EEG-inception: An accurate and robust end-to-end neural network for EEG-based motor imagery classification

Article *in* Journal of Neural Engineering · March 2021

DOI: 10.1088/1741-2552/abed81

CITATIONS
53

READS
279

3 authors:

Ce Zhang
Virginia Tech (Virginia Polytechnic Institute and State University)
20 PUBLICATIONS **131** CITATIONS
SEE PROFILE

Young-Keun Kim
Handong Global University
71 PUBLICATIONS **625** CITATIONS
SEE PROFILE

Azim Eskandarian
Virginia Tech (Virginia Polytechnic Institute and State University)
190 PUBLICATIONS **4,514** CITATIONS
SEE PROFILE

EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification

Journal of Neural Engineering

PAPER

EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification

Ce Zhang[1,*], Young-Keun Kim[2] and Azim Eskandarian[1]

[1] Mechanical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States of America
[2] Mechanical and Control Engineering, Handong Global University, Pohang, Gyeongsang, Republic of Korea
[*] Author to whom any correspondence should be addressed.

E-mail: zce@vt.edu

RECEIVED
26 October 2020

REVISED
18 February 2021

ACCEPTED FOR PUBLICATION
10 March 2021

PUBLISHED
25 March 2021

Keywords: brain–computer interface (BCI), electroencephalography (EEG), motor imagery (MI), neural network, time series data augmentation

Abstract

Objective. Classification of electroencephalography (EEG)-based motor imagery (MI) is a crucial non-invasive application in brain–computer interface (BCI) research. This paper proposes a novel convolutional neural network (CNN) architecture for accurate and robust EEG-based MI classification that outperforms the state-of-the-art methods. *Approach.* The proposed CNN model, namely EEG-inception, is built on the backbone of the inception-time network, which has showed to be highly efficient and accurate for time-series classification. Also, the proposed network is an end-to-end classification, as it takes the raw EEG signals as the input and does not require complex EEG signal-preprocessing. Furthermore, this paper proposes a novel data augmentation method for EEG signals to enhance the accuracy, at least by 3%, and reduce overfitting with limited BCI datasets. *Main results.* The proposed model outperforms all state-of-the-art methods by achieving the average accuracy of 88.4% and 88.6% on the 2008 BCI Competition IV 2a (four-classes) and 2b datasets (binary-classes), respectively. Furthermore, it takes less than 0.025 s to test a sample suitable for real-time processing. Moreover, the classification standard deviation for nine different subjects achieves the lowest value of 5.5 for the 2b dataset and 7.1 for the 2a dataset, which validates that the proposed method is highly robust. *Significance.* From the experiment results, it can be inferred that the EEG-inception network exhibits a strong potential as a subject-independent classifier for EEG-based MI tasks.

1. Introduction

Brain–computer interface (BCI) is a direct pathway to communicate between the human brain and external devices [1]. Electroencephalography (EEG) has become one of the most common brain activity recording methods because it is non-invasive and low-cost. An important EEG-based BCI study area is motor imagery (MI), which triggers neural activities by imagining the movement of the body (e.g. left-hand and right-hand movement) [2–4]. Decoding the correct MI neural activities allows patients with motor neuron diseases (e.g. stroke and Parkinson's disease) to rehabilitate partial body movement skills with external devices' assistance. Other popular EEG-based MI application include wheelchair control [5], robot arm operation [6], and quadcopter manipulation [7].

EEG-based MI signals are non-stationary, where the signal properties such as variance and mean change with time [8]. Furthermore, EEG signals have a low signal-to-noise ratio (SNR) due to numerous artifacts and noises [9]. Therefore, decoding the EEG-based MI signals requires advanced signal processing techniques and statistical learning algorithms. There have been many investigations on the EEG-based MI tasks classification, which can be categorized into conventional machine learning algorithms and deep neural network algorithms.

Conventional machine learning algorithms are composed of three steps: signal pre-processing, feature extraction, and feature classification. The signal pre-processing step's objective is to remove artifacts such as muscle/ocular movement and system noises. As for the feature extraction step, most of them are designed based on known knowledge and previous

experiences about human brain dynamics and patterns. For feature classification, popular linear classifiers, such as support vector machine (SVM), and linear discriminant analysis (LDA), are applied for MI classification. Since the EEG signals' SNR is low, the extracted features that based on known knowledge are usually covered by noises and artifacts. Furthermore, most popular classification algorithms are linear, which is not suitable for non-stationary signal classification. Therefore, using conventional machine learning algorithms results in lower MI classification accuracy. To overcome this, various deep learning based neural network models have been proposed in the literature for EEG-based MI classification.

The most common neural network architectures applied to the EEG-based MI research are convolutional neural network (CNN) and recurrent neural network (RNN). The CNN models employ one or several customized kernel matrices for MI features extraction at both the temporal and the spatial domains. Because of the non-linearity in feature extraction and the large number of trainable parameters, a properly designed CNN models achieves higher classification accuracy than most conventional machine learning algorithms. However, the CNN models' sizes are large, which results in higher computation load. Another popular neural network architecture is RNN. The RNNs use previous time-series signal output as the input, which allows the weights of each neuron to be shared across time. Compared with CNN models, most RNNs are designed for extracting temporal domain features only. However, the RNN models generally require less computation load to achieve a relatively high classification accuracy. In summary, neural network models can automatically extract and classify EEG-based MI features. Moreover, the classification accuracy is dramatically improved as compared to the most conventional machine learning algorithms.

However, the existing EEG-based MI neural network models still have several issues that require more extensive research. The first issue is further enhancing the classification accuracy. Even though the neural network-based classification accuracy is superior as compared to conventional machine learning algorithms, most existing methods accuracies still ranges from high 70% to low 80%, which are not practical for real-world applications. The second issue is the availability of EEG datasets. Since the EEG-based MI experiments on human subjects are complex processes, the available dataset size is too small. The last issue is developing a subject-independent model. Currently, EEG-based MI classifiers are subject-dependent due to variations of neuronal feedbacks among different subjects. It is necessary to develop subject-independent MI models to be readily applied to new subjects without acquisitions of training datasets.

Therefore, this paper proposes a new CNN network to increase the accuracy of MI-EEG signal classification and robustness to subject-dependency. The proposed network is called EEG-inception and uses several inceptions [10] and residual [11, 12] modules as its backbone. Also, to tackle the limitation in the training data, a new data augmentation of EEG signals is proposed, which increases the average accuracy by 3%.

The contributions of the EEG-inception model can be categorized as:

- A CNN-based classification achieving 88.6% in the average accuracy outperforming all other state-of-the-art methods for binary classes dataset and 88.4% for four classes dataset.
- A novel data augmentation method for EEG signal to reduce overfitting and further improve classification accuracy with small training data size.
- A High robustness subject-dependent dataset with a low standard deviation in classifying different subjects.
- High potential for a subject-independent EEG-based MI classification.

The rest of the paper is organized as follows. Section 2 illustrates recent related work in EEG-based MI studies. Section 3 explains the EEG-inception neural network and the novel EEG data augmentation method. Section 4 describes the experimental protocol of the open-source dataset that we used for evaluation. Section 5 presents the proposed EEG-inception model performance and the data augmentation effectiveness results. In section 6, we conclude the proposed algorithm results and propose potential new future works based on the EEG-inception neural network.

2. Related works

As mentioned in the first chapter, EEG-based MI classification algorithms can be separated into two categories: (a) conventional machine learning algorithms and (b) neural network-based algorithms. In this chapter, we are going to discuss the related work in each category each category.

2.1. Machine learning-based classification

For classical machine learning algorithms, the EEG signals decoding process includes signal pre-processing, feature extraction, and feature classification.

The objective of signal pre-processing is to remove noises and artifacts. Common pre-processing methods are independent component analysis (ICA) for ocular artifacts removal, canonical correlation

analysis for muscle artifacts removal, and bandpass filtering [13–15].

The MI feature extraction algorithms contain data analysis in time, frequency, and spatial domains. In the time domain, the EEG event-related synchronization and desynchronization (ERD/ERS) features are determined [16]. The frequency-domain analysis is often combined with the time domain through wavelet transforms (WTs) or short-time Fourier transforms (STFT) [17–19]. According to Sadiq's work, empirical WT has also been applied for EEG-based MI classification [20, 21]. In the spatial domain, common spatial pattern (CSP), proposed by Pfurtscheller *et al*, is proved to be effective for EEG-based MI classification [22]. After the CSP algorithm had been developed, Ang has developed a new CSP-based filter bank (FBCSP) to improve the classification accuracy [23]. Based on the FBCSP, Zhang and Eskandarian have proposed a computationally efficient CSP algorithm to minimize the computing load [24]. Besides the CSP algorithm, spatial feature dimension reduction has also been proved to be an effective technique. Sadiq has applied different feature reduction algorithms such as principal component analysis, and ICA for MI feature extraction [25, 26].

As for feature classification, since the input signals are thoroughly processed, traditional classifiers such as *k*th nearest neighbor, LDA, and SVM have been proven successful for classification [27–30].

Conventional machine learning algorithms' advantages are (a) relatively simple algorithm implementation and (b) faster training period compared with neural network-based algorithms. However, since the conventional machine learning algorithms require manual feature extraction, only limited features can be extracted based on existing algorithms. Furthermore, it is known that the EEG-based MI classification performance heavily relies on feature extraction effectiveness. Therefore, the conventional machine learning algorithm classification results are relatively limited.

2.2. Neural network-based algorithms

Researchers have recently used neural network-based MI classification methods, which have shown better performance than conventional methods. The two popular neural network structures for EEG-based MI studies are CNN and RNN.

2.2.1. CNN architecture

The CNN structure can be categorized as one-dimensional CNN (1D-CNN) and multi-dimension CNN (2D/3D-CNN).

The idea of 1D-CNN is to convolve and extract the time series or frequency domain EEG signal features. Li *et al* proposed a novel multi-layers 1D-CNN neural network architecture (CP-MixedNet) for MI classification [31]. In their study, 44 channels of raw EEG signals were employed as input, and multiple 1D-CNN layers were used for spatial, temporal feature extraction. The classification accuracy results indicate that the CP-MixedNet outperforms traditional machine learning algorithms such as FBCSP. Besides convolving the time-series signals, EEG frequency amplitude is also another important input. Tabar *et al* proposed taking STFT based time-frequency amplitude as input for 1D convolution and a stack auto-encoder architecture for classification [32]. They have compared the CNN combined with the auto-encoder neural network (CNN-SAE) with the CNN architecture and other state-of-the-art algorithms. The results show that the CNN-SAE network average accuracy is around 77.6%, which is better than the 1D-CNN architecture and SVM classification. However, the proposed auto-encoder classifier has eight layers with the neuron number ranging from 900 to 2, which causes excessive computational load and long training periods. Miao *et al* have developed a deep 1D-CNN architecture to extract spatial and frequency features for MI feature extraction [33]. According to their comparison, their proposed model is around 10% higher than the CSP-based feature extraction methods.

Since each 1D-CNN layer can only convolve and extract EEG features in one dimension (time, frequency, or spatial), some researchers have applied 2D/3D-CNN to extract features from multiple dimensions at the same time. For EEG-based MI studies, most multi-dimension CNN structure ideas are borrowed from object detection algorithms such as AlexNet [34]. Lee *et al* proposed an end-to-end CNN for multiclass MI classification [35]. Their proposed ERA-CNN model is inspired by the hierarchical deep CNN developed for visual recognition. The ERA-CNN implements multiple 2D-CNN with kernel matrix size ranges from 36 to 288. The proposed neural network achieves around 66% for seven-class MI classification. Inspired by visual recognition, some researchers have converted the EEG time series classification problem to an image classification problem. Yang *et al* have proposed a CNN model with a multiple dimensional kernel matrix [36]. They have converted the EEG signals into an image as input and compared the 2D and 3D kernel matrix feature extraction and classification performance. According to the experiment results, the 2D kernel matrix generally exhibits higher accuracy than the 3D kernel matrix.

2.2.2. RNN architecture

For the RNN model, long-short term memory (LSTM), inspired by natural language processing, is one of the most popular EEG-based MI classification structures. Since raw EEG signals are in time series, the time domain's correlation can be used for MI class prediction. Wang *et al* have employed

the LSTM to achieve robust classification on EEG-based MI [37]. In the proposed LSTM model, the input signals are the normalized raw EEG data. After one-dimensional aggregate approximation and channel selection, the EEG signals are fed into the LSTM model. The proposed algorithm number of parameters is only 746, and the average classification accuracy for group I dataset is 76.47%. Even though the classification accuracy is not the best compared with other state-of-art algorithms, the small size of model parameters shows that the computational load is much lower than other neural networks. Besides directly feeding time-series EEG signal to the LSTM model, Jeong *et al* designed a CNN and LSTM combination model for MI decoding and a robotic arm control [38]. The EEG signals feature are extracted through a pre-trained CNN, then the extracted features are predicted by LSTM networks. Based on their two experiments, the robotic arm moving towards correct directions success rates are 0.6 and 0.43, respectively.

Based on the previous works on neural network study, we conclude that (a) the EEG-based MI classification accuracy needs to be further improved for real-world application, and (b) an effective subject independent algorithm is still not well developed.

3. Proposed network architecture

This section introduces the architecture of the proposed EEG-inception model and the data augmentation method. The overall workflow of the proposed EEG-inception model is shown in figure 1.

3.1. EEG-inception neural network

The proposed EEG-inception network's backbone is an ensemble of multiple inception modules and residual modules, as shown in figure 1. The inception module is inspired by image classification. It extracts features from both the depth (series extraction) and the width (parallel extraction). Deeper feature extraction can be achieved by adding more inception modules, while a wider feature extraction can be obtained by applying more convolutional kernel matrices in each inception module. The residual module, initially inspired by the ResNET, is used to diminish the 'vanishing gradient problem' caused by deeper layers and activation functions.

The proposed network comprises of six inception modules (one initial inception modules and five intermediate inception modules) and two residual modules, as illustrated in figure 1. Details about the inception and residual module functions are described below.

3.1.1. Inception module

The proposed algorithm inception module can be categorized into 'initial inception module' and 'intermediate inception module'. Both modules contain a bottleneck layer, multiple convolutional layers, a pooling layer, a batch normalization, and an activation function (figures 2(a) and (b)). The difference between the 'initial' and the 'intermediate inception module' is the bottleneck layer. In this section, we explain the details about every block in the inception module accordingly.

The bottleneck layer, inspired by network in network [39], is a $[1 \times 1]$ kernel matrix originally designed for decreasing computation load by reducing the depth of the input data dimension. It is proved to be a useful technique for feature reduction in the computer vision object detection field. However, for our initial inception module, we inverse the bottleneck layer by increasing the input data dimension from N to M (3 to 12 for the binary dataset, and 22 to 48 for the four-classes dataset). The reason to increase the dimension is that the provided EEG samples are one-dimensional time-series signals with a limited number of channels (three for binary classes and twenty-two for four-classes). Thus, the feature map size is too small for effective feature extraction. Therefore, the objective of inversing the bottleneck layer at the initial inception module is to increase the number of trainable weights for deeper feature extraction. For the intermediate inception module, the bottleneck layer is applied for decreasing the data dimension for lower computation load. The intermediate inception module bottleneck layer reduces the data dimension from $[4 \times M]$ to M. The details about the choice of the data dimension reduction size are discussed in the ablation study in section 5.

The convolutional layer is designed for extracting parallel features based on multiple time length. By adding the number of convolutional kernel matrices, EEG features can be extracted in parallel with various time lengths by selecting different kernel sizes. In this study, we have employed three 1D convolutional matrices for the binary dataset with kernel sizes of $[25 \times 1]$, $[75 \times 1]$, and $[125 \times 1]$, respectively (figure 2(a)). For the four-classes dataset, we have used five 1D convolutional matrices with the kernel size of $[25 \times 1]$, $[75 \times 1]$, $[125 \times 1]$, $[175 \times 1]$, and $[225 \times 1]$, respectively (figure 2(b)). The reason to choose the kernel sizes to be the times of 25 is that the provided data sample rate is 250 Hz. Thus, the extracted features from the EEG signal are based on 0.1 s, 0.3 s, and 0.5 s, etc. Furthermore, applying more kernel matrices for the four-classes data is because the class number is higher than the binary classes, which need more parallel features from varied time lengths to improve the classification accuracy.

The pooling layer is designed for features downsampling. In the proposed algorithm, the max-pooling is applied to select the maximum feature from the input data with a kernel size of 25 (0.1 s). After the pooling layer, a 1D convolutional layer with a $[1 \times 1]$ kernel matrix is applied to enlarge the depth of the data dimension to increase the number of learning parameters.

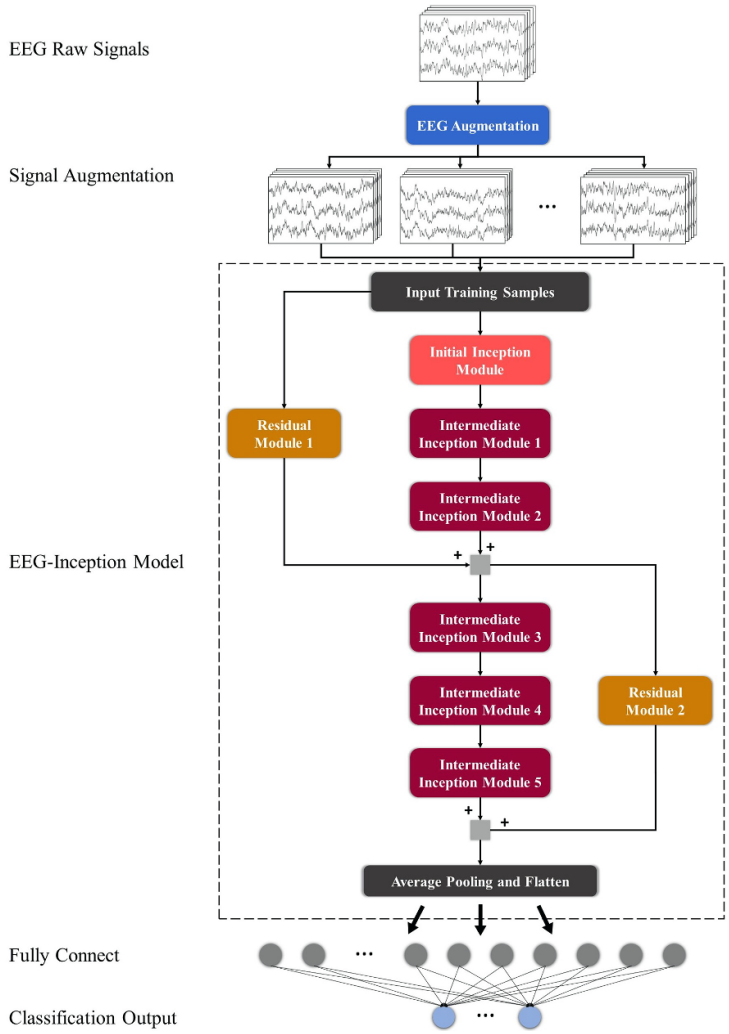

Figure 1. Proposed algorithm overall scheme. The proposed algorithm contains a novel data augmentation method, an EEG-inception model, and classification.

The batch normalization layer is a standard normalization algorithm that prevents overfitting and decrease training steps [40].

At the end of an inception module, a ReLU activation function is employed to preserve the features of the positive values. Compared with the traditional sigmoid and the 'tanh"()' activation function, the ReLU function overcomes the 'vanishing gradient problem', allowing the model to gain higher accuracy.

3.1.2. Residual module

According to figure 1, the residual module is applied after every three inception modules. The residual module is a convolutional layer with a kernel size of 1×1, and the equation is

$$y = \mathcal{F}(x_0, \{W_i\}) + x \quad (1)$$

where x_0 is the input vector, $\mathcal{F}(x_0, \{W_i\})$ represents the residual mapping from the input vector, x is the output layer from the inception module, and y is the summation between the residual layer output and the inception module output. He *et al* prove that the residual learning framework can effectively solve the learning degradation problem caused by deeper layers [34]. Thus, in our EEG-inception module, we implement a residual module for every three inception modules to prevent learning degradation. For both residual modules kernel matrices, the input and output dimension agrees with the inception modules' input and output for dimension consistency.

3.2. Data augmentation

The objective of data augmentation is to minimize the overfitting issue by transforming the training data to an extend the data size. EEG-based MI data size is usually small due to the lengthy and challenging experiment [41]. Thus, it is necessary to conduct data augmentation processing for EEG-based MI classification. Unlike standard computer vision data augmentation, EEG signals are non-stationary, which cannot be rotated, stretched, or scaled because these methods change the time-series signals properties. Therefore, noise addition is one of the best

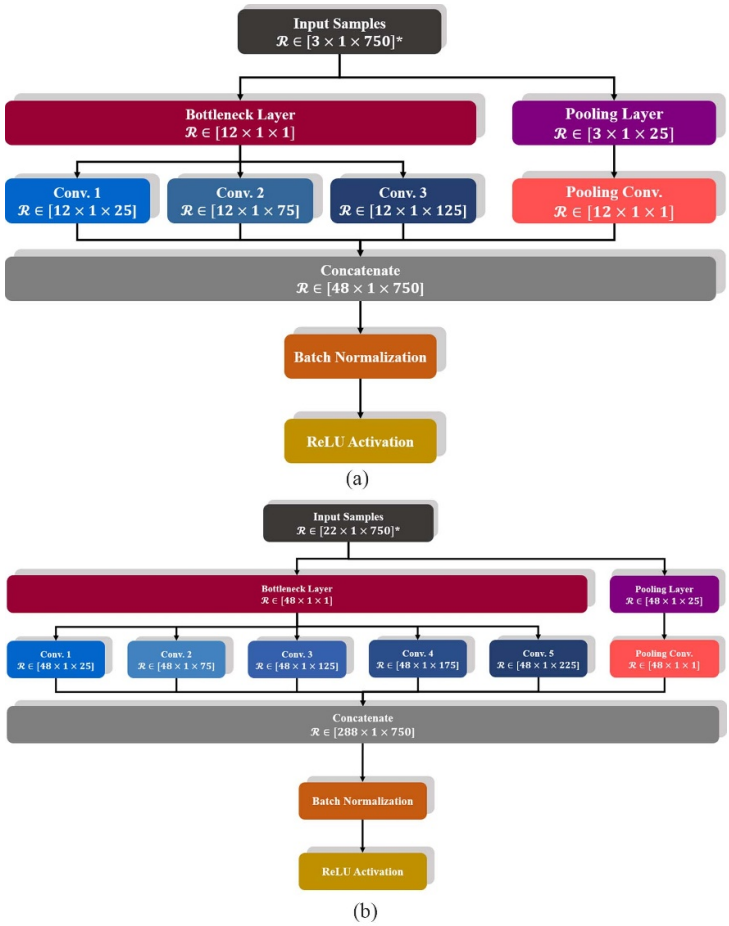

Figure 2. (a) Proposed EEG-inception neural network overall architecture for the binary-classes dataset. *The initial inception module input data depth is 3 while the intermediate inception module data depth is 48 due to concatenate from the previous inception module. (b) Proposed EEG-inception neural network overall architecture for the four-classes dataset. *The initial inception module input data depth is 22 while the intermediate inception module data depth is 288 due to concatenate from the previous inception module.

methods to expand the data size [42]. According to Muthukumaraswamy's work, most brain activities exist in a frequency range from 0 to 100 Hz, and the frequency above 100 Hz can be considered as artifacts and noises [43]. Thus, our proposed data augmentation method extracts the above-100 Hz signal from one trial, then apply it to another trial, as presented in figure 3.

According to figure 3, we have designed an 8th order Butterworth high-pass filter with a cut-off frequency of 100 Hz to extract the noise candidates at first. Then, the original EEG signals are used to subtract their noise candidates. Finally, noise candidates extracted from another trial are added, as shown in equation (2)

$$S_{\text{aug}}(i) = S_0(i) - S_n(i) + S_n(k) \tag{2}$$

where $S_{\text{aug}}(i)$ is the augmented signal for the ith trial, $S_0(i)$ is the ith trial original signal, S_n is the noise candidates, and k represents a random trial number. The benefit of the proposed data augmentation method is that it can increase the data size by $(n-1)$ times at most. However, increasing the data size to $(n-1)$ times than the original signal can cause potential signal repetition and overfitting issues. Therefore, in our study, we increase the training data size by three times larger than the original training size for the binary dataset and six times larger for the four-classes dataset. The augmented signal comparison with the original signal is shown in figures 4 and 5.

3.3. Ablation study on layer depth

An ablation study, commonly applied in neuroscience research, is employed for artificial neural network performance analysis [44]. The objective is to investigate the change in classification accuracy with different numbers of neural network layers or different features. In this study, we have tested our neural network by varying the depth of the convolutional kernel matrix. The convolutional kernel matrix dimensions were varied from 6 to 64 for the binary-classes dataset and 24 to 84 for the four-classes dataset. According to our assumption,

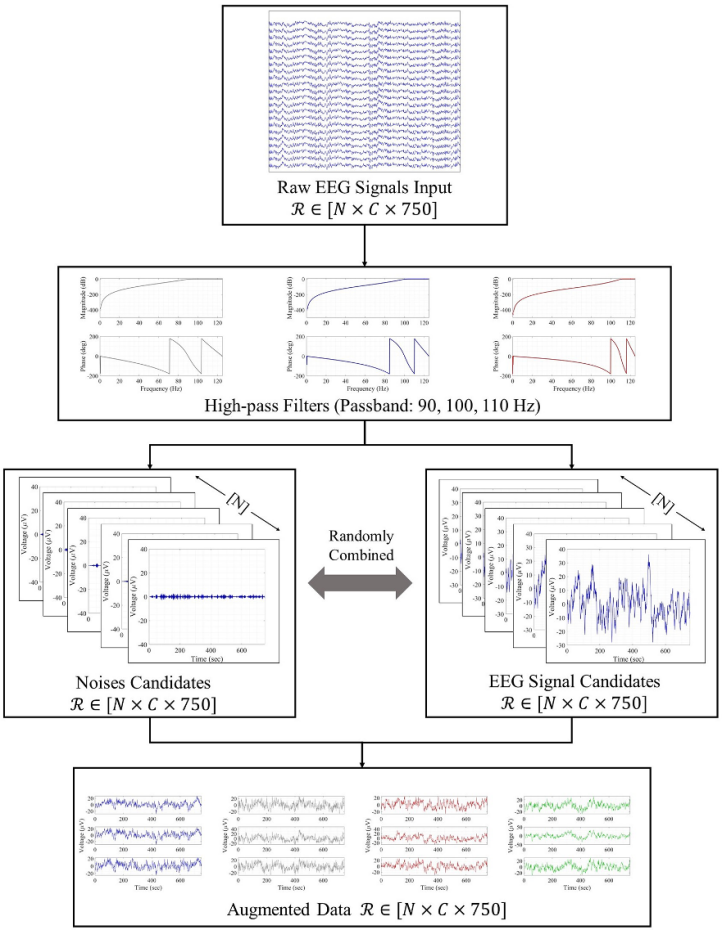

Figure 3. Data augmentation scheme. The extracted noise candidates are combined randomly with the signal candidates.

the computation time should increase because of the increasing convolutional kernel matrices depth. Besides, the classification accuracy trend is unpredictable because the larger size of trainable parameters might cause the overfitting problem.

4. Dataset and experiment protocol

4.1. Dataset description

To compare the performance with other state-of-the-art algorithms, we have used two publicly available datasets from the BCI Competition IV, namely dataset 2a and dataset 2b [45]. Both datasets contain EEG and EOG signals with a sampling frequency of 250 Hz from nine subjects. For dataset 2a, subjects were required to perform four classes (left hand, right hand, feet, and tongue) MI, and for dataset 2b, subjects were asked to perform binary classes MI tasks (left hand and right hand). Figure 6 presents the localizations of the EEG channels for both datasets 2a and 2b. Twenty-two channels are provided by the dataset 2a, and three channels are provided by the dataset 2b.

For the binary-classes dataset, each subject has conducted five sessions where the first two sessions are MI without results feedback (MI w/o feedback), and the last three sessions are MI with results feedback (MI w/feedback). Figure 7 shows the paradigm of the MI w/o feedback for one trial. During the MI w/o session, the subjects are asked to perform 60 trials of MI tasks per-class, 120 trials in total. At each trial, the MI period is 3 s. The scheme for the MI w/feedback session of one trial is shown in figure 8. In the MI w/feedback session, the subjects are asked to perform 80 trials of MI tasks per-class, 160 trials in total. At each trial, the MI period is around 4 s. Moreover, the MI w/feedback session evaluates the MI performance for every trial where the green smiley face represents a correct MI task while the sad red face indicates a wrong task.

For the four classes dataset, each subject has conducted two sessions without results feedback. The experiment procedures are similar to the two classes dataset, as shown in figure 9. The four classes dataset contains 72 trials per class (288 trials in total). At each trial, the MI period is 3 s.

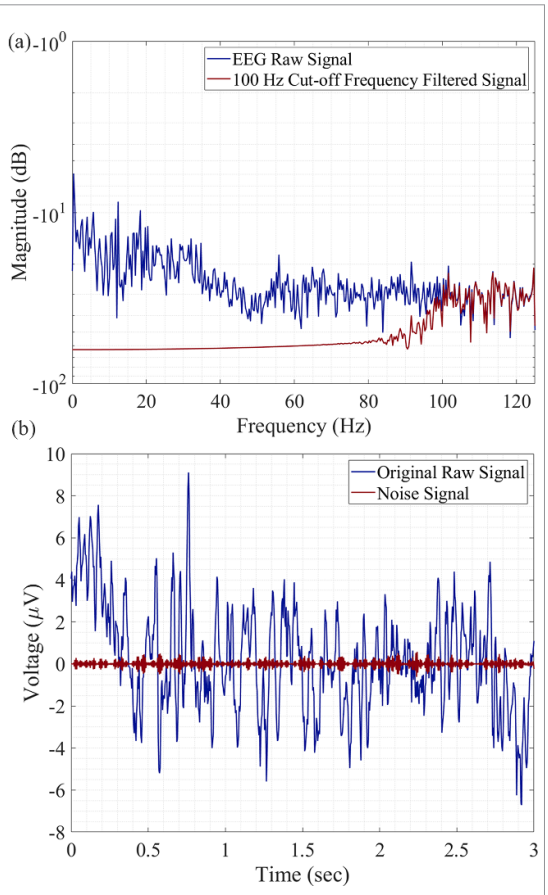

Figure 4. (a) Frequency domain of raw signal and extracted noises. The extracted noise candidate is at a cut-off frequency of 100 Hz. (b) Time domain of raw signal and extracted noises.

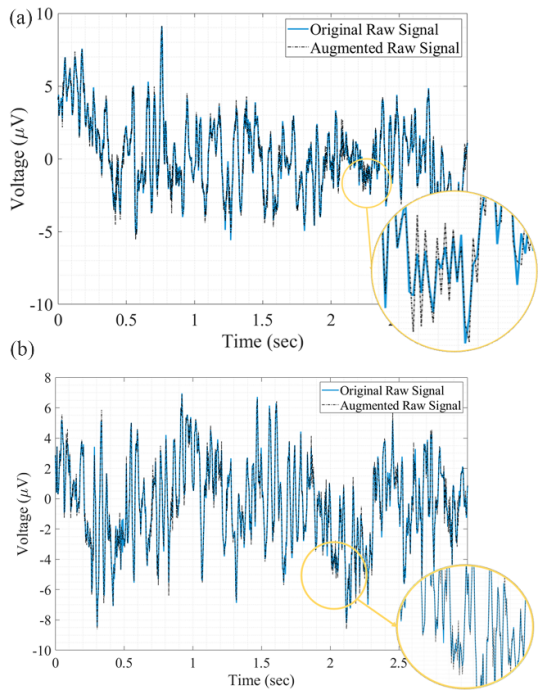

Figure 5. (a) Original and augmented signal for subject 1 trial 1. (b) Original and augmented signal for subject 4 trial 4.

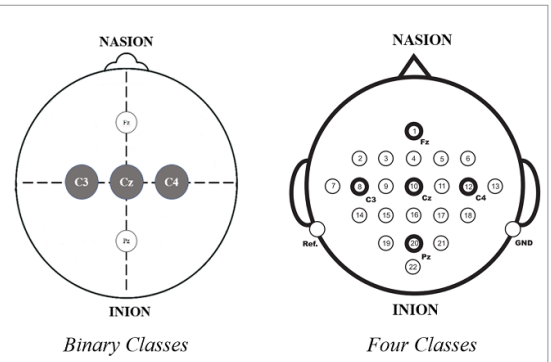

Figure 6. Dataset 2a and 2b channels localization. The 2a dataset contains 22 channels while the 2b dataset contains three channels.

4.2. Experiment protocol

Since both dataset 2a and 2b sizes are too small for neural network model training and evaluation, some modifications to the input signal are necessary. For dataset 2b, we doubled the number of trials in the last three sessions by selecting the first three-second imagery period as one trial, and the last three-second imagery period is another trial, as shown in figure 8. Thus, by combining all sessions, the total number of MI trials is 1200 for every subject, and the length of each trial is 3 s, which are used as input signals. For dataset 2a, since the MI period is three seconds fixed, we only applied more augmented signals for training.

To ensure the correctness of input training data, rejected trials labeled by the dataset are removed. Therefore, the total number of trials is varied based on different subjects. In the dataset, we set the training and testing data ratio around 3:1, and the details about the training and testing samples for every subject are presented in table 1.

5. Results and discussions

The EEG-inception model is evaluated through an open-source dataset 2a and 2b from the BCI competition IV. The proposed algorithm testing results are presented from five aspects: (a) the proposed model training process, (b) the ablation study outcome, (c) the data augmentation method effectiveness, (d) the proposed algorithm accuracy comparison with other state-of-art algorithms, and (e) a preliminary examination of the proposed EEG-inception model for subject-independent study.

5.1. Training process

The EEG-inception neural network model is written with Python programming language-based Pytorch platform [46]. To achieve faster computation speed, the model is trained using an Nvidia GeForce RTX 2080Ti graphics card. During the training process, a backpropagation method is employed for the neural network weight update. Since the input signals are non-stationary and the model is relatively complex,

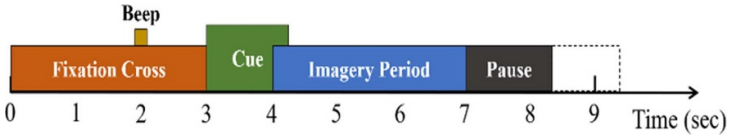

Figure 7. MI binary class w/o feedback session paradigm for one trial.

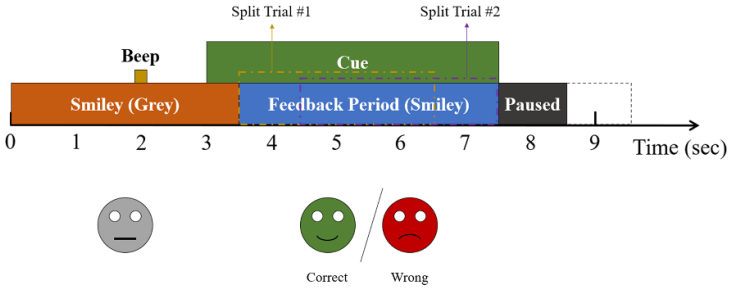

Figure 8. MI binary class w/feedback session paradigm for one trial.

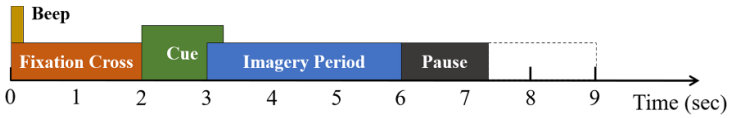

Figure 9. MI four classes session paradigm for one trial.

we use adaptive moment estimation (Adam) as the optimizer with a general learning rate of 0.005. The Adam equation is shown in equation (3)

$$\Theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \in}}\hat{m}_t \tag{3}$$

where θ is the updated parameters, m_t and v_t are the first and second-moment gradient, respectively, η is the general learning rate, and $\in$ is a smoothing term. The loss is calculated by the binary class cross-entropy loss function.

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(\mathrm{p}(\mathrm{y_i})) + (1 - y_i) \cdot \log(1 - p(y_i)) \tag{4}$$

where y is the label, $p(y)$ is the predicted label probability, and N is the total sample size. In our experiment, the training iteration were set to 100 to ensure the model converges, and the batch size is set to be 32 for less calculation memory.

5.2. Ablation study

As stated before, our ablation study is conducted by varying the convolution kernel matrix depth. Table 2 summarizes the number of model parameters and model size with different kernel matrix depths for binary-classes model and four- classes model. When increasing the layer depth, both model parameters and the model size are linearly increasing. Figure 10 presents the model classification accuracies and computation time results with varied kernel matrix depths for dataset 2b. Similar behavior can also be observed on the dataset 2a results. With increasing kernel

Table 1. Training/testing samples summary for all subjects at both binary class and four classes dataset.

	Acceptance rate (%)	Train w/o augmentation	Train w/ augmentation	Test
Binary classes dataset				
S1	75.33%	678	2034	226
S2	79.17%	712	2136	238
S3	72.83%	655	1965	219
S4	97.17%	874	2622	292
S5	89.25%	803	2409	268
S6	77.58%	698	2094	233
S7	79.83%	718	2154	240
S8	74.50%	670	2010	224
S9	78.50%	706	2118	236
Four classes dataset				
S1	86.46%	498	2988	140
S2	86.28%	497	2982	139
S3	84.27%	488	2928	137
S4	76.56%	441	2646	123
S5	84.03%	484	2904	135
S6	67.71%	390	2340	110
S7	86.46%	498	2988	138
S8	83.51%	481	2886	135
S9	78.13%	450	2700	127

Table 2. Model parameters and size comparison with different layer depths.

Binary classes dataset			Four classes dataset		
Layer depth	Total parameters	Total size (MB)	Layer depth	Total parameters	Total size (MB)
6	51 386	5.23	24	22 36 468	8.6
12	204 002	10.83	36	50 21 356	19.2
16	361 986	14.77	48	89 17 348	34.1
24	812 930	23.18	60	139 24 444	53.2
32	14 43 842	32.27	72	200 42 664	76.5
64	57 67 170	75.49	84	272 71 948	1040

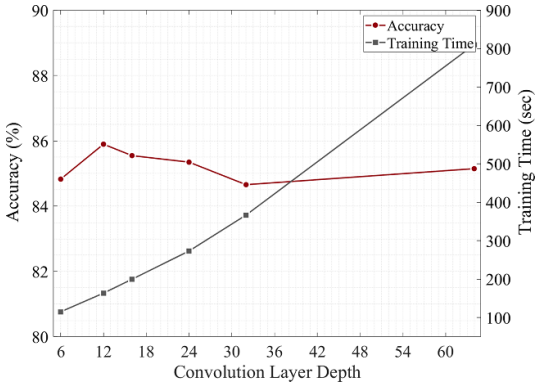

Figure 10. Training time and classification accuracy with different convolution layer depth.

matrix depths, the computation time follows an exponentially increasing trend while the classification accuracy is unpredictable. The classification accuracy is not guaranteed to improve by simply increasing kernel matrix depth because a deeper kernel matrix causes the model overfitting. Based on this result, the final convolution kernel matrix depth is 12 for dataset 2b and 48 for dataset 2a. The total model parameters are over 20 000 for the binary classes model and over eight million for the four classes model, and the total model sizes for binary-classes and four-classes models are 10.83 and 34.10 megabytes, respectively. To better understand the proposed inception-EEG model, the binary-classes inception EEG model summary is tabulated in table 3.

5.3. Effect of data augmentation

The effectiveness of the proposed data augmentation method can be evaluated from the accuracy convergence speed and the classification accuracy improvement.

For accuracy convergence speed, the proposed data augmentation method is faster than the dataset without augmentation. As shown in figure 11, the classification accuracy with the data augmentation method converged after 10–20 iterations, while the one without augmentation method did not converged until 40–60 iterations. Since the proposed data augmentation method randomly combined the noise candidates with the signal candidates, the randomness of the training samples is improved, but the signal temporal pattern features and frequency domain features are still well preserved. Thus, increasing the randomness of the training samples and the dataset size leads to a faster convergence speed than the training without using the proposed augmentation method.

For classification results, the overall accuracy for the proposed data augmentation method is 3%–4% higher than the without data augmentation training for both datasets. According to table 4, the data augmentation method's average accuracy is 2.8% higher for the binary-classes dataset and 3.6% higher for the four-classes dataset. By observing the binary-classes dataset S2–S5, and the four-classes dataset S1, S5, and S6, we find that the proposed data augmentation method can significantly improve the classification accuracy when the original accuracy is low. It is known that the main reason for low classification accuracy is artifacts caused by ocular, muscle, and other miscellaneous situations. For the 'poor-performed' subjects, the noises artifacts are usually more inconsistent than the 'well performed' subjects. Therefore, the proposed data augmentation method effectively extracts certain noise artifacts and randomly combined them with the MI signals, which increases the number of noisy MI samples for better model learning and higher testing classification accuracy.

5.4. Comparison with state-of-the-art algorithms

We have compared our EEG-inception neural network with six popular EEG-based binary MI classification algorithms [47–52] for both the binary-classes dataset and four-classes dataset. To ensure a fair comparison, all the algorithms use the BCI Competition IV 2b and 2a dataset, and the results are shown in tables 5 and 6, respectively. For both datasets, we evaluate the proposed EEG-inception network through classification accuracy, computation time, and the standard deviation of classification accuracy among all subjects. In the remaining of this section, binary-classes dataset results are discussed at first, and the four-classes dataset results are illustrated then.

Table 3. EEG-inception neural network architecture summary for binary classes dataset and four classes dataset.

Layer	Output shape	Param #
Initial inception module	[48, 750]	32 628
Intermediate inception module_1	[48, 750]	33 708
Intermediate inception module_2	[48, 750]	33 708
Residual module_1	[48, 750]	288
Intermediate inception module_3	[48, 750]	33 708
Intermediate inception module_4	[48, 750]	33 708
Intermediate inception module_5	[48, 750]	33 708
Residual module_2	[48, 750]	2448
Average pooling	[48, 750]	—
Linear	[1, 2]	98
Initial inception module		
MaxPooling-1d	[3, 750]	—
Convolution-1d	[12, 750]	48
Convolution-1d	[12, 750]	48
Convolution-1d	[12, 750]	3612
Convolution-1d	[12, 750]	10 812
Convolution-1d	[12, 750]	18 012
BatchNormalization-1d	[48, 750]	96
ReLU activation	[48, 750]	—
Intermediate inception module		
MaxPooling-1d	[3, 750]	—
Convolution-1d	[12, 750]	48
Convolution-1d	[12, 750]	48
Convolution-1d	[12, 750]	3612
Convolution-1d	[12, 750]	10 812
Convolution-1d	[12, 750]	18 012
BatchNormalization-1d	[48, 750]	96
ReLU activation	[48, 750]	—
Residual module_1		
Convolution-1d	[48, 750]	192
BatchNormalization-1d	[48, 750]	96
ReLU activation	[48, 750]	—
Residual module_2		
Convolution-1d	[48, 750]	2352
BatchNormalization-1d	[48, 750]	96
ReLU activation	[48, 750]	—
Parameters summary		
Total parameters		204 002
Trainable parameters		204 002

5.4.1. Binary-classes dataset results

As for average classification accuracies, the proposed EEG-inception neural network is the highest among all state-of-art algorithms for the binary-classes dataset, which is 1% higher than the HS-CNN and almost 10% higher than the FBCSP algorithm. For single-subject classification accuracy examination, the EEG-inception model exhibits better classification accuracies for 'poorly performed' subjects. Figure 12 shows the EEG-inception model accuracy comparison results with the top three state-of-art algorithms. According to figure 12, #2 and #3 subjects exhibit the worst classification accuracies for all algorithms. The proposed EEG-inception model maintains the classification accuracy at around 80%, which is approximately 12.5% higher than the other state-of-art algorithms. The leading causes of #2 and #3 low classification accuracies are unclear neuronal features and artifact contamination. Therefore, based on the #2 and #3 classification results, the proposed EEG-inception algorithm can extract more effective features when subject brain signals are covered by noises and artifacts.

The computation time for the EEG-inception network is 0.0187 s for one sample. Instead of adding convolutional layers in series, the EEG-inception network applies several convolutional layers in parallel to extract MI features from different time lengths. Therefore, the EEG-inception network is more computationally efficient compared with other deep CNNs.

Besides the classification accuracy and computation time, we compared the average standard deviation (std. dev.) among all subjects. The proposed EEG-inception neural network std. dev. is 5.5, which is 35.14% less than the HS-CNN algorithm and 58.31% less than the average std. dev. among all state-of-art algorithms. Such low std. dev. of the proposed algorithm indicates that the proposed algorithm is robust to all subjects, which has the potential for subject-independent learning.

5.4.2. Four-classes dataset results

The EEG-inception network classification accuracy for the four-classes dataset is 88.39%, which ranked 2nd among all other state-of-art algorithms (figure 13) [49, 51, 53–56]. The only algorithm that surpasses the EEG-inception network for multiclass classification is the HS-CNN algorithm. The main reason that EEG-inception is lower than the HS-CNN algorithm is because of the small training data size. Compared with the binary-classes dataset, the original four-classes dataset training sample is around 35% smaller. The EEG-inception applies five parallel convolutional kernel matrices, which requires a more extensive training dataset to learn the kernel matrices parameters for better feature extraction performance. With the improvement of the EEG-based MI experiment protocol, creating a larger EEG-based MI dataset is feasible. Therefore, we believe that the EEG-inception network is more suitable for larger dataset training, and the classification accuracy can be further improved.

As for computation time, the EEG-inception takes 0.0215 s for testing one sample. The reason for the longer computation time than the binary-classes dataset is because the four-classes dataset contains

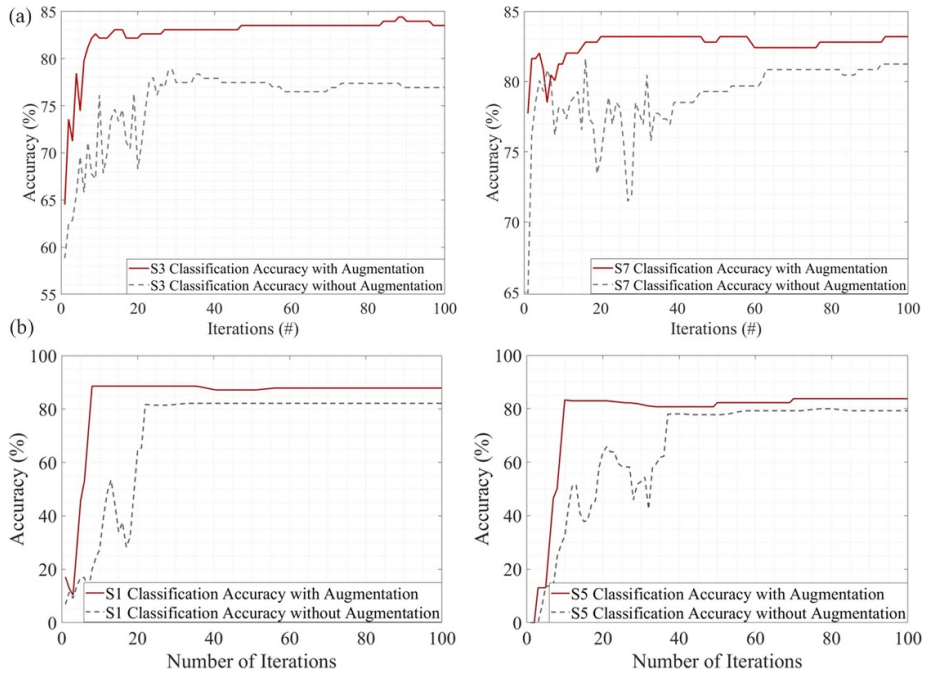

Figure 11. (a) Selected subjects binary-classes dataset classification accuracy curve comparison. (b) Selected subjects four-classes dataset classification accuracy curve comparison.

Table 4. Classification accuracy comparison with augmentation and non-augmentation.

	Binary classes dataset		Four classes dataset	
	InceptionTime w/augmentation	InceptionTime w/o augmentation	InceptionTime w/augmentation	InceptionTime w/o augmentation
S1	87.20%	84.51%	89.61%	81.52%
S2	79.79%	77.31%	80.01%	78.68%
S3	84.19%	77.17%	96.17%	94.09%
S4	96.32%	95.21%	81.26%	80.48%
S5	94.06%	93.66%	83.76%	79.66%
S6	89.27%	88.19%	81.20%	76.98%
S7	82.98%	81.90%	94.75%	91.47%
S8	90.63%	89.45%	98.28%	91.36%
S9	92.80%	84.55%	90.50%	89.17%
Average	88.58%	85.77%	88.39%	84.82%
Std. dev.	5.5	6.46	7.06	6.59

Table 5. Binary classes dataset classification accuracy comparison with state-of-art methods.

	WT-Isomap [46]	Boltzmann machine [47]	EMD-MI [48]	RSMM [49]	HS-CNN [50]	Bi-spectrum [51]	EEG-inception
S1	84.6	81	62.8	72.5	80.5	77	**87.2**
S2	66.3	65	67.1	56.4	70.6	64.5	**79.79**
S3	62.9	66	**98.7**	55.6	85.6	61	84.19
S4	95.8	**98**	88.4	97.2	94.6	96.5	96.32
S5	89.2	93	96.3	88.4	**98.3**	82	94.06
S6	**97.9**	88	75.3	78.7	86.6	84.5	89.27
S7	82.1	82	72.2	77.5	**89.6**	75	82.98
S8	86.3	94	87.8	91.9	**95.6**	91	90.63
S9	**97.1**	91	85.3	83.4	87.4	87	92.8
Average	84.69	84.22	81.54	77.96	87.64	79.83	**88.58**
Std. dev.	12.72	11.94	12.74	14.57	8.48	11.73	**5.5**

Table 6. Four classes dataset classification accuracy comparison with state-of-art methods.

	DFFS [52]	EMD-MI [48]	MEMD-Riemann [53]	Adaptive-MI [54]	R-CSP [55]	HS-CNN [50]	EEG-inception
s1	63.69	66.7	**91.49**	90.28	88.89	90.07	89.61
s2	61.97	63.9	60.56	54.17	51.39	**80.28**	**80.01**
s3	91.09	77.8	94.16	93.75	96.53	**97.08**	96.17
s4	61.72	63.2	76.72	64.58	70.14	**89.66**	81.26
s5	63.41	72.2	58.52	57.64	54.86	**97.04**	83.76
s6	66.11	70.1	68.52	65.28	71.53	**87.04**	81.2
s7	59.57	64.6	78.57	65.2	81.25	92.14	**94.75**
s8	62.84	76.4	97.01	90.97	93.75	**98.51**	**98.28**
s9	84.46	77.1	93.85	85.42	**93.75**	82.31	90.5
Average	68.32	70.2	79.93	73.84	78.01	**91.57**	**88.39**
Std. dev.	11.29	5.92	14.99	15.72	17.01	**6.49**	**7.06**

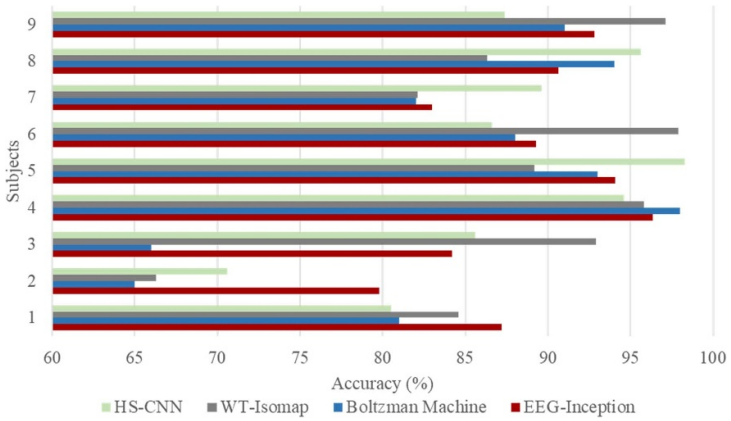

Figure 12. Binary class classification accuracy comparison with the top three state-of-art algorithms.

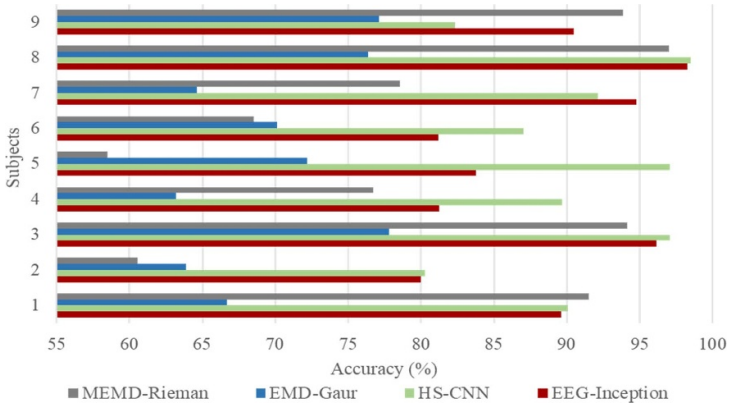

Figure 13. Four classes classification accuracy comparison with the top three state-of-art algorithms.

22 EEG channels, and the number of parallel convolutional matrices is increased from 3 to 5. Thus, the computation time is increased. However, this computation time is still fast enough for real-time processing.

The EEG-inception network std. dev. among all subjects is 7.06 (table 6). Similar to the classification accuracy results, the std. dev. for EEG-inception network can be further decreased with a larger dataset. However, such low std. dev. still shows that EEG-inception is robust to all subjects and has the potential for subject-independent learning.

5.5. Preliminary subject-independent classification examination

Since the EEG-inception neural network std. dev. is the smallest compared with other state-of-art algorithms, it has the potential to be used for subject-independent classification. These results are

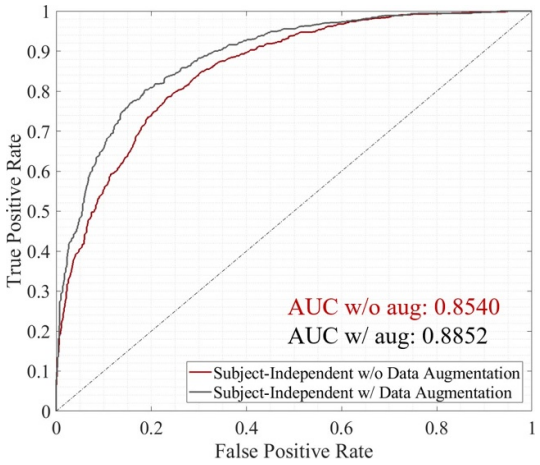

Figure 14. ROC curve for subject-independent at binary classes classification.

Table 7. Confusion matrix for binary classes subject-independent classification.

Binary classes dataset subject independent classification			
Actual Predicted	Left	Right	Accuracy
Left	998	406	71.08%
Right	85	687	88.99%
Avg. accuracy	77.44%	Kappa	0.55
F1-score	0.737	Recall	0.628

Table 8. Confusion matrix for four classes subject-independent classification.

Four classes dataset subject independent classification					
Actual Predict	Left	Right	Foot	Tongue	Accuracy
Left	253	32	3	5	86.35%
Right	60	214	6	12	73.29%
Foot	79	57	159	16	51.13%
Tongue	74	51	11	152	52.78%
Avg. accuracy	65.88%		Kappa		0.544
F1-score	0.655		Recall		0.657

preliminary because the EEG-inception subject-dependent model is directly applied without further fine-tuning and adjustment. Figure 14 and tables 7 and 8 present the ROC curve and confusion matrices for the subject-independent classification for both datasets, respectively.

For the binary-classes dataset, the classification accuracy is 77.5%, which is 12.5% lower than the EEG-inception subject dependent model but close to the RSMM and Bi-spectrum algorithms. Furthermore, according to the ROC curve, we find that the proposed data augmentation method is still effective in the subject-independent study where the AUC result is 0.03 higher than the w/o augmentation method.

For the four-classes dataset, the classification accuracy is 65.88%, which is 22% lower than the EEG-inception subject-dependent model. Since the Soft-Max equation ranges from 0 to 1, it is acceptable to see classification accuracy drop when the number of target classes are increased from two to four due to the smaller probability range boundary for each class. Furthermore, the dataset size for the four-classes dataset is not large enough for subject-independent classification. Thus, the four-classes dataset subject-independent classification accuracy is lower than the binary-classes dataset subject-independent results. However, 65.88% still shows the strong potential that the EEG-inception network can be applied for subject-independent classification with further modification.

6. Conclusions and future works

In this paper, we presented a novel data augmentation method and modified the inception time neural network for EEG-based MI, namely EEG-inception. The data augmentation method can diminish the overfitting issue caused by the small data size. The proposed EEG-inception neural network achieves an accuracy of 88.58% for the binary-classes dataset and 88.39% for the four-classes dataset. Moreover, the std. dev. among all subjects is 5.5 and 7.1 for the binary-classes dataset and four-classes dataset, respectively. The low std. dev. represents the proposed algorithm's features extraction is robust among all subjects. We have also conducted a preliminary examination for subject-independent analysis, in which the classification accuracy is 77.5% and 65.9% for binary-classes and four-classes, respectively.

In the future, we will further investigate the specificity and generality of each layer from the EEG-inception neural network. After finding the specificity and generality of the EEG-inception network, we are going to fine-tune the EEG-inception model so that it fits for the EEG-based MI subject independent classification task.

Data availability statement

No new data were created or analyzed in this study.

ORCID iDs

Ce Zhang https://orcid.org/0000-0002-4344-5259
Young-Keun Kim https://orcid.org/0000-0002-5373-1128
Azim Eskandarian https://orcid.org/0000-0002-4117-7692

References

[1] Krucoff M O, Rahimpour S, Slutzky M W, Edgerton V R and Turner D A 2016 Enhancing nervous system recovery

through neurobiologics, neural interface training, and neurorehabilitation *Front. Neurosci.* **10** 584
[2] Pfurtscheller G and Aranibar A 1977 Event-related cortical desynchronization detected by power measurements of scalp EEG *Electroencephalogr. Clin. Neurophysiol.* **42** 817–26
[3] Pfurtscheller G, Brunner C, Schlögl A and Lopes Da Silva F H 2006 Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks *NeuroImage* **31** 153–9
[4] Padfield N, Zabalza J, Zhao H, Masero V and Ren J 2019 EEG-based brain-computer interfaces using motor-imagery: techniques and challenges *Sensors* **19** 1423
[5] Yu Y, Zhou Z, Liu Y, Jiang J, Yin E, Zhang N, Wang Z, Liu Y, Wu X and Hu D 2017 Self-paced operation of a wheelchair based on a hybrid brain-computer interface combining motor imagery and P300 potential *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 2516–26
[6] Rakshit A, Konar A and Nagar A K 2020 A hybrid brain-computer interface for closed-loop position control of a robot arm *IEEE/CAA J. Autom. Sin.* **7** 1344–60
[7] Khan M J, Zafar A and Hong K eds 2016 Hybrid EEG-NIRS based active command generation for quadcopter movement control *2016 Int. Automatic Control Conf. (CACS)* (*9–11 November 2016*) (https://doi.org/10.1109/CACS.2016.7973909)
[8] Zhao H, Zheng Q, Ma K, Li H and Zheng Y 2020 Deep representation-based domain adaptation for nonstationary EEG classification *IEEE Trans. Neural Networks Learn. Syst.* **32** 1–11
[9] Haumann N T, Parkkonen L, Kliuchko M, Vuust P and Brattico E 2016 Comparing the performance of popular MEG/EEG artifact correction methods in an evoked-response study *Comput. Intell. Neurosci.* **2016** 7489108
[10] Fawaz H I, Lucas B, Forestier G, Pelletier C, Schmidt D, Weber J, Webb G I, Idoumghar L, Muller P-A and Petitjean F 2019 InceptionTime: finding alexnet for time series classification (arxiv: 1909.04939)
[11] He K, Zhang X, Ren S and Sun J eds 2016 Deep residual learning for image recognition *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (*27–30 June 2016*) (https://doi.org/10.1109/CVPR.2016.90)
[12] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2015 Going deeper with convolutions *2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 1–9
[13] Zhang C and Eskandarian A 2020 A survey and tutorial of EEG-based brain monitoring for driver state analysis *IEEE/CAA J. Autom. Sin.* (https://doi.org/10.1109/JAS.2020.1003450
[14] Akhtar M T, Mitsuhashi W and James C J 2012 Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data *Signal Process.* **92** 401–16
[15] Clercq W D, Vergult A, Vanrumste B, Paesschen W V and Huffel S V 2006 Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram *IEEE Trans. Biomed. Eng.* **53** 2583–7
[16] Neuper C, Müller-Putz G R, Scherer R, Pfurtscheller G 2006 Motor imagery and EEG-based control of spelling devices and neuroprostheses *Progress in Brain Research* vol 159, ed C Neuper and W Klimesch (Amsterdam: Elsevier) pp 393–409
[17] Yang H, Guan C, Chua K S, Chok S S, Wang C C, Soon P K, Tang C K and Ang K K 2014 Detection of motor imagery of swallow EEG signals based on the dual-tree complex wavelet transform and adaptive model selection *J. Neural Eng.* **11** 035016
[18] Hsu W-Y and Sun Y-N 2009 EEG-based motor imagery analysis using weighted wavelet transform features *J. Neurosci. Methods* **176** 310–8
[19] Bashar S K and Bhuiyan M I H 2016 Classification of motor imagery movements using multivariate empirical mode decomposition and short time Fourier transform based hybrid method *Eng. Sci. Technol. Int. J* **19** 1457–64
[20] Sadiq M T, Yu X, Yuan Z and Aziz M Z 2020 Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform *Electron. Lett.* **56** 1367–9
[21] Sadiq M T, Yu X, Yuan Z, Fan Z, Rehman A U, Li G and Xiao G 2019 Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform *IEEE Access* **7** 127678–92
[22] Pfurtscheller G, Guger C and Ramoser H eds 1999 EEG-based brain-computer interface using subject-specific spatial filters *Engineering Applications of Bio-Inspired Artificial Neural Networks 1999* (Berlin: Springer) (https://doi.org/10.1007/bfb0100491)
[23] Kai Keng A, Zheng Yang C, Haihong Z and Cuntai G eds 2008 Filter bank common spatial pattern (FBCSP) in brain-computer interface *2008 IEEE Int. Joint Conf. on Neural Networks* (IEEE World Congress on Computational Intelligence) (*1–8 June 2008*) (https://doi.org/10.1109/IJCNN.2008.4634130)
[24] Zhang C and Eskandarian A eds 2020 A computationally efficient multiclass time-frequency common spatial pattern analysis on EEG motor imagery *2020 42nd Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* (*20–24 July 2020*) (https://doi.org/10.1109/EMBC44109.2020.9176705)
[25] Sadiq M T, Yu X and Yuan Z 2021 Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interfaces *Expert Syst. Appl.* **164** 114031
[26] Sadiq M T, Yu X, Yuan Z, Zeming F, Rehman A U, Ullah I, Li G and Xiao G 2019 Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces *IEEE Access* **7** 171431–51
[27] Shang-Lin W, Chun-Wei W, Pal N R, Chih-Yu C, Shi-An C and Chin-Teng L eds 2013 Common spatial pattern and linear discriminant analysis for motor imagery classification *2013 IEEE Symp. on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)* (*16–19 April 2013*) (https://doi.org/10.1109/CCMB.2013.6609178)
[28] Ma Y, Ding X, She Q, Luo Z, Potter T and Zhang Y 2016 Classification of motor imagery EEG signals with support vector machines and particle swarm optimization *Comput. Math. Methods Med.* **2016** 4941235
[29] Chatterjee R and Bandyopadhyay T eds 2016 EEG based motor imagery classification using SVM and MLP *2016 2nd Int. Conf. on Computational Intelligence and Networks (CINE)* (*11–11 January 2016*) (https://doi.org/10.1109/CINE.2016.22)
[30] Bhaduri S, Khasnobish A, Bose R and Tibarewala D N eds 2016 Classification of lower limb motor imagery using K nearest neighbor and Naïve-Bayesian classifier *2016 3rd Int. Conf. on Recent Advances in Information Technology (RAIT)* (*3–5 March 2016*) (https://doi.org/10.1109/RAIT.2016.7507952)
[31] Li Y, Zhang X, Zhang B, Lei M, Cui W and Guo Y 2019 A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding *IEEE Trans. Neural Syst. Rehabil. Eng.* **27** 1170–80
[32] Tabar Y R and Halici U 2017 A novel deep learning approach for classification of EEG motor imagery signals *J. Neural Eng.* **14** 016003
[33] Miao M, Hu W, Yin H and Zhang K 2020 Spatial-frequency feature learning and classification of motor imagery EEG based on deep convolution neural network *Comput. Math. Methods Med.* **2020** 1981728
[34] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks *Proc. 25th Int. Conf. on Neural Information Processing Systems* vol 1 (Lake Tahoe, NV: Curran Associates Inc) pp 1097–105

[35] Lee B, Jeong J, Shim K and Lee S, eds 2020 Classification of high-dimensional motor imagery tasks based on an end-to-end role assigned convolutional neural network *ICASSP 2020–2020 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (*4–8 May 2020*) (https://doi.org/10.1109/ICASSP40776.2020.9054359)
[36] Yang T, Phua K S, Yu J, Selvaratnam T, Toh V, Ng W H, Ang K K and So R Q eds 2019 Image-based motor imagery EEG classification using convolutional neural network *2019 IEEE EMBS Int. Conf. on Biomedical & Health Informatics (BHI)* (*19–22 May 2019*) (https://doi.org/10.1109/BHI.2019.8834598)
[37] Wang P, Jiang A, Liu X, Shang J and Zhang L 2018 LSTM-based EEG classification in motor imagery tasks *IEEE Trans. Neural Syst. Rehabil. Eng.* **26** 2086–95
[38] Jeong J, Shim K, Kim D and Lee S 2020 Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals *IEEE Trans. Neural Syst. Rehabil. Eng.* **28** 1226–38
[39] Lin M, Chen Q and Yan S 2014 Network in network *CoRR*. (arxiv: 1312.4400)
[40] Ioffe S and Szegedy C. 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift (arxiv: 1502.03167)
[41] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A and Yger F 2018 A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update *J. Neural Eng.* **15** 031005
[42] Wang F, Zhong S-H, Peng J, Jiang J and Liu Y eds 2018 Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. *MultiMedia Modeling 2018* (Berlin: Springer) (https://doi.org/10.1007/978-3-319-73600-6_8)
[43] Muthukumaraswamy S D 2013 High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations *Front. Hum. Neurosci.* **7** 138
[44] Meyes R, Lu M, Puiseau C and Meisen T 2019 Ablation studies in artificial neural networks (arxiv: 1901.08644)
[45] Tangermann M *et al* 2012 Review of the BCI Competition IV *Front. Neurosci.* **6** 55
[46] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library (arxiv: 1912.01703)
[47] Li M-A, Zhu W, Liu H-N and Yang J-F 2017 Adaptive feature extraction of motor imagery EEG with optimal wavelet packets and SE-isomap *Appl. Sci.* **7** 390
[48] Lu N, Li T, Ren X and Miao H 2017 A deep learning scheme for motor imagery classification based on restricted Boltzmann machines *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 566–76
[49] Gaur P, Pachori R B, Hui W and Prasad G eds 2015 An empirical mode decomposition based filtering method for classification of motor-imagery EEG signals for enhancing brain-computer interface *2015 Int. Joint Conf. on Neural Networks (IJCNN)* (*12–17 July 2015*) (https://doi.org/10.1109/IJCNN.2015.7280754)
[50] Zheng Q, Zhu F and Heng P 2018 Robust support matrix machine for single trial EEG classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **26** 551–62
[51] Dai G, Zhou J, Huang J and Wang N 2020 HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification *J. Neural Eng.* **17** 016025
[52] Shahid S and Prasad G 2011 Bispectrum-based feature extraction technique for devising a practical brain-computer interface *J. Neural Eng.* **8** 025014
[53] Luo J, Feng Z, Zhang J and Lu N 2016 Dynamic frequency feature selection based approach for classification of motor imageries *Comput. Biol. Med.* **75** 45–53
[54] Gaur P, Pachori R B, Wang H and Prasad G 2018 A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry *Expert Syst. Appl.* **95** 201–11
[55] Raza H, Cecotti H, Li Y and Prasad G 2016 Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface *Soft Comput.* **20** 3085–96
[56] Lotte F and Guan C 2011 Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms *IEEE Trans. Biomed. Eng.* **58** 355–62