Article:
Zhang, J and Li, K orcid.org/0000-0001-6657-0522 (Cover date: August 2023) A multi-view CNN encoding for motor imagery EEG signals. Biomedical Signal Processing and Control, 85. 105063. ISSN 1746-8094

https://doi.org/10.1016/j.bspc.2023.105063

A Multi-View CNN Encoding for Motor Imagery EEG Signals

Jiayang Zhang[a], Kang Li*[a]

[a]School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, UK

Abstract

Objective: The convolution neural network (CNN) has gained lots of attentions recently in decoding electroencephalogram (EEG) signals in Motor Imagery (MI) Brain-Computer Interface (BCI) designed for improving stroke rehabilitation strategies. However, the extremely non-linear, nonstationary nature of the EEG signals and diversity among individual subjects results in the overfitting of a CNN model and limits its learning ability. In this study, a densely connected convolutional network with multi-view inputs is proposed. Methods: First, different data subsets from the original EEG signals are created as the CNN model inputs through bandpass filters applied to the EEG signals to generate multiple frequency sub-band signals based on brain rhythms. Then, temporal and spatial features are captured based on the whole frequency band and the filtered sub-band signals, respectively. Further, two dense blocks with multi-CNN layers, which connect each layer to every other layer in the feed-forward path, are used to enhance the model learning capabilities and strengthen information propagation. Finally, a concatenation fusion method is used to integrate the extracted features and a fully connected layer for finalizing the classification. Results: The proposed method achieves an average accuracy of 75.16% on the public Korea University EEG dataset which consists the EEG signals of 54 healthy subjects for the two-class motor imagery tasks, higher than other state-of-the-art deep learning methods. Conclusion: The proposed method effectively extracts much richer motor imagery information from the EEG signals in the BCI system and improves the classification accuracy.

Keywords: Motor Imagery, Electroencephalography, Convolutional Neural Network,

1. Introduction

The brain-computer interface (BCI) systems are designed to capture the signals corresponding to neural processes from the brain and support communication way between the brain and the external equipment without using the common neuron muscular mechanism [1]. BCI technology can be used to assist the rehabilitation of people suffering from devastating neuron logical disorders and restore practical motor control after the brain disorders such as stroke, by interpreting the brain signals relating to the state of the brain activities [2, 3]. The BCI technology can also help to improve the muscle control of patients by increasing the effectiveness of rehabilitation programs and supporting impaired muscle control [4].

Electroencephalography (EEG), which reflects various electrical activities of the brain, has been widely used in BCI systems due to its high temporal resolution, cost-effectiveness, portability, and noninvasive nature [5]. Motor imagery (MI), as one of the most commonly used EEG paradigms, allows disabled people to self-regulate brain signals through voluntary modulation rather than external stimulus [6]. When a person is imagining or executing motor behavior, the related motor cortex on the brain scalp generates the corresponding MI responses with massive neuron activity [7]. Through decoding MI-EEG signals, the gap induced by stroke between motor intention and sensory feedback of motor movements is bridged which leads to swift motor functional recovery.

For MI-EEG decoding, feature extraction and classification are the two most important steps. Conventional machine learning methods like random forest (RF), linear discriminant analysis (LDA), support vector machine (SVM), and neural networks (NN) were adopted as various classifiers for the MI-EEG decoding tasks. In 2014, Bentlemsan et al used RF to combine bagging for bootstrap aggregation and features that are selected randomly [8]. In research [9], Luo et al proposed a feature-selected approach namely the dynamic frequency feature selection (DFFS) with a RF classifier to decode MI-EEG signals. The calculation cost of the LDA classifier is low which is beneficial for use in BCI applications based on MI-EEG [10]. Chen et al used LDA to obtain the classification results using multiple frequency band signals and vote through probability summation [11]. Fu et al adopted regularized linear discriminant analysis (RLDA) to enhance the dimension of the diagonal elements of the scatter matrices to improve classification accuracy [12]. Support Vector Machine (SVM) is another common classifier in BCI field. Islam et al used SVM on features with reduced dimensions obtained by adopting multi-band principal components analysis (PCA) for four-class classification problems [13]. However, the noisy and non-linear nature of EEG signals makes it difficult for the LDA classifier to

get excellent results. SVM produces better classification results but cannot deal with the multiclass problem and decode complex EEG signals effectively [10]. The neural network (NN) has been extensively used in the BCI field for providing a reasonable balance between accuracy and training speed. Sagee et al used a bayesian network (BN) for maximum probable channel selection and NN for feature classification [14]. Hamedi et al employed Radial Basis Function (RBF) neural networks to reduce training time while ensuring high accuracy [15]. The research aforementioned has to consider if the extracted features are suitable for the classifiers, which is also a challenge in MI-EEG decoding tasks.

Among the existing feature extraction methods, the common spatial pattern (CSP) and its variants are effective approaches in constructing optimal spatial filters that discriminate two classes of EEG measurements in MI-based BCI [16]. In [17], the common spatio-spectral pattern (CSSP) is proposed which adopted the technology of delay embedding to extend the CSP algorithm to the state space. Dornhege et al proposed a new approach namely the common sparse spectral-spatial pattern (CSSSP) that optimizes both the spatial filter and the spectral filter together to enhance the difference between multi-channel EEG signals [18]. In 2007, Novi et al decomposed the EEG signals into sub-bands using a filter bank and got a score from each sub-band after using CSP [19]. The final decision is decided on the scores based on different sub-bands. Based on the fusion of different sub-band scores, Ang et al [20] proposed a method namely filter bank common spatial pattern (FBCSP) to choose the optimal features based on CSP by several band-pass filters with different band ranges. Higashi et al proposed a discriminative filter bank CSP (DFBCSP) that considers the combination of finite impulse response filters and spatial weights. This method optimized the corresponding weights by a function which can be regarded as another variation of the CSP algorithm. [21]. However, feature extraction and feature classification are two separate steps, and how to effectively integrate the two steps to improve the overall classification accuracy is still a challenge that needs to be addressed in the machine learning field.

Deep learning (DL), as an implement for extracting abstract information embedded in data by applying multi-layered neural network structures and the end-to-end process, has been successfully applied to analyze brain signals in recent years [22–27]. In 2017, Schirrmeister et al proposed three different DL structures and used the cropped training strategy to increase training examples for the network [28]. Lawhern et al in 2018 purposed a novel method namely 'EEGNet' that learns features that can be applied to several BCI tasks and reduced parameter calculations [29]. Amin et al [30] proposed a multi-layer convolutional neural networks (CNN) algorithm which considered layers with different features and structures to improve classification accuracy. Freer et al [31] combined CNN with long short-term memory (LSTM) layers and achieved adequate results in distinguishing between different control classes. However, few deep learning models have considered the influence of frequency band range which varies from subject to subject and the reuse of features in a deep learning structure.

There exist two key challenges in frequency band selection in the data preprocessing step. The first is the difficulty to determine brain rhythms as they vary with time and with different subjects, the other is the diversity among different subjects. The brain rhythms reflect the functional states of different neuronal cortical networks. The most common frequency band used in the MI-EEG field is α rhythm [32] which is about 10 Hz and β rhythm which is around 20 Hz [33]. In the research [17], 7 - 30 Hz was selected as the α and β rhythm. In another study [34], it is shown that 26 Hz is the upper limit of the β rhythm. Besides that, θ rhythm (4-7 Hz) was also proved useful in decoding MI-EEG signals [35, 36]. It is clear that, the frequency band range for special brain rhythms is found to be different in these studies. Meanwhile, Novi et al [19] proved that the variation among different subjects is enormous. That is also the reason that FBCSP and other later methods focused on selecting the appropriate operational frequency band for extracting discrimination. In DL methods, if no band selection is performed before building the classifier model, redundant information and noise will lead to the overfitting problem and make it difficult to learn useful features. However, on the other hand if the band selection is performed, the differences among different subjects may have a great impact on the final classification result. All these have to be taken into account.

The other challenge is the lack of reusable feature maps in the deep learning model. In the computer vision field, many works have demonstrated that the CNN model with shorter connections between layers close to the input and the ones close to the output can be trained more efficiently and accurately, and these shorter connections can help reuse features in the previous layers [37]. ResNets [38] redesigned the layers that learn residual functions concerning the layer inputs and side signals from one layer to the next through connections. Larsson et al [39] proposed FractalNets that combine several parallel layer sequences with different numbers of CNN and maintain many short connections in the network. The densely connected network (DenseNet) [37] uses direct connections between any two layers with the same feature-map size. Short paths lead to feature map reuse which can ensure efficient information flow between layers in the network and improve model learning capabilities. Liu et al [40] connected the DenseNet with 3D CNN to decode MI-EEG. However, the 22 input channels were forced to be the size of 7×7 by padding with the mean of all the EEG signals which introduced lots of redundant information. Yu et al [41] proposed a model combining the attention mechanism with DenseNet. However, the model only used three channels (C3, C4, Cz) to generate the input images by Continuous

Wavelet Transform (CWT), which ignored massive spatial information provided by the other channels in the cortex.

To address the two outstanding issues in deep learning for EEG signal decoding in BCI systems, this paper proposes a novel end-to-end CNN architecture with multi-views of EEG signals based on the densely connected convolutional network. First, the MI-EEG signals are fed to a 3-order Butterworth filter with different frequency bands (1-5 Hz, 4-8 Hz, 8-13Hz, and 13-32 Hz) according to the brain rhythms. Then each of the four EEG sub-band signals and the raw data covering the whole frequency band is fed into a CNN model respectively. Then two CNN layers are used to capture temporal and spatial features. Next, these features pass two dense blocks with multi-CNN layers, which helps to connect each layer to every other in a feed-forward mode. In the dense block, the feature maps extracted by the preceding layers are used as inputs, and their feature maps are fed into the succeeding layers, which help reuse feature maps and reduce overfitting on tasks. The final extracted features from each kind of input signal are fused in a fully-connected layer and end with the softmax classifier.

The remainder of the paper includes four sections. The details of the proposed method including data description, preprocessing steps, the detailed structure and parametrization of the proposed CNN model are given in Section II. Experimental results and discussion are presented in Section III and Section IV respectively. In the end, Section V concludes the paper.

2. Methods

This section first introduces the public dataset used in the experiment. Then the preprocessing steps and the proposed method are given in detail. The structure and parameters are shown at the end.

2.1. Data description

1) The public Korean University (KU) dataset [42] which includes fifty-four healthy subjects (ages 24-35; 25 females) is the largest binary dataset so far available in the public domain. Every subject had 200 trials of data (100 trials for imaging the left hand and right hand respectively). The EEG signals were collected with 62 electrodes based on the standard international 10–20 system placement. The sampling rate was 1,000 Hz. To ensure a fair comparison with other methods, we downsample raw signals to 250 Hz. 20 electrodes from the region related to motor function were selected (shown in Fig.1). The channel selection is based on the previous model [6, 43] which performed well on this dataset. The black fixation cross on the center of the monitor lasted for 3 seconds for subjects to prepare for the MI task. Next, the subject was asked to image the MI task for 4s according to the left or right arrow that appeared on the monitor and relaxed after a blank screen. We only intercept 4 s data for the MI tasks

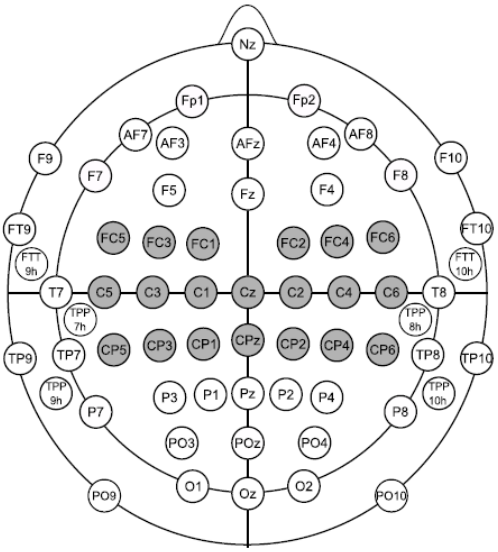


Figure 1: EEG electrodes position (KU dataset) [6]. The EEG electrodes shown in gray were used in the proposed model.

for the subsequent processing. The 10-fold cross-validation (CV) method was adopted to check the performance and robustness of our proposed model. 8 folds are used for training, 1 fold for testing and the remainder for validation. Hence, for each subject's model, a total of 160 training trials and 20 validation and testing trials were administered.

2) BCI Competition IV 2a (BCIC-IV-2a) dataset [44] consists of recordings from 9 healthy subjects performing 4 different motor imagery tasks: left-hand, right-hand, both-foot, and tongue. The signals were acquired using 22 EEG electrodes with a sampling frequency of 250 Hz and were bandpass filtered between 0.5 Hz and 100 Hz, as well as notch filtered at 50 Hz. Two sessions were recorded on different days for each subject, with each session comprising 288 trials. So there is an inherent drift in the statistical distributions between the two sessions. We used the first session for training and the second session for testing to validate the performance of the proposed model.

2.2. Preprocessing

We use a 3-order Butterworth filter to obtain multi-views of EEG signals based on different brain rhythms. We select four sub-bands with 1 Hz overlap, which are δ rhythm (1-5 Hz), θ rhythm (4-8 Hz), α rhythm (7-13 Hz) and β rhythm (12-32 Hz). As mentioned in section I, earlier work has already shown that the signals from these sub-bands can better decode MI-EEG signals and identify the intentions of the subjects. However, suitable boundaries of sub-bands for MI-EEG signals vary from person to person. To ensure that the feature diversity of each individual is properly learned, raw signals with the overall band covering the full frequency spectrum are also fed into the proposed model. Therefore, altogether five types of EEG signals are used as inputs including specific

information related to the motor imagery and the diversity among subjects.

2.3. The proposed model

The schematic of the proposed model is shown in Fig 2. Each view of specific MI-EEG signals is processed using the same deep-learning structure. Therefore, the final fully connected layer receives the same size of features through these five parallel structures.

2.3.1. Definitions

Define the raw EEG signals as $E = (X_i, Y_i)|i = 1, 2, ..., N$, where $X_i \in R^{C\times T}$ represents $i-th$ EEG trial with C channels and T samples. N is the total number of EEG signal trials. In our experiment, C equals 20 and T is 1000 because each trial has 4 s data with a 250 Hz sampling rate. Y_i is the matching label of X_i, which comes from the label set $M = \{m1 : right, m2 : left\}$.

2.3.2. Temporal and spatial block

As a time series, EEG signals contain abundant temporal features. Meanwhile, the different electrode positions on the scalp allow for spatial characterization of brain activities. Therefore, we use two CNN layers to extract temporal and spatial features which are the most important and commonly used in decoding MI-EEG tasks. First, a CNN layer with the kernel size of $1 \times k$ is adopted on each channel to perform a convolution over time. According to the experience of previous studies [29, 45] and ablation experiments on the kernel size selection, we allow $k = 64$ in the proposed model. Then, a depthwise CNN layer is applied across all channels. The number of the filter is set to be one so that signals from all channels at each time instant were compressed into one feature map. This approach facilitates a reduction in the number of trainable parameters and enables the efficient extraction of features [29]. No activation function intervenes between the two layers [28]. Then, the exponential linear unit (ELU) [46] and batch normalization [47] techniques are employed to mitigate the overfitting problem.

2.3.3. Dense block

Assume that the network has L layers, each of which implements a non-linear transformation $F_l(\cdot)$ where l is the index of the layer and the output of each layer is x_l. Traditional transition with a single connection between each CNN layer is:

$$x_l = F_l(x_{l-1}). \tag{1}$$

As the network becomes deeper with more layers, some useful features may also be filtered. Meanwhile, more parameters need to be optimized which leads to overfitting, especially for MI-EEG signals with a limited number of subjects. To address this problem, we refer to [37] and build direct connections from any layer to all subsequent layers. The detailed structure is shown in Fig 2. In a dense block, the lth layer receives the feature maps of all preceding layers, and the activation function is:

$$x_l = F_l([x_0, x_1, ..., x_{l-1}]). \tag{2}$$

where $[x_0, x_1, ..., x_{l-1}]$ are the feature maps from the preceding CNN layers $[l0, l1, ..., l-1]$. Each CNN layer adopts ELU as activation, followed by batch normalization and dropout techniques. If a CNN layer produces k new feature maps, the lth layer has $k_0 + k \times (l-1)$ inputs, where k_0 is the number of feature maps from the input. The k is called the growth rate of the network which reflects how much new information a CNN learned and contributed to the proceeding layer. The short paths from preceding layers to proceeding layers are greatly increased to $\frac{L(L+1)}{2}$ whereas traditional architecture only has L ones in a L layers network. Through the dense block, each layer has access to all the preceding feature maps which enhances the flow of information and feature reuse.

In summary, in the proposed model, the dense block receives an input consisting of 40 feature maps. To mitigate the risk of overfitting, we set k=7, and use 4 CNN layers in each block. Following the first CNN layer, we concatenated the 7 extracted feature maps combined with the original 40 feature maps to form a new input that is then fed into the subsequent CNN layer to learn 7 additional feature maps. This process is repeated before the third CNN layer, thereby allowing for the extraction of more informative and discriminative features. Each subsequent layer of CNN is not only connected to the previous layer, but also has connections to the outputs of all previous layers, enabling the establishment of shorter paths that help the flow of the information.

2.3.4. Fusion block

As mentioned earlier, in addition to the complete raw EEG data, four different inputs represent different views of the MI-EEG signals. These sub-band datasets are based on brain rhythms related to motor imagery tasks. While the raw signals without filtering reflect the diversity of features that varies from person to person. After further learning by two dense blocks, the extracted features are sent into a 1×1 CNN layer to reduce the size of feature maps before fusion. Fewer maps help to improve computational efficiency and reduce fluctuations in the loss function trajectory along the training. Finally, features obtained through all parallel branches are fed to a fully connected layer with a softmax classifier.

Using the single-band input can reduce a number of computational parameters, as demonstrated in classical models such as ConvNet [28] and EEGNet [29]. However, this approach often ignores the impact of differences across different frequency bands. When using multi-branch input that divides EEG into different frequency bands, the learned features need to be integrated before being fed into the classifier. The previous models such as HS-CNN [45] used a simple fully connected layer to fuse the information.

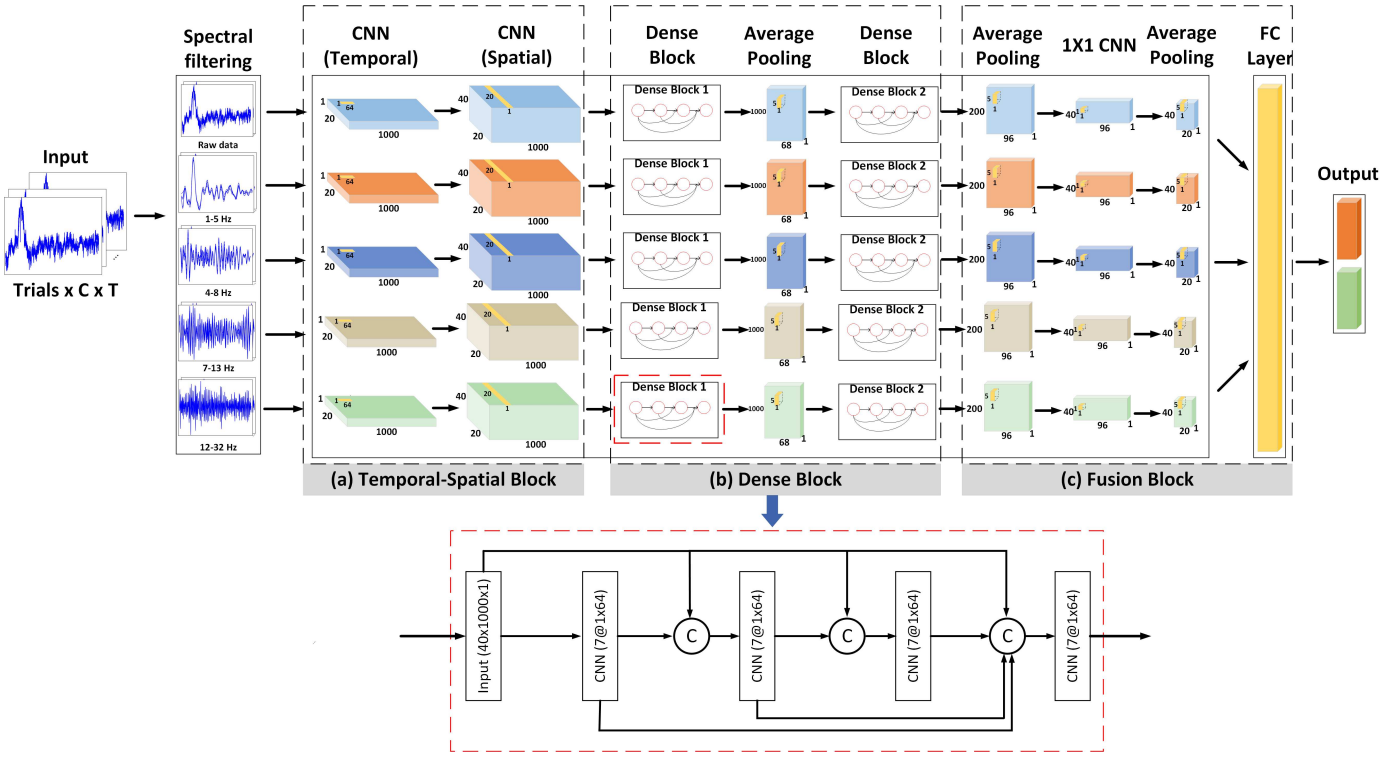


Figure 2: The overview of the proposed model structure. The structure is divided into three blocks: (a) Temporal-Spatial Block; (b) Dense Block; (c) Fusion Block. In the Dense Block, 7@1 × 64 means each CNN layer in the dense block has 7 filters with the size of (1 × 64). C means the concatenation procedure.

In contrast, our model has multiple layers for each branch, leading to numerous calculated feature maps. As a result, we need to reduce the dimensionality through one-dimensional CNN and pooling procedures to prevent overfitting and reduce the model complexity.

2.3.5. Training

The cross-entropy function is selected as a loss function which calculates the distance between the probability distribution of the neural network prediction values y_p and the true labels y_t [48]:

$$L\left(y_p, y_t\right) = -\sum_m y_{p,m} \log y_{t,m}. \tag{3}$$

where m is the index of y. The optimizer is Adam [49] and learning rate $lr = 0.0001$. The training takes 1000 epochs for each fold in the CV with 16 batches per epoch. The early stopping technology was used to save the best weights during each fold. The training step ended after checking if the validation loss value decreased for the last 150 epochs. After reaching the threshold, the model with the best weights produces the classification results of the test fold.

2.3.6. Baseline models

We use two traditional machine learning methods (CSP [16] and FBCSP [20]) and three state-of-the-art deep learning architectures (Shallow ConvNet [28], Deep ConvNet [28], and EEGNet 8-2 [29]) as baseline models to demonstrate the effectiveness of our proposed method. Since the baseline models use different datasets in the initial research, we have used the best parameters of these models and ensure a fair comparison. The details of the baseline models are described as follows:

1. CSP: The basic principle of the CSP algorithm is to find the optimal set of spatial filters for mapping data by using the diagonalization of the matrix. In this way, the difference between the variance values of the two tasks is maximized, thus gaining a feature vector with high discrimination. The methods CSSP, CSSSP and DFBCSP mentioned earlier all use CSP as the kernel algorithm.
2. FBCSP: FBCSP is also a successful algorithm commonly used in the BCI field. After extracting the features through CSP, FBCSP used a feature selection method to automatically select discriminative pairs of frequency bands. According to [20], we decompose EEG signals into nine frequency bands with a bandwidth of 4 Hz from 4 to 40 Hz through a Chebyshev filter. The classifier is the SVM with the default kernel radial bias function (RBF).
3. ConvNet: ConvNet included the shallow and deep two structures. Shallow ConvNet is a DL model with only two CNN layers and an average pooling layer, which achieved a better performance than FBCSP on the public dataset BCI competition IV 2a [50] and

high-gamma dataset [28]. Deep ConvNet included a temporal and spatial filter which are similar to the head of the Shallow ConvNet. Then the layer was followed by several convolution-max-pooling blocks and a fully-connected layer with a softmax classifier. It extended the choice of learning parameters and optimization plans, and has achieved excellent results [51].

4. EEGNet 8-2: Based on the Shallow ConvNet, this model adopted a separable CNN layer after extracting temporal-spatial features, which ensured the quality of the classification results with reduced calculation cost.
5. FBCNet: FBCNet divides the raw EEG signals into several frequency bands. Then the depthwise CNN layer was used to extract spatial features. After that, the model employed a variance layer to compute the temporal variance of the time series. Finally, the features are fused in a fully-connected layer and ended with the softmax. The model referred to the principle of FBCSP and achieved the best classification result on the Korean public MI dataset[43]

3. Results

The computer used in this experiment had 8 Intel cores Intel processors and 16 GB RAM. GTX 2080 GPU with 8 GB memory was used for training and testing EEG data. Keras was used for building the proposed model and the baseline models. The results and statistical analysis of the proposed model are reported in this section, including the performance comparison with other baseline models.

3.1. Overall performance

The averaged classification accuracy of different methods is shown in table I. The results from different methods are 64.69% (±15.25), 63.50% (±19.09), 67.81% (±17.55), 61.83% (±16.89), 68.96% (±17.17), 73.44% (±13.28), 75.16% (±14.01) for CSP, FBCSP, EEGNet 8-2, Deep ConvNet, Shallow ConvNet, FBCNet and our proposed method in KU dataset, respectively. The proposed method achieves 1.72% higher than the best result among the baseline models. To better validate the results, we use statistical significance tests including an Analysis of Variance (ANOVA) test for the multiple comparison tests and paired t-tests between each baseline method and the proposed method. In an ANOVA test, the proposed method significantly exceeds the others [$F = 5.054, p < 0.001$]. The results of paired t-test are CSP [$t_{(53)} = -6.074, p < 0.001$], FBCSP [$t_{(53)} = -6.220, p < 0.001$], EEGNet 8-2 [$t_{(53)} = -5.931, p < 0.001$], Deep ConvNet [$t_{(53)} = -8.226, p < 0.001$], Shallow ConvNet [$t_{(53)} = -5.788, p < 0.001$] and FBCNet [$t_{(53)} = -1.875, p < 0.01$]. The performance comparison for individual subjects based on the scatter plots is presented in Fig.3. There is a significant improvement in classification accuracy for most of the subjects. The proportions of subjects who had better results in the proposed model than in the baseline models are 81.4% (44 of 54), 85.2% (46 of 54), 79.6% (43 of 54), 87.03% (47 of 54), 77.8% (42 of 54), and 55.5% (30 of 54) for CSP, FBCSP, EEGNet 8-2, Deep ConvNet, Shallow ConvNet and FBCNet, respectively. In the BCI IV 2a dataset, while FBCNet exhibited 0.26% higher than the proposed model, our model demonstrated notable classification performance when compared with other baseline models.

Table 1: Comparison of average classification accuracy (%) for different methods.

	KU Dataset	BCI IV 2a Dataset
CSP [16]	64.69 (15.25)	54.01 (12.77)
FBCSP [20]	63.50 (19.09)	65.79 (14.21)
EEGnet 8-2 [29]	67.81 (17.55)	67.81 (17.55)
Deep ConvNet [28]	61.83 (16.89)	65.34 (13.54)
Shallow ConvNet [28]	68.96 (17.17)	68.96 (14.28)
FBCNet [43]	73.44 (13.28)	**72.71 (14.67)**
Proposed model	**75.16 (15.03)**	72.45 (14.10)

Table 2: COMPARISON OF AVERAGE CLASSIFICATION ACCURACY(%) FOR METHODS BASED ON DIFFERENT RHYTHMS.

Algorithm	δ	θ	α	β	Overall
CSP[16]	52.93	53.81	64.50	63.29	54.39
EEGNet 8-2[29]	55.71	**54.09**	**66.63**	67.34	69.69
Deep ConvNet[28]	54.19	51.07	62.56	61.32	65.68
Shallow ConvNet[28]	53.68	52.53	64.38	**68.57**	64.16
Proposed method	**57.43**	54.07	66.52	67.47	**73.52**

3.2. Model performance on different sub-bands

Features extracted from different sub-bands have different impacts on the final classification accuracy. We tested different methods based on both the four commonly used sub-bands related to brain rhythms in MI-EEG decoding tasks and the raw data covering the whole frequency spectrum. Our proposed model fuses extracted features through five paralleled structures with the same parameters. Therefore, to validate the effect of different sub-bands on the proposed model, only one paralleled structure without the fusion step is used in this experiment. The comparison results based on the KU dataset are shown in Table 2. On the θ and α sub-bands, EEGNet 8-2 achieved 54.09% and 66.63% respectively which are the best results. The Shallow ConvNet method performed well on the β sub-band. The proposed methods also produced good classification results based on different sub-bands, especially having much higher accuracy in decoding the raw MI-EEG signals over all frequency bands.

Fig 4 shows the accuracy of the proposed method based on different brain rhythms for each subject. The classification results based on α and β rhythms are higher than other cases for most subjects, which further confirms the findings reported in the previous research [32][33]. θ and δ rhythms also contain useful MI information which can

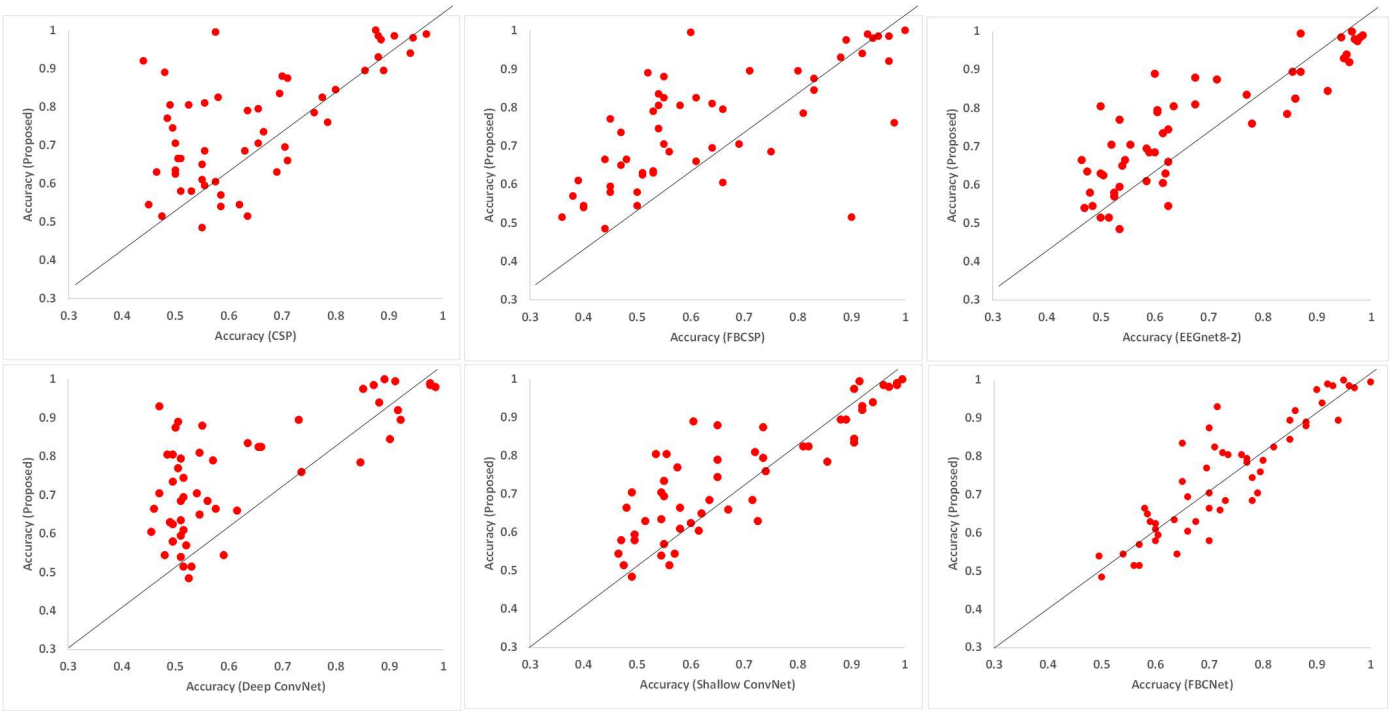


Figure 3: Scatter plot of individual classification performance. The horizontal axis represents the classification accuracy from baseline methods (CSP, FBCSP, EEGNet 8-2, Deep ConvNet, Shallow ConvNet and FBCNet), and the vertical axis represents the classification accuracy from our proposed method.

significantly improve the accuracy of some subjects. Furthermore, the proposed model extracts the most comprehensive information from the raw MI-EEG data covering the whole frequency spectrum without applying any filtering, and has achieved the best classification results for 33 out of 54 subjects.

3.3. Effect of hyper-parameters

The kernel structure of the proposed model is the dense block. To improve the model learning capacity, we test the effect of different numbers of feature maps on the classification performance based on all subjects in the KU dataset. The results are shown in Fig 5(a). The highest accuracy is achieved when the CNN filter in a dense block learns 7 feature maps each time. Fewer feature maps limited the information learned while too many ones lead to the overfitting problem. Besides that, the number of feature maps in the final 1×1 CNN layer also plays an important role. The influence on the model is shown in Fig 5(b). When the dimension of feature maps is compressed to 20, the model has the best performance. From the results we can find that the difference between different numbers of feature maps is not huge. However, if there is no 1×1 CNN layer in the fusion block, the classification rate only reaches about 68% which is similar to other excellent methods. One possible reason is that there are many extracted features from each sub-band signal and the raw data comprising all bands. Without decreasing feature maps by 1×1 CNN layer, a large number of parameters with redundant information will be calculated in the fully connected layer and in the final softmax layer which will have a negative impact on the final classification results.Besides that, we also test the influence of different activation functions (Fig 6). The ELU function performs best with the highest accuracy and runs fast.

3.4. Effect of fusion of sub-bands and the overall band

The classification results for methods based on different rhythms are shown in table II. Although the accuracy of the proposed model on the raw EEG signals covering the full frequency spectrum (overall band) is good enough, the combination with subset signals of specific brain rhythms can achieve better performance. As shown in Fig 7, without combining with the features from the raw data covering the full spectrum (whole bands), the results of the combination of only sub-bands covering brain rhythms δ, θ, α and β are no more than 68%. When we introduced the features from the overall band, the accuracy significantly improved. The proposed model with all sub-bands and the overall band together achieves 75.16% while the combination of the overall band with (α, β) and (θ, α, β) reaches 74.42% and 74.94% respectively.

4. Discussions

4.1. Comparison of different methods

Traditional machine learning methods included feature extraction and classification steps. Inappropriate combination of the feature extraction methods and classifiers

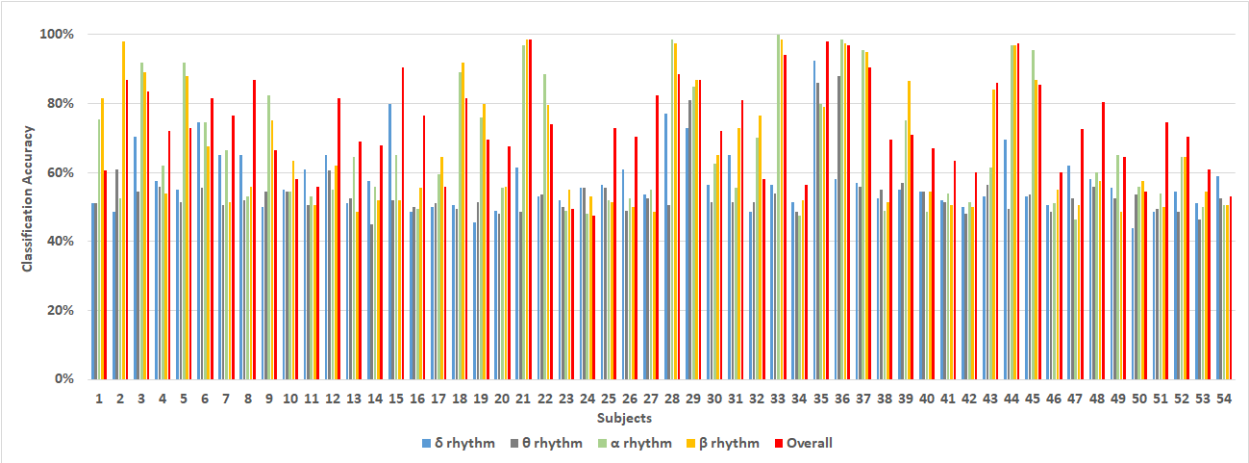
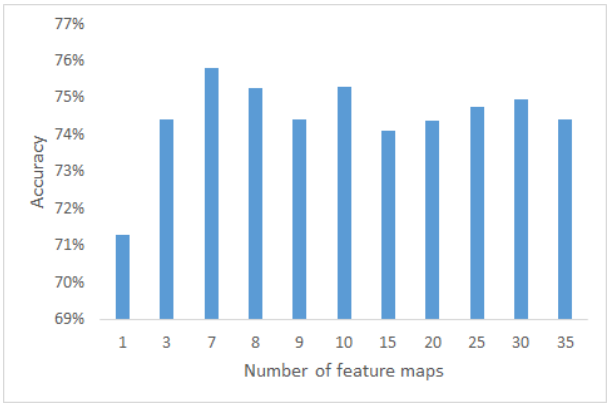
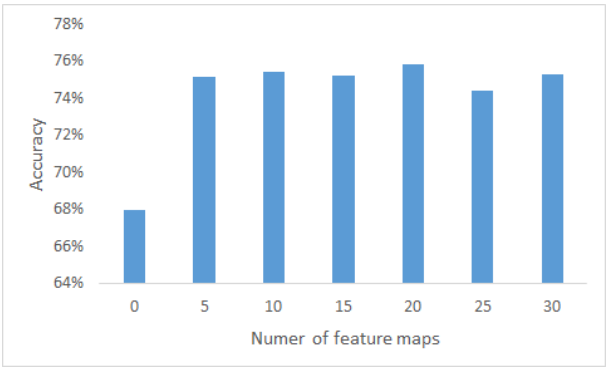


Figure 4: The average accuracy of the proposed method using a 10-fold CV based on different brain rhythms.



(a)



(b)

Figure 5: The effect of numbers of feature maps of the proposed model. (a) The feature maps come from the CNN filters in the dense block. (b) The feature maps come from the 1×1 CNN filters in the fusion block.

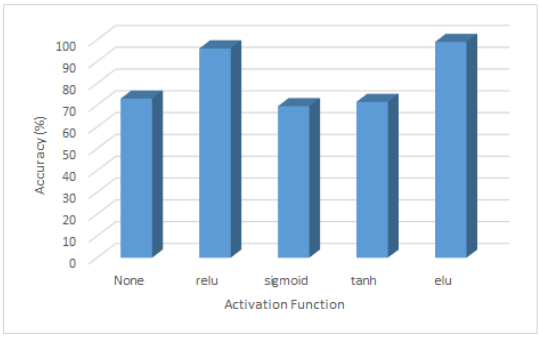


Figure 6: The effect of activation functions on example subject

leads to poor classification results. Deep learning, which has an end-to-end projection, shows great performance in decoding MI-EEG tasks [52]. Our proposed model uses different MI-EEG representations based on various sub-bands and raw data covering the overall band as inputs. The proposed CNN structure improves the model performance through feature reuse and fusion technology. In Fig 3, we show the comparison results with other methods on each subject. For most subjects, the classification accuracy has been significantly improved based on our proposed method, even more than 50% than some benchmark models. The number of subjects whose accuracy is over 80% is 11 for CSP, 15 for FBCSP, 16 for EEGNet 8-2, 12 for Deep ConvNet, 17 for Shallow ConvNet, 16 for FBCNet and 22 for our model, which further verifies the superiority and robustness of the proposed method. Additionally, we utilize the t-SNE [53] to achieve full visualization of the learned features from different methods. The t-SNE algorithm was used on the last fully connected layer. All inputs of the t-SNE are reshaped to trials $\times$ features to show the feature distribution in a two-dimension space. For a fair comparison, the extracted features of all methods are taken from the same subject (Fig 9). Compared with the other three baseline CNN model, our proposed model was able to extract more discriminative features through the dense blocks after learning temporal-spatial information from MI-EEG signals.

4.2. Analysis of fusion of features from sub-bands

To learn useful information related to motor imagery, the signals will pass a filter whose frequency bands are as-

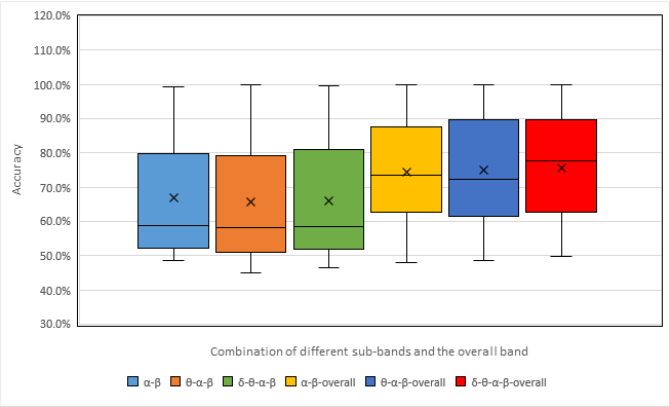


Figure 7: The classification rate of different combinations of sub-bands and the overall band based on the proposed model.

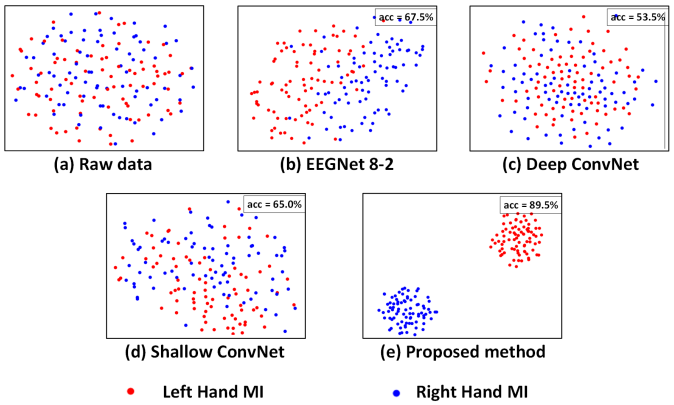


Figure 8: The feature map of the sixth subject obtained by various methods in 2-D embedding based on t-SNE. Part (a) is the distribution of the raw EEG data. Parts (b), (c), (d) and (e) show the distribution of extracted features in trained EEGNet 8-2, Deep ConvNet, Shallow ConvNet and the proposed method. The proposed method achieved 89.5% classification results, whereas EEGNet 8-2, Deep ConvNet and Shallow ConvNet resulted in 67.5%, 53.5% and 65.0% respectively.

sociated with the brain rhythms (commonly use α and β rhythm). The Previous study has shown the significant impact of frequency selection on the final classification results [17]. However, unclear EEG rhythm boundaries and differences in optimal frequency bands for each individual make it difficult to build an effective model. In table II, the same method gives completely different results in decoding MI-EEG signals on different rhythms. These methods tend to give about 10% higher classification results on the α and β rhythm than on the θ and δ rhythm. Our proposed method does not show better performance than other baseline methods on these four commonly used rhythms. However, our method produces much better results on the overall frequency band which reaches 73.52%. Although using signals covering the whole frequency spectrum may introduce more redundant information and noise, it however ensures that all motor imagery information and individual diversity among different subjects. The baseline models such as the EEGNet

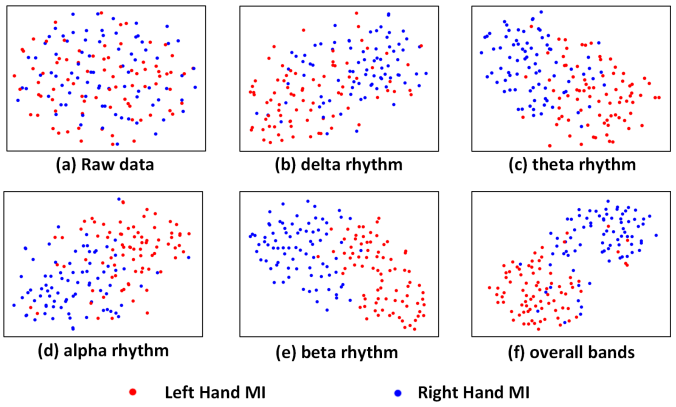


Figure 9: The feature map of the sixth subject with various inputs in 2-D embedding based on t-SNE. Part (a) is the distribution of the raw EEG data. Parts (b), (c), (d), (e) and (f) show the distribution of extracted features from δ rhythm, θ rhythm, α rhythm, β rhythm and the overall bands. The proposed method achieved 89.5% classification on this subject.

8-2 and Deep ConvNet also proved that the classification accuracy on the overall bands is higher than on the α or β rhythm. Fig 9 clearly reveals the significant improvement of the classification accuracy after fusing the features from the overall frequency band. We again use t-SNE before the final fully connected layer to show the distribution of the extracted features from different inputs. Except for the features extracted from δ rhythm, other ones can be distinguished clearly through the t-SNE visualization on the sixth subject.

4.3. Analysis of dense block

Compare with EEGNet 8-2, Deep and Shallow ConvNet, our proposed model has the same structure in the first and second CNN layers to extract temporal-spatial features from EEG signals. The difference is that we have further introduced two dense blocks to make the whole structure deeper. The deeper CNN layers can usually learn more abstract and high-level features which help to improve the model performance. However, a complex model also leads to the overfitting problem especially on limited data such as MI-EEG signals. Further, the non-linear and nonstationary characteristics of EEG may let deeper CNN layers learn more noise and redundant information rather than useful information embedded in the signals. Therefore, previous studies in the literature usually adopted a structure with only one or two CNN layers [52]. From the results listed in table I, EEGNet 8-2 and Shallow ConvNet with fewer CNN layers than Deep ConvNet perform much better. However, fewer CNN layers may limit the learning capability of a deep learning model. In [28], Schirrmeister et al combined ResNet structure to decode MI-EEG tasks, but the performance was worse than traditional methods like FBCSP. One possible reason is that ResNet evaluates the difference between the output of one layer and the input in the preceding layer. But such information is not

suitable to be fed into the following layers because simple addition and subtraction can lose useful EEG information, since the EEG signals are nonstationary and include a mass of noise signals. Our proposed model learns all features in the preceding layers instead of the difference between the inputs and outputs. In the dense blocks, the features learned by any of the CNN layers are connected by all subsequent layers which encourage information flow and feature reuse over the whole model. In the softmax classifier, the outputs are not only influenced by the features fed from the latest layer which may include redundant information due to the overfitting problem but also affected by the feature maps fed from all other preceding layers, which avoid unnecessary information loss as signals pass across through different layers. From table II, it is evident our proposed model can extract more valuable features from the overall frequency bands, while reducing the effects of noise, achieving a better trade-off between model complexity and model learning capability.

4.4. Limitations and future work

Our proposed method achieved the best classification results on the KR MI-EEG dataset, but there are still several limitations that deserve future investigations. First, the whole structure combines five sets of inputs and extracts features separately from each of the five sets of input signals. Although we use average pooling layers and 1×1 CNN to decrease calculated feature maps, it took a substantial amount of time to train the complex model. From Fig 4, we can find that good classification results are achieved for a few subjects only using the δ rhythm. If we can select the most relevant rhythm during training and prune the redundant layers and neurons, the model compactness can be improved. Therefore, future work is to identify the most suitable sub-bands and adopt neuron pruning methods to reduce the model size. Secondly, our model is based on within-the-subject like existing approaches in the literature, but it also implies that we have to build a specific model for each subject. In reality, building a cross-subject model is practically more valuable, especially for stroke patients. Saving much time to train a model can effectively help patients to be much more quickly engaged with the rehabilitation exercise. To achieve this objective, techniques such as transfer learning and adaptive learning will be adopted in future work. Finally, our model uses both the temporal and spatial features of MI-EEG signals and the selection of channels is primarily based on previous studies and experience. However, different subjects may have different motor imagery areas. Therefore, another future work will introduce automatic channel selection methods to improve our model adaptability.

5. Conclusion

In this paper, we propose a novel DL architecture based on the densely connected CNN for the recognization of MI tasks. Our proposed model uses both filtered MI-EEG signals based on four commonly used brain rhythms and the overall frequency band as inputs. The network first extracts temporal and spatial features using the first two CNN layers. Then, two dense blocks connect each CNN layer to all the rest layers in a feed-forward mode to further learn discriminative MI-EEG information. The dense block encourages feature reuse and strengthens information propagation. Next, average pooling layers and 1×1 CNN layer help to reduce computation and avoid the overfitting problem. Finally, the fully connected layer fuses the extracted feature from different inputs and ends with a classifier. The fused features include special motor imagery information based on different brain rhythms and also consider individual diversity among subjects without finding the optimal sub-bands. Both the classification accuracy and the distribution of extracted feature maps have demonstrated the superiority of the proposed method in the decoding MI-EEG tasks when compared with benchmark models, achieving an average accuracy of 75.16% on the public Korea University EEG datasets, higher than other state-of-the-art deep learning methods.

References

[1] G. Pfurtscheller, G. R. Müller-Putz, R. Scherer, C. Neuper, Rehabilitation with brain-computer interface systems, Computer 41 (10) (2008) 58–65.

[2] U. Chaudhary, N. Birbaumer, A. Ramos-Murguialday, Brain–computer interfaces for communication and rehabilitation, Nature Reviews Neurology 12 (9) (2016) 513–525.

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

[4] K. K. Ang, C. Guan, Brain-computer interface in stroke rehabilitation, Journal of Computing Science and Engineering 7 (2) (2013) 139–146.

[5] X. Ma, S. Qiu, W. Wei, S. Wang, H. He, Deep channel-correlation network for motor imagery decoding from the same limb, IEEE Transactions on Neural Systems and Rehabilitation Engineering 28 (1) (2019) 297–306.

[6] O.-Y. Kwon, M.-H. Lee, C. Guan, S.-W. Lee, Subject-independent brain–computer interfaces based on deep convolutional neural networks, IEEE transactions on neural networks and learning systems 31 (10) (2019) 3839–3852.

[7] J. Decety, D. H. Ingvar, Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation, Acta psychologica 73 (1) (1990) 13–34.

[8] M. Bentlemsan, E.-T. Zemouri, D. Bouchaffra, B. Yahya-Zoubir, K. Ferroudji, Random forest and filter bank common spatial patterns for eeg-based motor imagery classification, in: 2014 5th International conference on intelligent systems, modelling and simulation, IEEE, 2014, pp. 235–238.

[9] J. Luo, Z. Feng, J. Zhang, N. Lu, Dynamic frequency feature selection based approach for classification of motor imageries, Computers in biology and medicine 75 (2016) 45–53.

[10] S. Aggarwal, N. Chugh, Signal processing techniques for motor imagery brain computer interface: A review, Array 1 (2019) 100003.

[11] C.-Y. Chen, C.-W. Wu, C.-T. Lin, S.-A. Chen, A novel classification method for motor imagery based on brain-computer interface, in: 2014 International joint conference on neural networks (IJCNN), IEEE, 2014, pp. 4099–4102.

[12] R. Fu, Y. Tian, T. Bao, Z. Meng, P. Shi, Improvement motor

imagery eeg classification based on regularized linear discriminant analysis, Journal of medical systems 43 (6) (2019) 1–13.
[13] M. R. Islam, T. Tanaka, M. S. Akter, M. K. I. Molla, Classification of motor imagery bci using multiband tangent space mapping, in: 2017 22nd International Conference on Digital Signal Processing (DSP), IEEE, 2017, pp. 1–5.
[14] G. Sagee, S. Hema, Eeg feature extraction and classification in multiclass multiuser motor imagery brain computer interface u sing bayesian network and ann, in: 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), IEEE, 2017, pp. 938–943.
[15] M. Hamedi, S.-H. Salleh, A. M. Noor, I. Mohammad-Rezazadeh, Neural network-based three-class motor imagery classification using time-domain features for bci applications, in: 2014 IEEE region 10 symposium, IEEE, 2014, pp. 204–207.
[16] G. Pfurtscheller, C. Neuper, Motor imagery and direct brain-computer communication, Proceedings of the IEEE 89 (7) (2001) 1123–1134.
[17] S. Lemm, B. Blankertz, G. Curio, K.-R. Muller, Spatio-spectral filters for improving the classification of single trial eeg, IEEE transactions on biomedical engineering 52 (9) (2005) 1541–1548.
[18] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K.-R. Müller, Optimizing spatio-temporal filters for improving brain-computer interfacing, Advances in Neural Information Processing Systems 18 (2005).
[19] Q. Novi, C. Guan, T. H. Dat, P. Xue, Sub-band common spatial pattern (sbcsp) for brain-computer interface, in: 2007 3rd International IEEE/EMBS Conference on Neural Engineering, IEEE, 2007, pp. 204–207.
[20] K. K. Ang, Z. Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (fbcsp) in brain-computer interface, in: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), IEEE, 2008, pp. 2390–2397.
[21] H. Higashi, T. Tanaka, Simultaneous design of fir filter banks and spatial patterns for eeg signal classification, IEEE transactions on biomedical engineering 60 (4) (2012) 1100–1110.
[22] S. M. Mousavi, A. Asgharzadeh-Bonab, R. Ranjbarzadeh, Time-frequency analysis of eeg signals and glcm features for depth of anesthesia monitoring, Computational Intelligence and Neuroscience 2021 (2021).
[23] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghoushchi, S. Anari, M. Naseri, M. Bendechache, Brain tumor segmentation based on deep learning and an attention mechanism using mri multi-modalities brain images, Scientific Reports 11 (1) (2021) 1–17.
[24] R. Ranjbarzadeh, S. Jafarzadeh Ghoushchi, S. Anari, S. Safavi, N. Tataei Sarshar, E. Babaee Tirkolaee, M. Bendechache, A deep learning approach for robust, multi-oriented, and curved text detection, Cognitive Computation (2022) 1–13.
[25] S. Baseri Saadi, N. Tataei Sarshar, S. Sadeghi, R. Ranjbarzadeh, M. Kooshki Forooshani, M. Bendechache, Investigation of effectiveness of shuffled frog-leaping optimizer in training a convolution neural network, Journal of Healthcare Engineering 2022 (2022).
[26] R. Ranjbarzadeh, N. Tataei Sarshar, S. Jafarzadeh Ghoushchi, M. Saleh Esfahani, M. Parhizkar, Y. Pourasad, S. Anari, M. Bendechache, Mrfe-cnn: multi-route feature extraction model for breast tumor segmentation in mammograms using a convolutional neural network, Annals of Operations Research (2022) 1–22.
[27] R. Ranjbarzadeh, S. Dorosti, S. Jafarzadeh Ghoushchi, S. Safavi, N. Razmjooy, N. Tataei Sarshar, S. Anari, M. Bendechache, Nerve optic segmentation in ct images using a deep learning model and a texture descriptor, Complex & Intelligent Systems 8 (4) (2022) 3543–3557.
[28] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for eeg decoding and visualization, Human brain mapping 38 (11) (2017) 5391–5420.
[29] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, B. J. Lance, Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces, Journal of neural engineering 15 (5) (2018) 056013.
[30] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, M. S. Hossain, Deep learning for eeg motor imagery classification based on multi-layer cnns feature fusion, Future Generation computer systems 101 (2019) 542–554.
[31] D. Freer, G.-Z. Yang, Data augmentation for self-paced motor imagery classification with c-lstm, Journal of neural engineering 17 (1) (2020) 016041.
[32] H. H. Jasper, H. L. Andrews, Electro-encephalography: Iii. normal differentiation of occipital and precentral regions in man, Archives of Neurology & Psychiatry 39 (1) (1938) 96–115.
[33] H. Jasper, W. Penfield, Electrocorticograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus, Archiv für Psychiatrie und Nervenkrankheiten 183 (1) (1949) 163–174.
[34] O. Mokienko, L. Chernikova, A. Frolov, P. Bobrov, Motor imagery and its practical application, Neuroscience and Behavioral Physiology 44 (5) (2014) 483–489.
[35] M. Ahn, H. Cho, S. Ahn, S. C. Jun, High theta and low alpha powers may be indicative of bci-illiteracy in motor imagery, PloS one 8 (11) (2013) e80886.
[36] L. R. Trambaiolli, P. J. Dean, A. M. Cravo, A. Sterr, J. R. Sato, On-task theta power is correlated to motor imagery performance, in: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, 2019, pp. 3937–3942.
[37] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
[38] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[39] G. Larsson, M. Maire, G. Shakhnarovich, Fractalnet: Ultra-deep neural networks without residuals, arXiv preprint arXiv:1605.07648 (2016).
[40] T. Liu, D. Yang, A densely connected multi-branch 3d convolutional neural network for motor imagery eeg decoding, Brain Sciences 11 (2) (2021) 197.
[41] Z. Yu, W. Chen, T. Zhang, Motor imagery eeg classification algorithm based on improved lightweight feature fusion network, Biomedical Signal Processing and Control 75 (2022) 103618.
[42] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, S.-W. Lee, Eeg dataset and openbmi toolbox for three bci paradigms: an investigation into bci illiteracy, GigaScience 8 (5) (2019) giz002.
[43] R. Mane, N. Robinson, A. P. Vinod, S.-W. Lee, C. Guan, A multi-view cnn with novel variance layer for motor imagery brain computer interface, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, pp. 2950–2953.
[44] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, et al., Review of the bci competition iv, Frontiers in neuroscience (2012) 55.
[45] G. Dai, J. Zhou, J. Huang, N. Wang, Hs-cnn: a cnn with hybrid convolution scale for eeg motor imagery classification, Journal of neural engineering 17 (1) (2020) 016025.
[46] D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), arXiv preprint arXiv:1511.07289 (2015).
[47] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, PMLR, 2015, pp. 448–456.
[48] R. Zhang, Q. Zong, L. Dou, X. Zhao, A novel hybrid deep learning scheme for four-class motor imagery classification, Journal of neural engineering 16 (6) (2019) 066004.
[49] D. P. Kingma, J. Ba, Adam: A method for stochastic optimiza-

tion, arXiv preprint arXiv:1412.6980 (2014).
[50] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, G. Pfurtscheller, Bci competition 2008–graz data set a, Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology 16 (2008) 1–6.
[51] K. Zhang, N. Robinson, S.-W. Lee, C. Guan, Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network, Neural Networks 136 (2021) 1–10.
[52] A. Craik, Y. He, J. L. Contreras-Vidal, Deep learning for electroencephalogram (eeg) classification tasks: a review, Journal of neural engineering 16 (3) (2019) 031001.
[53] L. Van Der Maaten, Accelerating t-sne using tree-based algorithms, The Journal of Machine Learning Research 15 (1) (2014) 3221–3245.