Subject-Aware User State Classification with Deep Learning Models: An Exploratory Study

Haram Kwon[1], Jaehoon Choi[1], Jin Woo Choi[2], Sungho Jo[1,3]
[1]School of Computing, KAIST, Republic of Korea
[2]Information & Electronics Research Institute, KAIST, Republic of Korea
[3]shjo@kaist.ac.kr

***Abstract*—Inter-subject variability is one of the critical issues that hinder electroencephalogram-based brain-computer interfaces from wide usage. Recent studies that aimed to tackle such problems utilize deep learning methods such as domain adaptation to have their feature extractor learn domain-invariant features. As such approaches employ user state classification as well as subject classifiers to contribute toward domain-invariant feature extractions, considering the performance on both the state classification and subject identification for designing the feature extraction model may further benefit such approach. Thus in this work, we aim to improve widely used convolutional neural network-based feature extractors by enhancing subject identification accuracy while preserving user state classification. Along with our approach of using multi layer perceptron, we trained and evaluated our method using the visual imagery dataset and the speech imagery dataset collected from five participants. By training with EEG dataset of one paradigm and evaluating with the other dataset, our proposed feature extraction method achieved higher subject identification accuracy than the baseline models, while preserving their user state classification performance.**

***Keywords*—Brain-computer interface, Electroencephalography, Speech imagery, Visual imagery, Subject independent**

I. Introduction

Brain-computer interface (BCI) enables direct communication between a computer and a brain by using signals from the brain [1]. Among various signals utilized in BCIs, a non-invasive approach such as electroencephalography (EEG) is widely used due to its accessibility. Various EEG paradigms have been well-explored for potential applications in device control depending on their intended purpose. Reactive BCI paradigms, such as P300 or steady-state visual-evoked potential (SSVEP), decode user's intention based on brain signals generated as a response to an external stimulus, while active BCI paradigms, such as motor imagery (MI) [2], visual imagery (VI), and speech imagery (SI) [3], employ brain activations during different mental tasks that do not require an external stimulus. These EEG paradigms are used for various purposes such as quadcopter control [4], controlling home appliances [5], spellers [6], and playing chess [7]. Different paradigms are used in multiple studies as each of them has its own distinctive feature that can be made distinguishable through signal analysis [8].

Despite a variety of choices in EEG paradigms, one issue remaining in BCI systems is the data scarcity problem. This is a difficult issue to overcome since the problem occurs due to the fact that EEG collection is a burdensome task for the user and that massive EEG data collection is not feasible. Allowing for subject-independent decoding of EEG is one approach amongst many in resolving the issue [9], [10]. The idea is that enabling the use of EEG data from others will vastly increase the amount of EEG in training and reduce the need for the actual user to collect EEG data repetitively. However, the classification accuracy coming from the subject independent approach is yet lower than subject dependent approach. One approach to improving cross-subject classification involves the use of domain adversarial neural network (DANN), a well-known method used to achieve domain adaptation with convolutional neural network-based models such as ShallowConvNet or EEGNet as a feature extractor. It aims to train the model to extract domain-invariant features with the use of a gradient reversal layer [11]. This approach involves the user-state classifier as well as the domain classifier in order to consider user-state-related and subject-related features respectively. Previous works involving DANN in EEG-related studies trained user-state classifiers with only task-related knowledge; however, given that the classifier can efficiently learn both task and subject-related features, providing additional information regarding the subjects during training may result in a faster and more separable feature space when used with domain adaption.

In this work, the possibility of a user-state classifier that can learn both task and subject-related features is explored. Existing deep learning models, such as ShallowConvNet and EEGNet, utilize several convolutional layers as a feature extractor and a single dense layer as a classification layer, which may not be deep or complex enough. We propose using an additional multi-layer perceptron (MLP) as the classification layer instead; an additional block of MLPs is applied to a pre-trained feature extractor block and fine-tuned to better learn the subject and user-state related features simultaneously. The models are trained and evaluated with a hybrid paradigm dataset containing visual imagery (VI) and speech imagery (SI), using labels containing both subject and task-related information and evaluated in terms of subject identification performance and paradigm task classification performance. Our proposed method showed higher cross-paradigm subject identification accuracy when compared to basic ShallowConvNet and EEGNet while maintaining the user state classification performance.

2023 11th International Winter Conference on Brain-Computer Interface (BCI) | 978-1-6654-6444-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/BCI57258.2023.10078542

II. Methods

A. Data acquisition

Visual imagery (VI) and speech imagery (SI) EEG were collected from a total of five subjects, two males and three females, with an average age of 23.8, ranging from 23 to 26. Cues were displayed through a monitor while the participant remained seated. EEG was recorded in a dark soundproof room in order to minimize distractions when recording. EEG data were obtained at a sampling rate of 500Hz using BrainProducts' actiChamp and actiCAP. Having AFz as ground and Fz as a reference, a total of 27 electrodes (Fp1, F3, F7, FC5, FC1, C3, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, CP6, CP2, Cz, C4, FC6, FC2, F4, F8, Fp2) were used in both the VI and SI settings according to the international 10-20 system.

B. Paradigms and protocols

All five subjects participated in a data acquisition procedure that lasted for four days in total, where VI data were collected for two consecutive days and SI data were also collected for the other two consecutive days. For both VI and SI, two sessions of EEG recording took place each day, yielding a total of four sessions of recording for each paradigm. In between two sessions conducted in a single day, a ten-minute break was given.

A protocol of a single trial in VI EEG recording is illustrated in Figure 1a. The trial starts with a two-second rest by showing a big cross, followed by a four-second long visual cue-providing period. A visual cue is displayed in the monitor as an animation that shows a particular movement of a mass drone over a 4-second span. Then another four seconds are given for rest in order to prevent the recorded EEG from being influenced by the visual stimuli given as cue. Then for the next four seconds, participants are required to visualize the animation they have seen through the visual cue while staring at the screen. Four different types of labels were collected based on the type of animation that the participant visualized in each trial, where four types of animation are spread out, split, fall in, and rest. Visual cue shows an animation of drones spreading out in a circle for spread out, splitting into two groups of drones towards the outer direction for split, drones moving towards a center circle from being spread out for fall in, and drones remaining stationary for rest class. One session consists of a total of 60 trials, 15 trials per label, and the entire VI dataset of a subject contained 240 four-second-long EEG data.

The procedure for recording a single trial of SI EEG is represented in Figure 1b. The trial begins with a cue that indicates the word that the participant is required to speak out in mind for two seconds. Then one-second interval is given afterward, followed by a two-second-long speech imagination period. The participants are required to imagine speaking out a word displayed in the cue period. One-second interval and speech imagination period is then repeated three more times, imagining a speech of the word given in the cue period. Once the participant imagined the speech of a word given in the cue

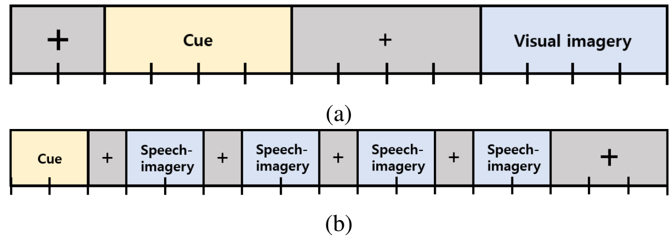

Fig. 1: (a) A single trial illustration of visual imagery protocol. Each line represents one second. Cue is provided in the form of animation of a mass drone movement (b) A single trial illustration of speech imagery protocol. Each line represents one second

period four times, a two-second long interval is given before the participant moves on to receive the next cue. Four different words were provided as cues in SI, each of them being ba, ku, he, and li. One session consisted of a total of 240 trials, 60 trials per label, and the entire SI dataset of a subject contained 960 two-second long EEG data.

C. Data preprocessing

SI and VI datasets were preprocessed in an identical manner. 60Hz notch filter was applied to both datasets which was later band-pass filtered between 1Hz and 90Hz. This is to keep the high-frequency band that is known to include SI-related patterns. Data augmentation was applied to the VI dataset by cropping four-second long EEG windows in the VI dataset into 1.5-second long windows with a stride of 0.1 seconds. Since the length of each trial of SI is 2 seconds, SI had to be augmented with a narrower stride in order to provide a sufficient number of samples. Thus, two-second long EEG windows in the SI dataset were cropped into 1.5-second long windows with a stride of 0.05 seconds.

D. Proposed model architecture

The proposed method adds multi layer perceptron in between the feature extraction network and the dense layer of the deep learning based models. The multi layer perceptron used in the proposed method is composed of four dense layers with ReLU activations in between each dense layer. The dimension of the dense layer narrows down and returns to the original size at the end. With ShallowConvNet, the multi layer perceptron receives an input sequence with the length of 760, which gradually narrows down to 512 and 256 and increases the size again to 512 and 760. Output dimensions of dense layers in EEGNet are smaller than that of ShallowConvNet. The multi layer perceptron added to the end of EEGNet receives an input sequence with a length of 288, which is narrowed down to 256 and 128, and then gradually increases the size to 256 and 288.

E. Training and evaluation

The focus of the evaluation is on whether the model can produce generalized subject-specific features while preserving user state-related features. Evaluation of the proposed model was designed to demonstrate the capability to learn both subject-specific and user state related features. VI and SI dataset were used as a training set and testing set respectively, and vice versa. Test sets were used for subject identification performance evaluation; this was to measure how well the model learned features related only to the subject identity and not the paradigm. VI and SI datasets that were initially labeled as 4-class data were relabeled as 20-class data when used as a training dataset, where each label was assigned to each visual or speech imagery label from each subject (4 paradigm labels $\times$ 5 subjects = 20). Each dataset was divided into train, valid, and test sets as stated below. Sessions recorded on the first day were assigned to train set, EEG from the first session of day 2 to valid set, and the second session of day 2 was assigned to test set. EEG of a paradigm not used in the training was relabeled as a 5-class dataset to have each label represent the subject and was used in evaluating subject identification performance. When evaluating 4-class or 5-class dataset with the model which classifies 20 classes, the output of the model was relabeled into a corresponding label of a given dataset.

The output dimension of the dense layer of ShallowConvNet and EEGNet was set to 20 for models to learn both subject-related features and user state-related features. Each model that applied the proposed MLP will be referred to as MLP-ShallowConvnet and MLP-EEGNet. The order of both experiments proceeded as follows. The original deep learning models were first trained either with 20-class SI or VI dataset. After the first training, the dense layer of the trained model was replaced with multi layer perceptron (MLP) that classifies 20-class EEG to build a proposed model. With its feature extraction network frozen, only the MLP was fine-tuned using the same 20-class EEG dataset used in the previous training.

The performance of the proposed model was compared against ShallowConvNet and EEGNet, two widely used EEG deep learning models in EEG domain adaptation. MLP-ShallowConvNet and MLP-EEGNet were also trained for comparison with the baseline model. With both VI and SI dataset band-pass filtered between 1Hz and 90Hz after applying 60Hz notch filter, the batch size was set to 32. Adam optimizer was used to train models with a learning rate of 0.00001. ShallowConvNet was trained for 1000 epochs and EEGNet was trained for 2000 epochs as EEGNet tended to train slowly. Once the training was finished, the dense layer of ShallowConvNet and EEGNet was replaced with the proposed MLP to build a proposed model. Before fine tuning the model, all of the trained layers from the first training were frozen. This left only the newly replaced MLP to be trained in the fine tuning step. Fine tuning was done for 100 epochs with the same dataset used in the initial training. ShallowConvNet and EEGNet after the first training were considered baseline models and the classification performance of fine tuned MLP-ShallowConvNet and MLP-EEGNet were compared against each of its baseline models. Dataset of a paradigm not used in the training process was used to evaluate the 5-class subject identification performance of each model as well as the performance of task classification of the paradigm used during the training.

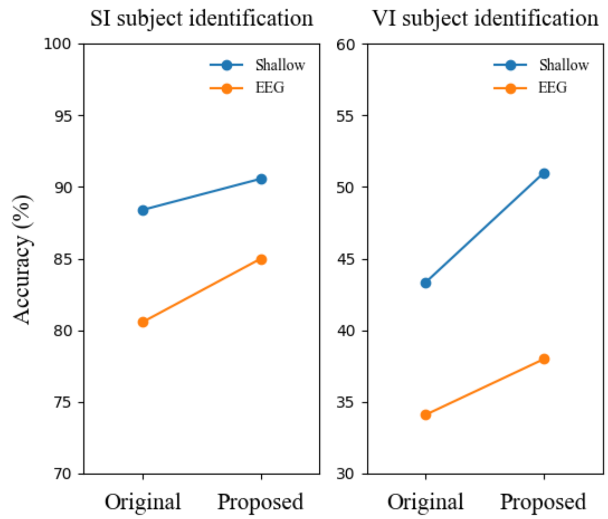

Fig. 2: Subject identification accuracy using SI (left) and VI dataset (right).

III. Results and Discussion

Figure 2 shows the result of subject identification done with speech imagery and visual imagery EEG. The plot on the left evaluation result of the subject identification performance of the models that were trained with the VI dataset using the SI dataset. The plot on the right evaluates the VI dataset subject identification performance of the model trained with SI dataset. The original model refers to ShallowConvNet and EEGNet, each denoted as Shallow and EEG in the plot, whereas the proposed model refers to MLP-ShallowConvNet and MLP-EEGNet.

The result indicates that cross-paradigm 5-class subject identification classification accuracy improved when the proposed model was used. When the model was first trained with VI dataset, SI subject identification accuracy of MLP-ShallowConvNet increased by 2.17% compared to ShallowConvNet while MLP-EEGNet has shown 4.42% improvement in accuracy in comparison to EEGNet. When the model was first trained with SI dataset, MLP-ShallowConvNet showed accuracy higher than that of ShallowConvNet by 7.64% and MLP-EEGNet has observed a 3.88% increase in accuracy from EEGNet.

While the proposed architecture improved subject classification accuracy, it is also crucial to ensure that the model is capable of classifying user-state as accurately as the baseline models. Thus, user-state accuracies of VI and SI by the baseline models and the proposed models were examined. Although the accuracies have decreased in general, the change was no larger than 0.94% and observed accuracies were above chance level (25%). Relatively low accuracies might have been from training the models on 20-class dataset, which may have affected the performance. In addition to that, it is likely that the bandpass filter range used in both datasets was not the optimal

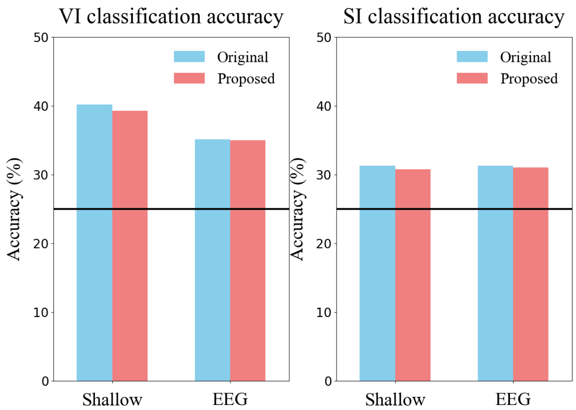

Fig. 3: The classification accuracy of 4-class SI (left) and VI (right) dataset. The black horizontal line marks the random chance level.

bandpass filter range for each paradigm since the same range of passband was applied to both VI and SI dataset.

It is generally known that multi-layer perceptron is advantageous over single-layered perceptron in the sense that it allows for the recognition of not linearly separable patterns [12]. The result implies that the proposed model may extract more useful subject-specific features than the baseline models while keeping the user-state classification accuracy. However, there remains an issue that is yet to be explained.

Figure 2 clearly shows that the subject identification accuracy differs when trained with different paradigms. The SI subject identification accuracies of the models trained with VI dataset ranged from 80% to 90% while the VI subject identification accuracies of the models trained with SI dataset ranged from 34% to 50%. Such a difference could have originated from the difference in activated regions when performing each paradigm. The difference in regions related to each paradigm could have affected the model to focus on specific regions, which may not provide useful information in the subject identification of certain paradigms. However, the reason why subject identification accuracy differs remains unknown and will need further investigation.

IV. Conclusion and Future Works

In this paper, a method of attaching multi-layer perceptron to ShallowConvNet and EEGNet has been proposed. Such an adjustment allowed each model to identify subjects using SI data with higher accuracy when trained with VI data and vice versa. This may reflect the enhanced ability to capture generalized subject-specific features present in a different paradigm as a result of the adjustment. Further investigation of the proposed method will need to be made to apply the model to the domain adaptation network in the future.

Acknowledgment

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00432).

References

[1] D. J. McFarland and J. R. Wolpaw, “Brain-computer interfaces for communication and control,” *Commun. ACM*, vol. 54, p. 60–66, may 2011.

[2] J. W. Choi, S. Huh, and S. Jo, “Improving performance in motor imagery bci-based control applications via virtually embodied feedback,” *Computers in Biology and Medicine*, vol. 127, p. 104079, 2020.

[3] N. Kaongoen, J. Choi, and S. Jo, “Speech-imagery-based brain–computer interface system using ear-eeg,” *Journal of Neural Engineering*, vol. 18, p. 016023, feb 2021.

[4] M. Wang, R. Li, R. Zhang, G. Li, and D. Zhang, “A wearable ssvep-based bci system for quadcopter control using head-mounted device,” *Ieee Access*, vol. 6, pp. 26789–26798, 2018.

[5] N. Kaongoen, J. Choi, and S. Jo, “A novel online bci system using speech imagery and ear-eeg for home appliances control,” *Computer Methods and Programs in Biomedicine*, vol. 224, p. 107022, 2022.

[6] Z. Lin, C. Zhang, Y. Zeng, L. Tong, and B. Yan, “A novel p300 bci speller based on the triple rsvp paradigm,” *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.

[7] J. W. Choi, E. Rho, S. Huh, and S. Jo, “An eog/eeg-based hybrid brain-computer interface for chess,” in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 129–134, IEEE, 2018.

[8] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, “A comprehensive review of eeg-based brain–computer interface paradigms,” *Journal of Neural Engineering*, vol. 16, p. 011001, jan 2019.

[9] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, “Subject-independent mental state classification in single trials,” *Neural Networks*, vol. 22, no. 9, pp. 1305–1312, 2009. Brain-Machine Interface.

[10] J. Lee, J. W. Choi, and S. Jo, “Subject-independent motor imagery eeg classification based on graph convolutional network,” in *Pattern Recognition* (C. Wallraven, Q. Liu, and H. Nagahara, eds.), (Cham), pp. 268–281, Springer International Publishing, 2022.

[11] X. Tang and X. Zhang, “Conditional adversarial domain adaptation neural network for motor imagery eeg decoding,” *Entropy*, vol. 22, no. 1, p. 96, 2020.

[12] J. Singh and R. Banerjee, “A study on single and multi-layer perceptron neural network,” in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 35–40, 2019.