Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom – August 03-05, 2023
Paper No. ICBES 126
DOI: 10.11159/icbes23.126

EEG Channel Selection Method for Subject-Independent Motor Imagery Classification using Shapley Additive exPlanations

Vishnupriya R[1], Neethu Robinson[2], Ramasubba Reddy M[1]
[1]Indian Institute of Technology Madras
Chennai 600036, India
vishnupriyaece94@gmail.com; rsreddy@iitm.ac.in
[2]Nanyang Technological University
50 Nanyang Avenue, Singapore
nrobinson@ntu.edu.sg

Abstract – Electroencephalography (EEG) is a non-invasive method for measuring the brain's electrical activity. Deep learning models, such as deep convolutional neural networks, have shown great promise in analyzing motor imagery EEG signals for tasks such as motor imagery classification. However, careful EEG channel selection is needed for optimal performance. Explainable artificial intelligence (XAI) methods provide a way to interpret and understand the predictions of deep learning models. This paper proposes shapely additive explanations (SHAP), an XAI method based on Shapley values for subject-independent motor imagery EEG channel selection. This study uses a Deep ConvNet trained on a 62-channel motor imagery EEG dataset from Korea University. We also compare the performance of the proposed SHAP-based channel selection method with other XAI-based channel selection methods. Our results show that the proposed methodology obtained 84.07% (± 11.84) classification accuracy with only 20 selected channels, similar to the baseline classification accuracy (84.46% ± 11.56) with all 62 channels. The comparison results show that SHAP-based EEG channel selection performs better than other XAI-based EEG channel selections in terms of model accuracy and provides a more interpretable and direct way of selecting the most important EEG channels. These results suggest that SHAP is a viable alternative for subject-independent motor imagery EEG channel selection methods using deep learning models.

Keywords**:** Electroencephalography (EEG), convolutional neural network (CNN), explainable artificial intelligence (XAI), motor-imagery, brain-computer interface (BCI)

1. Introduction

Brain-computer interfaces (BCIs) are an alternative way of translating human intentions into commands for controlling external devices. This helps people with neurological impairments communicate more effectively with the external world [1]. Motor imagery BCI (MI-BCI), which is the imagination of motor action without actual limb movement, has been used widely for practical applications [2]. Deep learning models based on convolutional neural networks, such as Deep ConvNet, has shown great promise in analyzing and classifying the motor imagery EEG signal [3]. It has been discovered that the most informative channels for generating event-related de-synchronization (ERD) and event-related synchronization (ERS) differ between subjects [4]. Thus, one challenge with subject-independent EEG-based deep learning models is carefully selecting EEG channels to achieve optimum performance.

Most recently, explainable artificial intelligence (XAI) methods are gaining popularity in the BCI field because they help to increase trust and transparency in the decision-making process of a deep learning model. XAI methods can help determine which EEG channels are most important for the model's prediction. In the previous study, A. Nagarajan *et al.* investigated layer-wise relevance propagation (LRP), an XAI method, for EEG channel selection in subject-independent MI-EEG deep learning models. Identifying common subset of channels that produces optimal accuracy across all subjects is mentioned as the limitation of the study [5]. To address the foregoing, this study proposes a novel shapely additive explanation (SHAP), a different XAI method for subject-independent MI-EEG channel selection. Using the proposed method, we have identified a common subset of channels and evaluated its performance across all subjects. We also perform a comparative study between SHAP and LRP to investigate the effectiveness of these two XAI methods.

2. Methodology

The MI dataset reported by M. Lee et al. is used for evaluating the proposed method [6]. This dataset contains the EEG signal of 54 subjects while they performed two-class MI tasks (left-hand and right-hand imagined movement). EEG signals were recorded at a sampling rate of 1000 Hz using 62 Ag/AgCl electrodes. Each subject underwent two data recording sessions on different days, with training and testing phases in each session. Each phase included 100 trials per class, for a total of 400 trials. The experiment begins with a 3s rest period to prepare subjects for the MI task. The subject then performed the corresponding MI task for 4s while following the visual cue. After completing each task, the screen remained blank for 6s. For each trial, 4s MI tasks are segmented from continuous EEG signals and further down-sampled to 250 Hz. For the baseline study, all 62 channels were used. We used the state-of-the-art Deep ConvNet model proposed by Schirrmeister et al. as our baseline model [3]. The Deep ConvNet architecture is formed by four convolution max-pooling blocks and a fully connected SoftMax classification layer. The first block, in particular, contains temporal and spatial filters for handling the input EEG signals. For each convolutional max-pooling block, batch normalization and dropout are added.

2.1. Proposed Methodology: Shapely values-based EEG channel selection

Shapely Additive exPlanations (SHAP) is an XAI method that provides a way to explain the predictions made by deep learning models. It is based on the concept of Shapely values from cooperative game theory. In the context of deep learning, Shapely values can be used to determine the contribution of each feature to the prediction made by the model. To calculate Shapely values for each feature, SHAP generates subsets of features, calculates the contribution of each feature for each subset, and then computes the average contribution of each feature across all possible subsets. The resulting Shap values represent the importance of each feature in the model prediction for a given instance. The formula for calculating the feature importance of a feature i is given as follows [7]:

$$\varphi_i \;=\; \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(\,|N|-\,|S|-1)!}{|N|!} \big(v(S \,\cup\, \{i\}) -\, v(S)\big) \qquad (1)$$

Where, φ_i is the Shapely value for the feature i, N is the set of all features, S is the subset of features excluding the feature i , $v\,(S)$ is the model output when only features in subset S are used as model input, $v\,(S \,\cup\, \{i\})$ refers to the model's predicted output when feature i is added to the subset of feature S. We used the DeepExplainer function in SHAP python library to calculate the Shapley values.

In our proposed methodology, we utilized SHAP for selecting subject-independent MI-EEG channels using Deep ConvNet. The first step is training the baseline Deep ConvNet model using the training and validation dataset. The training process involves adjusting the network weights to minimize the error between the predicted and the actual outputs. The DeepExplainer in the SHAP library needs to be initialized with the trained model and a background dataset. The background dataset is used to compute each EEG channel's Shapely values. The DeepExplainer function returns the Shapley values for each trial and both classes separately. Then, compute the mean Shapely values across all trails for both classes. Sort the mean Shapley values in descending order and rank the channels. Then, select the top 20 channels with the highest Shapley values and train the Deep ConvNet model again for motor imagery classification. Compare the classification accuracy of the trained Deep ConvNet using the top 20 channels and the baseline model with all 62 MI-EEG channels. The block diagram for the proposed method is shown in Fig.1.

3. Experiments

We use the leave-one-subject-out cross-validation (LOSO-CV) method to evaluate the subject-independent classification. The model is trained with all data except the target subject. Based on the previous studies, data from all 53 subjects are split randomly into 85% training and 15% validation data, and the subject-independent model is trained. The following hyperparameter values were used for each model training: The Adam optimizer is used to minimize the loss function, the batch size is fixed as 16, and an early stopping strategy is used to avoid overfitting. As mentioned in the literature [3], a two-stage training strategy for training the Deep ConvNet is followed; after two stages of training, the best model with minimum validation loss is saved and evaluated on the test dataset, which is the last 100 trials (session 2 of Day 2) of the target subject's data. The proposed SHAP-based channel selection method uses the same data division and training process.

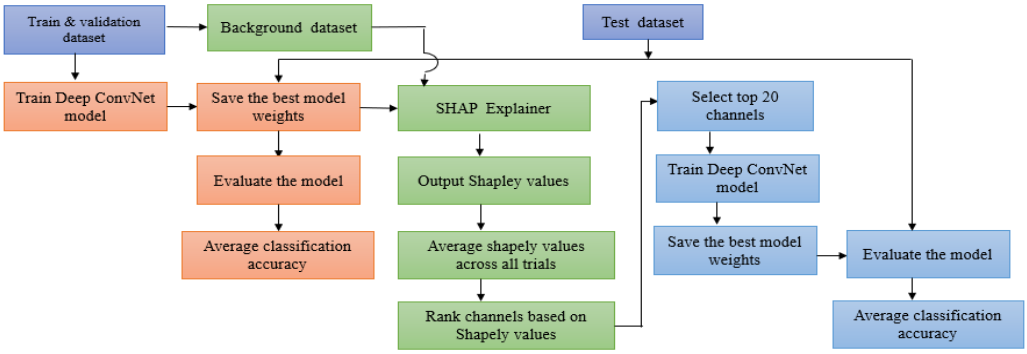

Fig. 1: Block diagram for Shapley based MI-EEG channel selection method

Additionally, for the background data required for the DeepExplainer function,100 randomly selected trials from the training dataset were used. All training process was carried out on an NVIDIA Tesla V100 GPU with 32 GB GPU of memory.

To better understand the efficacy of SHAP and LRP as MI-EEG channel selection methods, we compared the proposed SHAP-based channel selection method with the LRP-based channel selection method proposed in [5],which uses the same dataset and Deep ConvNet model.

4. Results and discussion

The average classification accuracy of the baseline subject-independent model and the proposed methodology using LOSO-CV is shown in Table 1. It is observed that the baseline model has an average classification accuracy of 84.46% (± 11.39) across all 54 subject models, which is similar to those mentioned in the literature [8]. The average classification accuracy using the top 20 MI-EEG channels selected by the proposed SHAP-based method is 84.07% (± 11.84%), with *p-value* = 0.616, indicating no significant difference between the performance of the baseline model with all 62 channels and the proposed method with top 20 selected channels. This result shows that using only 20 channels selected by the proposed SHAP-based channel selection method achieves comparable performance to the baseline model while also reducing the baseline model's complexity.

Table 1: Average classification accuracies (mean ± standard deviation)

Baseline model accuracy (62 channels)	**SHAP-based channel selection method (top 20 channels)**
84.46% (± 11.39%)	84.07% (± 11.84%) **

***P-value*=0.616, indicating no significant difference in accuracy between the baseline and the SHAP-based channel-selection method

We also identified the common subset of channels, by selecting the most frequently selected channels across all 54 subject models. As a common subset of channels, 21 channels selected more than 20 times by the subject-independent models are considered. The average classification accuracy for LOSO-CV, using the selected subset of channels is 84.37% (± 11.85), with *p-value* = 0.889, indicating no significant difference between the baseline accuracy. Thus, the proposed subject-independent channel selection method, produces optimal accuracy with the selected subset of channels. For better visualization, the frequency of the channels selected across all the models and the identified common subset of channels are shown in Fig. 2(a) and 2(b), respectively. The Fig. 2(a) demonstrates that in addition to the channels selected from the motor cortex region, our proposed method also selected channels from the frontal and the visual areas of the brain. These brain regions also involve during MI activity in addition to the motor cortex region.

We also compared the average classification accuracy for LOSO-CV using the top 20 channels selected by our proposed method with the 20 motor channels mentioned in the literature [5][6]. The 20 motor channels are shown in Fig. 2(c). The average classification accuracy obtained using 20 motor channels is 81.72% (± 12.61), which is ~ 2% less than our proposed method's performance. We also compared our proposed SHAP-based method with the performance of the LRP-based MI-EEG channel selection method mentioned in [5]. The performance using the top-20 channels selected by LRP is

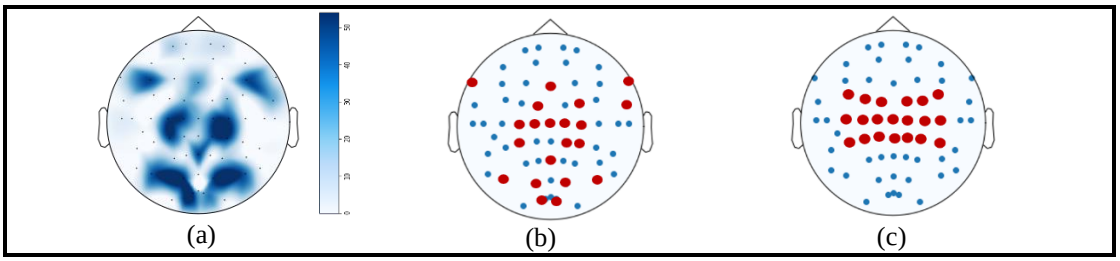

Fig. 2: In part (a), frequency of channels selected across all 54 subject models by the proposed method. In part (b), the topomap of the common subset of channels. Part (c), visualizes 20 motor channels

is 83.19% (±11.60) which is lesser than the performance of the proposed method's accuracy of 84.07%. Compared to the literature, LRP requires a minimum of the top 24 channels to meet the baseline model accuracy. In contrast, our proposed SHAP-based method is meeting the baseline accuracy with the top 20 channels itself. Thus, from the above results, SHAP is preferred over LRP for MI-EEG channel selection.

5. Conclusion

One of the challenges with subject-independent EEG-based deep learning models is the careful selection of EEG channels to achieve optimal performance. To overcome this issue, a novel Shapley value-based channel selection method is proposed. The results show that, with only top-20 selected channels, the proposed method produces comparable accuracy to the baseline model, which uses all 62 channels. There is still a scope for further improvement in the performance of the proposed method after adaptation. The common subset of channels which produces optimal accuracy across all subject- independent models are identified. The performance of the proposed method with the LRP-based channel selection method are compared. The results show that the proposed method outperforms the LRP-based channel selection method in terms of classification accuracy. Thus, SHAP can be more useful in the application of MI-EEG channel selection.

References

[1] Blankertz B, Tangermann M, Vidaurre C, Fazli S, Sannelli C, Haufe S, Maeder C, Ramsey LE, Sturm I, Curio G, Mueller KR., "The Berlin brain-computer interface: Non-medical uses of BCI technology," *Front. Neurosci.*, vol. 4, no. DEC, pp. 1–17, 2010, doi: 10.3389/fnins.2010.00198.

[2] Pfurtscheller, G., Neuper, C., Flotzinger, D. and Pregenzer, M., "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, 1997, doi: 10.1016/S0013-4694(97)00080-1.

[3] Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W. and Ball, T., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017, doi: 10.1002/hbm.23730.

[4] Yang, H., Guan, C., Wang, C.C. and Ang, K.K., "Maximum dependency and minimum redundancy-based channel selection for motor imagery of walking EEG signal detection," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 1187–1191, 2013, doi: 10.1109/ICASSP.2013.6637838.

[5] Nagarajan, A., Robinson, N. and Guan, C., "Relevance-based channel selection in motor imagery brain-computer interface," *J. Neural Eng.*, vol. 20, no. 1, 2023, doi: 10.1088/1741-2552/acae07.

[6] Lee, M.H., Kwon, O.Y., Kim, Y.J., Kim, H.K., Lee, Y.E., Williamson, J., Fazli, S. and Lee, S.W., "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *Gigascience*, vol. 8, no. 5, pp. 1–16, 2019, doi: 10.1093/gigascience/giz002.

[7] Lundberg, S.M. and Lee, S.I., "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, 2017.

[8] Zhang K, Robinson N, Lee SW, Guan C., "Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network," *Neural Networks*, vol. 136, pp. 1–10, 2021, doi: 10.1016/j.neunet.2020.12.013.