



EEG-based classification combining Bayesian convolutional neural networks with recurrence plot for motor movement/imagery

Wenqie Huang, Guanghui Yan*, Wenwen Chang, Yuchan Zhang, Yueting Yuan

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, 730070 China



ARTICLE INFO

Keywords:
 Electroencephalogram (EEG)
 Real execution
 Motor imagery (MI)
 Deep learning (DL)
 Bayesian convolutional neural networks (BCNNs)
 Recurrence plots (RPs)
 Feature extraction
 Classification

ABSTRACT

Electroencephalogram (EEG)-based Motor imagery (MI) is a key topic in the brain-computer interface (BCI). The EEG-based real execution and motor imagery multi-class classification tasks are also crucial, but only a few kinds of literature research it. In addition, classification accuracy still has room for improvement, and the inter-individual variability problems in BCI applications need to be solved. To address these issues, we developed a novel model (RP-BCNNs) that combines the recurrence plot (RP) and Bayesian Convolutional Neural Networks (BCNNs). First, we employ an RP computation for preprocessed EEG signals of each channel and merge all RPs of all channels into one based on the weighted average method. Then, we feed the RP features into BCNNs to classify 2-class, 3-class, 4-class, and 5-class on real/imaginary movements classification tasks. The results show that the RP-BCNNs model outperforms the state-of-the-art methods, achieving average accuracies of 92.86%, 94.12%, 91.37%, 92.61% for real movements and 94.07%, 93.77%, 90.54%, 91.85% for imaginary movements. Our findings suggest that combining complex network methods with deep learning can improve the classification performance of EEG-based BCI systems (e.g., motor imagery, emotion recognition, and epileptic seizure classification).

1. Introduction

Motor Imagery (MI) is a primary paradigm in the field of electroencephalogram (EEG) based brain-computer interface (BCI), which does not rely on the traditional brain information output pathways but uses engineering techniques to become a bridge between the brain and the external devices, providing a better and quality life for the people with disabilities [1] or dyskinesia [2,3]. EEG-based BCI systems can realize the output of control instructions by monitoring the changes of scalp level signals and extracting corresponding motion patterns in the real execution/motor imagery process. The most critical advantages of EEG are non-invasive, high temporal resolution, simplicity, and portability compared with other existing neuroimaging modalities [4], e.g., electrocorticography (ECoG), electromyography (EMG), or electrooculography (EOG). Thus, EEG-based BCI systems have been most widely applied in BCI applications in recent years, driving many researchers to focus on EEG-based MI, e.g., control for wheelchairs [5] or robot arms [6,7], etc. Subjects were asked to perform real execution or MI movements during the MI experiment, producing the ERS/ERD phenomenon in the sensorimotor cortex [8,9], where the EEG signals were recorded

with the EEG amplifier.

Due to the dynamism and low signal-to-noise ratio (SNR) of the EEG, it is challenging to classify various real execution/motor imagery tasks correctly. Several studies have investigated that specialized preprocessing was used for EEG signals to extract relevant features followed by a classifier [10,11]. The Common Spatial Patterns (CSP) method, an algorithm for capturing the appropriate MI features, has been most commonly used in those works. Lotte et al. [10] presented the CSP with Tikhonov regularization and the weighted Tikhonov regularization, outperforming CSP by nearly 10% in median classification accuracy and lead to more neurophysiologically relevant spatial filters for the MI task. Aghaei et al. [11] proposed the separable common spatio-spectral patterns to extract discriminant spatio-spectral EEG features in MI-BCIs, achieving encouraging results compared with the filter-bank CSP method. However, those traditional methods lose the correlation information between different channel EEG series (i.e., different brain regions), which have not achieved satisfactory results (there is room for improvement in accuracy). Therefore, it is necessary to consider new feature extraction methods that can connect EEG time series between different brain regions, which can detect more valuable and detailed

* Corresponding author at: School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China.

E-mail address: yanguanghui@mail.lzjtu.cn (G. Yan).

motion-related information from EEG signals.

Considering EEG signals are non-linear, dynamical, and complex, several studies have shown that the hidden dynamical features in EEG signals can be explored by using non-linear analysis methods. The recurrence of phase spatial trajectories is one of the most critical basic features of the dynamical system, which can be applied to describe the behavior of the system in phase space. Thus, many researchers use recurrence based methods to explore dynamical systems, such as dynamical time series [12–15]. Introduced by Eckmann et al. in 1987 [16], the recurrence plot (RP), a binary symmetric square matrix, describes natural time correlation information (i.e., encodes the times when two states are neighbors in phase space) and usually performs better compared with other complex network methods of time series. Hence, it drives us to extract typical RP features from EEG signals and feed them into the appropriate classifier to classify various real execution/motor imagery tasks.

Recently, deep learning (DL) methods have been introduced into EEG-based classification tasks. In [17], Shen et al. combined recurrent neural networks and convolutional neural networks and employed the stacked random forest model to enhance classification performance. In [18], Dose et al. proposed the DL model using a convolutional neural network for learning generalized features and dimension reduction to conduct multi-class classification tasks. In [19], Sun et al. proposed an EEG-based identification system using a 1D-convolutional LSTM neural network, achieving excellent results from 109 subjects MI-EEG data. In [20], Hou et al. presented a novel approach combining source EEG source imaging and convolutional neural networks to decode four-class MI EEG signals, obtaining encouraging accuracy and improving the generalization performance by CNN techniques. In [21], Huang et al. proposed a novel model that combines the local reparameterization trick and convolutional neural network (LRT-CNN), handling the challenge of individual variability and achieving satisfactory results for four-class classification tasks. Those successful researches encourage us to utilize the DL model as a classifier for EEG-based classification tasks. In [22], Fan et al. proposed a novel algorithm that combines a Filter Band Combination (FBC) module and a Multi-View structure to preserve as many frequency and time domain features as possible, thereby improving the decoding performance of MI EEG signals. In [23], a spatio-spectral feature representation method was proposed, and its features were classified using a 3DCNN-LSTM model. This approach improves the decoding performance of MI EEG signals while preserving as much feature information as possible in both spatial and frequency domain. But there is room for improvement in the decoding performance. In [24], Kim et al. presented a novel deep learning model based on Riemannian geometry. The model learned the Riemannian barycenter for each class and penalizes the distances between matrices and their respective class centers. It also normalized the distribution of symmetric positive definite (SPD) matrices. Experimental results demonstrated that the model was able to simultaneously reduce intra-class distances and enlarge inter-class distances. To sum up, due to the lack of correlation information between brain regions, a large number of training parameters and deeper network depth, the application of DL techniques in EEG-based field should be further studied and developed.

Several researchers have demonstrated that combining the RP method and DL techniques can capture the advantages of both. Zeng et al. [25] combined gray recurrence plot (GRP) and densely connected convolutional network (DenseNet) to identify seizures and epilepsy from single-channel, long-term EEG signals, achieving excellent classification accuracy and giving excellent computational efficiency. Khodabakhshi et al. [26] presented two different recurrence-based chaotic schemes (the Poincaré map function and recurrence plots (RPs)) to analyze the long-term dynamics of EEG signals associated with state space (SS) trajectory of the time series. Their experimental results show that combining non-linear approaches with the one capable of producing occipital activity can improve performance. Meng et al. [27] proposed

the convolutional neural network of recurrence plot method, which can enhance the classification accuracy of MI EEG and can identify left and right movements accurately. To sum up, by calculating RP features from EEG signals as the input of the neural network, we can effectively keep correlation information between brain regions, reduce the input dimension, improve the training efficiency and enhance the accuracy.

In this paper, we develop a novel approach, combining the recurrence plot (RP) and Bayesian Convolutional Neural Networks (BCNNs) for real execution/motor imagery classification. Real execution refers to the physical act of moving a limb, while motor imagery refers to the mental rehearsal of a movement without any physical movement. To fuse the correlation information between all brain regions, we first employ an RP computation for preprocessed EEG signals of each channel. Then we assign a set of weight factors based on the normal distribution to RPs of all EEG channels and merge all RPs into one according to the weighted average method, reducing the complexity of the original EEG signals. Furthermore, we feed the RP features into Bayesian Convolutional Neural Networks (BCNNs) to fully utilize RP's natural time correlation information and promote the performance of real execution/motor imagery classification. The schematic diagram of the real execution/motor imagery classification system is shown in Fig. 1. Although prospective studies have shown encouraging results, they did not systematically explore the EEG classification of real movements. Moreover, we think it is necessary to study the problem of inter-individual variability in EEG-based applications. Therefore, based on the proposed method, the following two issues will be addressed: (1) Classification w.r.t. real/imaginary movements. (2) Common inter-individual variability problems in EEG-based applications.

The rest of this paper is organized as follows: Section 2 introduces the definition and mathematics. In Section 3, the proposed architecture are presented. Section 4 shows the experiment and results with respect to real/imaginary movements. Section 5 mainly discusses the proposed method, and Section 6 is the conclusion of this paper.

2. Preliminaries

We begin by describing the Recurrence plot (RP), Bayes by Backprop, and Convolutional Neural Networks (CNNs). Then we describe how Bayes by Backprop can be efficiently applied to CNNs.

2.1. Recurrence plot (RP)

Recently, many researchers have viewed the brain as a complex network (dynamical system) [28,29]. Hence, it is necessary to find a multifaceted and powerful method to study a dynamical system. The recurrence plot (RP), a matrix of dots in an $N \times N$ square to reveal the nonlinear dynamical systems represented by time series, was introduced by Eckmann et al. [16]. The principle of creating an RP is to embed time series into phase space and then calculate the distance among trajectory states. If the distance is greater than the threshold distance, black dots will be plotted in RP; otherwise, white dots will be plotted. In other words, RP is a recurrence matrix of 2D structure, which is constructed through the trajectory moving to its previous states.

Given a specific time series of length $N \{x_i, i = 1, \dots, N\}$, the phase space vectors $\vec{x}_i (i = 1, \dots, N)$ can be represented. The distance between trajectory states can be calculated by calculating the following matrix:

$$R_{i,j} = \Theta\left(\epsilon - \|\vec{x}_i - \vec{x}_j\|\right), i, j = 1, 2, \dots, N \quad (1)$$

$$\Theta(z) = \begin{cases} 1, & z \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where R is the recurrence matrix, $\|\cdot\|$ is a norm (Euclidean norm was used in this study), ϵ is the threshold distance, and $\Theta(\cdot)$ is the Heaviside function as shown in Eq. (2). Based on Eq. (1) and Eq. (2), the RP can be

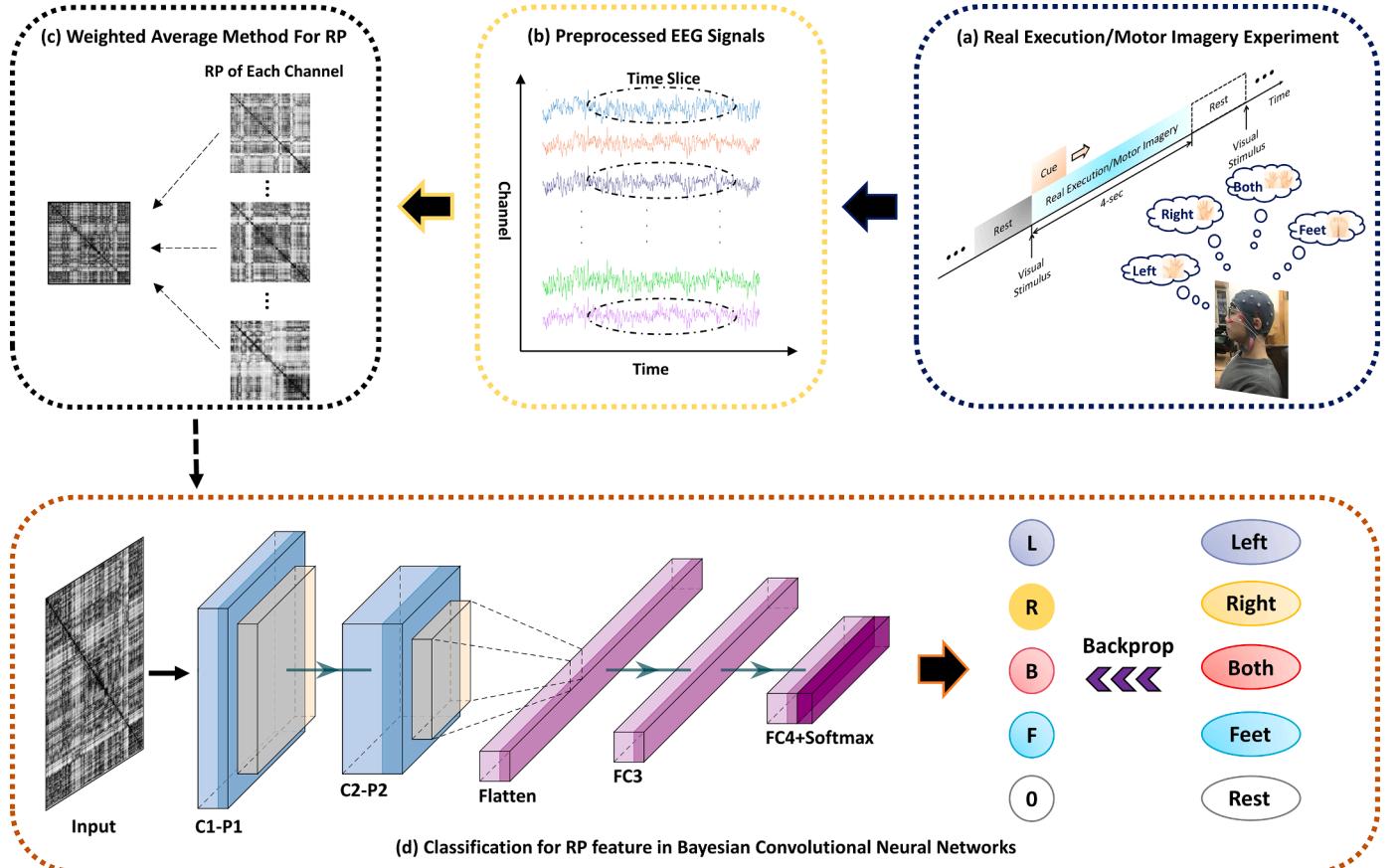


Fig. 1. The Schematic paradigm of the real execution/motor imagery classification system. (a): a participant performs the real execution or motor imagery experiment, and the EEG signals will be recorded on a device with 64 EEG channels. (b): Preprocess the EEG signals, such as filtering, ICA processing, time slice segmentation, etc. (c): The recurrence plot (RP) of each channel EEG signal is calculated, and use the weighted average method to fuse the RP features for all channel. (d): The RP feature is used as the input of Bayesian Convolutional Neural Networks with two convolutional layers (C1 and C2), two pooling layers (P1 and P2), and two fully connected layers (FC3 and FC4) to classify the 2-class (L, R or B, F), 3-class (L, R, rest or B, F, rest), 4-class (L, R, B, F), or 5-class (L, R, B, F, rest) classification tasks.

described by matrix R because the values of the matrix R are zero and one, corresponding to white and black dots in RP. In addition, considering different dynamical systems correspond to different ϵ , there is no public standard for how to determine the threshold ϵ . In this study, we set a series of parameters (generate a threshold every 0.05 from 0.1 to 1) to analyze the RP structure of different movements in two datasets. The results show that the best RQA analysis results can be obtained when the

threshold value is 0.2, and the RP structure of both datasets can reveal a prominent substructure. Section 4.4 shows the RQA analysis results when the threshold value is 0.2. Hence, 0.2 was used for the real execution and motor imagery system finally.

Based on the literature [30], we can analyze the RPs of three prototypical systems, namely of the chaotic Rössler system (Fig. 2(a)), of the EEG-based MI system (Fig. 2(b)), of uniformly distributed noise (Fig. 2

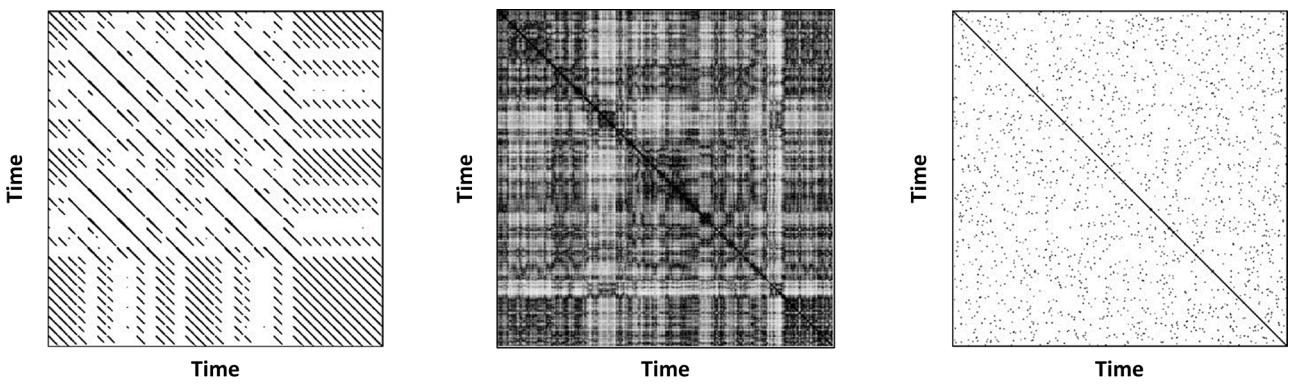


Fig. 2. Recurrence plots of three systems. (b): RP of the one channel EEG signal. To analyze RP of the EEG-based MI system, note that (a) and (c) are from the literature [30].

(c)). The RP of the EEG-based MI system is between the chaotic Rössler system and the uniformly distributed noise, which indicates the EEG-based MI system is traceable and can be studied. It means that RP can reflect the characteristic structure and similarity rule and reveal the internal structure of EEG-based signals. Therefore, we can utilize RP to analyze EEG signals and extract time-related information between brain regions.

For experimental details, we first calculate the recurrence matrix based on Eq. (1) and Eq. (2) as an RP for preprocessed EEG signals of each channel. Then we assign a set of weight factors based on the normal distribution to RPs of all EEG channels and merge them into one (i.e., weighted average method). Therefore, an RP fused by the weighted average method can be used as the input feature of the neural networks.

2.2. Bayes by Backprop

Bayes by Backprop [31,32] is a variational inference (VI) method to learn the posterior distribution $p(W|D)$ of the weight W of the neural network. Since the number of parameters in the weight of the neural network is very large and the function form is not suitable for precise integration, the variational approximation is used instead of the Monte Carlo method to find the likelihood Bayesian posterior distribution.

According to Bayes' principles,

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)} \quad (3)$$

where $p(\theta)$ is the prior distribution. Try to find the posterior distribution of the model parameter θ given some data X . Actually,

$$p(X) = \int p(X, \theta)d\theta \quad (4)$$

This integral makes the whole thing intractable, and it's natural to approximate it.

Suppose posterior distribution $q_\theta(W)$ is the density on the parameter set W , which is parameterized by θ . We're going to use the KL-divergence to reduce the difference between $q_\theta(W)$ and true posterior $p(W|D)$ by adjusting the parameters,

$$\begin{aligned} KL(q_\theta(W) \| p(W|D)) &= \int q_\theta(W) \ln \frac{q_\theta(W)}{p(W|D)} dW \\ &= \int q_\theta(W) \ln \frac{q_\theta(W)p(D)}{p(D|W)p(W)} dW \\ &= \int q_\theta(W) \left[\ln \frac{q_\theta(W)}{p(W)} - \text{Inp}(D|W) + \text{Inp}(D) \right] dW \quad (5) \\ &= E_q \left[\ln \frac{q_\theta(W)}{p(W)} - \text{Inp}(D|W) \right] + \text{Inp}(D) \\ &= KL(q_\theta(W) \| p(W)) - E_q[\text{Inp}(D|W)] + \text{Inp}(D) \\ &= -F(q_\theta) + \text{Inp}(D) \end{aligned}$$

Thus, the KL-divergence between the approximate distribution and the real distribution can be approximated by the approximate logarithmic likelihood of $F(q_\theta)$. The stochastic gradient descent method [31] is used to approximate the likelihood term with small batches of data,

$$\mathcal{L}(W, \theta) = -\frac{N}{M} \sum_{i=1}^N E_q[\text{Inp}(D_i|W)] + KL(q_\theta(W) \| p(W)) \quad (6)$$

where $D_i \subset D$, and each subset has size M .

After passing through a subset D_i , backpropagation is applied to update the model parameters. Blundell et al. [32] proposed a method of approximate inference in the bayesian neural network. The method uses the reparameterization technique to show how to find the unbiased estimator of the desired derivative. The specific form of its expected derivative is shown below,

$$\begin{aligned} \frac{\partial}{\partial \theta} E_{q(W|\theta)}[f(W, \theta)] &= \frac{\partial}{\partial \theta} \int f(W, \theta) q(W|\theta) dW \\ &= \frac{\partial}{\partial \theta} \int f(W, \theta) q(\epsilon) d\epsilon \\ &= E_{q(\epsilon)} \left[\frac{\partial f(W, \theta)}{\partial W} \frac{\partial W}{\partial \theta} + \frac{\partial f(W, \theta)}{\partial \theta} \right] \end{aligned} \quad (7)$$

Here $W = t(\theta, \epsilon)$, and $t(\theta, \epsilon)$ is a definite function. $q(\epsilon) d\epsilon = q(W|\theta) dW$, $f(W, \theta)$ can be viewed as the independent variable of expectation, and it is defined in Bayes by Backpropagation algorithm as below [32].

$$f(W, \theta) = \ln \frac{q_\theta(W)}{p(W)} - \text{Inp}(D|W) \quad (8)$$

Here, we assume Gaussian posterior $\theta = (\mu, \sigma)$, and $\sigma = \text{Softplus}(\rho)$ is used to ensure that the standard deviation parameter is always non-negative. Then we reparameterize the distribution of $W \sim N(\mu, \sigma^2)$,

$$W = t(\theta, \epsilon) = \mu + \sigma \odot \epsilon, \text{ where } \odot \text{ is pointwise multiplication.} \quad (9)$$

2.3. Convolutional neural networks

The Convolutional neural network (CNN) is a deep neural network that uses convolution operation to realize weight sharing, which can learn and extract features from input data, thus improving the classification performance. A CNN model consists of the input layer, convolutional layer, pooling layer, fully connected layer and output layer.

In the convolutional layer, the feature value can be calculated as:

$$b^{[l]} = g(a^{[l]} * w^{[l]} + \text{bias}) \quad (10)$$

where $a^{[l]}$ is the input of the l^{th} layer by $a \in R^l$ (note that $a^{[0]} = x_{\text{input}}$), $w^{[l]}$ is the weight matrix of the l^{th} layer by $w \in R^{l \times O}$, $b^{[l]}$ is the output of the l^{th} layer by $b \in R^O$, $*$ denotes the convolution operator, and $g(\cdot)$ denotes the activation function. Eq. (3) means that after the weight and feature are convoluted and added with the bias, they propagate to the next layer after the activation operation and until they reach the last convolutional layer.

In addition to the convolutional layer, the pooling layer is often used after the convolutional layer to reduce the representation size of the model by taking the average pooling or max pooling operation, which can increase the calculation speed, improve the extracted features' robustness, and reduce the model complexity. In our proposed model, max pooling was utilized because it can retain the characteristics of the input data to the greatest extent. Mathematically, the pooling operation can be defined as:

$$p^{[l]} = \text{pooling}(b^{[l]}) \quad (11)$$

where $\text{pooling}(\cdot)$ denotes the pooling operation, $b^{[l]}$ is the feature after activation operation, and $p^{[l]}$ is the result after pooling operation. The effect of the max pooling operation is the filter extracts features and then maintains the maximum value.

The fully connected layer, often followed by the convolutional layer and pooling layer, is used to integrate the previous feature information. In detail, the fully connected layer converts the features of the last convolutional layer into a vector, and finally, separates specific categories through the softmax operation. The softmax value can be defined as:

$$\text{Softmax}(\hat{y}_i) = \frac{e^{\hat{y}_i}}{\sum_j e^{\hat{y}_j}} \quad (12)$$

where \hat{y}_i represents the predicted value of i^{th} class after softmax operation, n denotes the numbers of class. In order to make the predicted

value \hat{y}_i closer to the real value y_i , we need to select the optimal model parameters by minimizing the loss function. The cost function can be defined as:

$$J = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(\hat{y}_i, y_i) \quad (13)$$

where J represents the cost value of the loss function (Hence we need to minimize this value), m denotes the training set size, and $\mathcal{L}(\cdot)$ is the loss function that describes the loss between the predicted value \hat{y}_i and the real value y_i . In this study, stochastic gradient descent (SGD) was used to optimize our proposed model.

2.4. Bayes by Backprop for convolutional neural networks

In this section, we consider how Bayes by Backprop can be efficiently applied to CNNs. Following [33,34], we place a prior distribution over the weight on the convolutional and fully connected layers and approximately integrate it with posterior distributions, then compute the intractable posterior probability distributions $p(W|D)$ by applying Bayes by Backprop, as described in Section 2.2. Hence, it will cause deep thought that a probability distribution over the weights implies a probability distribution over the activations. Based on the study [33,35], to transform the global uncertainty in the weight into a local uncertainty, which is independent between different examples and more accessible to sample, the local reparameterization trick was applied to yield a computationally and statistically efficient gradient estimator on the convolutional and fully connected layers. The expressions for the forward pass through a convolutional layer and a fully connected can be represented as:

ensure a positive non-zero variance $\alpha_{ijhw}\mu_{ijhw}^2$, we took the logarithm of $\alpha_{ijhw}\mu_{ijhw}^2$ in our experiment.

So far, we know that after applying the local reparameterization trick to the activation function, two parameters need to be learned through backpropagation, i.e., mean μ_{ijhw} and variance $\alpha_{ijhw}\mu_{ijhw}^2$. In fact, we only need to learn one parameter per convolution operation because the variance includes α_{ijhw} and μ_{ijhw}^2 . Therefore, the mean μ_{ijhw} needs to be learned in the first convolution operation and the α_{ijhw} needs to be learned in the second convolution operation. In this way, the learning process is like CNN updated by frequentist inference with point estimation, which reduces a certain amount of calculation.

3. Proposed architecture

Having liberated from the complex theories of Recurrence Plot (RP), Variational Inference (VI), Bayes by Backprop, and Convolutional neural network (CNN), in this section, we propose a model of combining RP and Bayesian Convolutional neural networks (RP-BCNNs). The proposed model's architecture diagram is shown in Fig. 3. Given preprocessed EEG signals, the recurrence plots (RPs) can be calculated based on Eq. (1) and Eq. (2). Then feed the RPs features into Bayesian Convolutional Neural Networks (BCNNs) to classify the 2-class (L, R or B, F), 3-class (L, R, rest or B, F, rest), 4-class (L, R, B, F), or 5-class (L, R, B, F, rest) classification tasks. In this study, a 640×640 RP is calculated from one channel (20 channels in total), so we obtain 20 RPs (640×640). Then 20 RPs were fused into one RP by the weighted average method. By doing so, the correlation information between different brain regions from 20-channel EEG signals can be fully utilized for the classification phase. For the convenience of calculation, a row vector (1×640) in the

$$\text{Convolutional layer, } b_j = A_i * \mu_i + \varepsilon_j \odot \sqrt{A_i^2 * (\alpha_i \odot \mu_i^2)}, \text{ where } q(w_{ijhw}|D) = N(\mu_{ijhw}, \alpha_{ijhw}\mu_{ijhw}^2) \quad (14)$$

$$\text{Fullyconnectedlayer, } b_j = \sum \mu_{ij} a_i + \varepsilon_j \sqrt{\sum \alpha_{ij} \mu_{ij}^2 a_i^2}, \text{ where } q(w_{ijhw}|D) = N(\mu_{ijhw}, \alpha_{ijhw}\mu_{ijhw}^2)$$

where b is the output vector, $q(w_{ijhw}|D)$ is the variational posterior probability distribution belongs to $N(\mu_{ijhw}, \alpha_{ijhw}\mu_{ijhw}^2)$, $*$ denotes the convolution operator, \odot represents the component-wise multiplication, A is the receptive field, a is the input vector, i and j are the input and output layers, h and w are the height and width of given filter, and $\varepsilon \sim N(0, 1)$ is sampled independently for each object in a mini-batch. Significantly, to

weighted average RP is used as the input of BCNNs. Based on the literature [20], we convert the vector of 1×640 dimension into 32×20 dimension as the input of BCNNs. In BCNNs (see Fig. (b)), using the probability distribution as weights for classification tasks is more reasonable than single point-estimates based on the study [33,34]. Meanwhile, the Softplus function, a smooth approximation of ReLU, was used in the proposed model because it has the subtle and essential

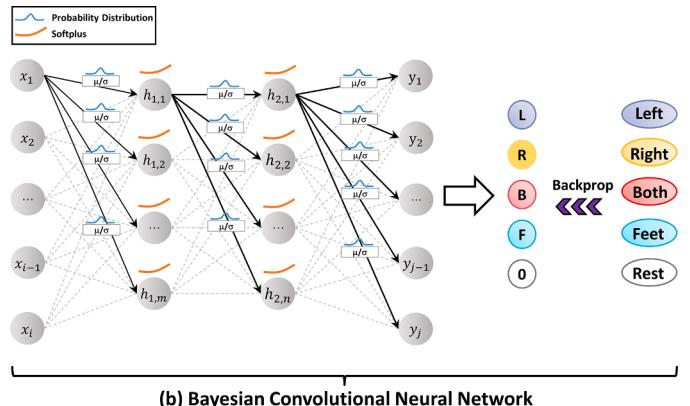
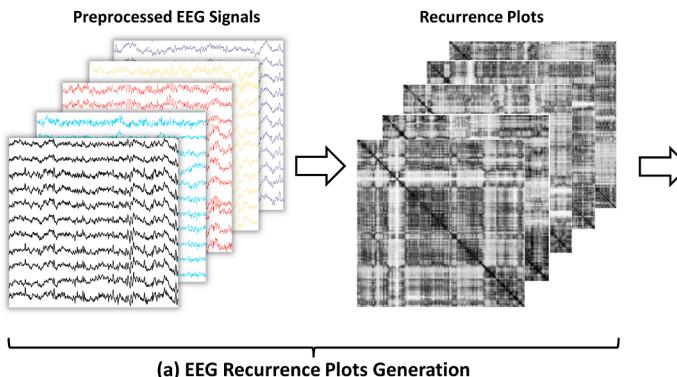


Fig. 3. The architecture of the proposed RP-BCNNs. (a): RPs features are calculated from the preprocessed EEG signals. (b): Note that in Bayesian Convolutional Neural Networks, the weight is represented by probability distribution rather than point estimation. The softplus function is used in the hidden layer.

advantage that the Softplus function never becomes zero when $x \rightarrow -\infty$, whereas the ReLU will. The Softplus function can be defined as:

$$\text{Softplus}(x) = \frac{1}{\beta} \cdot \log(1 + \exp(\beta \cdot x)), \text{ where } \beta \text{ is by default set to 1.} \quad (15)$$

On the other hand, BCNNs are more robust in the over-fitting phenomenon and can learn well from small datasets. Meanwhile, the average value is calculated on many models during the training phase by using a prior probability distribution to integrate the parameters, which provides a regularization effect for the network to prevent over-fitting.

4. Experiment and results

We will elaborate the experimental results from the following aspects: [Section 4.1](#) introduces the description of two datasets. In [Section 4.2](#), the cross-validation are presented. [Section 4.3](#) describes several statistical metrics. [Section 4.4](#) shows the RQA results with respect to real/imaginary movements. [Sections 4.5 and 4.6](#) shows the classification results with respect to real/imaginary movements. [Section 4.7](#) introduces 2-class, 3-class, 4-class, 5-class classification in real and imaginary movements. [Section 4.8](#) shows the comparison with state-of-the-art methods.

4.1. Description of datasets

The proposed method mentioned above has been applied to the PhysioNet dataset¹ [36] and GigaDB dataset² [37], which contain real and imaginary execution of movements.

The PhysioNet dataset was recorded from 109 subjects at 64 EEG channels with a sampling rate of 160 Hz per second (4-second of each trial). The experimental paradigm of each trial is shown in [Fig. 4](#). Each subject performed three two-minute runs of four different MI tasks (Task 1: real execution of left/right fist, Task 2: imaginary execution of left/right fist, Task 3: real execution of both fists/feet, Task 4 imaginary execution of both fists/feet). Only 25 subjects' trials for the real and imaginary movements were used in this work. Note that all data from 25 subjects were preprocessed, i.e., splitting data, filtering, and removing artifacts (ICA). The following subsets were created from 25 subjects' data.

- **2-class:** Real or imaginary movements of opening and closing left/right fist (Real L/R or imaginary L/R) or opening and closing both fists/feet (Real B/F or imaginary B/F).
- **3-class:** To study the EEG changes of subjects at the resting state, where the subject does not perform any movements, we have designed the 3-class experiment. Those three classification tasks are real/imaginary movements of opening and closing left/right fist, and the resting state or opening and closing both fists/feet, and the resting state, i.e., L, R, and rest tasks or B, F, and rest tasks.
- **4-class:** Real or imaginary movements of opening and closing left/right fist, and opening and closing both fists/feet, i.e., L, R, B, F tasks.
- **5-class:** To study the EEG changes of all the executions from subjects during the experiment, we combined all the real or imaginary movements with the resting state to design the 5-class experiment, i.e., L, R, B, F, and rest tasks.

The GigaDB dataset contains EEG data from 52 subjects (S1-S52). There exist 64 channels in the data. The dataset was collected from two different real execution or MI tasks: left hand and right hand (each task lasts for three seconds). The EEG data were sampled at 512 Hz. For more detail, please refer to [37]. In this work, only 20 subjects were used for testing the proposed method.

According to the literature [38], when selecting the EEG channels, we considered two factors: retaining channels with high information content (motor cortex region) and reducing computational complexity. Therefore, 20 EEG channels (FC-1/2/3/4/5/6, C-1/2/3/4/5/6/z, CP-1/2/3/4/5/6/z) (Please note that these 20 channels are selected in both datasets) in the motor cortex region were selected from a total of 64 channels for actual execution and MI tasks.

4.2. Cross-validation

To estimate the proposed model's stability, accuracy, and reliability, we divided the data into the training data (90%) and the test data (10%). Furthermore, the training data was separated into five 80/20 cross-validation splits, i.e., only 80% of the training data (training set) were trained for each time, and the remaining 20% of the training data (validation set) were used for validation. Then, we averaged the results of each time (five times in total) and explored the optimal model as a global measure, which ensured the test data (test set) were more applicable and proved the results were reliable and robust. The result of 5-fold cross-validation can be defined as:

$$\text{Result} = \frac{1}{5} \sum_{k=1}^5 \text{result}^{(k)} \quad (16)$$

To train the optimal model for real execution/motor imagery classification tasks, we selected the optimal network parameters based on the imaginary 2-class (L, R) classification tasks, as shown in [Fig. 5](#). The five subjects mean accuracies from six different frameworks are 90.48%, 94.68%, 95.35%, 95.37%, 94.87%, and 92.74%, respectively. The accuracy will not change significantly by increasing the convolutional layer, the pooling layer, and the fully connected layer after the framework has two convolutional layers, two pooling layers, and two fully connected layers. Furthermore, we added batch normalization for each convolutional, pooling, and fully connected layer except for the last fully connected layer, avoiding network imbalance problems during training. Considering the complexity of the model, we finally selected the CBP-CBP-FB-F framework to train the rest of the classification tasks. More implementation details are shown in [Table 1](#). We utilized the Softplus activation function for training 100 epochs for each experiment, which alleviated the gradient vanishing problem. Finally, Stochastic Gradient Descent (SGD) with 1024 batch sizes was optimized by the Adam optimizer with a 0.01 learning rate.

4.3. Evaluation metrics

Several statistical metrics were calculated for evaluating the performance of the proposed model. For clearness, the numbers of the true positive, true negative, false positive, and false negative are defined as TP, TN, FP, and FN, respectively. Based on the above values, accuracy (ACC), Cohen's kappa coefficient (Kappa), precision, recall, and F1-score, are applied to appraise the performance of the proposed model. Those metrics can be defined as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (17)$$

$$Kappa = \frac{ACC - P_e}{1 - P_e}, \text{ where } P_e \text{ is the hypothetical probability of agreement.} \quad (18)$$

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

$$F1Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (21)$$

¹ publicly available at <https://archive.physionet.org/pn4/eegmmidb/>

² publicly available at <http://gigadb.org/dataset/100295>

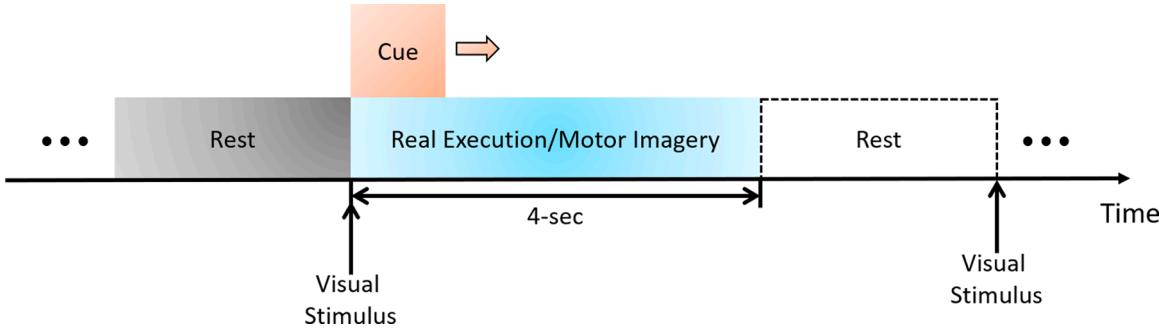


Fig. 4. The experimental paradigm of each trial in the PhysioNet dataset. The subject is cued to execute the real execution or MI task for four seconds while the cue appears, then rest until the next trial start.

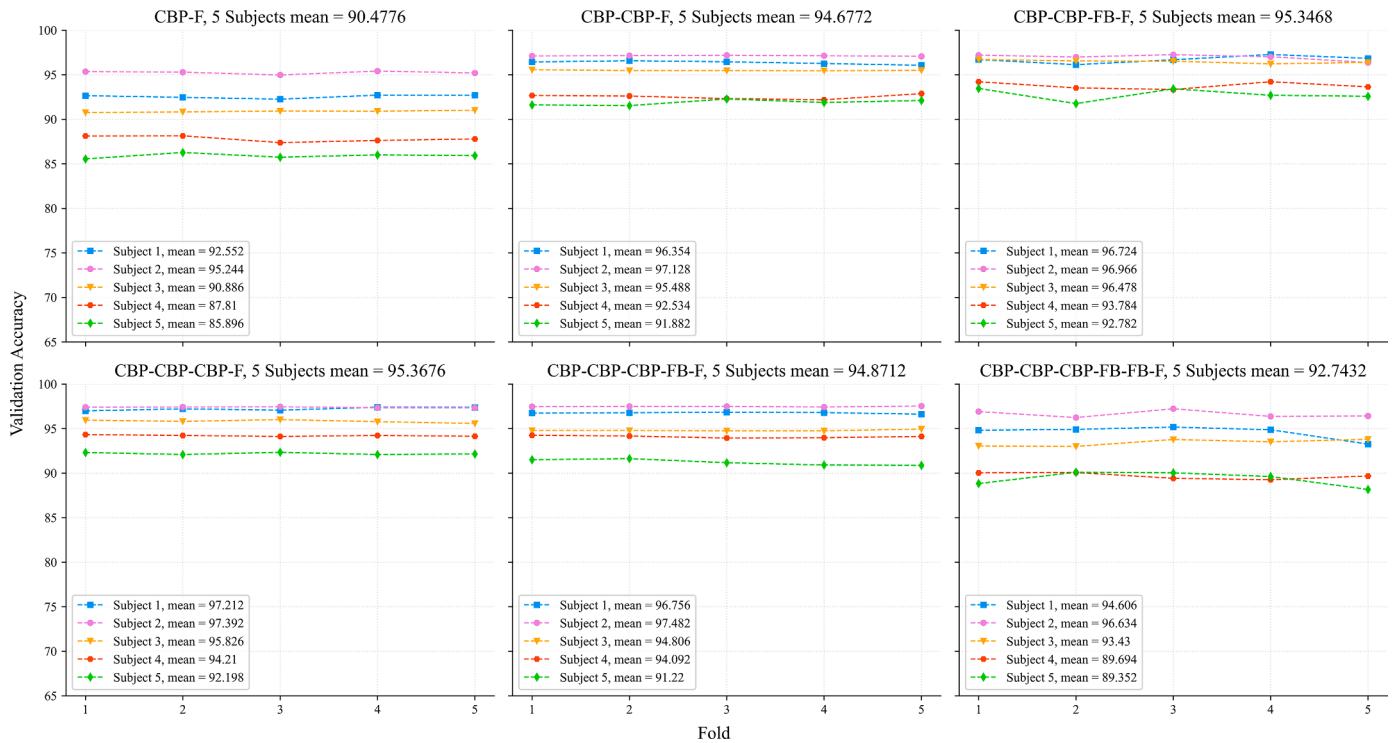


Fig. 5. Five subjects (S1-S5) validation accuracy of 5-fold cross-validation for different frameworks. C: Convolutional layer. B: Batch normalization. P: Pooling layer. F: Fully connected layer. For clarity, take CBP-CBP-FB-F as an example, two convolutional layers with batch normalization, two pooling layers with batch normalization, one fully connected layer with batch normalization, and one fully connected layer.

Table 1
The implementation details of the presented architecture.

	Activation shape	Activation Size	# Parameters
Input:	(32, 20, 1)	640	0
C1($f = 3, s = 1, p = 1$)	(32, 20, 32)	20,480	320
P1($f = 3, s = 2, p = 0$)	(15, 9, 32)	4320	0
C2($f = 3, s = 1, p = 1$)	(15, 9, 64)	8640	9248
P2($f = 3, s = 2, p = 0$)	(7, 4, 64)	1792	0
FC3	(64, 1)	64	114,752
FC4 + Softmax	(2, 1)	2	130

C1: the first convolutional layer, P1: the first pooling layer, C2: the second convolutional layer, P2: the second pooling layers, FC: fully connected layer; f denotes filter size, s denotes stride length, p denotes padding length; the activation shape of the convolutional layer and the pooling layer can be defined as: $(input_width, input_length, input_channel)$. # Parameters: the number of parameters. For the specific calculation method, please refer to Literature [21].

4.4. RQA w.r.t. real movements and imaginary movements

Recurrence quantification analysis (RQA) allows us to characterize the dynamics of the RP features and compare different motor movement/imagery tasks. For this purpose, we employed 2 RQA measures [30], i.e. recurrence rate (RR) and determinism (DET). The RR measures the density of recurrence points in the RP and the DET is introduced as a measurement for the predictability of the dynamic system. The larger value of DET implies that the RP are more 'deterministic'. Here in the MI-EEG dynamic systems, the higher 'deterministic' means that different movement patterns represented by RP have longer continuous functional network sequences that repeat at different times, i.e., there exist RP structure representing movement patterns. We used the t-test to assess whether the mean value of RR and DET measures was different between the MI and real execution groups in the PhysioNet dataset and GigaDB dataset. Fig. 6 shows that there are significant differences between MI and real execution in the two datasets. The group differences

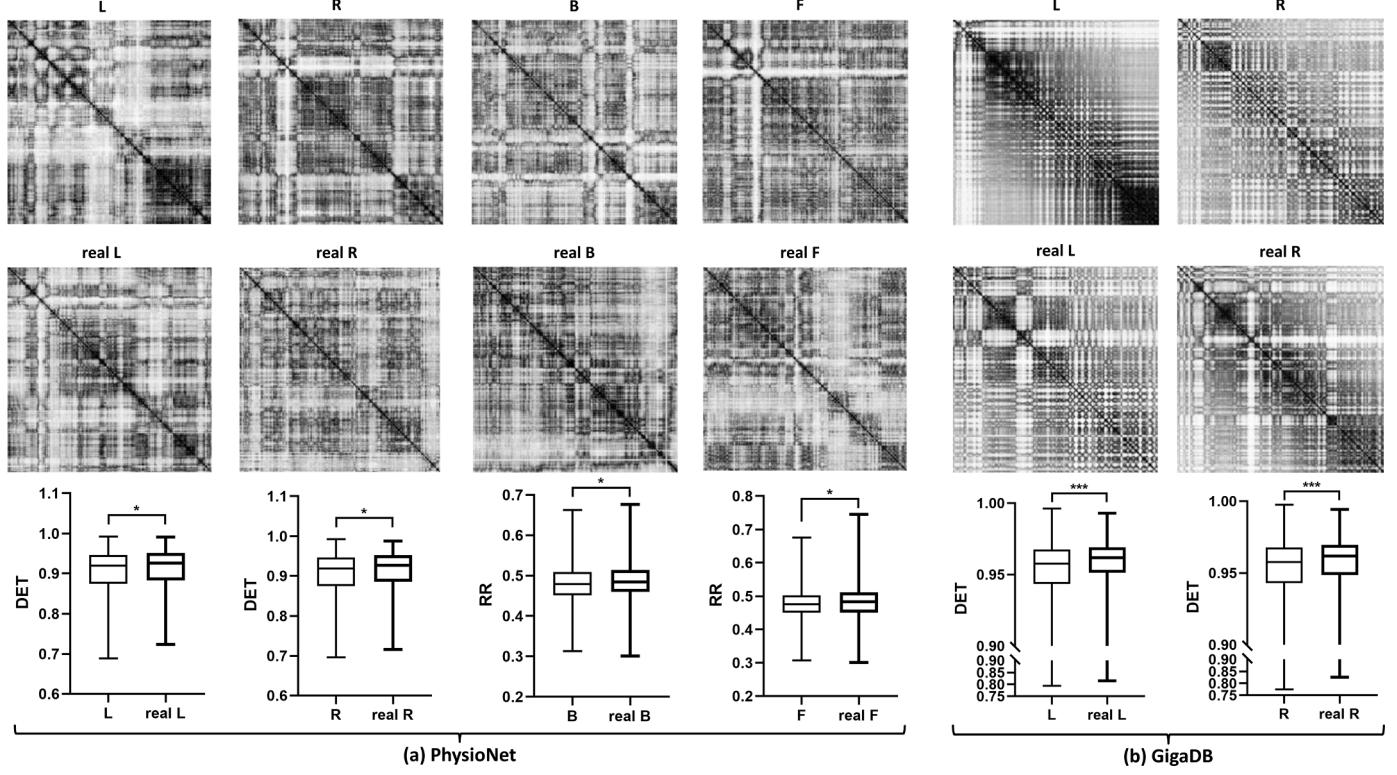


Fig. 6. RQA of RP features on MI and real execution tasks in PhysioNet and GigaDB datasets. *: $p < 0.05$, ***: $p < 0.001$.

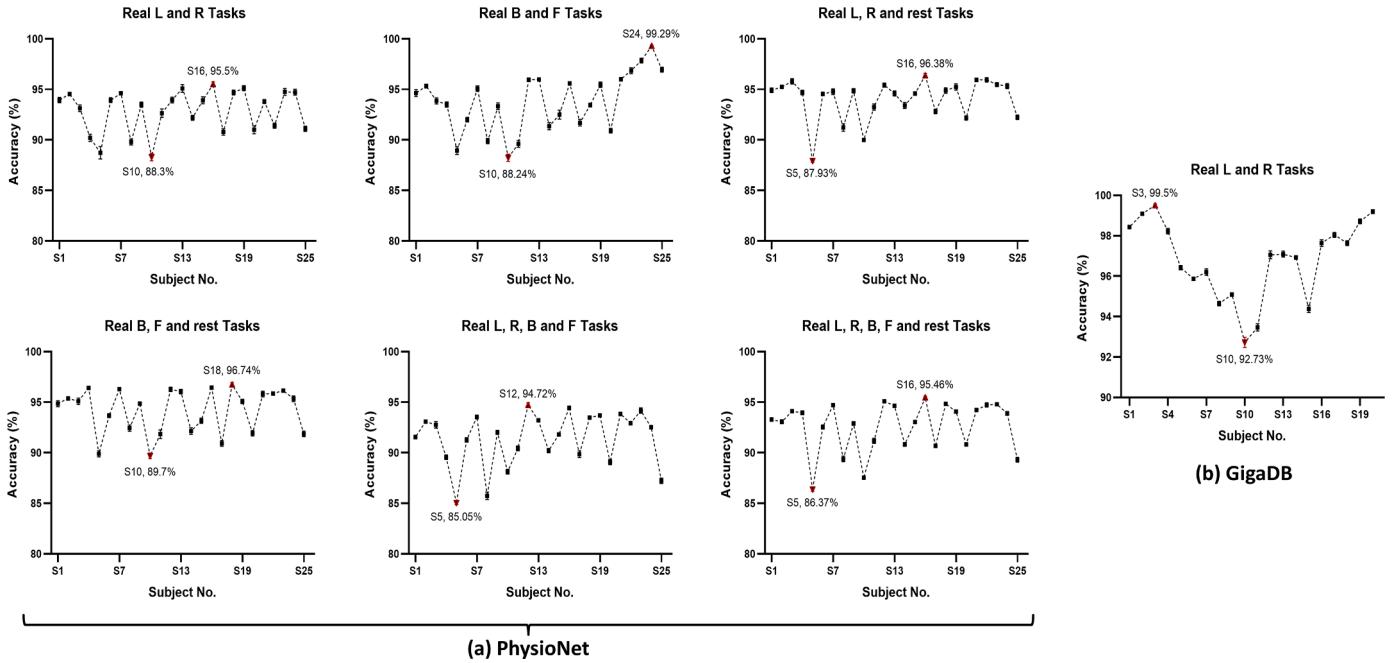


Fig. 7. The accuracy of two datasets in six real execution tasks. Note that ten times results were tested for each subject in each classification task.

observed reveal that RP structures may be used to characterize whether individuals perform MI or real execution tasks.

4.5. Classification w.r.t. real movements

To verify that the proposed model can handle EEG changes of two cases (real and imaginary movements), we have carried out classification experiments on both real and imaginary movements. Due to the

inter-individual variability in the EEG-based application, we tested the accuracy of each subject separately, and each result was tested ten times. For real movements situation, Fig. 7 presents the accuracy of 2-class (Real L, R or Real B, F), 3-class (Real L, R, rest or Real B, F, rest), 4-class (Real L, R, B, F), and 5-class (Real L, R, B, F, rest) from the PhysioNet dataset and 2-class (Real L, R) from the GigaDB dataset. On average, the accuracy of real L, R tasks, real B, F tasks, real L, R, rest tasks, real B, F, rest tasks, real L, R, B, F tasks, real L, R, B, F, rest tasks

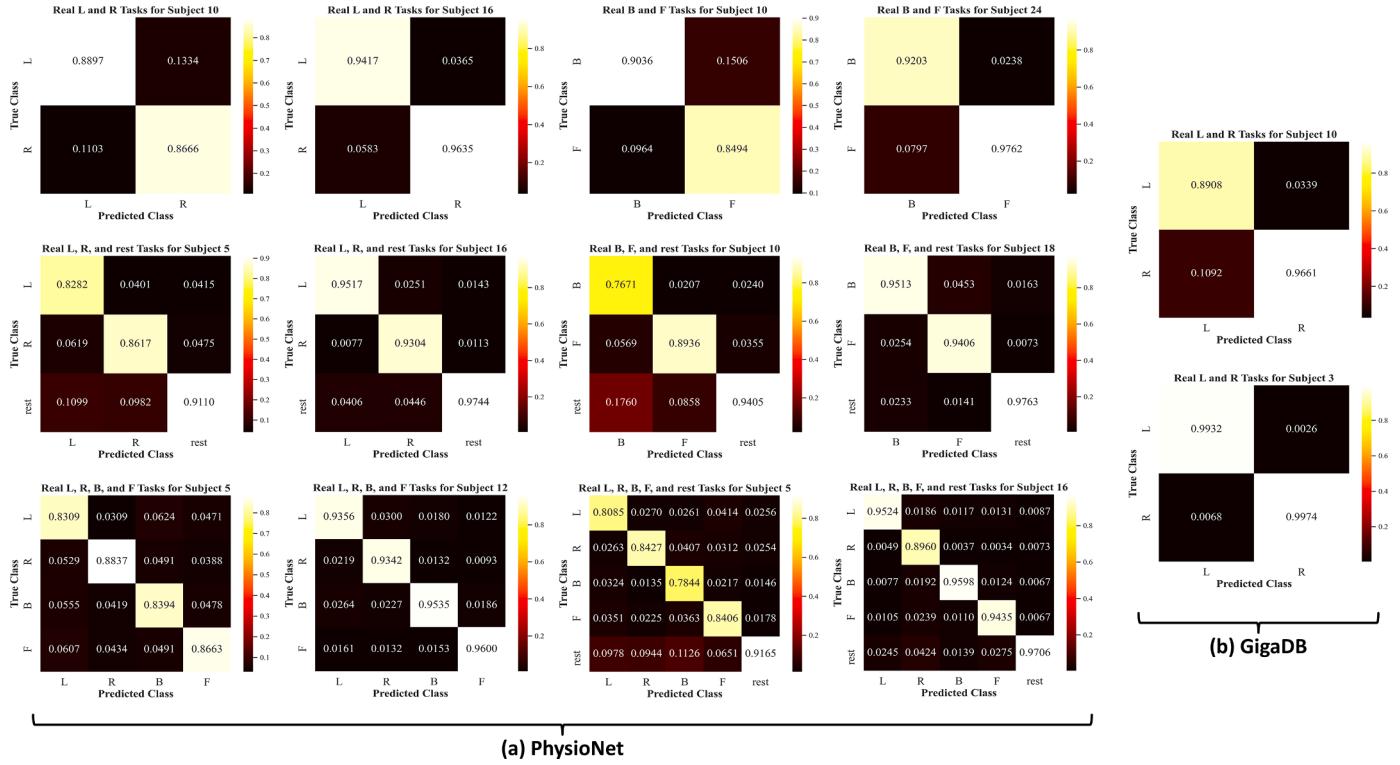


Fig. 8. The confusion matrices of real execution tasks on two datasets for different subjects. Each classification task has two confusion matrices calculated by the subjects who obtain the lowest and highest overall accuracy.

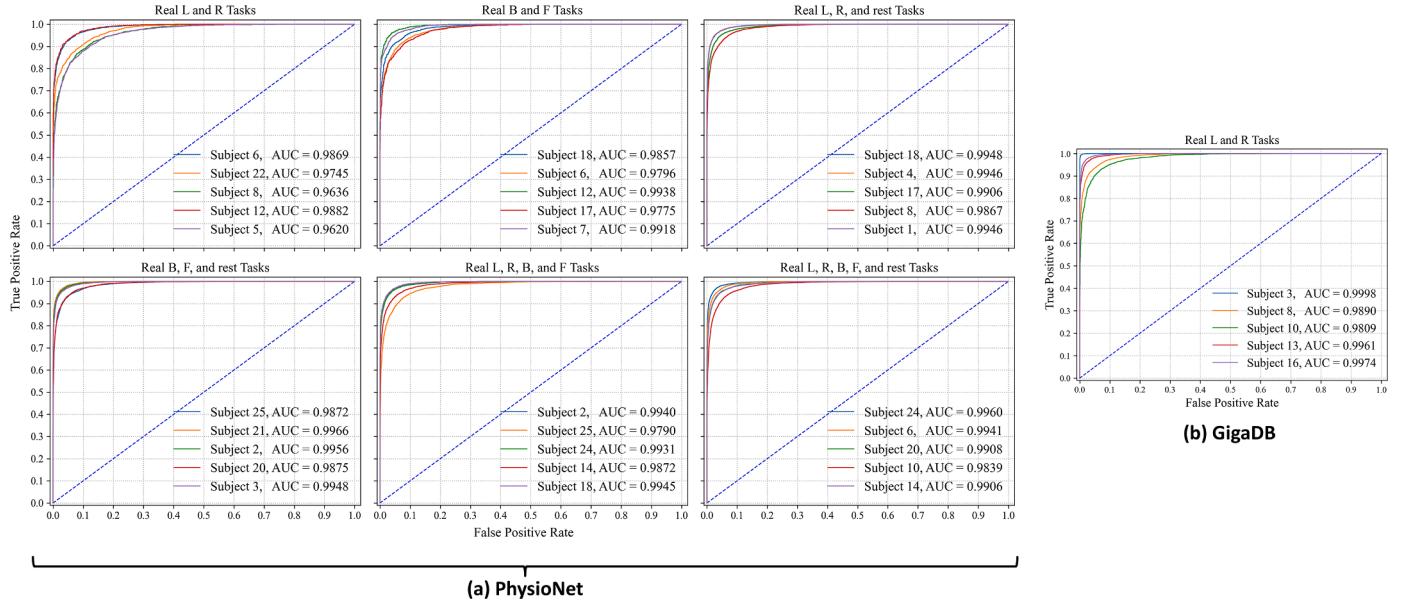


Fig. 9. ROC and AUC of 5 Subjects for two datasets. Note that five subjects in the different real execution tasks were randomly selected.

from the PhysioNet dataset and real L, R tasks from the GigaDB dataset are **92.85%**, **92.86%**, **94.06%**, **94.17%**, **91.37%**, **92.61%**, and **96.82%**, respectively. The results show that the accuracy in all cases remains above 85%, which is very impressive.

As shown in Fig. 7, we can observe the overall accuracy of two datasets, but the accuracy of a single movement is equally important. Thus, we calculated the confusion matrices of the subjects who obtained the lowest and highest overall accuracy in each task, revealing the accuracy of different movements in each task (see Fig. 8). The results show that the overall accuracy is more than 85% not only in some

Table 2
Evaluation metrics in real L, R tasks .

Subject	S6	S22	S8	S12	S5	Mean \pm SD
ACC (%)	93.96	90.52	89.24	93.96	88.96	91.33 \pm 2.21
Kappa (%)	87.9	81.09	78.52	87.87	77.94	82.66 \pm 4.39
Precision (%)	94.09	91.3	89.71	93.9	89.06	91.61 \pm 2.08
Recall (%)	93.9	90.68	89.37	93.97	89.02	91.39 \pm 2.15
F1 Score (%)	93.95	90.5	89.22	93.93	88.95	91.31 \pm 2.21

Table 3

Evaluation metrics in real B, F tasks.

Subject	S18	S6	S12	S17	S7	Mean±SD
ACC (%)	93.06	92.08	96.11	91.6	95.07	93.58±1.74
Kappa (%)	86.13	84.18	92.22	83.18	90.12	87.17±3.47
Precision (%)	93.13	92.24	96.11	91.62	95.23	93.67±1.73
Recall (%)	93.31	92.16	96.13	91.58	95.01	93.64±1.71
F1 Score (%)	93.05	92.08	96.11	91.59	95.06	93.58±1.74

Table 4

Evaluation metrics in real L, R, rest tasks.

Subject	S18	S4	S17	S8	S1	Mean±SD
ACC (%)	94.74	94.96	93.14	91.16	94.91	93.78±1.47
Kappa (%)	89.89	90.17	86.44	83.16	89.89	87.91±2.74
Precision (%)	92.33	93.14	92.23	87.6	95.08	92.08±2.46
Recall (%)	93.55	92.37	89.64	89.42	91.74	91.34±1.59
F1 Score (%)	92.9	92.72	90.87	88.28	93.31	91.62±1.87

Table 5

Evaluation metrics in real B, F, rest tasks.

Subject	S25	S21	S2	S20	S3	Mean±SD
ACC (%)	91.9	95.49	95.2	91.91	94.78	93.86±1.61
Kappa (%)	83.92	91.42	90.6	83.62	89.92	87.9±3.4
Precision (%)	90.48	93.04	94.26	92.02	92.67	92.49±1.24
Recall (%)	87.7	95.47	93.11	86.14	93.39	91.16±3.59
F1 Score (%)	88.76	94.18	93.65	88.73	93	91.66±2.41

classification tasks but also in some single movements, indicating that the proposed model works well.

Furthermore, we randomly picked five subjects to observe the classifier performance by calculating the Receiver Operating Characteristic (ROC) curve, Area Under ROC Curve (AUC), and five evaluation metrics. As shown in Fig. 9, the AUC values of five randomly selected subjects in each real execution task are more than 96%, proving that the proposed model with robustness and stability can adapt to different subjects in real execution tasks. As listed in Tables 2–7, the evaluation metrics of various tasks were calculated through Eq. (17)–(21). The experimental results show that five evaluation metrics of five randomly selected subjects from the PhysioNet dataset are more than 82% in real execution classification tasks, fully proving that the proposed model is robust and stable and can handle the individual variability problem in real execution tasks.

4.6. Classification w.r.t. imaginary movements

In this section, the results achieved by applying the proposed model to the imaginary movements are presented. The overall accuracy of different classification tasks from two datasets is shown in Fig. 10. On average, the accuracy of imaginary L, R tasks, imaginary B, F tasks, imaginary L, R, rest tasks, imaginary B, F, rest tasks, imaginary L, R, B, F tasks, imaginary L, R, B, F, rest tasks from PhysioNet dataset and imaginary L, R tasks from GigaDB dataset are **93.87%**, **94.28%**, **93.76%**, **93.77%**, **90.54%**, **91.85%**, and **96.58%**, respectively.

Furthermore, we can observe the confusion matrices to assess the performance of a single movement in the MI classification tasks through Fig. 11. The experimental results show that although the accuracy fluctuates slightly in the MI task, the overall accuracy is also more than 83%. In particular, the accuracies of a single imaginary movement in some MI tasks are also more than 99%, which indicates the proposed model works well for EEG MI decoding.

Next, in order to demonstrate the proposed model's performance of MI tasks further, we calculated ROC, AUC, and five evaluation metrics for various tasks. The ROC curve and the corresponding AUC were illustrated by Fig. 12, from which the AUC of six MI tasks from randomly picked subjects reaches 97% above, revealing that the proposed model

Table 6

Evaluation metrics in real L, R, B, F tasks.

Subject	S2	S25	S24	S14	S18	Mean±SD
ACC (%)	93.26	86.82	92.8	89.98	93.32	91.24±2.53
Kappa (%)	91.01	82.42	90.39	86.63	91.09	88.31±3.37
Precision (%)	93.35	87.39	92.88	90.08	93.33	91.41±2.35
Recall (%)	93.27	86.87	92.78	89.95	93.44	91.26±2.53
F1 Score (%)	93.27	86.88	92.8	90	93.32	91.25±2.51

Table 7

Evaluation metrics in real L, R, B, F, rest tasks.

Subject	S24	S6	S20	S10	S14	Mean±SD
ACC (%)	93.75	92.28	90.73	87.38	90.92	91.01±2.12
Kappa (%)	91.19	89.08	86.86	82.43	87.18	87.35±2.91
Precision (%)	91.84	91.17	89.33	85.42	89.67	89.49±2.23
Recall (%)	93.2	91.24	88.67	86.57	89.42	89.82±2.26
F1 Score (%)	92.5	91.15	88.82	85.89	89.51	89.57±2.24

achieves good effectiveness and robustness for EEG MI decoding. As listed in Table 8–13, the mean values of ACC, kappa, precision, recall, and F1 scores from all randomly picked subjects from the PhysioNet dataset are more than 85% in all MI tasks, which demonstrate the proposed model can decode the MI EEG characteristic well and is suitable for various subjects.

4.7. 2-class, 3-class, 4-class, 5-class classification in real and imaginary movements

As shown in Figs. 7 and 10, we intuitively observe that the accuracy on 4-class and 5-class fluctuates relatively up and down. In order to explore this phenomenon further, five subjects were selected for comparative experiments between different tasks. As illustrated by Fig. 13, the proposed model performs well at 2-class and 3-class with more than 90% average accuracy, while it performs relatively poorly on 4-class and 5-class (especially in imaginary movements), but more than 88% average accuracy yet.

To examine the results of the proposed model in more detail, to be clear, take the MI EEG data from S1 as an example (other subjects have similar results). As shown in Fig. 14, we extracted the prediction data in the classification process and used the Principal Component Analysis (PCA) method to present scatter plots for 2-class, 3-class, 4-class, and 5-class classification tasks to examine the classifier performance. The experimental results show that the proposed model can distinguish different features on 2-class, 3-class, and 4-class (partly overlap, which is the reason for the decline of accuracy, but it does not affect the overall classification results). The parts of class L, R, B, and F overlap with the rest state (see Fig. 14(f)), which increases the difficulty for classification and is also the reason for the relatively low accuracy. A similar observation has been documented for this phenomenon [18], probably because the subjects were resting without performing any movement during the MI process. In other words, if the EEG data characteristics are not prominent, the proposed model will classify the data as class rest.

4.8. Comparison with state-of-the-art methods

To verify the performance of the proposed method in terms of maximum accuracy (Max. ACC) and average accuracy (Avg. ACC), we compare our proposed method to other state-of-the-art methods, including Wavelet transform DNN [39], complex CSP equipped with the strong uncorrelating transform and SVM classifier (SUT-CCSP + SVM) [40], Phase information [41], the strong uncorrelating transform complex common spatial patterns (SUTCCSP) and random forest classifier [42], an iterative multiobjective optimization for channel selection (IMOCS) [43], RNNs-based method [44], CNNs-based method [18], scout EEG source imaging (ESI) with convolutional neural network

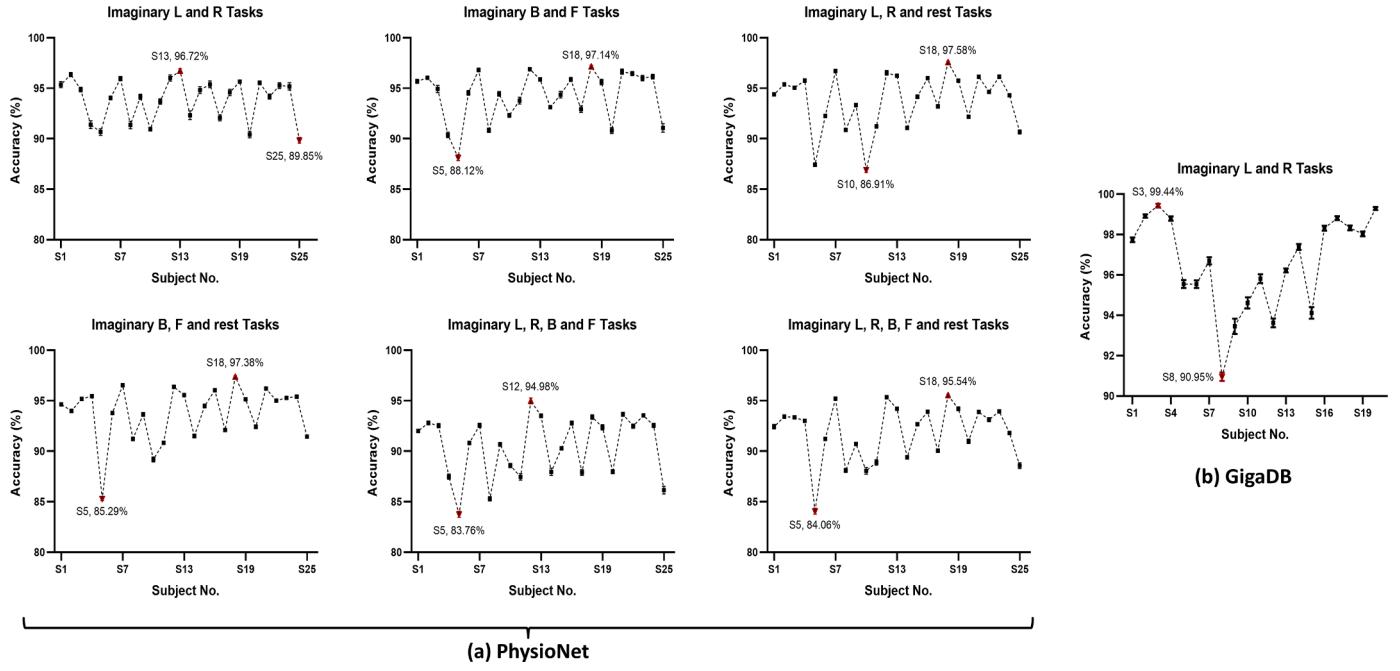


Fig. 10. The accuracy of two datasets in six MI tasks. Note that ten times results were tested for each subject in each classification task.

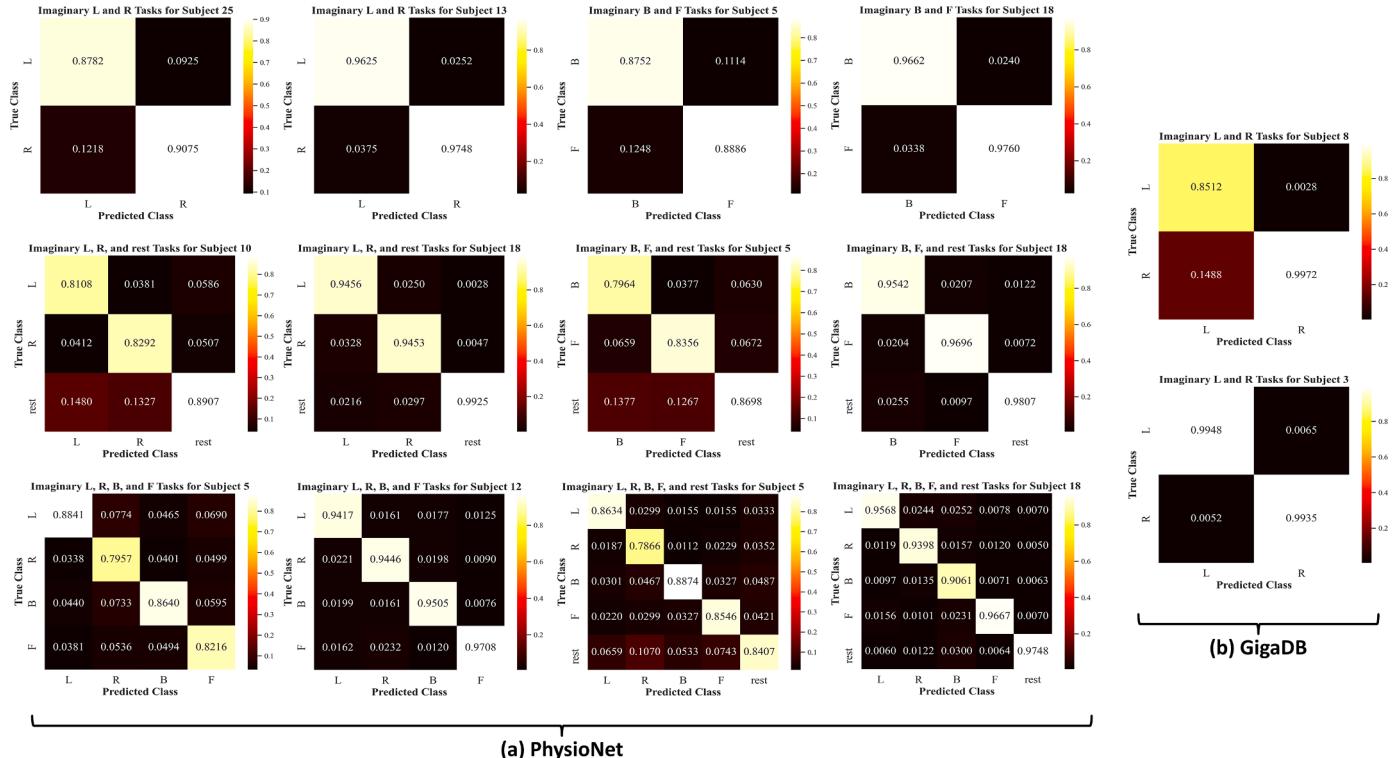


Fig. 11. The confusion matrices of MI tasks on two datasets for different subjects. Each classification task has two confusion matrices calculated by the subjects who obtain the lowest and highest overall accuracy.

(CNN) (ESI + CNNs) [20], and Graph-based Convolutional Recurrent Attention Model (G-CRAM) [45]. As listed in Table 14, all previous works performed the MI classification tasks except literature [41], however,

Loboda et al. [41] only carried out 2-class classification tasks, and the accuracy still needs to be improved. Furthermore, the literature mentioned above only performed 2-class, 3-class, or 4-class tasks but did

not conduct comprehensive research on 2-class, 3-class, 4-class, and 5-class classification tasks w.r.t. real/imaginary movements. To fill this gap, we performed the 2-class, 3-class, 4-class, and 5-class classification tasks w.r.t. real/imaginary movements. Note that all related works used the same data on the PhysioNet Dataset. In addition, we train the data of a single subject as a sample to verify the adaptability of our model to different subjects. In other words, the proposed model can handle the

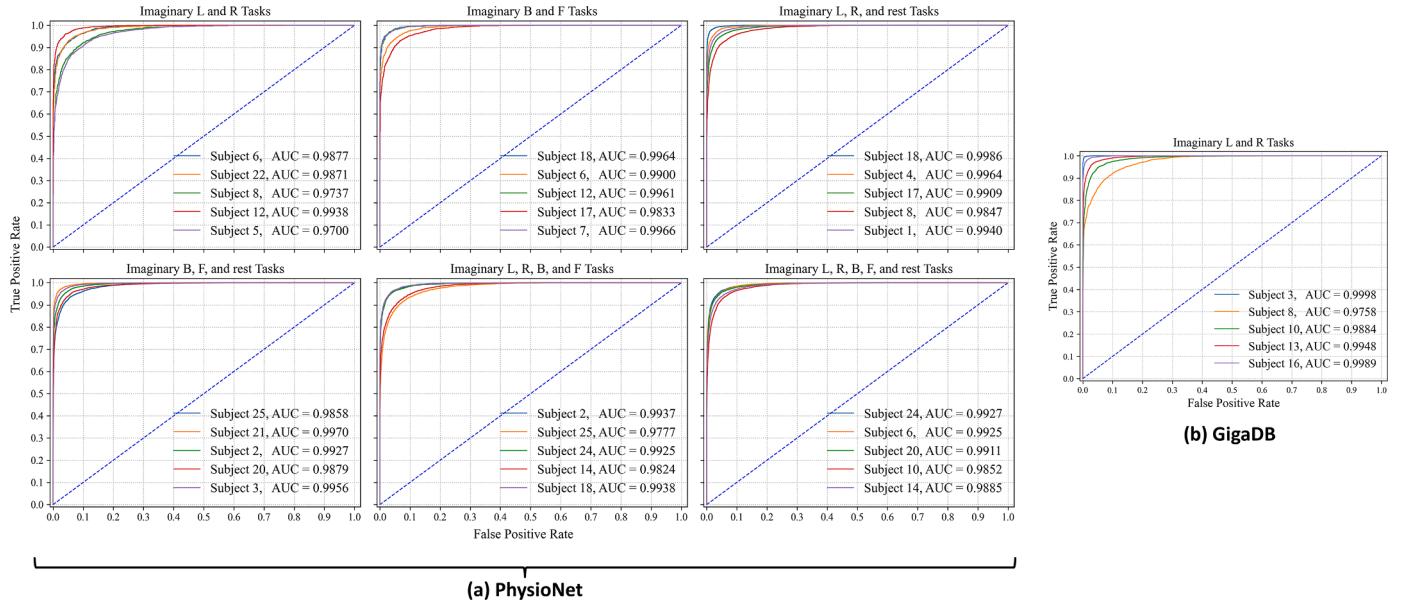


Fig. 12. ROC and AUC of 5 Subjects for two datasets. Note that five subjects in the different MI tasks were randomly selected.

Table 8

Evaluation metrics in MI L, R tasks.

Subject	S6	S22	S8	S12	S5	Mean ± SD
ACC (%)	93.75	93.85	90.87	95.94	90.28	92.94±2.09
Kappa (%)	87.45	87.71	81.72	91.82	80.44	85.83±4.19
Precision (%)	93.72	93.85	90.88	96.04	90.31	92.96±2.11
Recall (%)	93.73	93.87	90.84	95.83	90.16	92.89±2.1
F1 Score (%)	93.73	93.85	90.86	95.91	90.22	92.91±2.1

Table 9

Evaluation metrics in MI B, F tasks.

Subject	S18	S6	S12	S17	S7	Mean±SD
ACC (%)	96.84	94.41	96.88	93.09	96.6	95.56±1.54
Kappa (%)	93.68	88.81	93.75	86.15	93.19	91.12±3.09
Precision (%)	96.84	94.45	96.87	93.03	96.59	95.56±1.55
Recall (%)	96.85	94.38	96.88	93.17	96.62	95.58±1.53
F1 Score (%)	96.84	94.41	96.88	93.07	96.6	95.56±1.55

Table 10

Evaluation metrics in MI L, R, rest tasks.

Subject	S18	S4	S17	S8	S1	Mean±SD
ACC (%)	97.49	95.41	93.35	90.96	94.34	94.31±2.17
Kappa (%)	95.15	91.08	86.78	82.14	88.88	88.81±4.34
Precision (%)	96.09	93.13	92.71	88.99	93.45	92.87±2.27
Recall (%)	96.48	93.04	89.43	86.72	91.66	91.47±3.3
F1 Score (%)	96.28	93.05	90.97	87.8	92.52	92.12±2.77

Table 11

Evaluation metrics in MI B, F, rest tasks.

Subject	S25	S21	S2	S20	S3	Mean±SD
ACC (%)	91.41	96.25	93.99	92.36	95.23	93.85±1.78
Kappa (%)	83.21	92.6	88.15	85.02	90.65	87.93±3.46
Precision (%)	88.84	96.13	93.2	90.26	94.28	92.54±2.65
Recall (%)	87.84	94.09	91.06	89.05	93.1	91.03±2.36
F1 Score (%)	88.31	95.06	92.09	89.64	93.68	91.76±2.49

inter-individual variability problems.

Table 14 shows that the proposed method's classification performance is better than other comparison methods and the proposed method achieved 90% average accuracy above, which demonstrated

Table 12

Evaluation metrics in MI L, R, B, F tasks.

Subject	S2	S25	S24	S14	S18	Mean±SD
ACC (%)	93.06	86.6	93.02	87.81	93.11	90.72±2.9
Kappa (%)	90.73	82.11	90.69	83.73	90.81	87.61±3.87
Precision (%)	93.2	86.94	93.03	88.84	93.23	91.05±2.65
Recall (%)	92.96	86.58	93.01	87.72	93.16	90.69±2.91
F1 Score (%)	93	86.57	93.01	87.88	93.11	90.71±2.88

Table 13

Evaluation metrics in MI L, R, B, F, rest tasks.

Subject	S24	S6	S20	S10	S14	Mean±SD
ACC (%)	91.84	90.95	91.2	87.74	89.56	90.26±1.46
Kappa (%)	88.25	87.08	87.6	82.49	85.24	86.13±2.08
Precision (%)	92.95	91.11	89.31	87.8	88.84	90.0±1.82
Recall (%)	89.05	88.6	89.8	84.92	88.06	88.09±1.68
F1 Score (%)	90.87	89.75	89.5	86.23	88.34	88.94±1.57

that combining the RP and the BCNNs is effective for EEG-based real execution and motor imagery classification tasks. There are two reasons as follows: (1) the RP can quantify the correlation information between all brain regions and reduce the complexity of the original EEG signals. We utilize the RP features to characterize EEG signals because the RPs are easily obtained for the dynamical systems (the brain can be viewed as complex/dynamical systems). They describe important information about time scales, which can be easily interpreted and quite challenging to obtain in other cases. (2) In BCNNs, use probability distributions as weights rather than the single point-estimates, which is more robust to over-fitting and can easily learn from small datasets. In detail, the BCNNs model integrates the parameters by using a prior probability distribution, and the average value is calculated on many models during training, which provides regularization effects for the network to prevent over-fitting.

5. Discussion

In this study, we demonstrated that combining the recurrence plot (RP) and Bayesian Convolutional Neural Networks (BCNNs) can improve the performance of real execution/motor imagery classification

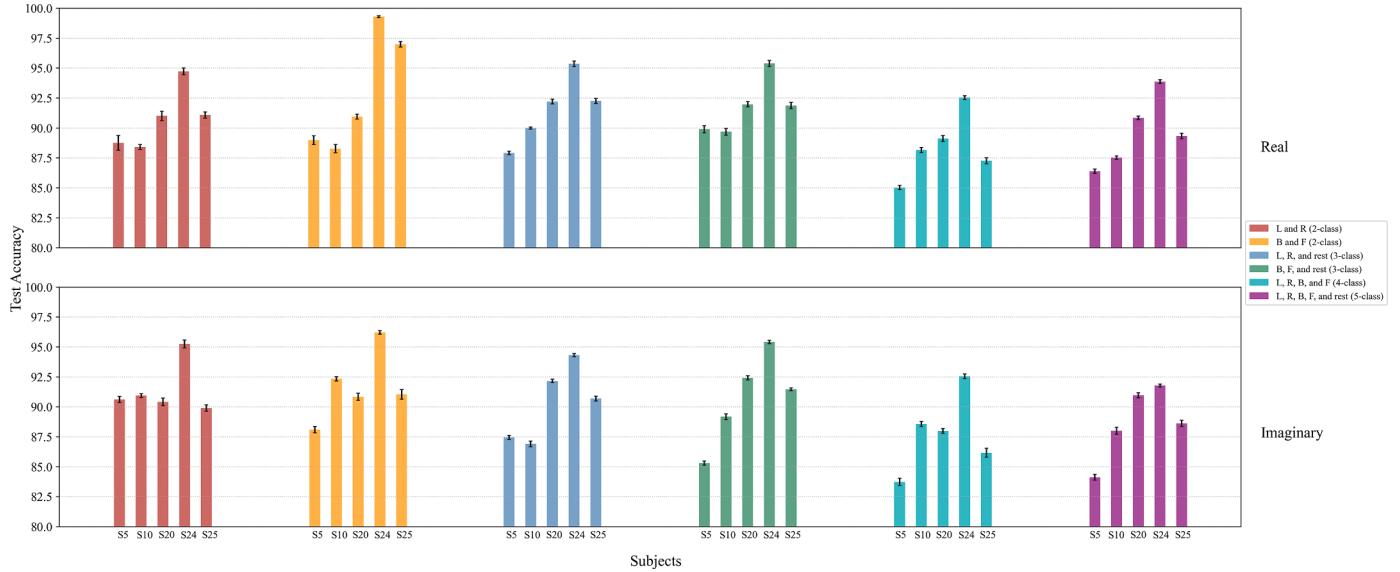


Fig. 13. Test accuracy of 2-class (L, R or B, F), 3-class (L, R, rest or B, F, rest), 4-class (L, R, B, F), 5-class (L, R, B, F, rest) Classification tasks from five subjects (S5, S10, S20, S24, S25) for real and imaginary movements.

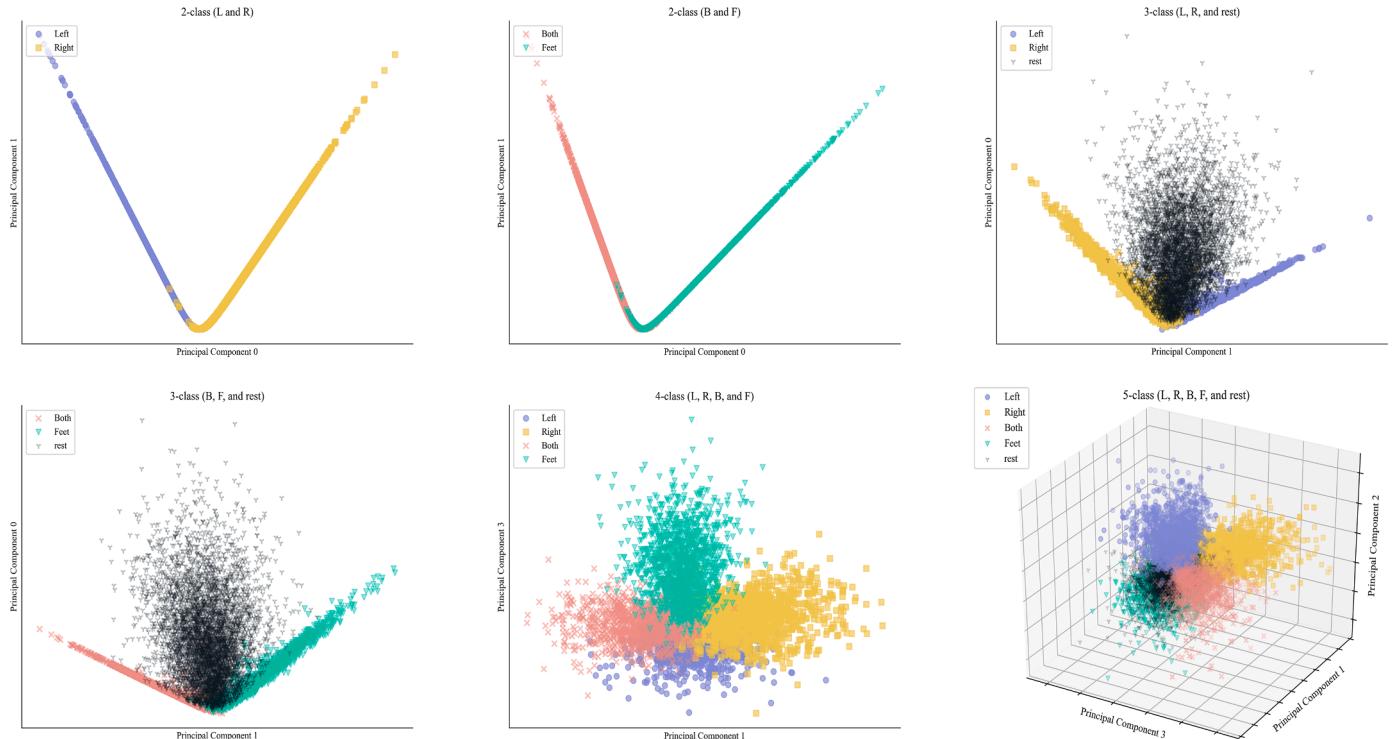


Fig. 14. PCA for 2-class, 3-class, 4-class, and 5-class tasks from S1 MI EEG data. (f): The 3D scatter diagram is used for display because the 2D scatter plot can't display the 5-class classification results well.

tasks and handle the common inter-individual variability problems in EEG-based applications. To clarify the findings, we performed the classification experiments w.r.t. real/imaginary movements. Note that a single subject was used for each experiment to test the generalization ability of the proposed model to different subjects. Specifically, we employed a recurrence plot (RP) computation for preprocessed EEG signals of each channel and merged all RPs into one based on the weighted average method, which can extract EEG signals correlation information between all brain regions well and reduce the complexity of the original EEG signals. Furthermore, Bayesian Convolutional Neural Networks (BCNNs), a method based on the Bayesian posterior inference,

is utilized to classify 2-class, 3-class, 4-class, and 5-class classification tasks through analyzing the RP features. Our results indicate that the proposed model can improve the performance of real execution/motor imagery classification and handle the challenge of individual variability between subjects based on the above method.

5.1. Comparison of 4-class and 5-class classification tasks

We have also shown in Section 4.7 that the proposed model achieves the accuracy of 5-class classification tasks in imaginary movements is unexpectedly higher than that in the 4-class classification task because it

Table 14

Comparison with state-of-the-art methods on the PhysioNet Dataset.

Method	Subjects	Training	Class Type	Max. ACC	Avg. ACC
Wavelet transform DNN [39]	4	Subject	MI 2-class	72.82%	68.21%
SUT-CCSP + SVM [40]	56	Group	MI 2-class	90%	72.37%
Phase information [41]	103	Group	Real 2-class	78.95%	–
		Group	MI 2-class	71.55%	–
SUT-CCSP Random forest [42]	24	Group	MI 2-class	–	80.05%
		Subject	MI 2-class	96.13%	–
IMOCS [43]	85	Group	MI 2-class	–	63%
	35	Group	MI 2-class	–	79.9%
RNNs [44]	12	Subject	MI 5-class	82.65%	68.2%
CNNs [18]	105	Group	MI 2-class	–	87.98%
		Group	MI 3-class	–	76.61%
		Group	MI 4-class	–	65.73%
ESI + CNNs [20]	10	Group	MI 4-class	–	94.5%
	3	Subject	MI 4-class	94.54%	94.07%
G-CRAM [45]	105	Group	MI 2-class	74.71%	–
Proposed method	25	Subject	Real 2-class	99.29%	92.86%
			Real 3-class	96.74%	94.12%
			Real 4-class	94.72%	91.37%
			Real 5-class	95.46%	92.61%
			MI 2-class	97.14%	94.07%
			MI 3-class	97.58%	93.77%
			MI 4-class	94.98%	90.54%
			MI 5-class	95.54%	91.85%

Training: training with a group of subjects or with individual subjects. **Class Type:** real execution or motor imagery. **Max. ACC:** maximum classification accuracy. **Avg. ACC:** average classification accuracy. Note that the Avg. ACC in our work is the average result of 10 times experiments, Avg. ACC in 2-class is the complete result of L, R tasks and B, F tasks, and Avg. ACC in 3-class is the complete result of L, R, rest tasks and B, F, rest tasks.

is reasonable that the accuracy decreases with the increase of classification task. There is also the case for the accuracy of Real 4-class and Real 5-class or MI 4-class and MI 5-class in Table 14. As shown in Fig. 13, The reason for the decline of the accuracy in MI tasks on 4-class and 5-class may be that the subjects are distracted during the EEG acquisition process. After all, the EEG of real movements is more accessible to collect compared to the MI process. Significantly, the accuracy of 4-class classification tasks in imaginary movements is lower than that of 5-class tasks. A reasonable explanation is that the EEG signals of the subjects in the rest state are the easiest to obtain because the subjects do not perform any task in this process, and there are very few artifacts. Therefore, excellent EEG signals can be collected to produce excellent RP features. Finally, the model can better distinguish the rest state, improving the overall accuracy in the classification stage. As shown in Fig. 14(f), the rest state is in the center of the four categories (L, R, B, F), and the proposed model can accurately distinguish the rest state. Hence, the proposed model can separate the five categories well. We think this phenomenon should be further studied in the future.

5.2. Limitation

Our study subjects were selected from the PhysioNet dataset and GigaDB dataset to test the proposed model, so it is unknown whether our model is applicable to other datasets, such as BCI Competition IV datasets 1 or BCI Competition III datasets IVa, which is publicly available^{3 4}. Hence, further studies with other datasets are needed. On the other hand, we fuse the RP features of 20 channels into one RP by the weighted average method, which may lose some time-series information, frequency and spatial domain features from EEG signals. A feasible approach is to treat the time-series of all channels as an M-dimensional real-valued time-series and use the multiplex network method to handle it [14], which is also an aspect of our follow-up research.

6. Conclusions

Taken together, this paper, for the first time, presented a novel model that combines the recurrence plot (RP) and Bayesian Convolutional Neural Networks (BCNNs), which decoded the nonlinear dynamic features of EEG signals and handle the inter-individual variability problems in real execution/motor imagery classification tasks. Experimental studies on the same subjects from the PhysioNet dataset and GigaDB dataset indicated that combining the RP features and BCNN classifier was beneficial for three reasons. First, since EEG signals contained various nonlinear dynamic information, creating a series of RP features was conducive to collecting as much EEG information as possible. Second, the recurrence quantification analysis (RQA) revealed that the RP structures may be used to characterize whether individuals perform MI or real execution tasks. Hence, the corresponding RP features effectively recognized the characteristics of different tasks. Third, the weight is represented by probability distribution rather than point estimation in the BCNNs, which is beneficial for learning input features. Therefore, the proposed RP-BCNNs model was suitable for learning nonlinear dynamic features.

The proposed method has several worthwhile future research directions. The first direction is to explore the nonlinear dynamic features in other EEG-based BCI fields, such as emotion recognition and epileptic seizure classification. Due to its generation scheme, the proposed RP feature representations are beneficial to conserving nonlinear dynamic information of task-related and can be directly applied to other BCI fields. Meanwhile, the BCNNs module would also be conducive to extracting the nonlinear dynamic features. The other research direction is that to develop any real-time BCI systems that have gained popularity in recent times and helps to promote the advancement of brain science. We speculate that the real execution/motor imagery online system combined with nonlinear dynamic feature extraction method can attract great attention of BCI community.

CRediT authorship contribution statement

Wenqie Huang: Methodology, Visualization, Investigation, Writing – original draft, Writing – review & editing. **Guanghui Yan:** Supervision, Visualization, Methodology, Resources, Funding acquisition.

³ <http://www.bbci.de/competition/iv/>

⁴ <http://www.bbci.de/competition/iii/>

Wenwen Chang: Supervision, Resources, Conceptualization, Writing – review & editing, Funding acquisition. **Yuchan Zhang:** Software, Formal analysis, Investigation, Data curation, Visualization. **Yueling Yuan:** Data curation, Formal analysis, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by (i) the National Natural Science Foundation of China under grant 62062049, (ii) the Science and Technology Project of Lanzhou City in China under grant 2021-1-150, (iii) the Special Funds for Guiding Local Scientific and Technological Development by The Central Government under grant 22ZY1QA005, (iv) the Science Project of Gansu Province under grant 21ZD8RA008, (v) the Open Project Program of the National Laboratory of Pattern Recognition (NLPR) under grant 202100020, and (vi) the Open Project of Cognitive EEG and Transcranial Electrical Stimulation Regulation from "neuracle" company under grant BRKOT-LZJTU-20220329C.

References

- [1] J.R. Wolpaw, Brain–computer interfaces. *Handb. Clin. Neurol.*, Elsevier, 2013, pp. 67–74.
- [2] G. Prasad, P. Herman, D. Coyle, S. McDonough, J. Crosbie, Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study, *J. Neuroeng. Rehabil.* 7 (2010) 1–17.
- [3] F. Aloise, F. Schettini, P. Aricò, L. Bianchi, A. Riccio, M. Mecella, F. Babiloni, D. Mattia, F. Cincotti, Advanced brain computer interface for communication and control, in: *Proc. Int. Conf. Adv. Vis. Interfaces*, 2010, pp. 399–400.
- [4] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, *Sensors* 12 (2012) 1211–1279, <https://doi.org/10.3390/s12020121>.
- [5] R. Zhang, Y. Li, Y. Yan, H. Zhang, S. Wu, T. Yu, Z. Gu, Control of a wheelchair in an indoor environment based on a brain-computer interface and automated navigation, *IEEE Trans. Neural Syst. Rehabil. Eng.* 24 (2016) 128–139, <https://doi.org/10.1109/TNSRE.2015.2439298>.
- [6] M.A. Silver, D. Ress, D.J. Heeger, Topographic maps of visual spatial attention in human parietal cortex, *J. Neurophysiol.* 94 (2005) 1358–1371, <https://doi.org/10.1152/jn.01316.2004>.
- [7] H. Wang, X. Dong, Z. Chen, B.E. Shi, Hybrid gaze/EEG brain computer interface for robot arm control on a pick and place task, in: *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, IEEE, 2015, pp. 1476–1479, <https://doi.org/10.1109/EMBC.2015.7318649>.
- [8] M. Tariq, P.M. Trivailo, M. Simic, Detection of knee motor imagery by Mu ERD/ERS quantification for BCI based neurorehabilitation applications, in: 2017 11th Asian Control Conf., 2017, pp. 2215–2219, <https://doi.org/10.1109/ASCC.2017.8287519>.
- [9] N. Brodu, F. Lotte, A. Lécuyer, Exploring two novel features for EEG-based brain–computer interfaces: multifractal cumulants and predictive complexity, *Neurocomputing* 79 (2012) 87–94, <https://doi.org/10.1016/j.neucom.2011.10.010>.
- [10] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, *IEEE Trans. Biomed. Eng.* 58 (2011) 355–362, <https://doi.org/10.1109/TBME.2010.2082539>.
- [11] A.S. Aghaei, M.S. Mahanta, K.N. Plataniotis, Separable common spatio-spectral patterns for motor imagery BCI systems, *IEEE Trans. Biomed. Eng.* 63 (2016) 15–29, <https://doi.org/10.1109/TBME.2015.2487738>.
- [12] N. Marwan, J.F. Donges, Y. Zou, R.V. Donner, J. Kurths, Complex network approach for recurrence analysis of time series, *Phys. Lett. A.* 373 (2009) 4246–4254, <https://doi.org/10.1016/j.physleta.2009.09.042>.
- [13] R.V. Donner, M. Small, J.F. Donges, N. Marwan, Y. Zou, R. Xiang, J. Kurths, Recurrence-based time series analysis by means of complex network methods, *Int. J. Bifurc. Chaos*. 21 (2011) 1019–1046, <https://doi.org/10.1142/S0218127411029021>.
- [14] L. Lacasa, V. Nicosia, V. Latora, Network structure of multivariate time series, *Sci. Rep.* 5 (2015) 15508, <https://doi.org/10.1038/srep15508>.
- [15] D. Eroglu, N. Marwan, M. Stebich, J. Kurths, Multiplex recurrence networks, *Phys. Rev. E.* 97 (2018) 12312, <https://doi.org/10.1103/PhysRevE.97.012312>.
- [16] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, *Europhys. Lett.* 4 (1987) 973–977, <https://doi.org/10.1209/0295-5075/4/9/004>.
- [17] Y. Shen, H. Lu, J. Jia, Classification of motor imagery EEG signals with deep learning models, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Springer Verlag, 2017: pp. 181–190. https://doi.org/10.1007/978-3-319-67777-4_16.
- [18] H. Dose, J.S. Möller, H.K. Iversen, S. Puthuserrypaday, An end-to-end deep learning approach to MI-EEG signal classification for BCIs, *Expert Syst. Appl.* 114 (2018) 532–542, <https://doi.org/10.1016/j.eswa.2018.08.031>.
- [19] Y. Sun, F.P.-W. Lo, B. Lo, EEG-based user identification system using 1D-convolutional long short-term memory neural networks, *Expert Syst. Appl.* 125 (2019) 259–267, <https://doi.org/10.1016/j.eswa.2019.01.080>.
- [20] Y. Hou, L. Zhou, S. Jia, X. Lun, A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN, *J. Neural Eng.* 17 (2020) 16048, <https://doi.org/10.1088/1741-2552/ab4af6>.
- [21] W. Huang, W. Chang, G. Yan, Z. Yang, H. Luo, H. Pei, EEG-based motor imagery classification using convolutional neural networks with local reparameterization trick, *Expert Syst. Appl.* 187 (2022), 115968, <https://doi.org/10.1016/j.eswa.2021.115968>.
- [22] Z. Fan, X. Xi, Y. Gao, T. Wang, F. Fang, M. Houston, Y. Zhang, L. Li, Z. Lü, Joint filter-band-combination and multi-view CNN for electroencephalogram decoding, *IEEE Trans. Neural Syst. Rehabil. Eng.* 1 (2023), <https://doi.org/10.1109/TNSRE.2023.3269055>.
- [23] W. Huang, W. Chang, G. Yan, Y. Zhang, Y. Yuan, Spatio-spectral feature classification combining 3D-convolutional neural networks with long short-term memory for motor movement/imagery, *Eng. Appl. Artif. Intell.* 120 (2023), 105862, <https://doi.org/10.1016/j.engappai.2023.105862>.
- [24] B.H. Kim, J.W. Choi, H. Lee, S. Jo, A discriminative SPD feature learning approach on Riemannian manifolds for EEG classification, *Pattern Recognit.* 143 (2023), 109751, <https://doi.org/10.1016/j.patcog.2023.109751>.
- [25] M. Zeng, X. Zhang, C. Zhao, X. Liu, Q. Meng, GRP-DNet: a gray recurrence plot-based densely connected convolutional network for classification of epileptiform EEG, *J. Neurosci. Methods*. 347 (2021), 108953, <https://doi.org/10.1016/j.jneumeth.2020.108953>.
- [26] M.B. Khodabakhshi, V. Saba, A nonlinear dynamical approach to analysis of emotions using EEG signals based on the Poincaré map function and recurrence plots, *Biomed. Eng. /Biomed. Tech.* 65 (2020) 507–520, <https://doi.org/10.1515/bmt-2019-0121>.
- [27] X.J. Meng, S. Qiu, S. Wan, K. Cheng, L. Cui, A motor imagery EEG signal classification algorithm based on recurrence plot convolution neural network, *Pattern Recognit. Lett.* 146 (2021) 134–141, <https://doi.org/10.1016/j.patrec.2021.03.023>.
- [28] M. Ahmadlou, H. Adeli, Functional community analysis of brain: a new approach for EEG-based investigation of the brain pathology, *Neuroimage* 58 (2011) 401–408, <https://doi.org/10.1016/j.neuroimage.2011.04.070>.
- [29] J. Zhang, W. Cheng, Z. Liu, K. Zhang, X. Lei, Y. Yao, B. Becker, Y. Liu, K. M. Kendrick, G. Lu, J. Feng, Neural, electrophysiological and anatomical basis of brain-network variability and its characteristic changes in mental disorders, *Brain* 139 (2016) 2307–2321, <https://doi.org/10.1093/brain/aww143>.
- [30] N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, *Phys. Rep.* 438 (2007) 237–329, <https://doi.org/10.1016/j.physrep.2006.11.001>.
- [31] A. Graves, Practical variational inference for neural networks, in: J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, K.Q. Weinberger (Eds.), *Adv. Neural Inf. Process. Syst.*, Curran Associates, Inc., 2011, in: <https://proceedings.neurips.cc/paper/2011/file/7eb3c8be3d411e8ebfab08eba5f49632-Paper.pdf>.
- [32] C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight uncertainty in neural networks, in: *Proc. 32nd Int. Conf. Mach. Learn. - Vol. 37*, JMLR.org, 2015, pp. 1613–1622.
- [33] K. Neklyudov, D. Molchanov, A. Ashukha, D. Vetrov, Variance networks: when expectation does not meet your expectations, *Int. Conf. Learn. Represent.* (2019). <https://openreview.net/forum?id=B1GAUs0ckQ>.
- [34] K. Shridhar, F. Laumann, M. Liwicki, Uncertainty estimations by Softplus normalization in Bayesian convolutional neural networks with variational inference, *ArXiv* (2019). <https://arxiv.org/abs/1806.05978>.
- [35] D.P. Kingma, T. Salimans, M. Welling, Variational dropout and the local reparameterization trick, *Adv. Neural Inf. Process. Syst.* (2015).
- [36] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J. E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet, *Circulation* 101 (2000) E215–E220, <https://doi.org/10.1161/01.CIR.101.23.e215>.
- [37] H. Cho, M. Ahn, S. Ahn, M. Kwon, S.C. Jun, Supporting data for “EEG datasets for motor imagery brain computer interface, GigaScience Database (2017), <https://doi.org/10.5524/100295>.
- [38] J.-S. Bang, M.-H. Lee, S. Fazli, C. Guan, S.-W. Lee, Spatio-spectral feature representation for motor imagery classification using convolutional neural networks, *IEEE Trans. Neural Networks Learn. Syst.* (2021) 1–12, <https://doi.org/10.1109/TNNLS.2020.3048385>.
- [39] M. Tolic, F. Jovic, Classification of wavelet transformed EEG signals with neural network for imagined mental and motor tasks, *Kinesiology* 45 (2013) 130–138.
- [40] C. Park, C.C. Took, D.P. Mandic, Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks, *IEEE Trans. Neural Syst. Rehabil. Eng.* 22 (2014) 1–10, <https://doi.org/10.1109/TNSRE.2013.2294903>.

- [41] A. Loboda, A. Margineanu, G. Rotariu, A. Mihaela, Discrimination of EEG-based motor imagery tasks by means of a simple phase information method, *Int. J. Adv. Res. Artif. Intell.* 3 (2014), <https://doi.org/10.14569/ijarai.2014.031002>.
- [42] Y. Kim, J. Ryu, K.K. Kim, C.C. Took, D.P. Mandic, C. Park, Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns, *Comput. Intell. Neurosci.* 2016 (2016), 1489692, <https://doi.org/10.1155/2016/1489692>.
- [43] V.S. Handiru, V.A. Prasad, Optimized bi-objective eeg channel selection and cross-subject generalization with brain-computer interfaces, *IEEE Trans. Human-Machine Syst.* 46 (2016) 777–786, <https://doi.org/10.1109/THMS.2016.2573827>.
- [44] X. Ma, S. Qiu, C. Du, J. Xing, H. He, Improving EEG-based motor imagery classification via spatial and temporal recurrent neural networks, in: *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2018, 2018, pp. 1903–1906, <https://doi.org/10.1109/EMBC.2018.8512590>.
- [45] D. Zhang, K. Chen, D. Jian, L. Yao, Motor imagery classification via temporal attention cues of graph embedded EEG signals, *IEEE J. Biomed. Heal. Informatics.* 24 (2020) 2570–2579, <https://doi.org/10.1109/JBHI.2020.2967128>.

Wenqie Huang is a doctor's degree candidate. He received a master's degree in computer software and theory from Lanzhou Jiaotong University in 2022. His research interests include deep learning, brain-computer interface, motor imagery, and classification.

Guanghui Yan (Senior Member, IEEE) received a Ph.D. degree from Northwestern Polytechnical University, Xi'an, in 2009. He is currently a Professor with the School of Electronics and Information Engineering, Lanzhou Jiaotong University. He has published more than 50 articles in journals and conferences. His current research interests include database theory and systems, the Internet-of-Things engineering and application, data mining, and complex network analysis. He is a member of the China Computer Federation (CCF).

Wenwen Chang received a Ph.D. degree in mechatronic engineering from Northeastern University, Shenyang, China, in 2019. He is currently an Assistant Professor with the School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China. His current research interests include brain-computer interaction, functional brain networks, pattern recognition, and their applications in cognitive science and engineering.

Yuchan Zhang is a master's degree candidate. She received a bachelor's degree in Software engineering from Taiyuan University of Technology in 2020. Her research interests include deep learning, brain-computer interface, emotion recognition, and classification.

Yueteng Yuan received the bachelor's degree from the School of Electronics and Information Engineering, Lanzhou Jiaotong University, in June 2020. Now she is the master student of Lanzhou Jiaotong University. Her current research interests include brain computer interface, deep learning, and virtual reality based BCI system.