2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Motor Imagery, Execution, and Observation Classification using Small Amount of EEG Data with Multiple Two-Class CNNs

Tomohiko Igasaki, Yugo Kuramura, and Junya Takemoto

Abstract—**This study attempted to classify a small amount of electroencephalogram (EEG) data on five states: four tasks involving right index-finger flexion (kinesthetic motor imagery, visual motor imagery, motor execution, and motor observation) and resting with eyes open. We employed a convolutional neural network (CNN) as a classifier and compared the classification accuracies of two types of CNNs: 1) a "single five-class CNN," which classified the aforementioned states with a single CNN and 2) "multiple two-class CNNs," wherein ten CNNs that classify pairs of states were combined. In addition, the classification accuracies were compared between two scenarios: one wherein the EEGs from all 19 scalp probe electrodes (19-channel EEG) were adopted as input data for the CNN, and the other wherein the EEGs of four regions closely related to the motor execution and observation of the index finger (4-channel EEG) were adopted. The classification accuracies of the single five-class CNN with 19- and 4-channel EEGs were 48.2 ± 5.9% and 46.6 ± 6.9%, respectively, and those of the multiple two-class CNNs with 19- and 4-channel EEGs were 52.8 ± 9.7% and 47.5 ± 9.4%, respectively. These results indicate the effectiveness of multiple two-class CNNs that utilize the EEGs of all scalp electrodes as input data for classifying motor imagery, execution, and observation, even in the case of the marginal dataset.**

I. Introduction

Neurorehabilitation based on brain plasticity has attracted considerable attention with regard to the rehabilitation of hemiplegic stroke patients with upper-limb motor disorders. One of the procedures of this type is the mental practice of motor imagery (MI). It is a method for recovering body functions by visualizing that the part of the body that cannot be moved owing to paralysis is being moved. Motor imagery can be classified into two categories: kinesthetic MI (KMI) and visual MI (VMI). KMI is the imagery of one's own physical activities via somatosensory information, such as joint movements and muscle output. Meanwhile, VMI is the imagery of one's own physical activities via visual or objective observation. It has been observed that the motor-related areas important for motor execution are more active during KMI and that the visual areas responsible for visual information processing are more active during VMI [1]. In addition, certain studies have classified VMI into first-person visual imagery and third-person visual imagery [2]. KMI or first-person visual imagery is considered more effective than third-person visual imagery for acquiring motor skills through imagery training [3]. Therefore, the objective and quantitative determination of which MI is being performed is a potential objective to characterize the best subject-specific MI method for optimizing the performance of individual subjects.

Recently, a method for classifying EEGs through machine learning during MI has been developed. Known as the brain–machine interface (BMI), this method classifies EEGs during the MI of the right hand, left hand, and foot through linear discriminant analysis and provides feedback to the patient by moving an avatar in the virtual space according to the classification results [4]. However, machine learning must be improved in terms of classification accuracy and online application. An approach for addressing this challenge involves the application of convolutional neural networks (CNNs) (which are effective owing to their suitability for end-to-end learning and their capability to learn the original waveform), whereby the need for feature extraction is circumvented [5].

Dose et al. applied a CNN to a large amount of EEG data from all 64 scalp electrode sites during (1) KMI for grasping the left hand and (2) right hand, (3) KMI for moving both feet, and (4) resting. They classified two ((1) and (2)), three ((1), (2), and (4)), and four ((1), (2), (3), and (4)) states with accuracies of 88.0%, 76.6%, and 65.7%, respectively [6]. Lun et al. compared the accuracy of the CNN classification of left- or right-hand grip KMI, both-feet KMI, and both-hands KMI using EEG from all 64 sites, with the accuracy of using EEG from only two sites (FC3 and FC4). The accuracies achieved in the two cases were 95.83% and 98.61%, respectively [9]. These studies used the same database and were conducted under conditions where large amounts of data (over 1500 EEGs recorded from 109 subjects) were secured.

In contrast, Sakamoto et al. extracted event-related (de)synchronization (ERD/ERS) features in the α and β bands of C3, C4, O1, and O2 from EEGs during KMI and VMI of right index-finger flexion. They obtained two classes using a support vector machine (SVM) and attained 82.5 ± 14.5% [7]. Tang et al. reported that when the ERD/ERS during left-and right-hand KMI was analyzed periodically to extract features and classified using CNN and SVM, the average accuracies were 86.41 ± 0.77% and 77.17 ± 7.69%, respectively [10]. In these studies, the number of subjects (less than three) and the number of classes (two) were small. Moreover, features were extracted from EEG and used as input to the classifier.

Although only a small amount of data can be acquired under conditions with limited resources and time, if EEG can be input to a classifier without feature extraction and if several states can be classified, it is likely to contribute to neurorehabilitation. Therefore, we aimed to classify the small amount of EEG data during the five states using a CNN. In this study, we compared the classification accuracies of a "single

T. Igasaki is with the Division of Information and Energy, Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto 860-8555 Japan (corresponding author to provide phone: +81-96-342-3613; fax: +81-96-342-3630; e-mail: iga@cs.kumamoto-u.ac.jp).

Y. Kuramura, and J. Takemoto are with the Department of Computer Science and Electrical Engineering, Graduate School of Science and Technology, Kumamoto University, Kumamoto 860-8555 Japan.

five-class CNN" and "multiple two-class CNNs." The latter classify five states using a combination of ten CNNs (each CNN classifies a pair of states). We also compared the classification accuracies when two types of input data were adopted: 1) the EEGs of all scalp electrode sites and 2) those of the electrode sites closely related to the motor execution and observation of the index finger.

II. Methods

A. Subjects

Ten healthy males aged 21–24 years participated in the study. The study was approved by the Ethics Committee of the Graduate School of Advanced Science and Technology, Kumamoto University. The subjects provided full informed consent for the study, in accordance with the Declaration of Helsinki.

B. Tasks

The subjects were asked to engage five states: two types of MI of right index-finger flexion (KMI and VMI), actual performance of this exercise (ME), observation of a video of this exercise (MO), and resting (NMI).

The subjects sat on a chair in an electromagnetically shielded room and were instructed to observe an LCD screen set on a tabletop 1 m in front of them.

The experimental protocol is illustrated in Fig. 1. Each trial lasts 60 s: –50 to –35 s is the break period, –35 to –15 s is the rest period prior to instruction (first rest period), –15 to –10 s is the instruction period, –10 to 0 s is the rest period after the instruction (second rest period), and 0 to 10 s is the task period. During the break period, the subjects were free to act as they pleased, while the corresponding image was displayed on the LCD screen. In the first and second rest periods, the LCD screen was blank, and the subjects were asked to rest with their eyes open. In the instruction period, the state to be engaged by the subject in the task section ("KMI," "VMI," "NMI," "ME," or "MO") was displayed on the LCD screen, and the subjects were asked to keep their eyes open and rest. In the task section, when the state to be engaged by the subject was "MO," a video of right index-finger flexion was displayed on the LCD screen. Otherwise, the screen remained blank, and the subjects engaged the state as instructed. The subjects were exposed to white noise (from a speaker installed on the tabletop) from the first rest period to the second and to pink noise during the task period. The variation from white to pink noise was adopted as a trigger for the subject to engage the required state.

The order of states was randomized. Furthermore, ten trials for each state were conducted per day, excluding the trials in which artifacts such as eye blinks or body movements were present in the recorded EEG. Trials were conducted until the number of trials for each state exceeded fifty.

C. Recordings

1) Psychological measurements

The movement imagery questionnaire-revised second edition (MIQ-RS) was administered for a subjective evaluation of the capability to perform motor imagery [8]. The MIQ-RS is a questionnaire that evaluates the KMI and VMI difficulty of seven exercises on a seven-point scale from 1 (very difficult) to 7 (very easy). After the experiment, the subjects were asked to describe the KMI and VMI difficulty of right index-finger flexion in an identical MIQ-RS format.

2) Physiological measurements

Electrodes were placed at 19 sites on the scalp and both earlobes, according to the international 10–20 system. EEG data were recorded using a linked earlobe reference. To verify the movement of the right index finger, surface electromyogram (EMG) data were measured by positioning the explore electrode on the flexor digitorum superficialis of the right hand and placing the reference electrode on the distal end of the radius.

Band-pass filters of 0.5–200 Hz and 50–3000 Hz were applied to the EEG and EMG, respectively. In addition, a notch filter of 60 Hz was applied to both the EEG and EMG. The sampling frequency was 1000 Hz. The total recording time was 40 s—from 5 s after the start of the first rest period to the end of the task period.

D. Convolutional Neural Network

The structure of the CNN adopted in this study is illustrated in Fig. 2. It is based on the CNN structure proposed by Does et al., which consists of an input, a first convolutional, a second convolutional, a pooling, a fully connected, and an

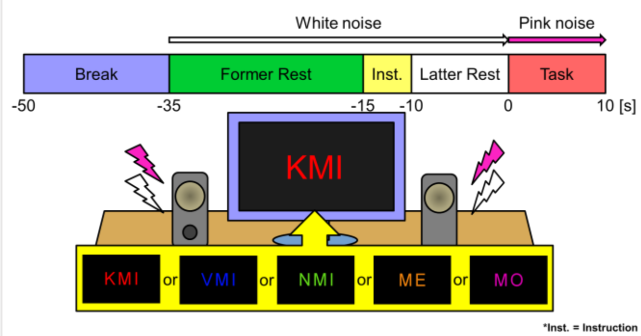

Figure 1. Experimental protocol. The colored bars represent periods. The state to be performed by the subject in the task section or a video of the right index-finger flexion is displayed on the LCD screen in the instruction section or the task section when the state is MO, respectively. White or pink noise is provided from a speaker in the first/second rest and instruction sections or the task section, respectively.

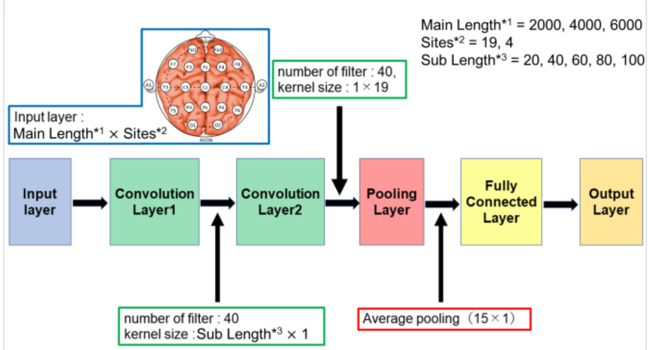

Figure 2. CNN architecture. Main length refers to the length of the input and sites to the number of EEG channels used. Convolutional layers 1 and 2 are for temporal and spatial convolution, respectively. The output layer is specified with softmax.

output layer [6]. The "data length × number of sites" of the EEG was input into the CNN. Then, the "state" was output from the CNN. We used 2000, 4000, and 6000 points from the start of the task period as the data length, and 19 (all the 19 sites) and 4 (C3, C4, O1, and O2, which are closely related to the motor execution and observation of the index finger) as the number of sites.

In the first convolutional layer, five kernels (20 × 1, 40 × 1, 60 × 1, 80 × 1, and 100 × 1) were adopted for temporal convolution with a one-point shift that implements zero paddings, and 40 feature maps were created.

In the second convolutional layer, 40 feature maps were developed via spatial convolution with a one-point shift using a 1 × 19 or 1 × 4 kernel (depending on the number of sites).

To reduce the number of features, a 15 × 1 kernel was adopted in the pooling layer to perform 15-point average pooling without overlapping. The reduced features were combined in the fully connected layer and then sent to the output layer, from which the classification results were output.

1) Single five-class CNN

We developed a CNN that classified the five states simultaneously for each subject. The CNN was trained and tested with 80% and 20% data, respectively, for each state.

2) Multiple two-class CNNs

First, for each subject, we developed ten CNNs, each of which classified a pair of states, i.e., a CNN each to classify "KMI and VMI," "KMI and NMI," ..., and "ME and MO."

Then, we input weighting data into each CNN and determined weights for each state of the CNN. For example, when J-weighted KMI data were input into the CNN that classified "KMI and VMI," we obtained the output for KMI, ${}^{KMI}_{VMI}O^{KMI}_{KMI_j}$, and VMI, ${}^{KMI}_{VMI}O^{VMI}_{KMI_j}$ $(j = 1,2,\cdots,J)$. The average output for KMI, i.e.,

$$ {}^{KMI}_{VMI}W^{KMI}_{KMI} = \frac{1}{J}\sum_{j=1}^{J} {}^{KMI}_{VMI}O^{KMI}_{KMI_j}, $$

and that for VMI, i.e.,

$$ {}^{KMI}_{VMI}W^{VMI}_{KMI} = \frac{1}{J}\sum_{j=1}^{J} {}^{KMI}_{VMI}O^{VMI}_{KMI_j}, $$

were used as the weights for KMI in the CNN. Similarly, ${}^{KMI}_{VMI}W^{KMI}_{VMI}$, ${}^{KMI}_{VMI}W^{VMI}_{VMI}$, ${}^{KMI}_{VMI}W^{KMI}_{NMI}$, ${}^{KMI}_{VMI}W^{VMI}_{NMI}$, ${}^{KMI}_{VMI}W^{KMI}_{ME}$, ${}^{KMI}_{VMI}W^{VMI}_{ME}$, ${}^{KMI}_{VMI}W^{KMI}_{MO}$, and ${}^{KMI}_{VMI}W^{VMI}_{MO}$ represent the weights for VMI, NMI, ME, and MO in the CNN. This process was also performed for the CNNs for classifying the other nine pairs of states.

The test data were then input to each CNN. Subsequently, we calculated the sum of the product of each output and each weight developed. For example, when the test data t were input into the CNN that classified "KMI and VMI," we obtained the output for KMI, ${}^{KMI}_{VMI}O^{KMI}_t$, and VMI, ${}^{KMI}_{VMI}O^{VMI}_t$. In this case,

$$ {}^{KMI}_{VMI}S^{KMI}_t = {}^{KMI}_{VMI}W^{KMI}_{KMI} \times {}^{KMI}_{VMI}O^{KMI}_t + {}^{KMI}_{VMI}W^{VMI}_{KMI} \times {}^{KMI}_{VMI}O^{VMI}_t $$

is the score of KMI in the CNN. Similarly, ${}^{KMI}_{VMI}S^{VMI}_t$, ${}^{KMI}_{VMI}S^{NMI}_t$, ${}^{KMI}_{VMI}S^{ME}_t$, and ${}^{KMI}_{VMI}S^{MO}_t$ are the scores for VMI, NMI, ME, and MO, respectively, in the CNN. This process was also performed for the CNNs for classifying the other nine pairs of states. Thereafter,

$$ S^{KMI}_t = {}^{KMI}_{VMI}S^{KMI}_t + {}^{KMI}_{NMI}S^{KMI}_t + {}^{KMI}_{ME}S^{KMI}_t + {}^{KMI}_{MO}S^{KMI}_t + {}^{VMI}_{NMI}S^{KMI}_t + {}^{VMI}_{ME}S^{KMI}_t + {}^{VMI}_{MO}S^{KMI}_t + {}^{NMI}_{ME}S^{KMI}_t + {}^{NMI}_{MO}S^{KMI}_t + {}^{ME}_{MO}S^{KMI}_t $$

was utilized to calculate the total KMI score. Similarly, the total scores S^{VMI}_t, S^{NMI}_t, S^{ME}_t, and S^{MO}_t for VMI, NMI, ME, and MO, respectively, were calculated. The highest total score was used as the classification result for the test data t.

With regard to the number of trials in each state, the CNNs were trained, weighted, and tested on 80%, 10%, and 10% of the data, respectively.

III. Results

The subjective evaluation of right index-finger flexion based on the MIQ-RS is presented in Table 1. All the subjects self-reported that VMI was more convenient than KMI. Subjects #1 and #7 provided the highest ratings.

Table 2 presents the classification accuracies of the single five-class CNN and multiple two-class CNNs for all the subjects. When the number of sites was 19, the classification accuracy of the multiple two-class CNNs was significantly higher than that of the single five-class CNN for all the subjects except Subject #2 ($p = 0.044$, paired t-test). In particular, for Subject #1, the classification accuracies of the single five-class CNN and multiple two-class CNNs were 62.3% and 73.1%, respectively (an improvement of over 10%). However, for Subject #2, the classification accuracies were 43.9% and 32.1%, respectively (a decrease of over 10%). In contrast, when the number of sites was four, the classification accuracy of the multiple two-class CNNs was higher than that of the single five-class CNN for half of the subjects, although the difference was negligible ($p = 0.604$, paired t-test).

With regard to the difference in the number of sites, the classification accuracy for 19 sites was higher than that for 4 sites with the single five-class CNN in the case of half of the subjects. However, the difference was negligible ($p = 0.094$, paired t-test). The classification accuracy for 19 sites was significantly higher than that for 4 sites in the case of the multiple two-class CNNs in all the subjects except Subjects #2, #8, and #10 ($p = 0.008$, paired t-test). In particular, for Subjects #1, #3, and #7, the classification accuracies for 4 sites were 61.5%, 35.7%, and 35.7%, respectively, whereas those for 19 sites were 73.1%, 46.4%, and 48.3% (an improvement of over 10%). The classification accuracies for Subjects #2, #8, and #10 were equal.

IV. Discussion

The highest and lowest classification accuracies observed in this study were 73.1% and 32.1%, respectively. Although the lowest classification accuracy was higher than the probability of correct classification by chance (20%) owing to the five-state classification, the average classification accuracy of 65.7% obtained by Dose et al. [6] was exceeded only for Subject #1. This occurred because the number of classes in this study was five, whereas that in the study by Dose et al. was four. Another reason is that the movement considered in this study was only right index-finger flexion,

TABLE I. SUBJECTIVE EVALUATION OF RIGHT INDEX-FINGER FLEXION BASED ON MIQ-RS

Subject	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
KMI	7	2	3	5	5	4	7	4	4	4
VMI	7	4	7	5	7	6	7	3	6	5

TABLE II. CLASSIFICATION ACCURACY OF ALL SUBJECTS IN THE SINGLE FIVE-CLASS CNN (SINGLE 5) AND MULTIPLE TWO-CLASS CNNS (MULTIPLE 2) [%]

CNN	Single 5	Multiple 2	Single 5	Multiple 2
Sites	19		4	
Subject				
#1	62.3	73.1	62.3	61.5
#2	43.9	32.1	40.4	32.1
#3	38.6	46.4	40.4	35.7
#4	43.9	48.3	38.6	35.7
#5	47.3	55.6	43.6	48.1
#6	48.3	55.2	46.6	51.7
#7	46.7	53.3	41.7	50.0
#8	50.0	56.7	51.7	56.7
#9	49.1	51.9	49.1	48.1
#10	51.9	55.6	51.9	55.6
Mean ± SD	48.2 ± 5.9	52.8 ± 9.7	46.6 ± 6.9	47.5 ± 9.4

and the corresponding electrode sites were localized around C3. In contrast, the motions in the study by Dose et al. were left- and right-hand grip and both-feet movements, and the corresponding electrode sites were located extensively around C4, C3, and Cz.

The classification accuracy of the single five-class CNN was lower than that of the multiple two-class CNNs for only Subject #2. This may be because Subject #2 experienced difficulties in both KMI and VMI (as revealed by the MIQ-RS results). The classification accuracy for Subject #2 was the second-lowest for the former CNN and the lowest for the latter. Therefore, we conclude that it is necessary to train such subjects to ensure correct motor imagery. However, Subject #1 self-reported that both KMI and VMI were convenient. Moreover, the classification accuracy was higher for the aforementioned subject, particularly for the latter CNN. For Subjects #8 and #10, the second- and third-highest classification accuracy were achieved. However, in the MIQ-RS, they self-reported their KMI and VMI to be 4 and 3 or 5, respectively. Thus, instructing such subjects to reduce the discrepancy with their subjective evaluations may be one of the solutions.

With regard to the difference in number of sites, Lun et al. compared the accuracy of the CNN classification of either left- or right-hand grip KMI, both-feet KMI, and both-hands KMI when EEG data from all sixty-four sites were used and that when EEG data from only two sites (FC3 and FC4) were used. The accuracies attained were 95.83% and 98.61%, respectively. This indicates the effectiveness of reducing the number of sites by selecting sites closely related to MI [9]. In contrast, in this study, the classification accuracy of both single five-class CNN and multiple two-class CNNs when all the 19 sites were used was higher than that when the 4 sites closely related to right index-finger flexion were adopted.

Tang et al. reported that when the ERD/ERS during left-and right-hand KMI was analyzed periodically to extract features and was classified using a CNN and SVM, the average accuracies were 86.41 ± 0.77% and 77.17 ± 7.69%, respectively [10]. When we incorporated Sakamoto et al.'s [7] method into SVM to classify the data used in this study, the average accuracy obtained was 44.1 ± 6.9%. This was lower than that of the single five-class CNN (46.6 ± 6.9%), which utilized EEG data from four sites. In this study, the original EEG waveform was used, and feature extraction was not performed. Furthermore, the number of data in this study (approximately 50 for each subject and each task) was insufficient.

V. CONCLUSION

In this study, only one weighting method (repeating a simple average equation) was examined. Because this method may not be optimal, it is necessary to determine an optimal weighting method in conjunction with other methods. In addition, because only one method was used to divide the training, weighted, and test data in this study, it is also necessary to utilize leave-one-out cross-validation and other methods to obtain reliable results. However, in such "small-data" (marginal dataset, far from "big-data") cases, two-class classification may yield a higher accuracy than multi-class classification can. This indicates the effectiveness of the multiple two-class CNN for the EEG-based MI/ME/MO classification proposed herein.

REFERENCES

[1] A. Guillot, C. Collet, V. A. Nguyen, F. Malouin, C. Richards, and J. Doyon, "Brain activity during visual versus kinesthetic imagery: an fMRI study," *Hum. Brain Mapp.*, vol. 30, no. 7, pp. 2157–2172, 2009.

[2] P. Holmes and C. Calmels, "A neuroscientific review of imagery and observation use in sport," *J. Mot. Behav.*, vol. 40, no. 5, pp. 433–445, 2008.

[3] K. R. Ridderinkhof and M. Brass, "How kinesthetic motor imagery works: a predictive-processing theory of visualization in sports and motor expertise," *J. Physiol. Paris*, vol. 109, pp. 53–63, 2015.

[4] Y. Hashimoto, J. Ushiba, A. Kimura, M. Liu, and Y. Tomita, "Change in brain activity through virtual reality-based brain-machine communication in a chronic tetraplegic subject with muscular dystrophy," *BMC Neurosci.*, vol. 11, no. 117, pp. 1–9, Sep. 2010.

[5] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, Aug. 2017.

[6] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Syst. Appl.*, vol. 110, pp. 532–542, Dec. 2018.

[7] K. Sakamoto, J. Takemoto, and T. Igasaki, "Classification of kinesthetic / visual motor imagery based on event-related desynchronization / synchronization," *J. Jpn. Biomag. Bioelecromag. Soc.*, vol. 30, no. 1, pp. 89, May 2017.

[8] M. Gregg, C. Hall, and A. Butler, "The MIQ-RS: a suitable option for examining movement imagery ability," *Evid. Based Complement. Alternat. Med.*, vol. 7, no. 2, pp. 249–257, Jun. 2010.

[9] X. Lun, Z. Yu, T. Chen, F. Wang, and Y. Hou, "A simplified CNN classification method for MI-EEG via the electrode pairs signals," *Front. Hum. Neurosci.*, vol. 14, no. 338, pp. 1–14, Sep. 2020.

[10] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, Feb. 2017.