, Z_{K+2}^{(n)}, \dots, Z_{K^*}^{(n)}, Z_{K^*+1}^{(n)}]$  by Eqs.(15);
8    Randomly generate new  $\hat{W}_h^{(n)}$ ,  $\hat{\beta}_h^{(n)}$ ;
9    Calculate  $H^{(n)} = [H_{M+1}^{(n)}, H_{M+2}^{(n)}, \dots, H_{M^*}^{(n)}, H_{M^*+G}^{(n)}]$  by
    Eq.(16);
10   Set  $A^{(n)} \triangleq [\hat{A}^{(n)} [Z_{K^*+1}^{(n)} H_{M^*+G}^{(n)}]]$ , calculate  $\hat{W}_o^{(n)}$  by
    Eqs.(17), (18) and (19);
11   While The validation error is not satisfied do
12     Add mapping and enhancement nodes and repeat
    5-10;
13   end
14 end

```

4 Experiments and results

4.1 Experimental setup

All experiments are carried out under the workstation, with 3.60 GHz Intel(R) Core i7—16 GB RAM NVIDIA T1000 GPU, CUDA 10.2 and cuDNN 7605. The model simulation will use Python 3.9 and Pytorch 1.10 in Spyder 5.1.5.

4.1.1 Parameter settings

In the evaluation experiments, the details of the key parameter settings are listed in Table 3. In particular, the number

of temporal feature filter D_1 and spatial feature filter D_2 are determined by repeatable experiments; more details are given in Section 4.5.

4.1.2 Evaluation approaches

In this paper, the effectiveness and plasticity of the proposed method are validated by fivefold cross-validation, as well as the stability evaluation employs the hold-out method (H–O) with a division ratio of 80%:20% (4:1), which ensures that the test datasets from previous sessions have not been involved in training iBLS.

Evaluation metrics include the test accuracy (Acc), standard deviation (Std), Cohen's Kappa coefficient (Kc), and forgetting rate (Fr) [59]. Acc is used to measure the recognition ability of the model for diverse tasks. Comparison of Std can deduce the generalization of the model for different individuals. Kc is commonly applied to evaluate the consistency between given labels and predictions, and it can be calculated by the confusion matrix:

$$Kc = \frac{P_o - P_e}{1 - P_e} \quad (20)$$

where P_o is the actual consistency probability and P_e is the randomized consistency probability. Fr is used to describe the memory of the updated model by new data to previous data. Suppose that the total of N sessions is given, Fr of the updated model for the n th session is summarized as [59]

$$Fr = \frac{Acc_n - Acc_{N,n}^*}{Acc_n} \times 100\% \quad (21)$$

where $n \in [1, N]$ and $Acc_{N,n}^*$ is the test accuracy of the model, which was trained on the N th new session, for the n th previous session. Acc_n is the test accuracy on the n th independent session. A lower Fr means that the model can maintain its performance for the old sessions after incorporating the new sessions. In contrast, a larger value is known as the catastrophic forgetting phenomenon.

4.2 Effectiveness evaluation on the initial session

A reliable baseline model is critical for continuously learning session-incremental MI-EEG. The effectiveness of the models is first evaluated on the initial session, specifically, the fivefold cross-validation results of ASiBLS are compared with the related methods, BLS [22] and EEGNet 8,2 (includes eight temporal filters and two spatial filters) [16], as listed in Tables 4 and 5.

Table 3 Details of the key parameter settings

Components	Parameters	Settings
Preprocessing	I_w	1 s
	I_s	0.2 s
	$\mathcal{N}$	11
CTS	k_1	250/4
	k_2	250/16
	D_1	8
	D_2	2
	D_3	8×2
	Dropout rate	0.5
	Epochmax	50
	Batch size	64
bBLS	No. of mapping nodes (a group)	[50,100]
	No. of enhancement nodes (a group)	[50,1000]
	Search step of K	5
	Search step of M	50
	K	[1, 10]
	M	1
	γ and μ	10^{-8}
	$\ell(\cdot)$	Linear
iCTS	$\xi(\cdot)$	Relu
iBLS	α	0.8
	K^* and M^*	[1, 10]

4.2.1 Evaluation on IV-2a

For the comparison results on IV-2a, as shown in Table 4, ASiBLS achieves an average *Acc.* of 79.28%, which is superior to BLS and EEGNet 8,2, especially for BLS. We further employ *Std.* to evaluate the inter-subject generalizability of our method; an average *Std.* of 6.31 is higher than BLS and EEGNet 8,2, which means that ASiBLS has

stronger generalization performance for individual variability. Meanwhile, ASiBLS obtains an average *Kc* of 0.72, which holds superior consistency for the predictions calculated and the actual value. In addition, the total training time of nine subjects and the test time of a sampling trial are also introduced to evaluate the model's computational complexity. Although ASiBLS requires more total time, it inherits the advantages of BLS and EEGNet 8,2, i.e., the decoding ability is significantly improved while a training time of 1028.79 s less than the combined time of BLS and EEGNet 8,2 is obtained.

4.2.2 Evaluation on IV-2b

To evaluate the generalizability of our method on different datasets, we again compare the evaluation results of ASiBLS with BLS, EEGNet 8,2 on the initial session of IV-2b, as summarized in Table 5. ASiBLS shows an average *Acc.* of 85.6%, *Std.* of 7.09%, and *Kc.* of 0.81; the decoding ability of ASiBLS to MI tasks is significantly improved than BLS and EEGNet 8,2. Although the rapidly updated BLS reveals obvious advancement in training and testing compute costs, its decoding ability and inter-subject generalizability need to be improved. To summarize, the effectiveness of ASiBLS in classifying MI-EEG is significantly enhanced than the related methods.

4.3 Plasticity evaluation

To evaluate the plasticity of the proposed methods, we compare the decoding accuracies of BLS, EEGNet 8,2, and ASiBLS for new sessions in both static and dynamic training scenarios. In the static training scenario, the model is independently trained and inferred independently

Table 4 Comparison of ASiBLS and BLS/ EEGNet 8,2 on the initial session of IV-2a

Subjects	BLS		EEGNet 8,2		ASiBLS	
	<i>Acc.</i> (%)	<i>Kc</i>	<i>Acc.</i> (%)	<i>Kc</i>	<i>Acc.</i> (%)	<i>Kc</i>
A1	73.78	0.6504	78.62	0.7149	83.47	0.7796
A2	56.07	0.4143	57.37	0.4316	68.73	0.5831
A3	69.12	0.5883	84.34	0.7912	86.29	0.8172
A4	68.16	0.5754	61.46	0.4861	75.30	0.6707
A5	63.09	0.5078	74.12	0.6550	79.73	0.7297
A6	72.40	0.6321	68.23	0.5763	70.64	0.6086
A7	54.07	0.3876	76.30	0.6840	81.29	0.7505
A8	63.94	0.5192	82.75	0.7700	83.18	0.7757
A9	57.26	0.4301	79.27	0.7236	84.85	0.7979
Avg. (%)	64.21	0.5228	73.61	0.6481	79.28	0.7237
<i>Std.</i> (%)	7.22	—	9.38	—	6.31	—
Training time (s)	411.56		926.55		1028.79	
Test time/trial (s)	0.831×10^{-3}		1.07×10^{-3}		4.80×10^{-3}	

Table 5 Comparison of ASiBLS and BLS/ EEGNet 8,2 on the initial session of IV-2b

Subjects	BLS		EEGNet 8,2		ASiBLS	
	Acc. (%)	Kc	Acc. (%)	Kc	Acc. (%)	Kc
B1	67.86	0.57	85.52	0.81	92.19	0.90
B2	54.98	0.40	63.13	0.51	72.77	0.64
B3	73.79	0.65	86.74	0.82	78.81	0.72
B4	75.38	0.67	86.75	0.82	91.36	0.88
B5	69.06	0.59	79.27	0.72	85.38	0.81
B6	70.86	0.61	82.16	0.76	89.42	0.86
B7	57.84	0.44	75.57	0.67	92.26	0.90
B8	59.96	0.47	68.91	0.59	79.06	0.72
B9	62.68	0.50	84.96	0.80	89.12	0.85
Avg. (%)	65.82	0.54	79.22	0.72	85.60	0.81
Std (%)	7.25	–	8.46	–	7.09	–
Training time (s)	379.08		534.29		710.55	
Test time/trial (s)	0.21×10^{-3}		0.84×10^{-3}		1.06×10^{-3}	

for each new session. Conversely, the dynamic training scenario involves the model trained on previous sessions, then updating it to infer subsequent sessions. The remaining sections follow similar principles. Comparison results on IV-2a with multiple MI tasks and IV-2b with multiple sessions are illustrated in Figs. 6 and 7, respectively.

4.3.1 Evaluation on IV-2a

In Fig. 6, the average *Acc.* of static ASiBLS only obtains 78.36%, while dynamic ASiBLS reaches 80.5%, which is increased by 2.14%. However, the continuous learning ability of BLS is limited, with the average accuracies of static and dynamic BLS being 62.57% and 63.19%, respectively, showing only a 0.62% improvement in accuracy. Dynamic EEGNet 8,2, which involves retraining on each new session, exhibits classification accuracy comparable to that of static EEGNet 8,2. Statistical results demonstrate that ASiBLS shows superior classification performance and continuous learning ability for new session data.

The overall classification performance of ASiBLS is given by calculating the average results of the initial and incremental sessions, where the average *Acc.* is 79.89%, the average *Std.* is 6.58, and *Kc.* is 0.73.

4.3.2 Evaluation on IV-2b

To evaluate the model's plasticity across multiple incremental sessions, the average accuracies of static and dynamic ASiBLS are compared in Fig. 7. The average results of dynamic ASiBLS are 1.27%, 0.74%, 1.60%, and 0.99% higher than static ASiBLS on the 2nd, 3rd, 4th, and 5th sessions, respectively. These results suggest that ASiBLS in

dynamically updating scenarios consistently outperforms it in static training scenarios for multiple sequential sessions. In other words, the innovatively introduced MIMC (relative to BLS) boosts the classification performance of ASiBLS by increasing the probability distribution of features between sessions.

For the overall classification performance of ASiBLS on the five continuous sessions, the average *Acc.* reaches 87.19%, with an average *Std.* of 7.24 and an average *Kc.* of 0.72.

4.4 Stability evaluation

Stability describes the method's ability to retain the memory for previous sessions after being updated by new sessions. The forgetting rates of ASiBLS are compared with BLS on IV-2a and IV-2b to evaluate the method's stability, as shown in Table 6 and Fig. 8.

4.4.1 Evaluation on IV-2a

For IV-2a with multiple MI tasks, in Table 6, the static *Acc.* refers to the test accuracy obtained after the model is independently trained on the initial session, and the dynamic *Acc.* is the tested accuracy of the model, which is updated by the 2nd session, for the initial session. Dynamic BLS and ASiBLS show 3.1% and 1.24% reductions in average accuracies than static BLS and ASiBLS, respectively. The average *Fr* of ASiBLS is lower than BLS by 3.32%. Although continuous learning weakens the model's generalization between subjects, as indicated by the standard deviations of 7.24 and 7.04, ASiBLS shows better robustness to inter-individual variability than the traditional BLS. Evaluation results reveal that ASiBLS improves the stability of the BLS-based method.

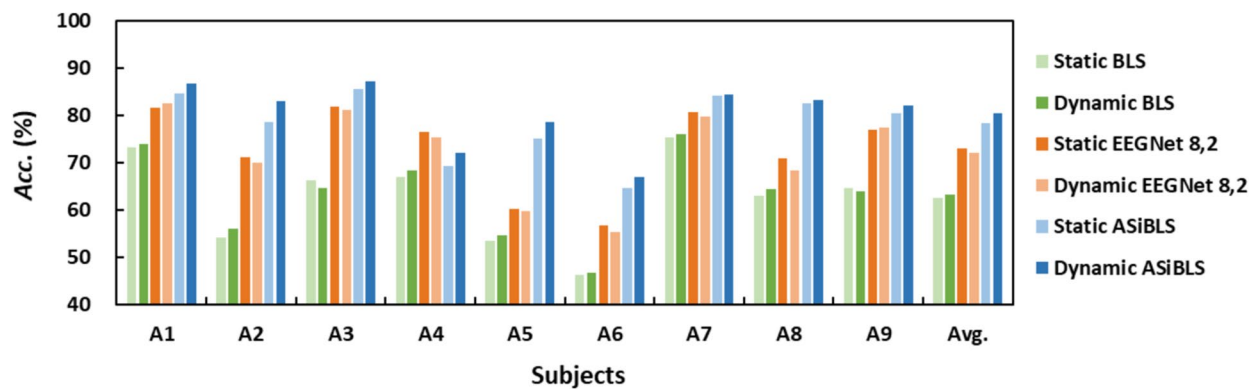


Fig. 6 Comparison of the accuracies of static and dynamic BLS, EEGNet 8,2, and ASiBLS on the 2nd session of IV-2a, where the static model shows that it performed training and inference indepen-

dently on the 2nd session, and the dynamic model means it is trained on the 1st session and then updated and tested on the 2nd session

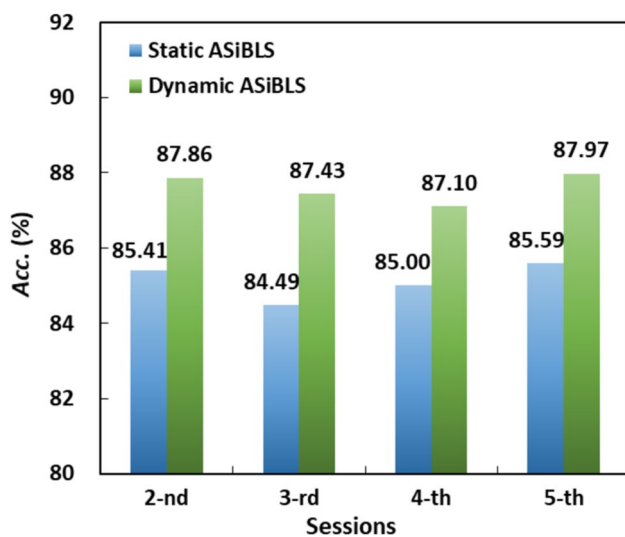


Fig. 7 Comparison of the average accuracies of static and dynamic ASiBLS on the multiple incremental sessions of IV-2b

4.4.2 Evaluation on IV-2b

For the IV-2b dataset, the *Fr* of the updated ASiBLS on each incremental session for the initial session is first given in Fig. 8a. The *Fr* of BLS exhibits a substantial increase with the progressive addition of new sessions, while the trend of increasing *Fr* in ASiBLS is slower. The lower *Fr* means that ASiBLS has a strong memory for old knowledge. Figure 8b–d are the *Fr* for the 2nd, 3rd, and 4th sessions, respectively. Although the variation of *Fr* of BLS is becoming slower when the newly incorporated sessions gradually decrease, it is still higher than ASiBLS. Comparison results prove that ASiBLS can alleviate the forgetting of old knowledge, especially since the advantage is more noticeable for multiple sequential sessions.

4.5 Impact of CTS

To explore the impact of the feature extractor on model performance, we compare ASiBLS without CTS, with CTS, and EEGNet as the feature extractor on the initial session; comparison results are presented in Table 7. ASiBLS without CTS, i.e., the filtered data is directly fed into BLS, shows the worst performance among the three feature extractor settings, and the average *Acc.* is only 64.21%. Moreover, only a *Std.* of 7.22% and a *Kc* of 0.52 mean that the generalization and consistency of ASiBLS without a CTS setting are also poor. When EEGNet is used as the feature extractor, the number of temporal feature filters D_1 is preferably set to 8 rather than 4. A large number of temporal filters can improve the feature extraction ability of the feature extractor. When D_1 and D_2 are set to 8 and 2, respectively, ASiBLS with CTS achieves the best classification performance among all feature extractor configurations. Overview, CTS is the most applicable and efficient method for extracting the spatial–temporal features from time-varying and multi-channel correlated MI-EEG.

4.6 Impact of MIMC

To further evaluate the contribution of MIMC, the *Acc.* and *Fr* of ASiBLS with and without MIMC for the 1st session and 2nd session are compared in Fig. 9, respectively. ASiBLS with MIMC outperforms ASiBLS without MIMC in terms of higher accuracy and slower forgetting for previous sessions. MIMC helps improve ASiBLS' continuous learning ability.

Furthermore, the two-dimensional probability distributions of the temporal–spatial features are visualized using t-distributed stochastic neighbor embedding (t-SNE), a common method for high-dimensional data visualization.

Table 6 Comparison of the test accuracies (%) and forgetting rates (%) of BLS and ASiBLS for the initial session of IV-2a, where “Static Acc.” is the test accuracy obtained after the model is independently trained on the initial session

Subjects	BLS			ASiBLS		
	Static Acc	Dynamic Acc	Fr	Static Acc	Dynamic Acc	Fr
A1	73.78	71.35	3.30	83.47	82.91	0.67
A2	56.07	53.69	4.25	68.73	66.33	3.49
A3	69.12	67.21	2.77	86.29	85.85	0.51
A4	68.16	64.99	4.65	75.30	74.28	1.35
A5	63.09	59.36	5.91	79.73	77.93	2.25
A6	72.40	68.36	5.59	70.64	67.98	3.77
A7	54.07	51.11	5.47	81.29	81.04	0.30
A8	63.94	59.87	6.36	83.18	82.37	0.97
A9	57.26	54.04	5.62	84.85	83.67	1.39
Avg	64.21	61.11	4.88	79.28	78.04	1.56
Std	7.22	7.24	1.23	6.31	7.04	1.27

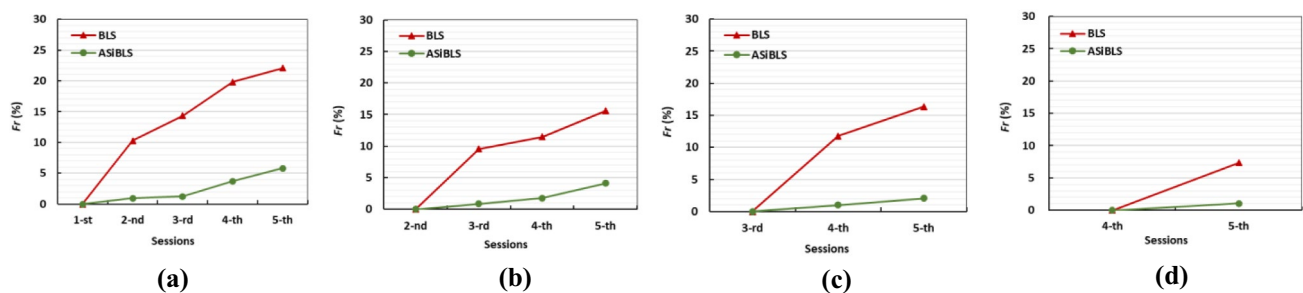


Fig. 8 Comparison of the average forgetting rates of BLS and ASiBLS for each previous session of IV-2b, where the x-axis represents the model being updated by the current session, and (a), (b), (c), and (d) show the forgetting rates for the 1st, 2nd, 3rd, and 4th sessions, respectively

The temporal–spatial feature distributions for subjects A1 and A6, who have the highest and lowest average accuracy, respectively, are shown in Fig. 10. CTS with MIMC demonstrates a greater overlap in feature distributions between the 2nd session and the initial session for both subjects. In contrast, the features extracted by CTS without MIMC are more dispersed. MIMC makes vital contributions to improving similarity in the probability distribution of features.

In addition, in Fig. 11, the number of node groups in feature mapping and enhancement layers is compared between ASiBLS with and without MIMC. It is shown that ASiBLS with MIMC requires fewer incremental nodes than ASiBLS without MIMC. This reduction in additional nodes indicates an enhancement in continuous learning ability and a decrease in computational complexity.

4.7 Comparison with the related methods

To evaluate the incremental learning abilities of ASiBLS, its average results across all continuous sessions are compared with several related methods. Comparison results on IV-2a and IV-2b are presented in Tables 8 and 9, respectively.

Table 7 Impact of CTS on the model performance, where D_1 and D_2 represent the number of temporal and spatial feature filters, respectively

Feature extractor configurations	D_1	D_2	Acc. (%)	Std. (%)	Kc
w/o CTS	—	—	64.21	7.22	0.52
EEGNet	4	2	72.96	9.71	0.66
	8	2	73.61	9.38	0.65
w/ CTS	4	2	74.37	7.28	0.64
	8	2	79.28	6.31	0.72
	16	10	78.42	6.95	0.72

Note: The black font is the best result

4.7.1 Comparison on IV-2a

In Table 8, the basic deep and broad networks BLS [22] and EEGNet 8,2 [16] achieve recognition accuracies of only 63.7 and 72.89%, respectively. MN-BLS [60], a BLS-based MI classification method, reaches only 63.15%. Cross-session transfer learning strategies noticeably boost the ability of DA-EEGNet [32] and GSAN [61] to decode MI tasks. Feature optimization method TWSB [33] and an improved 1D convolutional neural network EEG-FMCNN [34] are helpful

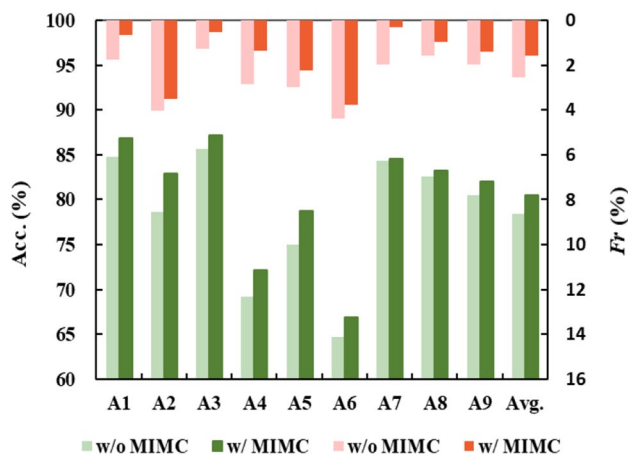
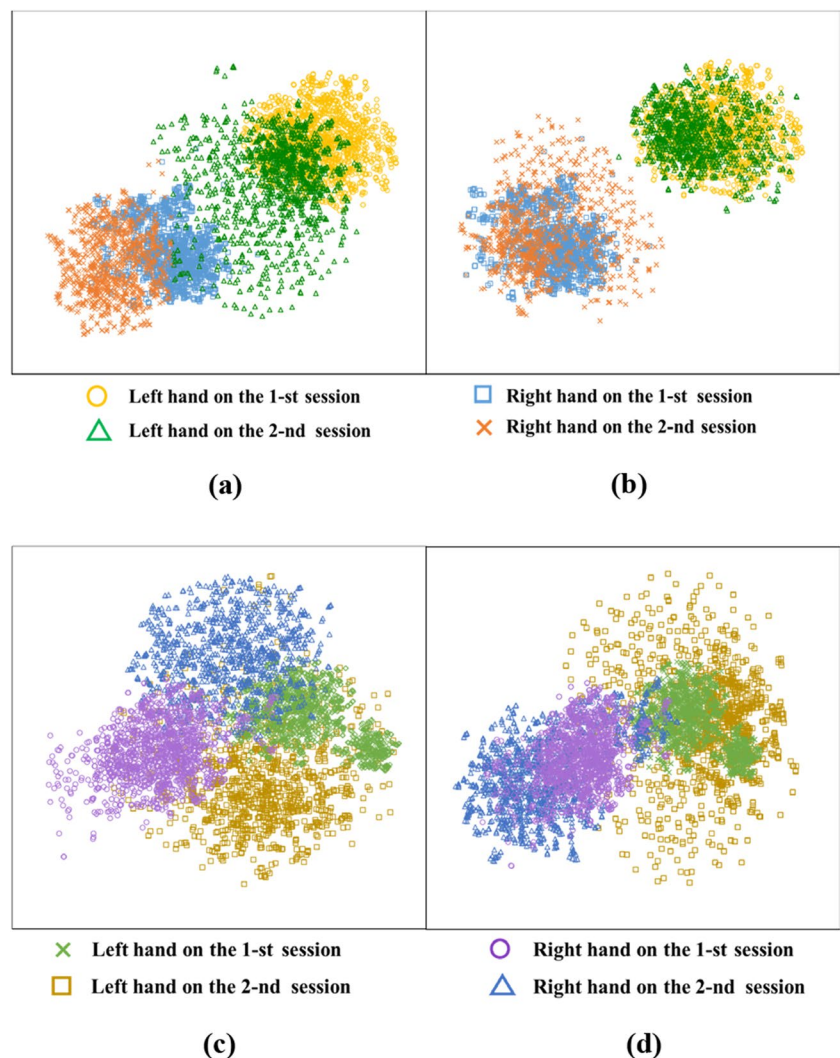


Fig. 9 Comparison of the test accuracies of ASiBLS with and without MIMC on the 2nd session of IV-2a (on the primary y-axis), and the forgetting rates for the 1st session (on the secondary y-axis)

Fig. 10 Visualization of the probability distributions of the temporal-spatial features for A1 and A6 on the 2nd session, where (a) ASiBLS without MIMC for A1, (b) ASiBLS with MIMC for A1, (c) ASiBLS without MIMC for A6, and (d) ASiBLS with MIMC for A6



for MI-EEG classification, but they still cannot satisfy the requirements of continuous decoding. M-FANet [35], IFB-CLNet [21], and TCN [53] combine the temporal and spatial properties of MI-EEG to achieve the accuracies of 79.28%, 78.79%, 79.26%, and 72.89%, respectively. ASiBLS obtains an average accuracy of 79.89% which is better than the related methods, especially for the BLS-based methods, a 16.19% improvement in accuracy over BLS. ASiBLS exhibits significantly superior inter-subject adaptability compared to other methods, except for BLS, as evidenced by the comparison of standard deviations. The higher *Acc.* and lower *Std.* of our method indicate an enhanced ability to accurately recognize MI tasks with incremental sessions, coupled with robust performance across different subjects.

4.7.2 Comparison on IV-2b

ASiBLS is further compared with several related methods on IV-2b, and the comparison results are shown in Table 9. SCW-CNN [62] introduced the Shannon complex wavelet

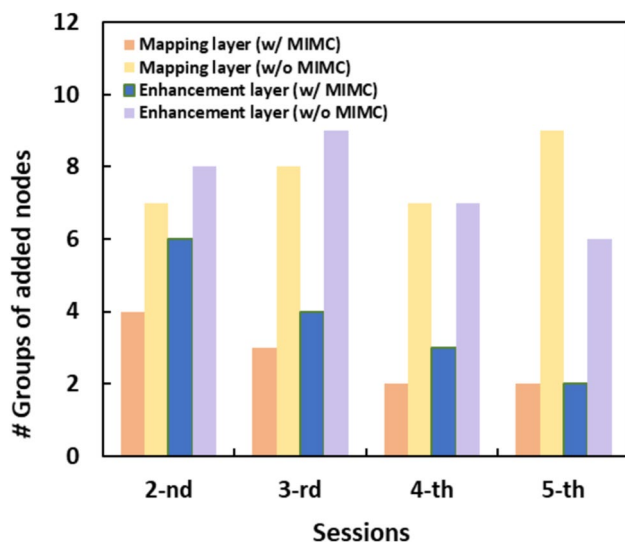


Fig. 11 Comparison of the number of node groups in the feature mapping and enhancement layers of ASiBLS with and without MIMC

to calculate the time–frequency matrix and then input it into CNN; the average classification *Acc.* is 85.2%. Based on the temporal–spatial domain features, MBGA-Net [2], MSHCNN [15], RGDDANet [54], and TCN [53] achieve accuracies of 86.99%, 85.25%, 85.19%, and 85.9%, respectively. MI-DABAN [63], a high-performance domain adversarial network designed, is proposed to reduce inter-subject variation and enhance classification accuracy in MI-EEG, achieving an average *Acc.* of 84.83%. Additionally, despite ASiBLS having a *Std.* of 7.24% and being surpassed by MBGA-Net, which does not support cross-session learning, our work still significantly outperforms other methods.

5 Discussion

5.1 Recognition performance

A high-performance decoding model is considered to have excellent recognition capabilities for MI tasks. ASiBLS's

Table 8 Comparison of ASiBLS with the related methods on IV-2a

Subjects	BLS	EEGNet 8,2	MN-BLS	DA-EEGNet	GSAN	TWSB	EEG-FMCNN	M-FANet	IFBCLNet	TCN	ASiBLS
A1	73.86	80.62	75.94	88.54	87.60	89.3	82.29	86.81	87.18	83.19	85.17
A2	56.07	63.69	41.67	69.09	59.76	66.9	53.47	75	58.65	69.97	75.83
A3	66.88	82.76	73.54	97.22	90.87	89.3	93.06	91.67	92.67	93.44	86.71
A4	68.26	68.36	59.72	72.20	80.38	69.3	72.22	73.61	78.07	70.97	73.71
A5	58.83	66.94	39.86	57.98	57.71	74.1	79.17	76.39	70.65	79.31	79.20
A6	59.60	61.72	43.02	59.72	51.04	60.1	71.18	61.46	60.46	67.28	68.78
A7	65.10	78.03	72.12	93.06	92.01	89.4	87.85	85.76	92.41	81.22	82.91
A8	64.15	75.52	81.98	86.80	87.08	88	83.68	75.69	82.28	84.66	83.20
A9	60.56	78.39	80.49	90.27	87.92	85.6	86.46	87.15	86.74	83.33	83.46
Avg	63.70	72.89	63.15	79.43	77.15	79.11	78.82	79.28	78.79	79.26	79.89
Std	5.53	7.79	17.42	14.85	16.22	11.55	11.85	9.38	12.91	8.40	6.58
Year	2018	2018	2019	2023	2024	2024	2024	2024	2024	2024	2024

Table 9 Comparison ASiBLS with the related methods on IV-2b

Subjects	SCW-CNN	MBGA-Net	MI-DABAN	MSHCNN	TWSB	RGDDANet	TCN	ASiBLS
B1	78.1	85.41	84.03	86.8	80.3	81.35	–	93.27
B2	64.3	79.55	64.43	77.94	70.7	61.07	–	72.54
B3	64.7	88.63	62.5	65.97	64.8	81.15	–	87.05
B4	97.2	91.75	98.61	97.97	95.3	97.5	–	92.78
B5	97.8	87.59	94.44	93.24	89.2	93.44	–	84.86
B6	93.1	83.96	88.19	88.88	87.4	90.94	–	90.75
B7	85.9	96.74	86.81	86.8	90.1	80.31	–	93.19
B8	92.5	94.37	94.79	82.89	89.5	89.79	–	79.08
B9	92.8	75.13	90.63	86.8	90.9	91.15	–	91.21
Avg. (%)	85.2	86.99	86.83	85.25	83.10	85.19	85.90	87.19
Std. (%)	13.15	6.52	12.97	9.20	10.25	10.87	–	7.24
Year	2022	2023	2023	2023	2024	2024	2024	2024

classification accuracy for the initial session MI-EEG has been validated on IV-2a and IV-2b. The recognition performance of classical BLS [22], EEGNet 8,2 [16], and ASiBLS are first compared in Tables 4 and 5. Student's *t*-test (*p*-value) is introduced to compare the superiority of our method. The average recognition accuracies of classical BLS and EEGNet 8,2 are lower than ASiBLS by 15.07% ($p < 0.01$) and 5.67% ($p < 0.05$) on IV-2a and by 19.78% ($p < 0.01$) and 6.38% ($p < 0.05$) on IV-2b, respectively. Our method significantly outperforms the related methods. The improved recognition accuracies support the reliability of our method and highlight its greater application potential. Moreover, the comparison results of Cohen's Kappa coefficients further validate the reliability of method. The standard deviations of BLS and EEGNet 8,2 are higher than ASiBLS by 0.91% and 3.07% on IV-2a and by 0.16% and 1.37% on IV-2b, respectively. The higher standard deviations suggest that the model is unstable for individuals with marked physiological differences; in other words, its generalization ability needs to be further promoted. In our work, the designed CTS, which serves as a feature extractor, sequentially captures the local temporal and spatial features of inputs, where the spatial features involve both dimensions in the EEG channel and the feature map channel. Unlike EEGNet, the novel CTS uses Conv2D instead of the separable convolutions; the spatial information can be learned from two dimensions simultaneously on spatial feature representation and achieves better recognition accuracy than EEGNet 8,2, as shown in Fig. 6. More importantly, the combined BLS further maps temporal and spatial features linearly and non-linearly from a global perspective, i.e., analyzing all input features in parallel. When the signals are directly inputted into BLS, only an $Acc. \pm Std.$ of $64.21\% \pm 7.22\%$ proves that its flat structure cannot characterize temporal and spatial properties in detail for high-dimensional MI-EEG. Combining CTS and BLS, which integrates the advantages of deep and broad learning, significantly improves the recognition ability of existing methods for high-dimensional, time-varying MI-EEG. It is noteworthy that the reliability of the incremental learning methods on the initial session (or pre-training phase) is crucial for accurately inferring sequential incremental tasks.

5.2 Continuous learning and self-updating capability

ASiBLS demonstrates superb plasticity in adapting to sequentially incorporated MI-EEG with the session through dynamic structural remodeling and parameter updating. Figures 6 and 7 suggest the comparison results for the continuous learning capabilities of BLS, EEGNet 8,2, and ASiBLS by introducing static and dynamic training mechanisms. As shown by the average *Acc.* in Fig. 6, the dynamic BLS and ASiBLS outperform the static BLS

and ASiBLS by 0.62% and 2.14%, respectively, and the static and dynamic EEGNet 8,2 are generally equal. Traditional EEGNet lacks incremental learning ability, so it needs to be retrained on each new session. Although BLS provides dynamic updating and broad expansion strategies for continuous learning, the flat structure is still a key limitation for learning new knowledge, resulting in the dynamic BLS showing a weaker *Acc.* improvement than the static BLS. Furthermore, for MI-EEG from multiple sessions of IV-2b, the dynamic ASiBLS outperforms the static ASiBLS for all sessions, and our work inherits the learning experience from previous sessions to improve the generalization of new knowledge.

According to the mutual information theory and incremental BLS, we first introduce MIMC into CTS to obtain iCTS, and then combine it with iBLS. Specifically, MIMC is constructed in Eq. (11), which is intended to maximize the probability distribution similarity of feature representations between adjacent sessions. This reasoning further is confirmed by the visualization of the feature space distribution in Fig. 10. Additionally, MIMC motivates iCTS to converge on new sessions without performing a thorough parameter update or adding exponentially more parameters. Figure 11 illustrates a comparison of the number of node groups in ASiBLS with and without MIMC. The model without MIMC requires significantly more additional nodes, which directly translates to higher computational costs. However, more parameters do not contribute to the recognition *Acc.*, as shown in Fig. 9, and the average *Acc.* of ASiBLS without MIMC is 2.14% lower than it with MIMC. Adding the new session data, which has a similar feature distribution as the old session MI-EEG, is equivalent to increasing the data size, thus the generalization of ASiBLS is improved. Consequently, MIMC is essential for facilitating continuous learning of ASiBLS and iBLS possesses a self-updating mechanism so that the performance of ASiBLS in learning new session EEG is guaranteed.

5.3 Memory retention in continuous learning

The human brain retains previously acquired knowledge even after learning new information, and an ideal neural network should exhibit similar behavior, effectively preventing catastrophic forgetting of prior knowledge. Table 6 and Fig. 8 summarize the accuracies of BLS and ASiBLS under dynamic and static training mechanisms, respectively, and their forgetting rates. The average *Fr* of BLS is 4.88%, while that of ASiBLS is just 1.56%. Although BLS has continuous learning capabilities, new parameters added compromise its ability to retain and adapt to prior knowledge. The mutual information theory is introduced into the loss function of CST as an MIMC term, constraining the parameter updates of CTS across sessions. Figure 9 compares the forgetting

rates of ASiBLS with and without MIMC for the initial session; the average *Fr* of the model with MIMC is 20% lower than without MIMC. More similar feature probability distributions mean that model parameters need only minimal revisions, so mitigating the forgetting of old knowledge by CTS updates.

5.4 Superiority to related methods

The decoding performance of ASiBLS with related methods is compared in Tables 8 and 9. For the IV-2a dataset, ASiBLS significantly outperforms BLS, EEGNet, MN-BLS [60], and GSAN [61] regarding recognition accuracies ($p < 0.05$). Although there are no significant differences between ASiBLS and DA-EEGNet [32] ($p < 0.05$), the average *Std.* of ASiBLS is 8.27% lower than it. Compared to the advanced TWSB, EEG-FMCNN [34], M-FANet [35], IFBCNet [21], and TCN [53], our work also improves the recognition accuracies by 0.78%, 1.07%, 0.61%, 1.1%, and 0.63%, respectively, and the standard deviations also have significant benefits. Similarly, the decoding performance of ASiBLS is also evaluated on the IV-2b dataset, as shown in Table 5. Compared to the benchmark BLS and EEGNet 8.2, ASiBLS improves the average accuracies by 19.78% and 6.38%, respectively, and reduces the standard deviations by 0.16% and 1.37%, respectively. The decoding performance of ASiBLS is significantly superior to the benchmark methods ($p < 0.05$). Furthermore, Table 9 gives a comparison of our method with advanced related methods. Although the significance distribution of ASiBLS for SCW-CNN [62], MSHCNN [15], TWSB [33], and TCN [53], which are deep learning methods based on temporal–spatial feature extraction, is $0.1 < p < 0.4$, the average accuracies are improved by 1.99%, 0.7%, 1.94%, 4.09%, and 1.29%, respectively. For the transfer learning methods across sessions, the accuracies of ASiBLS are 0.2%, 0.36%, and 2% better than MBGA-Net [2], MI-DABAN [63], and RGDDANet [54]. The boosted recognition accuracies imply that our work has a better decoding capability than related methods, which is crucial. Meanwhile, the improvement in standard deviations shows that ASiBLS also can be robust to inter-individual differences.

5.5 Limitations and future work

In our study, all evaluation experiments are subject independent. Although the generalization of ASiBLS has been demonstrated by subject-independent experimental results, higher computational costs are inevitable as the model is retrained for each subject. Our work achieves a substantial improvement on IV-2b, covering multiple sessions, but it only investigates the binary MI task. For the IV-2a dataset

with multiple multi-category MI tasks, the decoding capability of ASiBLS needs to be further improved.

In the future, we will optimize the BLS-based incremental learning method by introducing more feature optimization technologies instead of SAE, such as sparse Bayesian learning and feature selection masking. Moreover, the proposed method should not just be limited to session-incremental learning scenarios in MI-BCI; a generalized task- or subject-incremental learning approach in different EEG-based studies is our goal to pursue soon. Meanwhile, we will further pursue the computational speed of the decoding model to advance its online learning ability and advance the value of BCI technology in clinical rehabilitation applications.

6 Conclusion

In this paper, we propose an adaptive ASiBLS method based on mutual information constraints and temporal–spatial convolutions to recognize incremental MI-EEG across sessions. Inspired by the compact EEGNet, we design a novel compact temporal–spatial convolutional block (CTS), consisting of a temporal convolution block and a spatial convolution block, to effectively extract the temporal and spatial features of high-dimensional MI-EEG. The spatial convolution block uniquely combines depthwise Conv2D with Conv2D to capture both EEG channel and filter channel characteristics. The introduced baseline BLS (bBLS) efficiently classifies MI tasks in the initial session and serves as the foundation model. For incremental MI-EEG across sessions, mutual information maximization constraints (MIMC) are incorporated into the loss function of CTS, forming iCTS, which reduces the computational complexity of the incremental BLS (iBLS) while enhancing the plasticity and stability of ASiBLS. The superior performance of ASiBLS is validated using the public BCI Competition IV-2a and IV-2b datasets. In future work, we aim to further develop incremental learning methods to better meet the real-world needs of neuro-rehabilitation and contribute to broader clinical applications of BCI.

Acknowledgements The research was financially supported by the National Natural Science Foundation of China (No. 62173010). We would like to thank the provider of the dataset and all the people who have given us helpful suggestions and advice.

Author contribution Yufei Yang performed the data analysis and wrote the manuscript; Mingai Li performed reviewed and revised the manuscript; Linlin Wang performed linguistic embellishments.

Data availability All datasets are publicly available.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Majdi H, Azarnoosh M, Ghoshuni M, Sabzevari VR (2024) Direct lingam and visibility graphs for analyzing brain connectivity in BCI. *Med Biol Eng Comput* 1–16. <https://doi.org/10.1007/s11517-024-03048-5>
- Ma W, Wang C, Sun X, Lin X, Niu L, Wang Y (2023) MBGA-Net: a multi-branch graph adaptive network for individualized motor imagery EEG classification. *Compu Methods Prog Biomed* 240. <https://doi.org/10.1016/j.cmpb.2023.107641>
- Lu B, Huang X, Chen J, Fu R, Wen G (2024) Manifold attention-enhanced multi-domain convolutional network for decoding motor imagery intention. *Knowledge-Based Syst* 296. <https://doi.org/10.1016/j.knosys.2024.111904>
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, ... Hadsell R (2017) Overcoming catastrophic forgetting in neural networks. *Proc Nat Acad Sci* 114(13), 3521–3526. <https://doi.org/10.1073/pnas.1611835114>
- Li Z, Hoiem D (2017) Learning without forgetting. *IEEE Trans Pattern Anal Mach Intell* 40(12):2935–2947. <https://doi.org/10.1109/TPAMI.2017.2773081>
- Tian S, Li L, Li W, Ran H, Ning X, Tiwari P (2024) A survey on few-shot class-incremental learning. *Neural Netw* 169:307–324. <https://doi.org/10.1016/j.neunet.2023.10.039>
- Tian S, Li W, Ning X, Ran H, Qin H, Tiwari P (2023) Continuous transfer of neural network representational similarity for incremental learning. *Neurocomputing* 545:126300. <https://doi.org/10.1016/j.neucom.2023.126300>
- Sun G, Ji B, Liang L, Chen M (2024) CeCR: Cross-entropy contrastive replay for online class-incremental continual learning. *Neural Netw* 173:106163. <https://doi.org/10.1016/j.neunet.2024.106163>
- Fayek HM, Cavedon L, Wu HR (2020) Progressive learning: a deep learning framework for continual learning. *Neural Netw* 128:345–357. <https://doi.org/10.1016/j.neunet.2020.05.011>
- Yan S, Xie J, He X (2021) DER: Dynamically expandable representation for class incremental learning. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 3014–3023. <https://doi.org/10.1109/CVPR46437.2021.00303>
- He J (2024) Gradient reweighting: towards imbalanced class-incremental learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 16668–16677. <https://doi.org/10.1109/CVPR52733.2024.01577>
- Jodelet Q, Liu X, Phua YJ, Murata T (2023) Class-incremental learning using diffusion model for distillation and replay. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 3425–3433. <https://doi.org/10.1109/ICCVW60793.2023.00367>
- Lu Y, Yang L, Chen HR, Cao J, Lin W, Long S (2024) Federated class-incremental learning with dynamic feature extractor fusion. *IEEE Trans Mob Comput*. <https://doi.org/10.1109/TMC.2024.3419096>
- Zhang C, Tsang EC, Xu W, Lin Y, Yang L (2023) Incremental concept-cognitive learning approach for concept classification oriented to weighted fuzzy concepts. *Knowledge-Based Syst* 260:110093. <https://doi.org/10.1016/j.knosys.2022.110093>
- Tang X, Yang C, Sun X, Zou M, Wang H (2023) Motor imagery EEG decoding based on multi-scale hybrid networks and feature enhancement. *IEEE Trans Neural Syst Rehabil Eng* 31:1208–1218. <https://doi.org/10.1109/TNSRE.2023.3242280>
- Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ (2018) EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *J Neural Eng* 15(5):056013. <https://doi.org/10.1088/1741-2552/aace8c>
- Jiménez-Guarneros M, Alejo-Eleuterio R (2022) A class-incremental learning method based on preserving the learned feature space for EEG-based emotion recognition. *Mathematics* 10(4):598. <https://doi.org/10.3390/math10040598>
- Wang H, Xu L, Bezerianos A, Chen C, Zhang Z (2021) Linking attention-based multiscale CNN with dynamical GCN for driving fatigue detection. *IEEE Trans Instrum Meas* 70:1–11. <https://doi.org/10.1109/TIM.2020.3047502>
- Deng Z, Li C, Song R, Liu X, Qian R, Chen X (2023) Centroid-guided domain incremental learning for EEG-based seizure prediction. *IEEE Trans Instrum Meas*. <https://doi.org/10.1109/TIM.2023.3334330>
- Li Z, Tan X, Li X, Yin L (2024) Multiclass motor imagery classification with Riemannian geometry and temporal-spectral selection. *Med Biol Eng Comput* 1–13. <https://doi.org/10.1007/s11517-024-03103-1>. TBWS
- Wang W, Li B, Wang H, Wang X, Qin Y, Shi X, Liu S (2024) EEG-FMCNN: A fusion multi-branch 1D convolutional neural network for EEG-based motor imagery classification. *Med Biol Eng Comput* 62(1):107–120. <https://doi.org/10.1007/s11517-023-02931-x>
- Chen CLP, Liu ZL (2018) Broad learning system: an effective and efficient incremental learning system without the need for deep architecture/ *IEEE Trans. Neural Netw Learn Syst* 29(1):10–24. <https://doi.org/10.1109/TNNLS.2017.2716952>
- Issa S, Peng Q, You X (2021) Emotion classification using EEG brain signals and the broad learning system. *IEEE Trans Syst Man, Cybernetics: Syst* 51(12):7382–7391. <https://doi.org/10.1109/TSMC.2020.2969686>
- Gong X, Zhang T, Chen CP, Liu Z (2021) Research review for broad learning system: algorithms, theory, and applications. *IEEE Trans Cybernet* 52(9):8922–8950. <https://doi.org/10.1109/TCYB.2021.3061094>
- Guo J, Chen CP, Liu Z, Yang X (2024) Dynamic neural network structure: a review for its theories and applications. *IEEE Trans Neural Netw Learn Syst*. <https://doi.org/10.1109/TNNLS.2024.3377194>
- Fu R, Liang H, Wang S, Jia C, Sun G, Gao T, ... Wang Y (2024) Transformer-BLS: an efficient learning algorithm based on multi-head attention mechanism and incremental learning algorithms. *Exp Syst Appl* 238. <https://doi.org/10.1016/j.eswa.2023.121734>
- Huang T, Li H, Zhou G, Li S (2024) Stacking multi-view broad learning system with residual structures for classification. *Inf Sci* 669:120559. <https://doi.org/10.1016/j.ins.2024.120559>
- She Q, Zhou Y, Gan H, Ma Y, Luo Z (2019) Decoding EEG in motor imagery tasks with graph semi-supervised broad learning. *Electronics* 8(11):1273. <https://doi.org/10.3390/electronics8111273>
- Zhou Y, She Q, Ma Y, Kong W, Zhang Y (2021) Transfer of semi-supervised broad learning system in electroencephalography signal classification. *Neural Comput Appl* 33:10597–10613. <https://doi.org/10.1007/s00521-021-05793-2>
- Yang Y, Li M, Liu H, Li Z (2024) A session-incremental broad learning system for motor imagery EEG classification. *Biomed Signal Process Cont* 97. <https://doi.org/10.1016/j.bspc.2024.106717>
- Rebuffi SA, Kolesnikov A, Sperl G, Lampert CH (2017) icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* 2001–2010. <https://doi.org/10.1109/CVPR.2017.587>
- Chen R, Chen G, Liao X, Xiong W (2024) Class-incremental learning via prototype similarity replay and similarity-adjusted regularization. *Appl Intell* 1–16. <https://doi.org/10.1007/s10489-024-05695-5>
- Li X, Tang B, Li H (2024) AdaER: An adaptive experience replay approach for continual lifelong learning. *Neurocomputing* 572:127204. <https://doi.org/10.1016/j.neucom.2023.127204>
- Kang H, Choi DW (2024) Recall-oriented continual learning with generative adversarial meta-model. In *Proceedings of the AAAI Conf Artificial Intell* 38(12):13040–13048. <https://doi.org/10.1007/s10489-024-05695-5>
- He C, Wang R, Shan S, Chen X (2024) Introspective GAN: learning to grow a GAN for incremental generation and classification. *Pattern Recognit* 151. <https://doi.org/10.1016/j.patcog.2024.110383>

36. Yang H, He W, Shan Z, Fang X, Chen X (2024) Class incremental learning via dynamic regeneration with task-adaptive distillation. *Comput Commun* 215:130–139. <https://doi.org/10.1016/j.com-com.2023.12.030>
37. Mallya A, Davis D, Lazebnik S (2018) Piggyback: adapting a single network to multiple tasks by learning to mask weights. In *Proceedings of the European conference on computer vision (ECCV)* 67–82. <https://doi.org/10.48550/arXiv.1801.06519>
38. Du F, Yang Y, Zhao Z, Zeng Z (2023) Efficient perturbation inference and expandable network for continual learning. *Neural Netw* 159:97–106
39. Qazi MA, Almakky I, Hashmi AUR, Sanjeev S, Yaqub M (2024) Dynammo: dynamic model merging for efficient class incremental learning for medical images. In *Annual Conference on Medical Image Understanding and Analysis* 245–257. https://doi.org/10.1007/978-3-031-66955-2_17
40. Dhar P, Singh RV, Peng KC, Wu Z, Chellappa R (2019) Learning without memorizing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 5138–5146. <https://doi.org/10.48550/arXiv.1811.080>
41. Akmel F, Meng F, Wu Q, Chen S, Zhang R, Assefa M (2024) Class similarity weighted knowledge distillation for few shot incremental learning. *Neurocomputing* 584:127587. <https://doi.org/10.1016/j.neucom.2024.127587>
42. Serra J, Suris D, Miron M, Karatzoglou A (2018) Overcoming catastrophic forgetting with hard attention to the task. In *International conference on machine learning* 4548–4557. <https://doi.org/10.48550/arXiv.1801.01423>
43. Fu R, Wang Y, Jia C (2022) A new data augmentation method for EEG features based on the hybrid model of broad-deep networks. *Exp Syst Appl* 202:117386. <https://doi.org/10.1016/j.eswa.2022.117386>
44. Dasan E, Gnanaraj R (2022) Joint ECG–EMG–EEG signal compression and reconstruction with incremental multimodal autoencoder approach. *Circuits Sys Signal Process* 41(11):6152–6181. <https://doi.org/10.1007/s00034-022-02071-x>
45. García-Salinas JS, Torres-García AA, Reyes-García CA, Villaseñor-Pineda L (2023) Intra-subject class-incremental deep learning approach for EEG-based imagined speech recognition. *Biomed Signal Process Control* 81:104433. <https://doi.org/10.1016/j.bspc.2022.104433>
46. Deng Z, Mao T, Shao C, Li C, Chen X (2023) Domain incremental learning for EEG-based seizure prediction. *CAAI International Conference on Artificial Intelligence* 476–487. https://doi.org/10.1007/978-981-99-9119-8_43
47. Sun L, Zhang M, Wang B, Tiwari P (2023) Few-shot class-incremental learning for medical time series classification. *IEEE J Biomed Health Inform* 28(4):1872–1882
48. Duan T, Wang Z, Li F, Doretto G, Adjero DA, Yin Y, Tao C (2024) Online continual decoding of streaming EEG signal with a balanced and informative memory buffer. *Neural Netw* 176:106338
49. Graves A, Graves A (2012) Long short-term memory. *Supervised sequence labelling with recurrent neural networks*. 37–45. <https://doi.org/10.1007/978-3-642-24797-2>
50. Lu B, Wang F, Wang S, Chen J, Wen G, Fu R (2024) Improvement of motor imagery electroencephalogram decoding by iterative weighted sparse-group lasso. *Expert Syst Appl* 238:122286
51. Cao J, Li G, Shen J, Dai C (2024) IFBCLNet: Spatio-temporal frequency feature extraction-based MI-EEG classification convolutional network. *Bio-med Signal Process Control* 92:106092. <https://doi.org/10.1016/j.bspc.2024.106092>
52. Zeng L, Tang H, Wang W, Xie M, Ai Z, Chen L, Wu Y (2023) MAMC-Net: an effective deep learning framework for whole-slide image tumor segmentation. *Multimed Tools Appl* 82(25):39349–39369. <https://doi.org/10.1007/s11042-023-15065-x>
53. Zhang Y, Li P, Cheng L, Li M, Li H (2024) Attention-based multiscale spatial-temporal convolutional network for motor imagery EEG decoding. *IEEE Trans Consum Electron* 70(1):2423–2434. <https://doi.org/10.1109/TCE.2023.3330423>
54. Liu W, Guo C, Gao C (2024) A cross-session motor imagery classification method based on Riemannian geometry and deep domain adaptation. *Exp Syst Appl* 237. <https://doi.org/10.1016/j.eswa.2023.121612>
55. Brunner C, Leeb R, Müller-Putz G et al (2008) BCI Competition 2008–graz data set A, institute for knowledge Discovery (Laboratory of Brain-Computer Interfaces). *Graz Univ Technol* 16:1–6
56. Leeb R et al (2008) BCI Competition 2008–Graz data set B. *Graz University of Technology, Austria* 16:1–6
57. Jeon E, Ko W, Yoon JS, Suk HI (2021) Mutual information-driven subject-invariant and class-relevant deep representation learning in BCI. *IEEE Trans Neural Netw Learn Syst* 34(2):739–749. <https://doi.org/10.1109/TNNLS.2021.3100583>
58. Zaman A, Kumar S, Shatabda S et al (2024) SleepBoost: a multi-level tree-based ensemble model for automatic sleep stage classification. *Med Biol Eng Comput* 62:2769–2783. <https://doi.org/10.1007/s11517-024-03096-x>
59. Li D, Zeng Z (2023) CRNet: a fast continual learning framework with random theory. *IEEE Trans Pattern Anal Mach Intell* 45(9):10731–10744. <https://doi.org/10.1109/TPAMI.2023.3262853>
60. Zheng Y, Qin X, Xi Z, et al (2019) Mixed-norm based broad learning system for EEG classification, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 7092–7095. <https://doi.org/10.1109/EMBC.2019.8856666>
61. Li X, Tang X, Qiu S, Deng X, Wang H, Tian Y (2024) Subdomain adversarial network for motor imagery EEG classification using graph data. *IEEE Trans Emerg Topics Comput Intell* 8(1):327–336. <https://doi.org/10.1109/TETCI.2023.3301385>
62. Wang C, Wu Y, Wang C, Zhu Y, Wang C, Niu Y, ... Yu Y (2022) MI-EEG classification using Shannon complex wavelet and convolutional neural networks. *Appl Soft Comput* 130. <https://doi.org/10.1016/j.asoc.2022.109685>
63. Li H, Zhang D, Xie J (2023) MI-DABAN: a dual-attention-based adversarial network for motor imagery classification. *Comput Biol Med* 152. <https://doi.org/10.1016/j.combiomed.2022.106420>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Yufei Yang received her B.Sc. degrees in electrical engineering and automation from Tianjin University of Technology and Education, Tianjin, China, in 2018, and received the M.Sc. degree from Beijing Technology and Business University, Beijing, China, in 2021. She is currently pursuing her Ph.D. degree in Beijing University of Technology, Beijing, China. Her research interests include brain computer interface and incremental learning.



Linlin Wang received her B.Sc. degree and M.Sc. degree from Henan Normal University, Henan, China, in 2016 and 2019, respectively, and received her Ph.D. degree from Beijing University of Technology, Beijing, China, in 2024. Her research interests include brain computer interface, brain science, and artificial intelligence.



Mingai Li received her B.Sc. degree and M.Sc. degree from Daqing Petroleum Institute, Heilongjiang, China, in 1987 and 1990, respectively, and Ph.D. degree from Beijing University of Technology, Beijing, in 2006. She is a professor at the Faculty of Information Technology, Beijing University of Technology. Her research interests mainly include brain computer interface, artificial intelligence, pattern recognition, and rehabilitation robots.