IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 April 2023, accepted 4 June 2023, date of publication 12 June 2023, date of current version 19 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3285236

RESEARCH ARTICLE

Effects of Data Including Visual Presentation and Rest Time on Classification of Motor Imagery of Using Brain-Computer Interface Competition Datasets

KENTO SUEMITSU [1] AND ISAO NAMBU [2], (Member, IEEE)

[1]Department of Science of Technology Innovation, Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan

[2]Department of Electrical, Electronics and Information Engineering, Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan

Corresponding author: Kento Suemitsu (ksuemitsu@stn.nagaokaut.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (KAKENHI) under Grant 21H03480 and Grant 21K18304.

ABSTRACT Herein, we investigated the effects of using time segments, including visual presentation, motor imagery, and rest time, as training data in a brain-computer interface (BCI) competition. Using BCI Competition IV 2a and 2b, many researchers have attempted to create more robust classifiers with higher classification accuracy. Some studies have also used visual presentation time and rest time as training data. However, the use of training data outside of motor imagery makes comparisons of performance across models difficult, and may lead to models that are overfitted to the experimental environment. In addition, it is possible that brain activity other than motor imagery is involved in visual presentation. Hence, to examine the effects of the selection of training data, we compared several classifiers, including linear discriminant analysis (LDA), support vector machine, and convolutional neural networks (CNN), trained with data including visual presentation time and rest time, with data only during motor imagery. The results showed an improvement in performance when BCI Competition IV 2a and 2b data included visual presentation information in the training data. For the greatest improvement among participants, training data with visual presentation improved the accuracy by 13.44 % and 10.14 % in BCI Competition IV 2a (participant 9) for LDA and CNN, respectively; and by 8.38 % and 16.68 % in BCI Competition IV 2b (participant 3) for LDA and CNN, respectively. Training data that includes visual presentation information improves model performance, therefore, we recommend using only motor imagery time to train the model.

INDEX TERMS Brain-computer interfaces, convolutional neural networks, deep learning, electroencephalography, machine learning.

I. INTRODUCTION

Recently, a brain-computer interface (BCI) using electroencephalography (EEG) during motor imagery has been developed. This has enabled the manipulation of robotic arms in real-life applications [1], and virtual limbs and avatars in the virtual space [2], [3]. Consequently, the development of this technology may provide a new interface for patients with paralyzed limbs due to spinal injuries and cerebrovascular accidents.

There are two types of BCI: synchronous BCI [4], where participants are asked to perform a presentation task during an indicated period of time; and asynchronous BCI [5], where participants spontaneously control a specific EEG. In synchronous BCI, the experimental time schedule consists of a waiting period, task presentation period, motor imagery period, and rest period. Usually, the task presentation time is conducted before motor imagery time, and is necessary

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li.

to indicate the motor imagery that the participant should perform.

Jeong et al. visually presented a direction for 3 s, followed by 4 s of motor imagery to decode the direction of motor imagery [6]. Ofner et al. measured motor imagery in six different upper extremities and presented auditory information during trial measurements (beep-on). Two seconds after the presentation of the visual information, participants were asked to perform motor imagery [7]. In addition, Zolfaghari et al. aimed to decode the speed of upper limb motor imagery. They presented visual information for 12 s and had participants perform sustained motor imagery for 15.5 s after the presentation [8].

It is difficult to perform such a variety of experiments as they require specialist researchers, sophisticated equipment, participants, and facilities. Therefore, a BCI Competition IV was held in 2008, which provided high-quality neuroscience data to various researchers. The purpose of this competition was to find models with superior EEG classification performance and address the challenges of practical BCI systems. Several studies have used these datasets to evaluate classifier performance [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. The competition provided datasets with various practical challenges [22].

The BCI Competition IV 2a and 2b are well-known motor imagery datasets that are used for the evaluation of EEG analysis or classification models. In BCI Competition IV 2a, the training model is required to be more comprehensive in recognizing the four motor imageries. Moreover, the possibility of achieving more natural usability can be evaluated. BCI Competition IV 2b provides a dataset where two motor imageries are performed. The main feature of this dataset is that the time schedule and presentation contents differ between the sessions. Simply, a useful classifier for this dataset does not require recalibration.

Modern models used in classifying such motor imagery have various techniques, including deep learning, transfer learning, fuzzy models, and machine learning using linear discriminant models [11], [23], [24], [25]. To the best of our knowledge, from several reviews and recent papers [14], [24], [25], [26], [27], training models with high classification accuracies of 90 % or higher exist for BCI Competition IV 2a [10], [13], [14], [20] and IV 2b [10], [12], [16].

The most significant problem with studies using the data provided at BCI Competition IV is that they use temporal information, such as task presentation time, not motor imagery, as training data. In a recent study, temporal information, such as visual presentation time, which is not directly related to motor imagery, was used as training data. Our literature search revealed 10 studies on BCI Competition IV 2a [9], [11], [13], [15], [17], [19], [21], [28], [29], [30] and five on BCI Competition IV 2b [17], [18], [21], [28], [30] that used visual presentation time and rest time as training data and showed high classification accuracy. In addition, they were used in both machine and deep learning classifiers.

There are two concerns with the use of data other than motor imagery in competition data. One is the difficulty of comparing the performance of different models. Even if models show the same performance, when time segment used for training data differs, it becomes difficult to make easy comparisons across models. Our results show that the performance of the models changes as the time segment is varied. The second issue is the mismatch between the experimental and practical environments. In experiments with motor imagery, the experimental designer directs the task through visual presentations and cues. However, in the practical environment, there is no visual presentation information that would indicate the type of task by the experiment designer. Therefore, using visual presentation information as training data can result in overfitting the model to the experimental environment.

Moreover, an EEG with visual presentation information may contain information that classifies the type of task. Functional magnetic resonance imaging (fMRI) signals obtained by presenting images for 1 s can be used to classify images [31]. Furthermore, studies on the categorization of visual information, such as text, objects, and scenes, have been conducted using EEG with a high temporal resolution [32], [33], [34], [35].

These studies suggest that models with training data that includes visual presentation information and other information may affect performance. This can make it difficult to compare the performance of different models and may result in models that are not suitable for practical environments. Nevertheless, in the evaluation of the experimental environment, training data that includes visual presentation information has the potential to increase classification accuracy. Therefore, we compared the performance of models trained on data from different time segments.

Specifically, we created datasets for various time windows and widths in BCI Competition IV 2a and 2b. The model was trained using linear discriminant analysis (LDA), support vector machine (SVM), neural network (NN), and convolutional neural networks (CNNs) with created datasets.

In BCI Competition IV 2a and 2b, the results showed that including the visual presentation time in the training data improved the classification accuracy of classifiers, including LDA, SVM, NN, and CNNs.

II. RELATED WORKS

Various experimental paradigms have been investigated to analyze brain activity. We introduced processing tasks in the brain in the competition data other than motor imagery and classifiers in which various time segments are learned through machine learning and deep learning.

A. VARIOUS BRAIN ACTIVITIES MIXED IN COMPETITION DATA

During motor imagery, some of the motor cortex used in motor execution is activated, which is an important piece of information that contributes to discrimination, because the activity is similar to motor execution [36]. In BCI Competition IV 2a and 2b, we considered mixing different potential activities than during motor imagery because of the convenience of the experiment.

First, brain activity during the preparatory period before motor imagery is thought to be mixed. During this period, the higher motor cortex is activated by planning and preparing for exercise. In fact, it is possible to decode and discriminate motor preparation activity [37].

Second, brain activity also exists during visual presentation. Kay et al. and Kamitani et al. reported that visual information, such as images, can be discriminated from brain activity using fMRI [31], [38]. In particular, Kay et al. classified visual information (120 natural images) by analyzing voxels obtained by fMRI [31]. They used voxels from the presentation of a natural image for 1 s followed by a gray background at 3 s, and their model had a classification accuracy of up to 92 %. Studies have also been conducted on the temporal dynamics of brain activity in response to visual information. Harel et al. reported that P2, which is an event-related potential, peaked early (220 ms) after stimulus onset for scene images [39]. For classification of images as visual information, Spampinato et al. used a recurrent neural network (RNN) and CNN to classify 40 images of ImageNet presented during 0.5 s, achieving 85.4 % accuracy with EEG data from 0.04–0.48 s [32]. Zheng et al. proposed different deep learning models and achieved 97.13 % classification accuracy using the Spampinato et al. dataset [32], [33].

Thus, based on previous studies [31], [32], [33], [34], [35], [37], [38], the process of recognizing visual information, determining the kind of imagery to make, and preparing a plan takes place in the visual presentation in BCI Competition IV 2a and 2b. These results indicate that visual presentation time provides information that contributes to classification. This means that performance may vary depending on the time segment used for training data. In fact, our results show that including visual presentation information in the training data improves the performance of the models. This would make it difficult to compare performance across different models.

Furthermore, The experimental environment would include the experiment designer's intent to present the type of task [22], whereas the practical environment provides no information that presents the experimenter's task instructions necessary for evaluation.

Therefore, we recommend using only motor imagery time as training data to train the model.

B. TRADITIONAL MACHINE LEARNING AND DEEP LEARNING APPROACHES

In this section, we introduce different time segments, feature extraction methods, and classification methods used for machine learning and deep learning.

In machine learning, the feature extraction method using a common spatial pattern (CSP) showed the best accuracy in BCI Competition IV 2a and 2b [22]. Therefore, several improved CSP methods have been proposed. Ang et al. used Filter Bank CSP (FBCSP) in BCI Competition IV 2a and 2b, where the time segment of 0.5–2.5 s from visual presentation was used as training data and multiple bandwidth-limited signals were used for feature extraction with CSP, respectively [28]. Lotte et al. proposed Regularizing CSP in BCI Competition IV 2a, using the time segment 0.5–2.5 s from visual presentation as training data [29]. In BCI Competition IV 2a and 2b, Fang et al. divided the time segment 0.5–3.5 s from visual presentation to the end of motor imagery into six time windows, each of which was processed by filter banking [30].

An automatic time selection method was also proposed based on the idea that the optimal time segment for learning is participant-specific. In BCI Competition IV 2a, Ang et al. proposed an optimal spatial-temporal pattern (OSTP) [40] selection method that automatically selects the optimal time segment based on the signal-to-noise ratio-based mutual information calculation method [41].

Mahmoudi et al. automatically selected the optimal time segment out of three: 0.5–2.5 s, 1–3 s, and 1.5–3.5 s after the start of visual presentation. In addition, based on mutual information, they searched for the optimal time segment in BCI Competition IV 2a, which varies from 2.5–7 s with a window size of 2 s [42].

In deep learning, the neural network most commonly used for motor imagery classification is the CNN [14], [24], [25], [26], [27]. In the imaging field, AlexNet [43], Visual geometry Group Network (VGGNet) [44], GoogLeNet [45], and Residual NN (ResNet) [46] have been proposed. These networks were fitted to the input data through multiple nonlinear transformations by increasing the layer depth, and a high classification accuracy was obtained. However, EEG classification overfits because the amount of data is small, and the capacity of the model is large in deeper layers.

Therefore, CNNs with relatively shallow layers have been proposed. Here, methods to suppress overfitting include data augmentation [14], fine-tuning [47], and regularization, such as batch normalization [48] and dropout [49]. However, shallow layers may not learn enough features of the input data owing to their small capacity.

Dai et al. adapted multiple data augmentation to temporal data in the time segmentation 0.5–4 s from visual presentation to the end of motor imagery. In addition, the signals were divided using filter banks, and each was inputted in parallel to a different convolution layer. In addition, they predicted that there would be kernel sizes suitable for each participant, and proposed a hybrid convolution scale CNN (HS-CNN) [17] with multiple convolution layers with different kernel sizes in parallel. Roy et al. adapted two feature extraction methods and multiple data augmentation methods for temporal data during motor imagery. They proposed a Multi-Scale CNN (MSCNN) [11], which improves the HS-CNN and adds a pooling layer to the parallel units. Li et al. proposed a multi-level multi-scale feature fusion CNN (MLMSFF-CNN) [13] that has a parallel structure, such as a Multi-Scale CNN, but

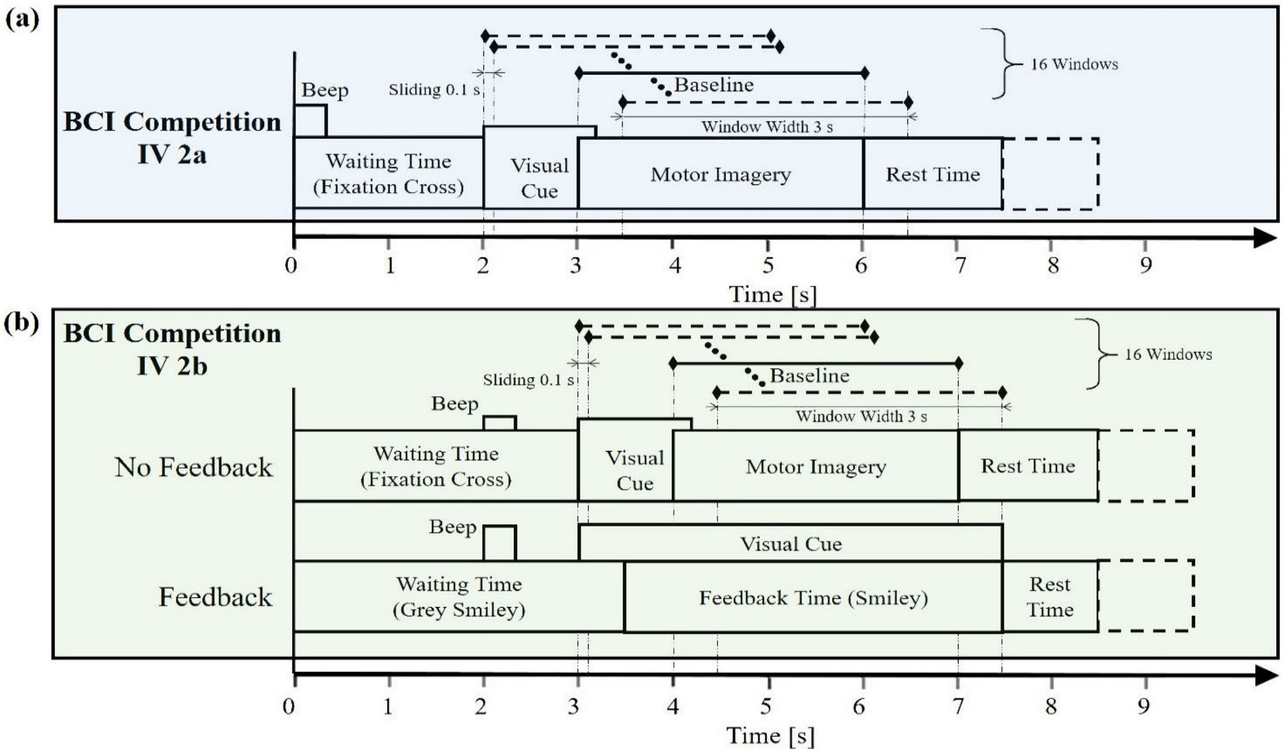


FIGURE 1. Timing scheme of the paradigm of (a) BCI Competition IV 2a and (b) BCI Competition IV 2b; and time windows in window width which is 3 s (dashed line with diamond-shaped ends in the figure). Window width are 2, 2.5, 3, and 3.5 s. Each time window which has one window width was slid in 0.1 s increments. Time windows with widths of 2, 2.5, 3, and 3.5 s produced 26, 21, 16, and 11 training data, respectively. The time window on motor imagery only was used as a baseline (3–6 s and 4–7 s) for comparing the results of the various time windows in BCI Competition IV 2a and 2b, respectively (line with diamond-shaped ends in the figure). (a) Each time window was slid between the start of the visual presentation and the rest time (2–6.5 s). (b) In no-feedback session, each time window was slid between the start of the visual presentation and the rest time (3–7.5 s). In addition, the fixation crosses are presented as visual information. In feedback sessions, gray smileys are presented as visual information. The first two sessions do not have feedback, the second three sessions have feedback. The first three sessions are provided as training data and the rest as test data.

extracts and concatenates the outputs of five convolutional blocks with different depths. They also prepared training data for multiple time intervals (from visual presentation, 0–4 s, 0.25–4.25 s, 0.5–4.5 s, 0.75–4.75 s, and 1–5 s), including visual presentation time, motor imagery time, and rest time.

In machine and deep learning-based classifiers, many studies used time segments other than during motor imagery as training data [9], [11], [15], [17], [18], [19], [21], [28], [29], [30]. Thus, various models have been proposed that use different time segments as training data, making simple comparison of performance between models difficult. Therefore, we recommend using only data during motor imagery for training.

III. METHODOLOGY

Our objective was to enable comparisons between models by using motor imagery as training data, and train models with data that closely resembles the practical environment. Thus, we examined the effects of using temporal information other than motor imagery as training data, which is currently employed in many proposed models. Therefore, we trained the machine learning and NN model at various time windows and various window widths on two datasets.

A. DATASETS

We used two competition data sets.

The first dataset is BCI Competition IV 2a [22], [50], [51]. Nine healthy participants performed left-hand (class 1), right-hand (class 2), both feet (class 3), and tongue (class 4) motor imagery tasks. Fig. 1 (a) shows the time schedule of the trial. Participants were presented with visual information supporting the type of motor imagery 2 s after the beep-on presentation. They were asked to perform a motor imagery task until the fixation cross disappeared from the screen (t = 6 s). In total, 288 trials were performed in one session, followed by two sessions. The experiment was conducted over two days, with data from the first day serving as the training data and data from the second day as the test data. EEG was measured using a 10–20 electrode system with 22 Ag/AgCl electrodes (Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, and POz) and three electrooculography (EOG) channels. The sampling frequency was 250 Hz, and a 50 Hz notch filter and 0.5–100 Hz bandpass filter were applied. The second dataset was BCI Competition IV 2b [22], [52]. Nine healthy participants were asked to perform left (class 1) and right-hand (class 2) motor imagery tasks. Fig. 1 (b) shows the time schedule of the trials. In the no-feedback session, a cross was presented for 3 s during the waiting period before motor imagery and visual information supporting the type of motor imagery was presented. One second after visual presentation, 3 s of motor imagery was performed. In the session with feedback, a gray smiley was presented during the waiting period before motor imagery. In 3–7 s, the visual information was presented. Participants were supported to move the smiley to the left or right in response to the cue, and from 3.5–7.5 s they received feedback with the smiley changing to green for the correct direction and red for the incorrect direction. During the feedback period, the smiley changed color depending on the direction of movement. The distance from the origin of the smiley was set according to the integrated classification output over 2 s. Participants were instructed to prolong their motor imagery during the visual presentation period. The EEG was measured using three bipolar electrodes (C3, Cz, and C4) and three EOG channels. The sampling frequency was 250 Hz, and a 50 Hz notch filter and 0.5–100 Hz bandpass filter were applied.

B. PREPROCESSING

We applied a third-order 0.5–30 Hz Butterworth bandwidth-limited filter to the two datasets. We extracted EEG of different time segments, as in Fig. 1, to test differences in classification performance by learning different temporal information: four time windows with widths of 2, 2.5, 3, and 3.5 s, each sliding in 0.1 s increments.

First, in BCI Competition IV 2a, each time window with one window width was slid from the visual presentation start time to the rest time (2–6.5 s). In time windows with widths of 2 s, 26 different time windows of 2–4 s, 2.1–4.1 s,…, 4.5–6.5 s were created. In time windows with widths of 2.5 s, 21 different time windows of 2–4.5 s, 2.1–4.6 s,…, 4–6.5 s were created. In time windows with widths of 3 s, 16 different time windows of 2–5 s, 2.5–5.1 s,…, 3.5–6.5 s were created. In time windows with widths of 3.5 s, 11 different time windows of 2–5.5 s, 2.1–5.6 s,…, 3–6.5 s were created. In total, there are 74 time windows.

Second, in BCI Competition IV 2b, each time window with one window width was slid from the visual presentation start time to the rest time (3–7.5 s). A total of 74 time windows were created, as well as in the BCI Competition IV 2a case.

C. TRADITIONAL MACHINE LEARNING APPROACHES

Our goal was to investigate how the classification accuracy of machine learning classifiers is affected by various time segments. We employed CSP for feature extraction with reference to the FBCSP [53], which showed the best results in BCI Competition IV 2a; the spatial filter $\boldsymbol{W}$ of the CSP computes spatial features with the best variance between two classes. In addition, FBCSP adapts CSP to each frequency band called a filter bank. Therefore, since frequency information is also important as a feature, FFT was used to convert time domains into frequency domains. Then, integral values were taken for each of alpha (8–13 Hz) and beta band (13–30 Hz) and these were used as feature values. These two feature extraction methods were classified by LDA [54] and SVM [55].

D. DEEP LEARNING

Our goal was to investigate how the classification accuracy of deep learning classifiers is affected by various time segments.

We employed CNN, which are often used as classifiers in deep learning for motor imagery [24], [25], [26], [27]. We created CNNs with two different input formats. As shown in Fig. 2, the structure is the same except for the convolutional layers. The one dimensional-CNN (1D-CNN) adopted a time $\times$ 1 input format, which maps the filter direction of the convolution layer to the channel data of the input data. The two dimesnsional-CNN (2D-CNN) adopts a time $\times$ channel input format. Accordingly, the number of input filters in the convolutional layer was one.

EEG is nonstationary and has covariate shifts that follow different probability distributions for training and test data [56]. To address this problem, the output of the convolution layer is inputted to a batch normalization layer [48] with regularization effects. The output of the batch normalization layer is the input to the pooling layer. While the often-used max pooling layer encodes distortions and translations in the input data, it discards a large amount of data and pools them non-continuously, which limits its generalization. Therefore, we adapted the fractional max pooling layer [57] with randomized pooling regions to reduce the size of the transmitted data and determine a potential suitable kernel size for the participants [17]. Finally, to reduce the number of calculations, smoothing was performed in the global average pooling layer [58] and inputted to the dense layer. The ReLu function [59] is often used as the activation function in motor imagery classification to suppress gradient loss; the ReLu function returns the input value if the input is positive but is otherwise zero. We used Softplus function [60], which smooths the zeros and linear parts of the ReLu function, in the hope of improving stability.

It is also important to investigate whether other neural networks perform well with the time segment. Therefore, we conducted the same experiment on neural networks with only fully-connected layers. We found a hidden layer size of 150, and used batch normalization [48] and dropout [49] used as regularization techniques. The ReLu function was used as the activation function; features from CSP and FFT were used as input data and trained; CSP-NN and FFT-NN, respectively.

E. ANALYISIS AND EVALUATION METHODS

We used the libraries and frameworks provided to simplify programming, using Python version 3. Feature extraction by CSP and classification by LDA were performed by MNE-Python version 1.2.3 [61]. SVM were performed by scikit-learn 1.2.2. NN and CNNs were trained on NVIDIA GeForce RTX 3070 with Cudatoolkit version 11 in Pytorch

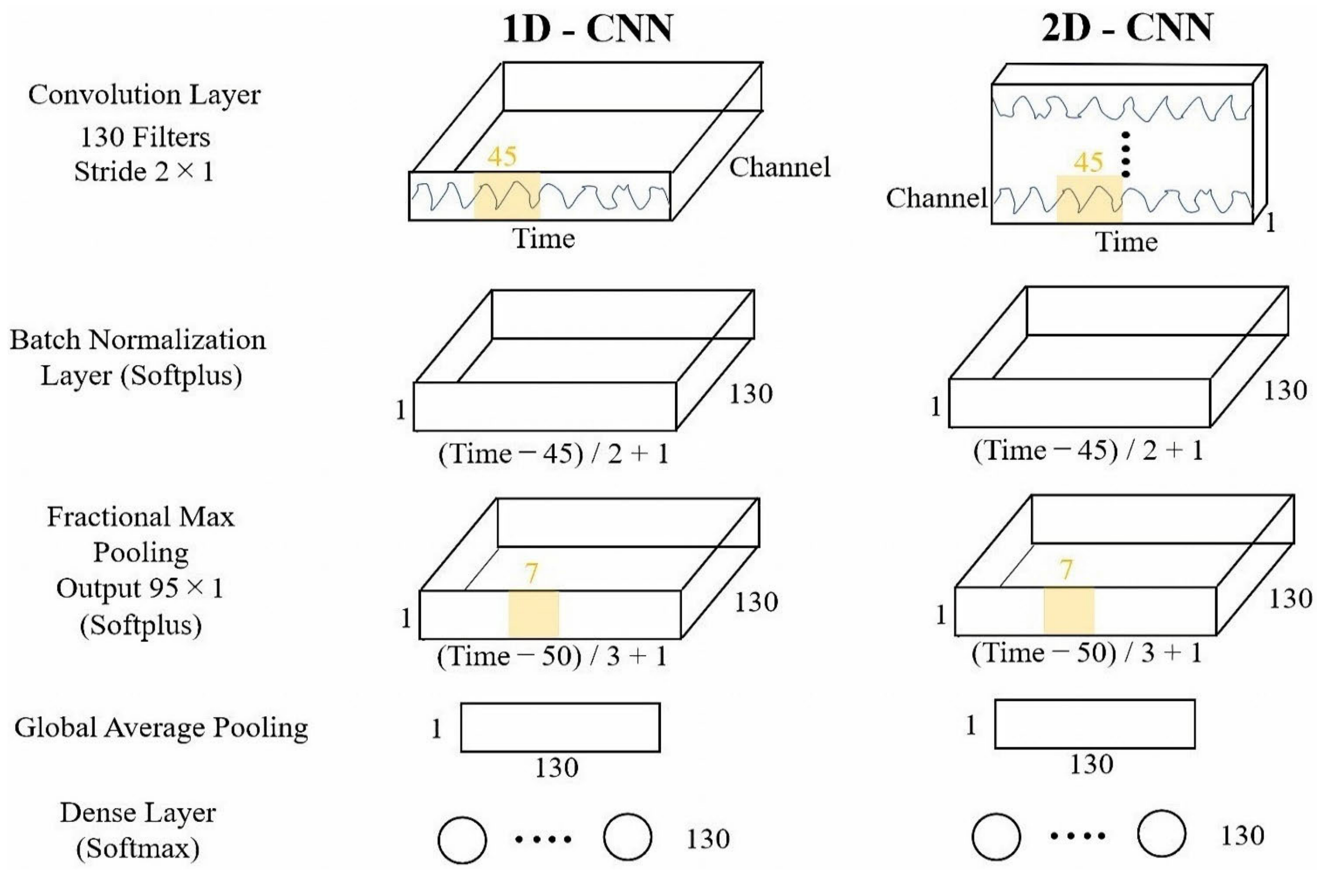


FIGURE 2. **Compositions of 1D-CNN and 2D-CNN used in our experiments. In 1D-CNN, the channel information was inputted as filter information in the convolution layer. In 2D-CNN, the measurement data was inputted as shown in the 2D image Time × Channel. The orange enclosure in the figure is kernel.**

version 1.7.0. Codes used in this study will be available at Code Ocean.

In the classifier using machine learning, features were extracted using CSP and FFT, and were classified using LDA and SVM (CSP-LDA, FFT-LDA, CSP-SVM, and FFT-SVM). In BCI Competition IV 2a, because there are four types of motor imagery, we used the one-versus-one (OVO) decomposition shcheme [62] to make it a two-class classification problem; six classifiers were created. The number of features was set to four in the CSP. In SVM, the regularization parameters and hyperparameters of Gaussian kernel were set to 1.0 and the inverse of the number of feature values, respectively. In the classifier using deep learning, we trained the NNs, 1D-CNN and 2D-CNN. Cross-entropy was used for the loss function and Adam [63] was used for the optimization function.

In BCI Competition IV 2a, in the CSP-LDA, FFT-LDA, CSP-SVM, and FFT-SVM, the evaluation method used data from the second day as test data, and the data from the first day were used as training data. In addition, in NN and CNNs, the first-day data were divided into training and validation data by five-fold-cross-validation. Validation data were used for early termination. In BCI Competition IV 2b, the first three sessions in CSP-LDA, FFT-LDA, CSP-SVM, and FFT-SVM were used as the training data, and the fourth and fifth sessions were used as the test data. In addition, in NN and CNNs, the first three sessions were divided into training and validation data using five-fold-cross-validation. The fourth and fifth sessions were used as test data.

In BCI Competition IV 2a and 2b, we set the baseline of accuracy using a time window with a window width of 3 s during motor imagery (3–6 s, 4–7 s) as the baseline in Fig. 1 (a) and (b), respectively. All classification accuracies calculated for all time windows differed from the baseline to compare the classification accuracy during motor imagery. In addition, due to the large number of combinations when verified, CSP-LDA, 1DCNN, and 2DCNN were used for detailed verification.

We classified BCI Competition IV 2a using feature extraction and machine learning. First, we calculated the classification accuracy of each participant's left- and right-hand motor imagery during a sliding window of 0.1 s in the range of 2 s to 6.5 s with a window width of 3 s using CSP-LDA. Second, we calculated the average classification accuracy across participants for left- and right-hand motor imagery in each time window with window widths of 2, 2.5, 3, and 3.5 s using

CSP-LDA. Third, OVO was used to calculate the average classification accuracy across participants for all classes in the 3 s window using CSP-LDA. For FFT-LDA, CSP-SVM, and FFT-SVM, we classified right- and left-handed motor imagery with a window with a width of 3 s, and calculated the average classification accuracy of participants. In deep learning, we first calculated the classification accuracy for each participant in various time windows, and the average classification accuracy across participants in various time windows with various window widths in the 1D-CNN when the window width was set to 3 s. Second, 2D-CNN, as in the 1D-CNN, we calculated the classification accuracy for each participant in fixed window widths and the average classification accuracy across participants in time windows with window widths. For CSP-NN and FFT-NN, the average classification accuracy of the participants was calculated with a window width of 3 s.

Next, we classified BCI Competition IV 2b. First, in the CSP-LDA, we calculated the classification accuracy of each participant during a sliding window of 0.1 s in the range 3–7.5 s with a window width of 3 s and calculated the average classification accuracy across participants during each time window with window widths of 2, 2.5, 3, and 3.5 s. Second, in the 1D-CNN as in the CSP-LDA, we calculated the classification accuracy in fixed window widths for each participant and the average classification accuracy across participants in time windows with window widths. Third, we calculated the classification accuracy of the 2D-CNN in the same manner as the 1D-CNN. For FFT-LDA, CSP-SVM, FFT-SVM, CSP-NN, and FFT-NN, the average classification accuracy of participants was calculated with the window width of 3 s.

IV. RESULTS

We extracted signals from different time segments of two BCI Competition IV 2a, 2b. We trained the classifiers using machine learning and deep learning to calculate accuracy. For more detailed results, see Supplemental Materials.

A. BCI COMPETITION IV 2A

The classification of the left- and right-hand motor images was performed in various time windows using LDA after feature extraction by CSP, and the classification accuracy of CSP-LDA was calculated when the window width was set to 3 s. Fig. 3 shows the results of the difference in classification accuracy between the sliding window (3 s) and baseline (4–7 s) in Fig. 1 (a).

First, from Fig. 3 (a), accuracies were higher when six participants used the time data at visual presentation in the range 2.1–2.3 s for training. In particular, accuracies at 2.4 s in participants 5, 7, and 9 were 10 % higher than during the motor imagery time (3 s). In contrast, for participant 4, the accuracy of the training data during visual presentation at 2.4 s was 12.07 % lower than during motor imagery time (3 s). In addition, the accuracy using the training data, including the time at rest decreased, with the exception of participant 5 (3.1–3.5 s).

Second, the results of the classification of motor imagery in Fig. 3(b) for various time windows show that the classification accuracy using the training data at the time of visual presentation (2–2.8 s) is higher than that at baseline (3–6 s), and the accuracy decreases as the time segment in the latter half of the task is used as training data.

Third, we created six different classifiers in the OVO for the four motor imagery classifications. Fig. 3(c), which shows the average classification accuracy for participants in the 3 s time window, indicates that the classification accuracy was better for the 2.5–2.8 s time window than for the motor imagery time window. In particular, the right vs. foot outperformed the results during motor imagery for all visual presentation times, with a maximum improvement of 5.56 %. It was also found that the accuracy decreased with time after motor imagery was used.

Now for the 1D-CNN and 2D-CNN with deep learning, Fig. 4 shows the results of the difference in classification accuracy between the sliding windows (3 s) and baseline (4–7 s) in Fig. 1 (a).

First, for the 1D-CNN, Fig. 4(a) shows the results of the difference between the classification accuracy of the sliding the time window, which was set to 3 s, and the classification accuracy during motor imagery (3–6 s). Fig. 4(a) shows that, with the exception of participant 2, the classifiers that used data during visual presentation for training had the highest accuracy. Furthermore, for most participants, the accuracy gradually increased between 2–2.5 s and gradually decreased between 2.5–3 s, as shown in Fig. 4(a). After 3 s, accuracy decreased. Participant 2, however, showed a steady increase in accuracy. Fig. 4(b), which shows the average classification accuracy of participants when classifying motor imagery in various time windows, indicates that, as with the CSP-LDA, the accuracy is higher for a wider time window. Furthermore, the classification accuracy was the highest when using the training data at the time of visual presentation, and the accuracy decreased as the temporal information in the latter half of the task became the training data.

Second, for the 2D-CNN, Fig. 4 (c) shows that, as with the 1D-CNN in Fig. 4 (a), with the exception of participant 2, the classifier that used the time data at visual presentation for training had the highest accuracy. Initially, during visual presentation, most participants showed a gradual increase in accuracy between 2–2.4 s and a gradual decrease in accuracy between 2.4–3 s. After 3 s, accuracy decreased. Figure 4(d), which shows the average classification accuracy of participants when classifying motor imagery in various time windows, shows that, as with the CSP-LDA, the training data at visual presentation had the highest accuracy in each window width. We found that as the time window slides from 2.5 s to the end of the slide range, the classification accuracy decreased.

Finally, Fig. 5 shows the difference between the baseline and the average classification accuracy across participants for all combinations in 0.1s increments from 2–6.5 s at a window width of 3 s. The classifier using machine learning

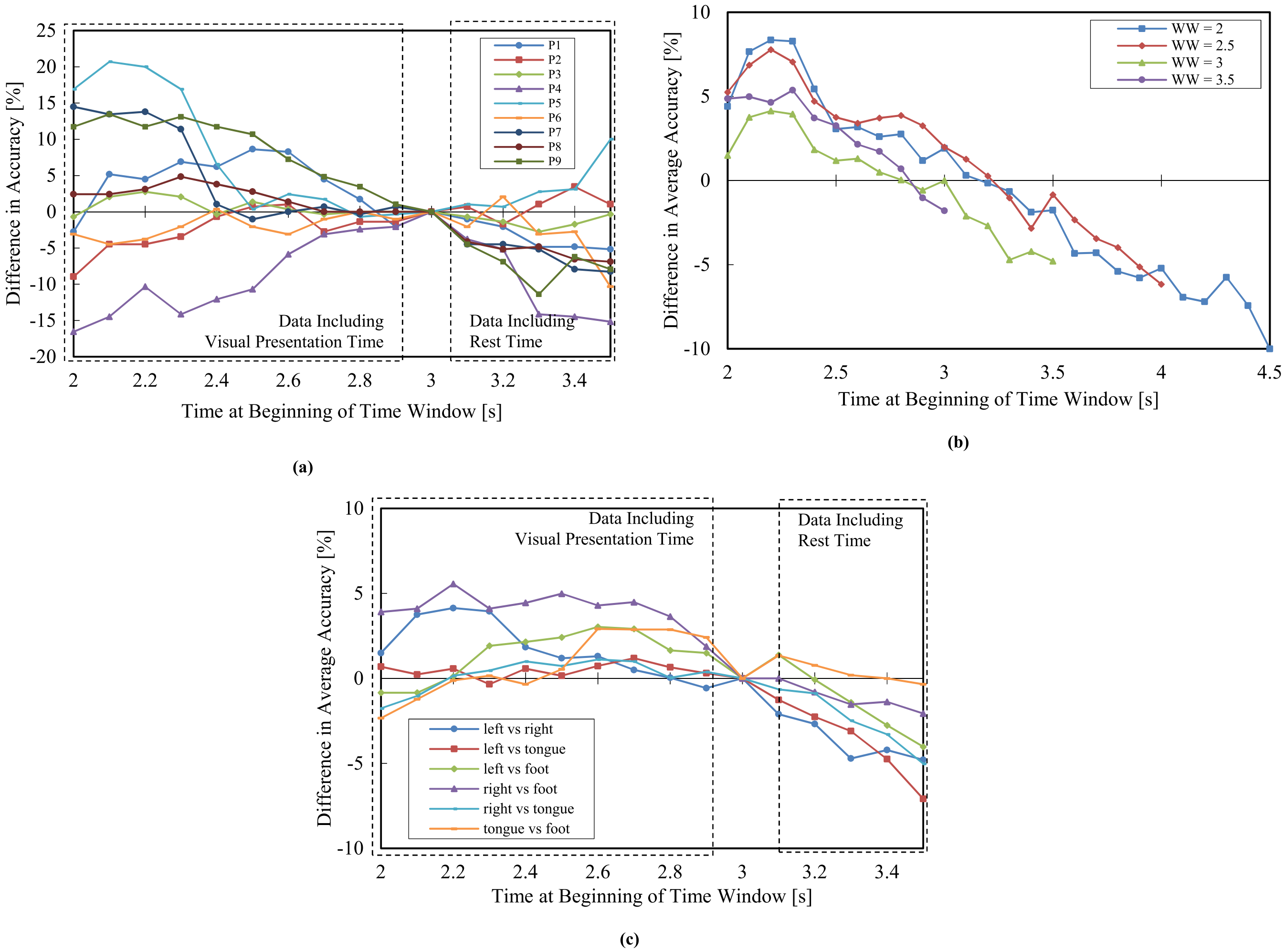


FIGURE 3. The differences in classification accuracy of CSP-LDA with BCI Competition IV 2a. The difference is based on the classification accuracy of the time data during motor imagery (3–6 s). Positive values indicate improved accuracy. (a) The differences in classification accuracy with each participant (P1-P9) when extracting time data of motor imagery of left-hand and right from 2–5 s to 3.5– 6.5 s in 0.1 s increments with a window width of 3 s. (b) The difference in average accuracy across nine participants when extracting time data of motor imagery of left- and right-hand in 0.1 s increments with various window widths (WW). (c) The difference in average accuracy across nine participants using OVO when extracting time data of four motor imageries from 2–5 s to 3.5–6.5 s in 0.1 s increments with a window width of 3 s.

was a two-class classification during right- and left-hand motor imagery.

Fig. 5 shows that accuracy improves with visually presented information for all verified models, and that data, including rest time, show a decrease in accuracy.

These results indicate that machine learning and NN-based classification is more accurate during visual presentation than during imagery, and less accurate for data that includes rest time

B. BCI COMPETITION IV 2B

In BCI Competition IV 2b, we used CSP-LDA, 1D-CNN, and 2D-CNN to classify motor imagery in various time windows. Fig. 6 (a), (c), and (e) show the difference in classification accuracy between the sliding window (3 s) and baseline (4–7 s) for each model. The differences in participants' average classification accuracies for various window widths and baselines (4–7 s) are also shown in Fig. 6 (b), (d), and (f).

First, for the CSP-LDA, the classification accuracies were calculated for various time window widths. Fig. 6 (a), which shows the difference between the classification accuracy at a window width of 3 s and that at motor imagery (4–7 s) in various time windows, shows that results for seven participants were more accurate when the signals at visual presentation were used for training. Participant 3 had more than 7 % higher accuracy. However, for participants 6 and 9, accuracies were more than 5 % lower when using the training data during visual presentation than during motor imagery. Next, the classification of motor imagery was trained usin data from various time segments with various time window widths. Fig. 6 (b) indicates that the narrower the time window, the higher the accuracy. In addition, the classification accuracy was highest when using the training data at the time of visual presentation and decreased as the time information in the latter half of the task was used as the training data.

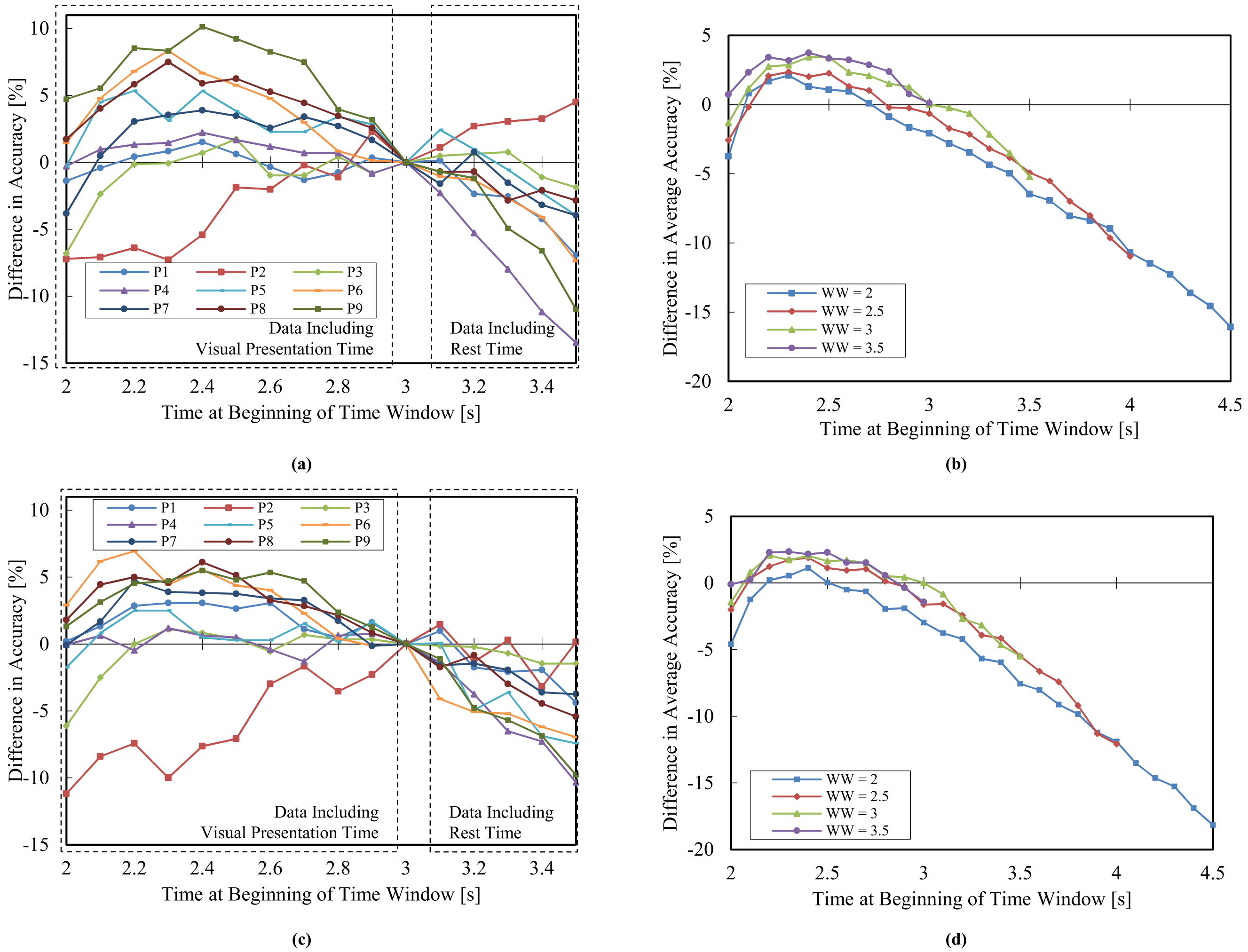


FIGURE 4. The difference in classification accuracy of our 1D-CNN and 2D-CNN with BCI Competition IV 2a. The difference is based on the classification accuracy of the time data during motor imagery (3–6 s). Positive values indicate improved accuracy; (a), (b) were used in 1D-CNN and (c), (d) were used in 2D-CNN. (a), (c) The differences in classification accuracy of 1D-CNN with each participant (P1-P9) when extracting time data from 2–5 s to 3.5–6.5 s in 0.1 s increments with a window width of 3 s. (b), (d) The difference in average accuracy across nine participants when extracting time data in 0.1 s increments with various window widths (WW).

Second, for the 1D-CNN, we calculated the classification accuracy in various time windows. Fig. 6 (c) shows that seven participants were more accurate when the signals at visual presentation (3.3–3.5 s) were used for training. In particular, the accuracy of participant 3 improved by 16.26 % (3.5 s). When the classification of motor imagery was trained on data from various time windows with various window widths, the accuracy also increased gradually in the interval from 3–3.4 s for all window widths, as shown in Fig. 6 (d). the accuracy then gradually decreased in the interval 3.4–5.5 s.

Third, for the 2D-CNN, the classification accuracy was calculated for various time window widths. As with the 1D-CNN, results were more accurate when the signals at visual presentation (3.3–3.5 s) were used for training. In particular, the accuracy of participant 3 improved by 17.13% (3.5 s). When the classification of motor imagery was trained on data from various time segments with various time window widths, the results were similar to those of the 1D-CNN.

Finally, Fig. 7 shows the difference between the baseline and the average classification accuracy across participants for all combinations considered in 0.1 s increments from 3–7.5 s at a window width of 3 s. Fig. 7 shows that accuracy improves with visually presented information for most of the models, and that data, including rest time, decreases in accuracy.

Therefore, we demonstrated that with traditional machine learning and deep learning, the accuracy improves when visual presentation information is incorporated into the training data.

V. DISCUSSION

We investigated the performance of the machine learning and deep learning classifiers on two datasets by varying the time window and window width used for training. The results

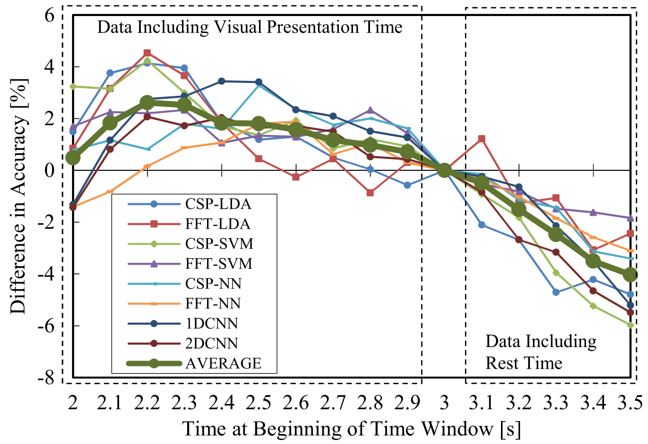


FIGURE 5. The difference in average accuracy across nine participants of with BCI Competition IV 2a when extracting time data from 2–5 s to 3.5–6.5 s in 0.1 s increments with a window width of 3 s. The difference is based on the classification accuracy of the time data during motor imagery (3–6 s). Positive values indicate improved accuracy.

showed that in BCI Competition IV 2a, both machine learning and deep learning models exhibited higher classification accuracy when the visual presentation time was included in the training data than when only motor imagery was included in the training data.

A. RESULT OF BCI COMPETITION IV 2A AND 2B

In BCI Competition IV 2a, classification accuracy was improved using training data that included visual presentation time (2–3 s). This may be due to brain activity that discriminates the presented visual information. Multiple studies have shown that visual information, such as images, can be discriminated from brain activity. It is important to note that the images are presented for a short period of time, less than 1 s [31], [32], [33], [34], [35], [39], which is similar to the visual presentation time of BCI Competition IV 2a. This suggests that the information contributing to the classification appears within 1 s. This finding supports our hypotheses.

In BCI Competition IV 2b, CSP-LDA showed similar results to BCI Competition IV 2a; Leeb et al. also used LDA to perform classification without feedback in multiple time window widths of 1, 1.5, 2, 2.5, and 3 s [52]. The results showed that, similar to our findings, the 2 s window width improved accuracy the most. However, although they did not specify which time segment they explored, the high performing time interval was 4.84–6.84 s on average for the participants, which differed from our results. While we used 0.5–30 Hz bandwidth information, Leeb et al. used the optimal bandwidth from 8–30 Hz range, with bandwidths of 2, 4, 6, and 8 Hz for a total of 72 bandwidths. Their bandwidth was much narrower than that in our study; furthermore, they did not use bandwidth information in the 0.5–8 Hz range. Therefore, it can be inferred that only the frequency band relevant for motor imagery was used, which would have improved performance at 4.84–6.84 s. Our results showed that performance improved when lower frequencies were included. Conversely, the results of this study in deep learning showed that classification accuracy was improved by using training data that included visual presentation time (3–4 s) as in CSP-LDA. In BCI Competition IV 2b, there are only three channels (C3, Cz, and C4), but information from the visual cortex may propagate, albeit attenuated, and contribute to classification.

In both datasets, accuracy was improved by training data that included visual presentation information. Here, Fig. 5 and 7 show that the average accuracy arоss participant is only a few percentage higher than that of the training data. However, this small difference is important for a fair comparison between the models.

A well-known competition in the image field is the ImageNet Large Scale Visual Recognition Challenge, which was held annually from 2010 to 2017. Recently, ImageNet has been used to evaluate models. In 2020, Noisy Student Training [64] and Big Transfer [65] used top-1 accuracy on ImageNet for evaluation, resulting in a 2.4% and 1.4% improvement in accuracy, respectively, compared to the previous state-of-the-art of 86.4% [66]. In BCI Competition IV 2a, Liu et al. showed a 2.79% improvement in accuracy over the previous method [19], and in BCI Competition IV 2b, Roy et al. claimed a 2.49% improvement in accuracy [11].

A comparison of these previous studies shows that a few percent difference between accuracies of models determines the superiority of the model. The small differences have a technical impact on the peripheral fields. Therefore, the results of this study are important for the realization of fair competition.

B. TRIAL AVERAGES OF TEMPORAL PROFILES AND SPATIAL TOPOGRAPHY OF BCI COMPETITION IV 2A AND 2B

We averaged each channel and each class in BCI Competition IV 2a and BCI Competition IV 2b to increase the validity of our hypothesis that there is information in visual information that contributes to classification and created the spatial topography, as shown in Fig. 8 and 9. In BCI Competition IV 2a, we used data from participant 9, whose accuracy improved in visual presentation time. In BCI Competition IV 2b, participant 3 whose accuracy improved in the visual presentation time. The results of the analyses, with and without feedback, are also presented.

In BCI Competition IV 2a, Fig. 8 shows that both right- and left-handed tasks show the largest amplitude at the time of visual presentation. Furthermore, the spatial topography at 2.25 s indicates that brain activity near the visual cortexwas active. Therefore, it is more likely that the classification model learns the visually presented image rather than the brain activity of motor imagery.

In BCI Competition IV 2b, we discuss left-handed motor imagery with participant 3, whose accuracy was improved by the use of visual presentation information. The difference between C3 and C4 was larger with feedback (Fig. 9 (b))

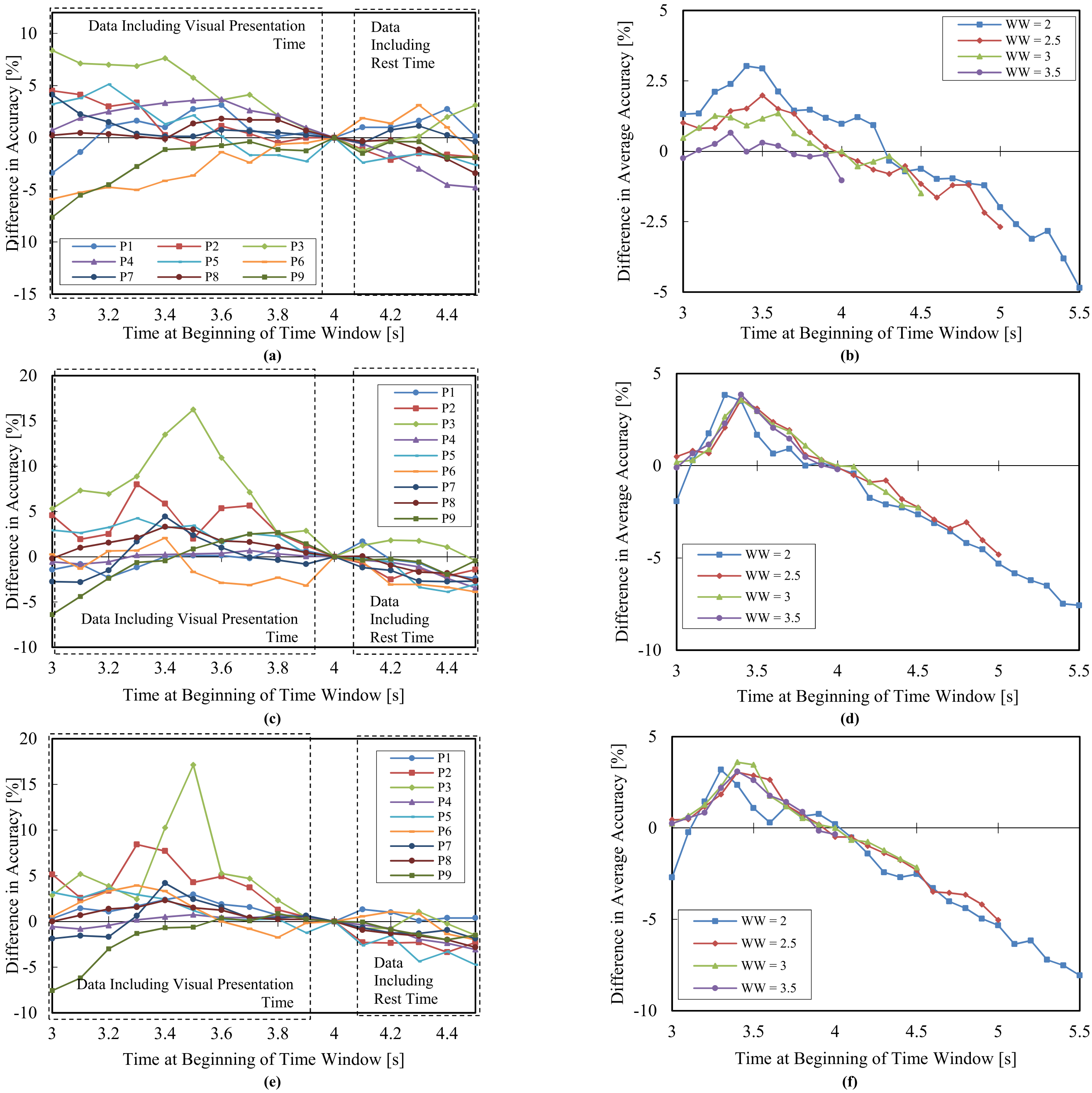


FIGURE 6. The difference in classification accuracy of CSP-LDA, 1D-CNN, and 2D-CNN with BCI Competition IV 2b. The difference is based on the classification accuracy of the time data during motor imagery (4–7 s). Positive values indicate improved accuracy. (a), (b) were used CSP-LDA, (c), (d) were used 1D-CNN, and (e), (f) were used 2D-CNN. (a), (c), and (e) The differences in classification accuracy of 1D-CNN with each participant (P1-P9) when extracting time data from 3–6 s to 4.5–7.5 s in 0.1 s increments with a window width of 3 s. (b), (d), (f) The difference in average accuracy across nine participants when extracting time data in 0.1 s increments with various window widths (WW).

than without feedback (Fig. 9 (a)) at the visual presentation time (3.5–4 s), suggesting that accuracy was improved at the visual presentation time. Furthermore, in the spatial topography at 3.5 s, the distribution was symmetrical for left- and right-handed motor imagery (Fig. 9 (c) and (d)), indicating that the features contributing to the classification were represented.

C. ANALYSIS OF VISUAL PRESENTATION TIME AND REST TIME OF BCI COMPETITION IV 2A AND 2B

We showed that visual information processing may contribute to classification. Therefore, we also examined whether training the model with only visual representations, motor images, and rest time intervals contributes to classification performance.

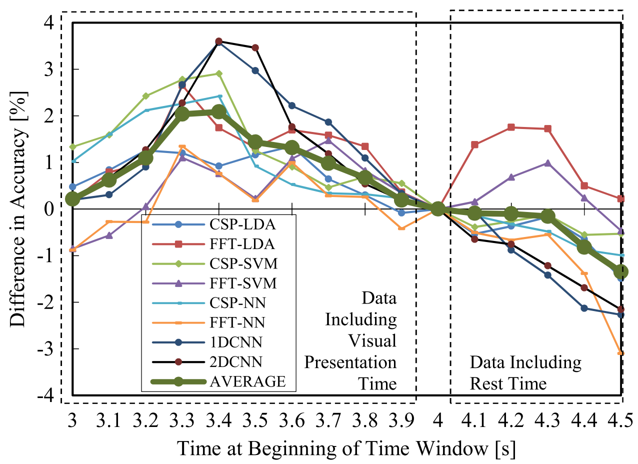


FIGURE 7. The difference in average accuracy across nine participants with BCI Competition IV 2b when extracting time data from 3–6 s to 4.5–7.5 s in 0.1 s increments with a window width of 3 s. The difference is based on the classification accuracy of the time data during motor imagery (4–7 s). Positive values indicate improved accuracy.

First, we evaluated the statistical significance compared to the baseline. In BCI Competition IV 2a, we set the null hypothesis that the hypothesized mean of the classification accuracy at baseline (3–6 s) and the hypothesized mean of the classification accuracy that was most improved using a window width of 3 s during the visual presentation were equal. In BCI Competition IV 2b, we set the null hypothesis that the hypothesized mean of the classification accuracy at baseline (4–7 s) and the hypothesized mean of the classification accuracy that was most improved using a window width of 3 s during visual presentation was equal. We performed a t-test to determine whether the null hypothesis could be rejected ($p < 0.05$).

TABLE 1. Results of t-test between the most accurate time window and baseline at 3 s time window width.

Dataset	Method	Time Window [s]	p value
BCI Competition IV 2a	CSP-LDA	2.7–5.7	0.001
	1D-CNN	2.4–5.4	0.048
	2D-CNN	2.4–5.4	0.189
BCI Competition IV 2b	CSP-LDA	3.6–6.6	0.062
	1D-CNN	3.4–6.4	0.037
	2D-CNN	3.4–6.4	0.014

As shown in Table 1, in BCI Competition IV 2a, significant differences were found in CSP-LDA and 1D-CNN. Significant differences were also observed. In BCI Competition IV 2b, significant differences were found in 1D-CNN and 2D-CNN.

Second, in BCI Competition IV 2a and BCI Competition IV 2b, respectively, we used EEG data with a short window width (1 s) during visual presentation time (2–3 s, 3–4 s), motor imagery (3–4 s, 4–5 s), and rest time (6–7 s, 7–8 s) for training LDA, 1D-CNN, and 2D-CNN and calculated participants' average classification accuracy. As shown in Fig. 10 and 11, all time segments, except for the BCI Competition IV 2b rest time (7–8 s), show performance that exceeds the chance level (50 %) by approximately 10 % or more. Therefore, we is observed that there is information contributing to classification in each time segment. For specific values, see Supplemental Materials.

D. EVALUATION OF DATA SIMILAR TO PRACTICAL ENVIRONMENTS BY LEARNING FROM EXPERIMENTAL ENVIRONMENTAL DATA

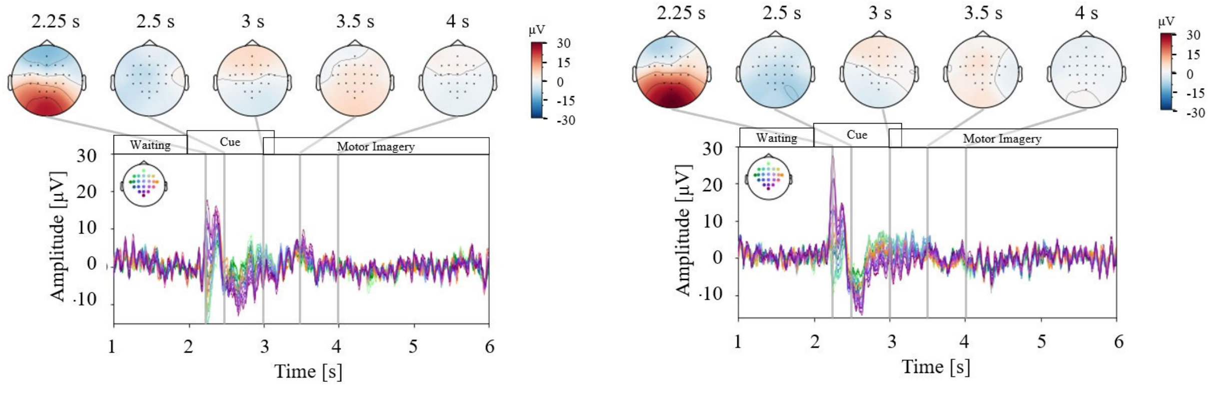
Fig. 10, 11, and Table 1 indicate that visual presentation information affects classifier performance. These findings suggest that using time segments that are not intervals of motor imagery can make it challenging to compare simple models. Furthermore, the visual presentation information is combined with class information that is mixed with the experiment designer's intention to direct the task content. Thus, the model may be overly fitted to the experimental environment. However, learning using data other than motor imagery, such as visual presentation information, mixed with the intention of the experiment designer may result in a more robust model, potentially allowing for the classification of data from practical environments with higher accuracy.

Therefore, we trained the classifier on data containing visual presentation information. We investigated whether the classifier could better recognize data from motor imagery segments, which are time segments which differ from training and are considered similar to the practical environment. In BCI Competition IV 2b, feedback information is given to the test data by the integrated classification accuracy of the LDA, which is trained on data 2 s prior to the time of feedback presentation [52]. In short, the feedback information is influenced by the data 2 s prior to that time schedule. Because of this, simple comparisons between training and test data at different time points are difficult to make. Therefore, we used only BCI Competition IV 2a. We assessed a window with a width of 3 s from 2–5 s to 2.9–5.9 s in 0.1 s increments for all classifiers tested in this study. Using the model trained with training data containing the visual presentation, we classified data in the motor imagery segment (3–6 s), which is data similar to the practical environment.

The results of data classification in motor imagery segment, as shown in Fig. 12, revealed that including with visual presentation information in training data led to higher accuracy (max 3.76%) when the visual presentation was not included in the CNN. However, combinations in which the classifier was adapted after feature extraction were more accurate (max 2.24%) when visual presentation was included. For more detailed results, see Supplemental Materials.

Since a convolutional layer is like a band-pass filter, because it captures periodic time change [17], [67], including time sequences, other than motor imagery, it would have reduced accuracy.

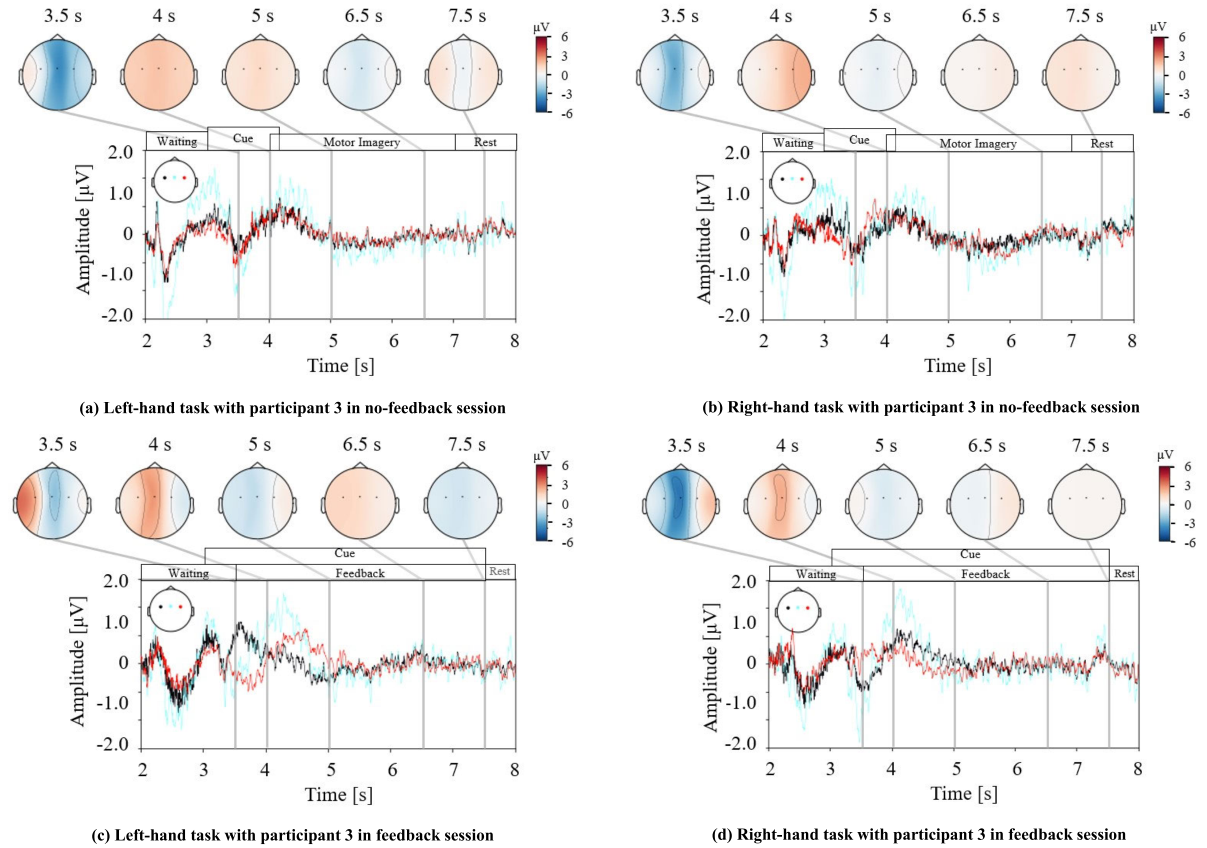
Alternatively, in the other combinations, the model does not take temporal connections into account, and it is thought that learning features are localized. However, among the other combinations, there are those for which the accuracy of the data is better for only the time of motor imagery. These two



(a)

(b)

FIGURE 8. Trial averages of temporal profiles and spatial topography with (a) left- and (b) right-hand tasks for participant 9 in the BCI Competition IV 2a. The signals with 22 colors in the figure show their respective channels corresponding to the upper left-brain map. The block above the waveform is the time schedule for the trial.



(a) Left-hand task with participant 3 in no-feedback session

(b) Right-hand task with participant 3 in no-feedback session

(c) Left-hand task with participant 3 in feedback session

(d) Right-hand task with participant 3 in feedback session

FIGURE 9. Trial averages of temporal profiles and spatial topography during right and left-handed tasks for participants 3 in the BCI data. (a), (b), (c), and (d) are results for participant 3. (a), (c) are during the left-hand task; (b) and (d) are during the right-hand task. (a) and (b) are sessions without feedback; (c) and (d) are sessions with feedback. The signals with three colors in the figure show their respective channels corresponding to the upper left-brain map. The block above the waveform is the time schedule for the trial.

results do not imply that the model becomes more robust when time segments other than motor imagery are included (i.e., when other task time is included). Instead, the results indicate that CNNs, which have been used by researchers in recent years and are adept at capturing temporal variation, exhibit lower accuracy when other time segments are included.

Fig.8-12 strengthen the evidence that visual presentation time and rest time have a significant impact on the performance of the model. Therefore, improving the validity of simplifying with performance comparisons between models and the validity using motor imagery segment, which is similar to the practical environment. In addition, the classification performance trained with the inclusion of these temporal data may show excessive results (BCI Competition IV 2a: [9], [11], [13], [15], [17], [19], [21], [28], [29], [30], BCI Competition IV 2b: [17], [18], [21], [28], [30]). It was also suggested that any classifier using a CNN could be less accurate when introduced to a practical environment.

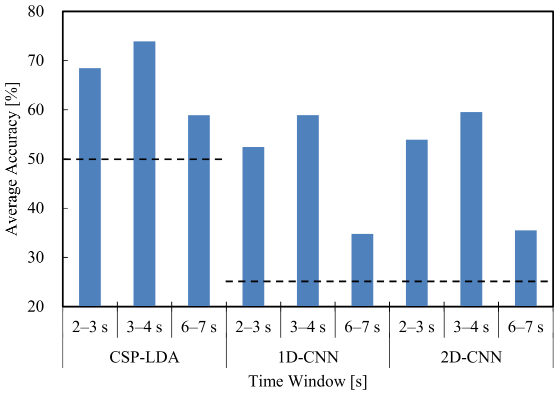


FIGURE 10. Comparison of classification average accuracy of CSP-LDA, 1D-CNN, 2D-CNN when using visual presentation time (2–3 s), motor imagery (3 –4 s), and rest time (6–7 s) as training data in BCI Competition IV 2a. Dashed lines represent chance levels.

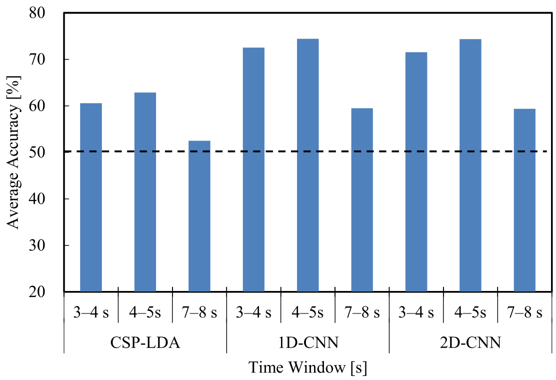


FIGURE 11. Comparison of classification average accuracy of CSP-LDA, 1D-CNN, 2D-CNN when using visual presentation time (3–4 s), motor imagery (4–5 s), and rest time (7–8 s) as training data in BCI Competition IV 2b. The dashed line represents chance levels.

However, in recent years, various classification models have been proposed, and not all of them fit our results. In BCI Competition IV 2a, Li et al. prepared training data for multiple time windows, including visual presentation time, imagery time, and rest time, of 0–4 s and 0.25–4.25 s after visual presentation, and trained them. Although not compared to training data only during motor imagery, the training data in the 2.75–6.75 s segment showed the greatest improvement in accuracy when the window width was 3 s. When the window width was 4 s, the highest accuracy was observed in the training data in the 3–7 s segment including rest periods, and unlike our results, the accuracy was better when rest time was included compared to when visual presentation was used.

However, in any model, the choice of time segment has a significant impact on the model. Therefore, we propose that the time segment used for training data are unified for purposes such as facilitating performance comparisons between models in competition data; this is applicable for the BCI competition IV 2a and 2b, but not intended to be generalizable to other datasets. However, learning information from other datasets that may not be relevant to motor imagery in the same way may cause problems, such as difficulties in comparing models and models that are too fitted to the experimental

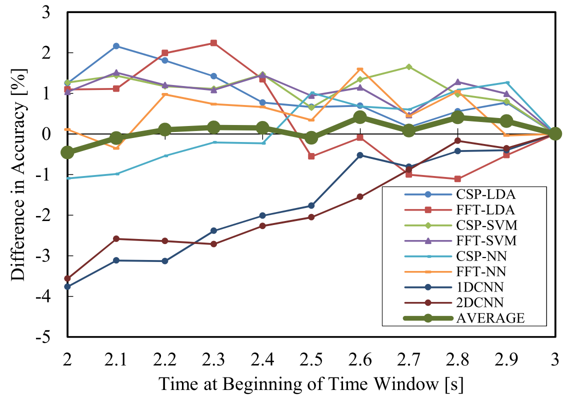


FIGURE 12. Evaluation of motor imagery data (3–6 s) in data trained with visual presentation information in BCI Competition IV 2a. In training data of visual presentation information, we used a window with a width of 3 s from 2–5 s to 2.9–5.9 s in 0.1 s increments.

environment. Thus, it is essential to exercise caution when referring to the time schedule in the dataset, considering gaps with the practical environment, and appropriately setting the segments used for training.

VI. CONCLUSION

We demonstrated that varying time segments for training data may significantly affect the model's performance in BCI Competition IV 2a and 2b. Using time segments other than motor imagery creates a challenge, as models are too well fitted to the experimental environment, making comparisons between models difficult. Therefore, we recommend that temporal information other than that during motor imagery cannot be used as training data. For models using machine learning and deep learning, for BCI Competition IV 2a and 2b, we indicated the possibility that visual presentation time improves classification performance. It was suggested that the models may be learning more visual information. In fact, activation of the visual cortex was observed from spatial topography during visual presentation. Related studies also support the hypothesis that visual information can be categorized. Moreover, training data that included areas other than motor imagery did not improve the classification accuracy of data that was similar to the practical environment. Summarily, the model proposed in recent years may exhibit an excessive classification performance. In other data sets, care may be required to refer to the time schedule and consider the gap between the dataset and the practical environment. Time segments should be chosen with caution to ensure fair competition and evaluation.

ACKNOWLEDGMENT

The authors would like to thank Editage (www.editage.com) for English language editing. KAKENHI are competitive research funds that are intended to significantly develop all scientific research (research based on the free ideas of the researcher), from basic to applied research in all fields. For more information about KAKENHI, visit (https://www.jsps.go.jp/english/e-grants/).

REFERENCES

[1] C. I. Penaloza and S. Nishio, "BMI control of a third arm for multitasking," *Sci. Robot.*, vol. 3, no. 20, Jul. 2018, Art. no. eaat1228, doi: 10.1126/scirobotics.aat1228.

[2] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Sci. Rep.*, vol. 6, no. 1, Dec. 2016, doi: 10.1038/srep38565.

[3] M. Sebastián-Romagosa, W. Cho, R. Ortner, N. Murovec, T. Von Oertzen, K. Kamada, B. Z. Allison, and C. Guger, "Brain computer interface treatment for motor rehabilitation of upper extremity of stroke patients—A feasibility study," *Frontiers Neurosci.*, vol. 14, Oct. 2020, Art. no. 591435, doi: 10.3389/fnins.2020.591435.

[4] L. Cao, B. Xia, O. Maysam, J. Li, H. Xie, and N. Birbaumer, "A synchronous motor imagery based neural physiological paradigm for brain computer interface speller," *Frontiers Hum. Neurosci.*, vol. 11, p. 274, May 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2017.00274

[5] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. D. R. Millán, "A brain-actuated wheelchair: Asynchronous and non-invasive brain–computer interfaces for continuous control of robots," *Clin. Neurophysiol.*, vol. 119, no. 9, pp. 2159–2169, Sep. 2008, doi: 10.1016/j.clinph.2008.06.001.

[6] J. Jeong, K. Shim, D. Kim, and S. Lee, "Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1226–1238, May 2020, doi: 10.1109/TNSRE.2020.2981659.

[7] P. Ofner, A. Schwarz, J. Pereira, and G. R. Muller-Putz, "Upper limb movements can be decoded from the time-domain of low-frequency EEG," *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0182578, doi: 10.1371/journal.pone.0182578.

[8] S. Zolfaghari, T. Y. Rezaii, S. Meshgini, A. Farzamnia, and L. C. Fan, "Speed classification of upper limb movements through EEG signal for BCI application," *IEEE Access*, vol. 9, pp. 114564–114573, 2021, doi: 10.1109/ACCESS.2021.3102183.

[9] J. Jin, H. Sun, I. Daly, S. Li, C. Liu, X. Wang, and A. Cichocki, "A novel classification framework using the graph representations of electroencephalogram for motor imagery based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 20–29, 2022, doi: 10.1109/TNSRE.2021.3139095.

[10] M. Li, J. Han, and L. Duan, "A novel MI-EEG imaging with the location information of electrodes," *IEEE Access*, vol. 8, pp. 3197–3211, 2020, doi: 10.1109/ACCESS.2019.2962740.

[11] A. M. Roy, "Adaptive transfer learning-based multiscale feature fused deep convolutional neural network for EEG MI multiclassification in brain-computer interface," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105347, doi: 10.1016/j.engappai.2022.105347.

[12] A. M. Roy, "An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces," *Biomed. Signal Process. Control*, vol. 74, Apr. 2022, Art. no. 103496, doi: 10.1016/j.bspc.2022.103496.

[13] M.-A. Li, J.-F. Han, and J.-F. Yang, "Automatic feature extraction and fusion recognition of motor imagery EEG using multilevel multi-scale CNN," *Med. Biol. Eng. Comput.*, vol. 59, no. 10, pp. 2037–2050, Oct. 2021, doi: 10.1007/s11517-021-02396-w.

[14] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *J. Neurosci. Methods*, vol. 346, Dec. 2020, Art. no. 108885, doi: 10.1016/j.jneumeth.2020.108885.

[15] X. Ma, D. Wang, D. Liu, and J. Yang, "DWT and CNN based multi-class motor imagery electroencephalographic signal recognition," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016073, doi: 10.1088/1741-2552/ab6f15.

[16] C. Zhang, Y.-K. Kim, and A. Eskandarian, "EEG-inception: An accurate and robust end-to-end neural network for EEG-based motor imagery classification," *J. Neural Eng.*, vol. 18, no. 4, Mar. 2021, Art. no. 046014, doi: 10.1088/1741-2552/abed81.

[17] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016025, doi: 10.1088/1741-2552/ab405f.

[18] K. Zhang, G. Xu, L. Chen, P. Tian, C. Han, S. Zhang, and N. Duan, "Instance transfer subject-dependent strategy for motor imagery signal classification using deep convolutional neural networks," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–10, Aug. 2020, doi: 10.1155/2020/1683013.

[19] X. Liu, Y. Shen, J. Liu, J. Yang, P. Xiong, and F. Lin, "Parallel spatial–temporal self-attention CNN-based motor imagery classification for BCI," *Frontiers Neurosci.*, vol. 14, Dec. 2020, Art. no. 587520. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2020.587520

[20] P. Li and C. Liu, "Research on recognition of O-MI based on CNN combined with SST and LSTM," *Optoelectron. Lett.*, vol. 18, no. 9, pp. 566–571, Sep. 2022, doi: 10.1007/s11801-022-2054-1.

[21] J. Bang, M. Lee, S. Fazli, C. Guan, and S. Lee, "Spatio-spectral feature representation for motor imagery classification using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 3038–3049, Jul. 2022, doi: 10.1109/TNNLS.2020.3048385.

[22] M. Tangermann, K. R. Muller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, and G. Nolte, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jan. 2012. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2012.00055

[23] C. F. Vega, J. Quevedo, E. Escandón, M. Kiani, W. Ding, and J. Andreu-Perez, "Fuzzy temporal convolutional neural networks in P300-based brain–computer interface for smart home interaction," *Appl. Soft Comput.*, vol. 117, Mar. 2022, Art. no. 108359, doi: 10.1016/j.asoc.2021.108359.

[24] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning for motor imagery EEG-based classification: A review," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102172, doi: 10.1016/j.bspc.2020.102172.

[25] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Apr. 2019, Art. no. 031001, doi: 10.1088/1741-2552/ab0ab5.

[26] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, vol. 35, no. 20, pp. 14681–14722, Aug. 2021, doi: 10.1007/s00521-021-06352-5.

[27] M. Li and D. Xu, "Transfer learning in motor imagery brain computer interface: A review," *J. Shanghai Jiaotong Univ.*, Aug. 2022, doi: 10.1007/s12204-022-2488-4.

[28] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2012.00039

[29] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011, doi: 10.1109/TBME.2010.2082539.

[30] H. Fang, J. Jin, I. Daly, and X. Wang, "Feature extraction method based on filter banks and Riemannian tangent space in motor-imagery BCI," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2504–2514, Jun. 2022, doi: 10.1109/JBHI.2022.3146274.

[31] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, "Identifying natural images from human brain activity," *Nature*, vol. 452, no. 7185, pp. 352–355, Mar. 2008, doi: 10.1038/nature06713.

[32] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah, "Deep learning human mind for automated visual classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4503–4511, doi: 10.1109/CVPR.2017.479.

[33] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, "Ensemble deep learning for automated visual classification using EEG signals," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107147, doi: 10.1016/j.patcog.2019.107147.

[34] I. Simanova, M. van Gerven, R. Oostenveld, and P. Hagoort, "Identifying object categories from event-related EEG: Toward decoding of conceptual representations," *PLoS ONE*, vol. 5, no. 12, Dec. 2010, Art. no. e14465, doi: 10.1371/journal.pone.0014465.

[35] Y. Wang, P. Wang, and Y. Yu, "Decoding English alphabet letters using EEG phase information," *Frontiers Neurosci.*, vol. 12, p. 62, Feb. 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2018.00062

[36] H. H. Ehrsson, S. Geyer, and E. Naito, "Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations," *J. Neurophysiol.*, vol. 90, no. 5, pp. 3304–3316, Nov. 2003, doi: 10.1152/jn.01113.2002.

[37] I. Nambu, N. Hagura, S. Hirose, Y. Wada, M. Kawato, and E. Naito, "Decoding sequential finger movements from preparatory activity in higher-order motor regions: A functional magnetic resonance imaging multi-voxel pattern analysis," *Eur. J. Neurosci.*, vol. 42, no. 10, pp. 2851–2859, Nov. 2015, doi: 10.1111/ejn.13063.

[38] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature Neurosci.*, vol. 8, no. 5, pp. 679–685, May 2005, doi: 10.1038/nn1444.

[39] A. Harel, I. I. A. Groen, D. J. Kravitz, L. Y. Deouell, and C. I. Baker, "The temporal dynamics of scene processing: A multifaceted EEG investigation," *eNeuro*, vol. 3, no. 5, pp. 1–18, Sep. 2016, Art. no. e0139-16.2016, doi: 10.1523/ENEURO.0139-16.2016.

[40] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs," *Pattern Recognit.*, vol. 45, no. 6, pp. 2137–2144, Jun. 2012, doi: 10.1016/j.patcog.2011.04.018.

[41] A. Schlögl, C. Neuper, and G. Pfurtscheller, "Estimating the mutual information of an EEG-based brain-computer interface," *Biomedizinische Technik/Biomed. Eng.*, vol. 47, nos. 1–2, pp. 3–8, 2002, doi: 10.1515/bmte.2002.47.1-2.3.

[42] M. Mahmoudi and M. Shamsi, "Multi-class EEG classification of motor imagery signal by finding optimal time segments and features using SNR-based mutual information," *Australas. Phys. Eng. Sci. Med.*, vol. 41, no. 4, pp. 957–972, Dec. 2018, doi: 10.1007/s13246-018-0691-2.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015, pp. 1–14.

[45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[47] X. Jia, Y. Song, and L. Xie, "Excellent fine-tuning: From specific-subject classification to cross-task classification for motor imagery," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104051, doi: 10.1016/j.bspc.2022.104051.

[48] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, 2015, pp. 448–456.

[49] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014.

[50] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008-graz data set A," Graz Univ. Technol., 2008, vol. 16, pp. 1–6.

[51] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *J. Neural Eng.*, vol. 3, no. 3, pp. 208–216, Sep. 2006, doi: 10.1088/1741-2560/3/3/003.

[52] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain–computer communication: Motivation, aim, and impact of exploring a virtual apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007, doi: 10.1109/TNSRE.2007.906956.

[53] K. K. Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Hong Kong, Jun. 2008, pp. 2390–2397, doi: 10.1109/IJCNN.2008.4634130.

[54] C.-Y. Chiu, C.-Y. Chen, Y.-Y. Lin, S.-A. Chen, and C.-T. Lin, "Using a novel LDA-ensemble framework to classification of motor imagery tasks for brain-computer interface applications," in *Proc. Int. Comput. Symp. (ICS)*, vol. 1, 2014, pp. 136–142, doi: 10.3233/978-1-61499-484-8-150.

[55] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998, doi: 10.1023/A:1009715923555.

[56] M. Spuler, W. Rosenstiel, and M. Bogdan, "Principal component based covariate shift adaption to reduce non-stationarity in a MEG-based brain-computer interface," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 129, Jul. 2012, doi: 10.1186/1687-6180-2012-129.

[57] B. Graham, "Fractional max-pooling," 2014, *arXiv:1412.6071*.

[58] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–10. [Online]. Available: https://openreview.net/forum?id=ylE6yojDR5yqX

[59] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.

[60] H. Zheng, Z. Yang, W. Liu, J. Liang, and Y. Li, "Improving deep neural networks using softplus units," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–4, doi: 10.1109/IJCNN.2015.7280459.

[61] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, Feb. 2014, doi: 10.1016/j.neuroimage.2013.10.027.

[62] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, Aug. 2011, doi: 10.1016/j.patcog.2011.01.017.

[63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015, pp. 1–15.

[64] Q. Xie, M. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves ImageNet classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10684–10695, doi: 10.1109/CVPR42600.2020.01070.

[65] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby, "Big transfer (BiT): General visual representation learning," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 491–507, doi: 10.1007/978-3-030-58558-7_29.

[66] H. Touvron, A. Vedaldi, M. Douze, and H. Jegou, "Fixing the train-test resolution discrepancy," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–11. [Online]. Available: https://papers.nips.cc/paper_files/paper/2019/hash/d03a857a23b5285736c4d55e0bb067c8-Abstract.html

[67] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Jul. 2018, Art. no. 056013, doi: 10.1088/1741-2552/aace8c.

KENTO SUEMITSU was born in Kagoshima, Japan, in 1998. He received the Bachelor of Engineering degree from the Nagaoka University of Technology, in 2021, where he is currently pursuing the Ph.D. degree in engineering. His research interests include deep learning and brain–computer interfaces.

ISAO NAMBU (Member, IEEE) received the Ph.D. degree in engineering from the Nara Institute of Science and Technology, in 2010. From 2012 to 2017, he was an Assistant Professor with the Nagaoka University of Technology, where he has been an Associate Professor, since 2017. His research interests include neuroimaging, brain–computer interfaces, and human auditory processing.