# Spatiotemporal Pooling on Appropriate Topological Maps Represented as Two-Dimensional Images for EEG Classification

Takuto Fukushima[1] and Ryusuke Miyamoto[2]

[1] Department of Computer Science, Graduate School of Science and Technology
[2] Department of Computer Science, School of Science and Technology
Meiji University, Kawasaki, Japan
{taku,miya}@cs.meiji.ac.jp

**Abstract.** Motor imagery classification based on electroencephalography (EEG) signals is one of the most important brain-computer interface applications, although it needs further improvement. Several methods have attempted to obtain useful information from EEG signals by using recent deep learning techniques such as transformers. To improve the classification accuracy, this study proposes a novel EEG-based motor imagery classification method with three key features: generation of a topological map represented as a two-dimensional image from EEG signals with coordinate transformation based on t-SNE, use of the InternImage to extract spatial features, and use of spatiotemporal pooling inspired by PoolFormer to exploit spatiotemporal information concealed in a sequence of EEG images. Experimental results using the PhysioNet EEG Motor Movement/Imagery dataset showed that the proposed method achieved the best classification accuracy of 88.57%, 80.69%, and 70.20% on two-, three-, and four-class motor imagery tasks in cross-individual validation.

**Keywords:** motor imagery classification · spatiotemporal pooling · topological map generation · electroencephalography (EEG) signals

## 1 Introduction

Brain-computer interfaces (BCIs) are expected to enable humans to use their thoughts as inputs to control various types of devices. Motor imagery classification is one of the most important BCI applications; it is a task of classifying the imagined body movements based on the brain waves generated when imagining a specific physical movement. Motor imagery classification can be used for medical applications, including rehabilitation for stroke patients [32] and assistance for individuals unable to walk [7], as well as nonmedical fields, applications, supporting unmanned vehicles [58] and personal authentication [61]. To make BCIs practical in such applications, the motor imagery classification accuracy needs to be improved [11].

Motor imagery classification can be performed using various approaches such as electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI) [3]. EEG-based methods have become the most popular because of their portability, low cost, and high temporal resolution. In general, estimating long-range temporal correlation is useful for extracting relevant information from EEG signals [47, 48] that exhibit long-range dependencies. EEG classification has been improved further by capturing both temporal and spatial features [3]. Recently, models utilizing a transformer [46], known for its strength in handling long-range dependencies, have been applied for EEG classification and achieved state-of-the-art results [13, 53, 54]. The effectiveness of a transformer in motor imagery classification tasks with EEG signals has been demonstrated [54].

t-CTrans [54] takes an array of raw EEG data as the input, captures spatial features using a convolutional neural network (CNN), adds channel information of the EEG with positional encoding, and captures temporal correlations using a transformer encoder. Its structure enables the effective capture of both spatial and temporal features with state-of-the-art performance. However, the accuracy cannot be improved further because of two problems: representation of input data and model structure for temporal feature extraction. The first problem is the one-dimensional representation of the spatial information in input EEG data; it may not prevent the CNN to capture spatial features effectively owing to the lack of electrode coordinate information. The second problem is that the current implementation adopts the same structure as that of transformer encoder used in NLP, as described in [46], and it may not extract the temporal correlation of EEG signals sufficiently as temporal features over several two-dimensional informations are important.

To solve the first problem, we aim to make the spatial features obtained from the input data more effective. The information inputted from EEG signals can be considered as an image by appropriately projecting the arrangement of EEG electrodes into a two-dimensional plane; recent computer vision methods, namely InternImage [51] for extracting spatial features, and MixUp [60] and CutMix [59] for data augmentation, can be applied for this purpose. This representation called a topological map maintains the locations of EEG electrodes in two-dimensional coordinates and keeps the observed values at these coordinates as pixel values [3] in a two-dimensional image.

To generate a topological map represented as a two-dimensional image, the electrode locations represented by three-dimensional coordinates must be converted appropriately to two-dimensional ones. Our proposed method introduces t-SNE [35], a popular statistical nonlinear dimensionality reduction technique, because t-SNE has the property of transforming similar high-dimensional data into close distances in low-dimensional data. t-SNE has been applied for feature extraction [28, 33]; our study is the first attempt to apply it for coordinate transformation.

By performing appropriate coordinate transformation, we can expect to more accurately extract spatial features using computer vision methods. This study

introduces InternImage [51] to improve the accuracy of spatial feature extraction beyond those of conventional methods. This is because even though InternImage is a CNN-based method, it achieves higher accuracy than transformer-based methods in certain tasks. Additionally, similar to transformer-based methods like Vision Transformer [12] and Swin Transformer [30], InternImage can incorporate advancements such as LayerScale [12] to achieve effective training even as the model structure grows larger.

The second problem is the inadequate handling of temporal information. t-CTrans merely use a transformer encoder that is not appropriate approach when images are arranged in a time series. Therefore, we incorporated the concept of PoolFormer [57], in which correlations are evaluated using multi-head attention to adequately handle the structure in which spatial features are arranged in a time series. In contrast, our proposed method treats spatial features as one-dimensional vectors, arranges them into a two-dimensional plane corresponding to the spatiotemporal features when sequenced over time, and applies 2D pooling to this plane. By performing pooling on spatiotemporal features in this way, the proposed method is expected to appropriately handle the characteristics of temporal changes in spatial features.

In summary, the features of the proposed method include appropriately transforming the input format, accurately capturing spatial features using InternImage, and introducing spatiotemporal pooling (ST-pooling) inspired by PoolFormer to adequately capture the temporal changes in spatial features. To evaluate the effectiveness of the proposed method, experiments were conducted using the PhysioNet EEG Motor Movement/Imagery Dataset [17, 42]. Comparative experiments were performed to assess the coordinate transformation and ST-pooling introduced by the proposed method; this enabled clarifying the contribution of each component to the improvement in accuracy.

The rest of this paper is organized as follows. Section 2 introduces related work about EEG classification. Section 3 describes a novel method to apply ST-pooling to topological maps generated from EEG signals. Section 4 describes the details of the experimental setup. Section 5 presents a comparison of the proposed method with conventional methods to demonstrate its effectiveness. Finally, Section 6 concludes this paper.

## 2 Related Work

This section describes related research on EEG classification and explains input representation, coordinate transformation, and classification model because they are important when considering input data as images to apply computer vision techniques.

### 2.1 Input representation

EEG signals are captured as time-series signals from multiple electrodes simultaneously. To treat these signals as a sequence of two-dimensional images,

they must be reshaped. Existing image-based analysis methods for transforming EEG signals using common spatial patterns or unique methods for feature extraction before model input [31, 34, 37, 62], directly inputting raw data as arrays [13, 18, 27, 54], applying transformations like short-time Fourier transform or continuous wavelet transform to produce spectral images [20, 45, 52, 55], and creating topological maps based on electrode locations [5, 27, 29, 63].

However, existing methods have various issues. Many of them adopt feature extraction before model input, and they may not capture relevant information effectively. Approaches that use positional encoding [13, 54] or aligned channels for convolution [18, 27] may extract spatial information inappropriately. Spatial feature extraction remains problematic with some methods attempt that use spectral images, where different representation are applied to embed channel locations in a two-dimensional images, including arranging channels vertically and horizontally in an image [20,45,52], and representing EEG channels as input image channels [55].

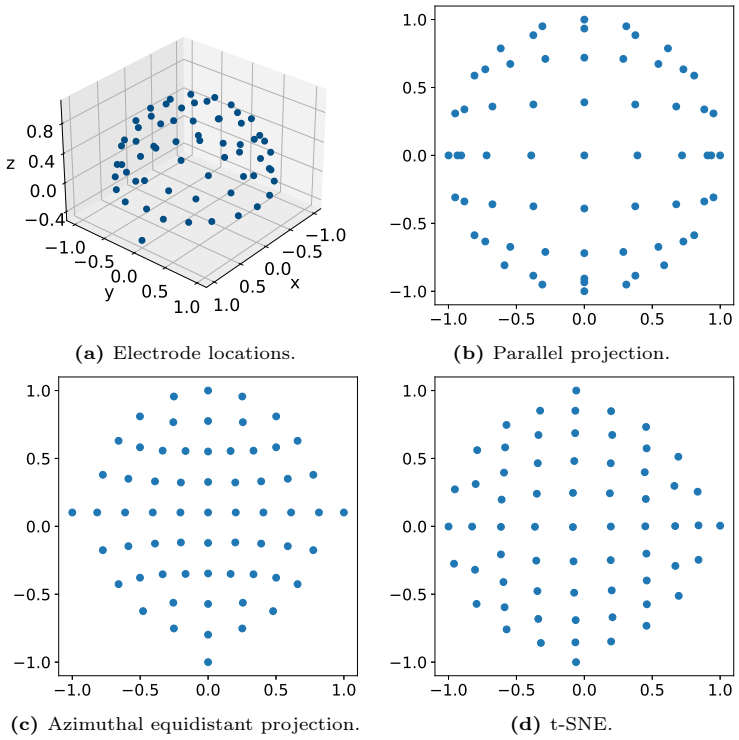
This study applies computer vision techniques that show remarkable performance in several deep learning tasks to topological maps represented as two-dimensional images and adopt as an input representation, where a pixel value and the coordinates represent an EEG signal and the electrode location, respectively.

### 2.2 Coordinate transformation

To represent input data as a sequence of images using topological maps, the electrode coordinates in three-dimensional space must be converted into two-dimensional coordinates. Existing methods employ manual determination [29,63] or azimuthal equidistant projection [5, 15, 44] for this conversion. Below, the use of t-SNE, the abovementioned method, and parallel projection for coordinate transformation is explained.

The parallel projection that is the simplest coordinate transformation is explained. This method involves simplifying the three-dimensional coordinates of EEG electrodes, represented in Euclidean coordinates such as $(x, y, z)$ by simply removing the $z-$component to project them onto a two-dimensional plane. The manual determination method is problematic because it requires redesigning when the number of EEG channels varies across datasets or when the resolution of the topological maps is changed. The azimuthal equidistant projection is commonly used for representing the world map, notably for depicting shortest flight paths. It accurately represents distances and angles from the center point but not distances between points far from the center; it may lose the relationships between electrodes that are inherently correlated. Therefore, our proposed method introduces t-SNE. t-SNE is a statistical method for visualizing high-dimensional data by performing nonlinear dimensionality reduction. It is designed such that similar data points are placed close to each other in a two-dimensional space, whereas dissimilar data points are positioned further apart.

Figure 1 shows the results of transformation to three-dimensional electrode locations. Figure 1b shows the result of applying parallel projection, where the




**(a)** Electrode locations. **(b)** Parallel projection.

**(c)** Azimuthal equidistant projection. **(d)** t-SNE.

**Fig. 1:** Three kinds of coordinate transformation that generates two-dimensional coordinates from three-dimensional ones of electrodes.

$z-$component is simply removed, leading to a concentration of electrode coordinates at the ends of the three-dimensional space where the vertical component is significant. Figure 1c shows the result of applying azimuthal equidistant projection, where the distortion increases with distance from the center. Figure 1d shows the result of applying t-SNE, where the electrodes are placed more evenly than in other methods.

### 2.3 Existing models for motor imagery classification based on EEG signals

Initially feature extraction methods with a support vector machine [39] and k-nearest neighbors [16] were predominantly used for motor imagery classification. However, similar to trends in image classification, improvements in deep learning accuracy have led to the adoption of methods based on deep learning that do not require feature extraction [9]. These methods utilize EEG data directly or without machine learning as an input to deep learning models. For example, there exist models that use only CNN [4, 25], only long short-term memory (LSTMs) [49], combinations of CNN and LSTM [56, 65]. CNN-based methods

are widely used for motor imagery classification; however they have limitations in adequately handling long-distance dependencies, and consequently, their accuracy can be improved to a limited extent [9,43]. Thus, similar to trends in NLP and general computer vision tasks, studies have started using transformer [46] for motor imagery classification following the introduction of the attention mechanism [54]. However, methods [23,54] focusing on the correlation of spatial features over time remain lacking, and their accuracy can be improved further.

## 3 ST-pooling with Appropriate Topological Maps

This study proposes a method for motor imagery classification based on EEG signals with three key features to improve the classification accuracy: generation of a topological map represented as a two-dimensional image from EEG signals, extraction of spatial features with InternImage, and ST-pooling inspired by PoolFormer.

### 3.1 Generation of a topological map

The primary feature of the proposed method is the generation of a topological map represented as a two-dimensional image from EEG signals. For this purpose, electrode locations that are originally expressed in three-dimensional coordinates must be appropriately represented in two-dimensional coordinates. Among the various methods described in Sec. 2, dimensionality reduction using t-SNE is adopted for coordinate transformation. It needs initial values to perform iterative computations. For determining the initial values, similar to the implementation in scikit-learn [38], the results of independent component analysis (ICA) on the set of three-dimensional electrode coordinates are used. However, whereas the original implementation intended for application to large datasets and therefore calculates singular values by performing singular value decomposition (SVD) on randomly selected samples during the ICA process, the implementation in this study performs SVD on all samples. This is because the number of samples is small, and acceleration through random sampling is unnecessary. Among the obtained singular values, the two largest ones are selected.

By applying t-SNE, the two-dimensional coordinates necessary for generating a topological map from the three-dimensional coordinates of the electrodes are obtained. These two-dimensional coordinates are not integer values. In addition, scaling is necessary to match the size of the input image to the subsequent processes. To obtain appropriate representations using two-dimensional images whose coordinates are integer values, interpolation similar to image interpolation is necessary: the nearest neighbor method was adopted because it showed the best accuracy in preliminary experiments. Finally, a topological map represented as a two-dimensional image from EEG signals is obtained: the number of channels is one, and the height $H$ and width $W$can be determined arbitrarily by scaling. This process is shown in Fig. 2.

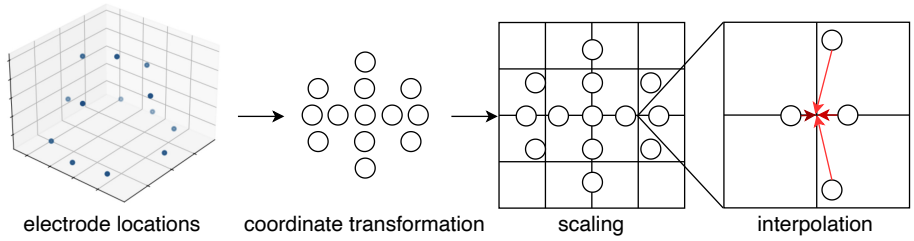



**Fig. 2:** Generation of a topological map.

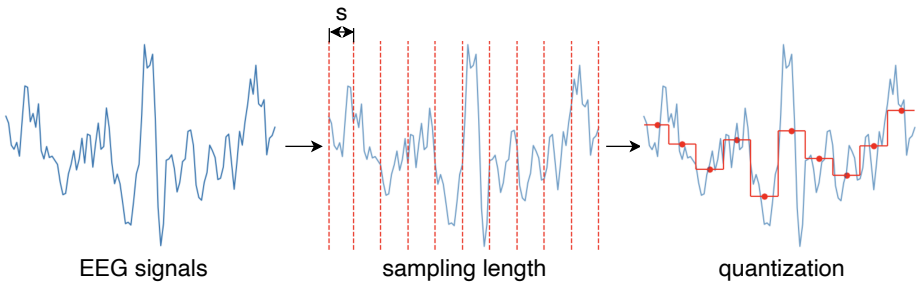



**Fig. 3:** Quantization process of EEG signals.

### 3.2 ST-pooling for EEG signals classification

By using the above process, EEG signals at a given moment can be treated as two-dimensional images. In this study, we propose a new model aimed at capturing the temporal changes in EEG signals with appropriate time-series images as inputs. However, considering the generally high sampling frequency of EEG signals, directly converting these signals into input images is not practical owing to the limitation of computational resources. Therefore, as shown in Fig. 3, quantization along the time direction is performed. In this process, the signals at each electrode are averaged over a fixed number of samples $s$, and this average is taken as the representative value for that interval.

We apply the proposed model to the spatiotemporal images generated in the described manner, as shown in Fig. 4. As shown on the left side of Fig. 4, the spatial features of each input frame are extracted through an InternImage process. However, to obtain the feature vectors used for subsequent ST-pooling, adaptive average pooling is applied followed by conversion into a one-dimensional format. This gives the following feature vector having length $L$:

$$L = 2^{stage-1} \times C_1, \tag{1}$$

where *stage* is the number of stages in InternImage, and $C_1$ is the parameter depending on the size of InternImage.

Positional encoding is performed to retain the temporal correlations of the feature vectors obtained from InternImage during the learning process. The widely used positional encoding in transformer-based methods, expressed by the following equation [46], is applied:

$$PE(n, 2l) = \sin\left(\frac{frame}{10000^{2l/L}}\right), and \tag{2}$$

$$PE(n, 2l+1) = \cos\left(\frac{frame}{10000^{2l/L}}\right), \tag{3}$$

where $n$ and $l$ are the frame number of an input sample and index of a one-dimensional feature vector, respectively.

Next, to extract spatiotemporal features, ST-pooling inspired by PoolFormer is performed. In this process, the spatial features outputted from InternImage are treated as one-dimensional vectors. These vectors are then arranged in chronological order for $N$ input frames, resulting in a two-dimensional plane, where $N$ is obtained by dividing the sampling length by $s$. The aim of this process is to extract spatio-temporal features by applying 2D pooling instead of applying multi-head attention applied as in previous transformer-based studies.

A previous study reported that Post-Norm leads to difficulties in training as models become deeper [50]; therefore, the proposed method, which employs a deep model structure in conjunction with InternImage, adopts Pre-Norm. As suggested in previous studies, to maintain the norm scale of the final output, a LayerNorm layer was added just before the final output. In addition to Pre-Norm, to address the challenges posed by the deep model structure, the LayerScale method that facilitates efficient learning in deep models, was introduced.

The flow of the proposed method with these processes is shown in Fig. 4. The input through the LayerNorm layer and then through a fully connected (FC) layer and the SoftMax function, resulting in the final output.

## 4 Experimental setup

This section describes the dataset and the approach to train the classifiers used in the evaluation.

### 4.1 Dataset

This study aims to minimize the user-dependent calibration incurred in constructing BCIs. Therefore, instead of within-individual validation, evaluations is conducted through cross-individual validation, where samples obtained from different subjects are used separately in training and testing sets.

PhysioNet EEG Motor Movement/Imagery Dataset and BCI Competition IV 2a [6] and 2b [26] are commonly used datasets for motor imagery classification. However, BCI Competition IV 2a and 2b have only nine subjects; this is too few samples for evaluating cross-individual validation. Therefore, the PhysioNet EEG Motor Movement/Imagery dataset was adopted.

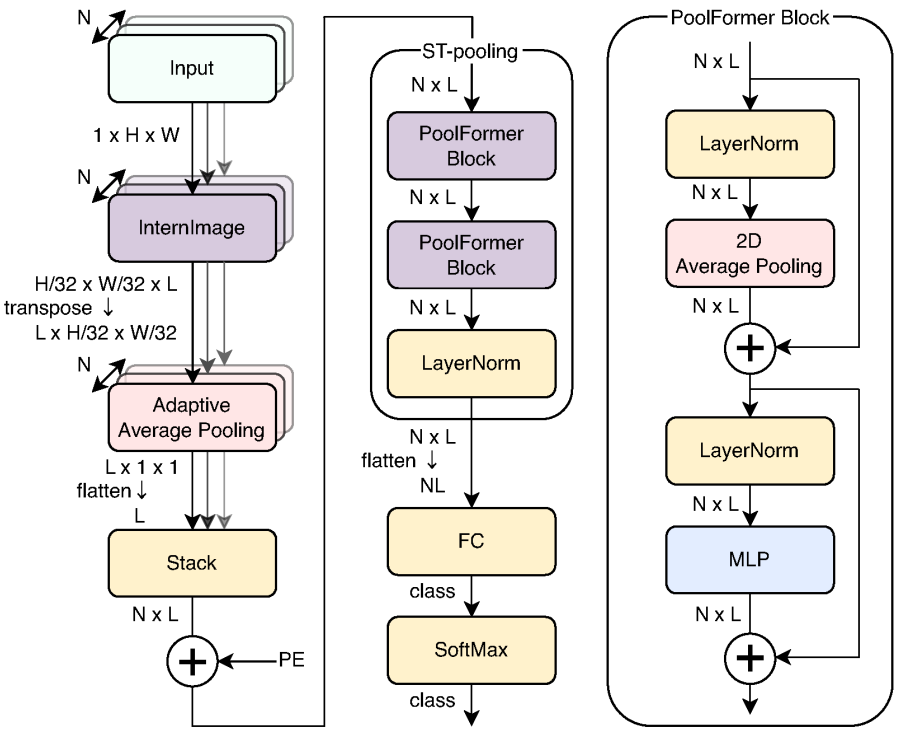



**Fig. 4:** Architecture of proposed model with ST-pooling. $N$, $H$, $W$, and $L$ means the number of input frames, height of topological maps, width of topological maps, and length of a feature vector from InternImage, respectively.

This dataset includes tasks involving motor imagery and actual movement. The EEG sampling frequency is 160Hz, and the electrode placement follows the 10-10 system. The number of subjects is 109, with each subject performing 14 trials in a single task session. The tasks are as follows:

- Baseline, eyes open (O)
- Three task runs for motor imagery of left fist (L)
- Three task runs for motor imagery of right fist (R)
- Three task runs for motor imagery of both feet (F)

Based on these labels as in previous studies [11, 54], the tasks were divided into three categories for accuracy measurement: four-class (L/R/O/F), three-class (L/R/O), and two-class (L/R) classification. However, owing to inaccuracies in the annotation data, the data from only 103 subjects, excluding subjects no. 38, 88, 89, 92, 100, and 104, were used in the experiments [40].

### 4.2 Cross-individual validation

Subjects were randomly divided into five blocks, with one block serving as the test data and the remaining four blocks used as training and validation data. The accuracy of the test data for all five blocks was determined, and their average was taken as the final accuracy. Off all subjects, 87.5% (i.e. 70% of total number of subjects) were randomly assigned to the training data, and the remaining subjects allocated to the validation data. The model that achieved the highest accuracy on the validation data was used for the test. This protocol is a standard one that has been used previously [40] except for model selection.

### 4.3 Data augmentation

Data augmentation is crucial for obtaining a good model. In this study, commonly used deep learning techniques such as MixUp [60] and CutMix [59] were applied. Furthermore, augmentation was performed by adding Gaussian noise to the EEG data used for creating topological maps. This is because noise can be introduced during the measurement of EEG signals. The noise used here has a mean of 0 and a standard deviation of 1, scaled down by $1e-4$.

RandAugment [10] was not adopted though it is known to be effective in image recognition tasks. This is because the data targeted in this study has the same format as images, in which values are arranged on a two-dimensional plane; however, the positional information of up, down, left, and right is shown as very important in motor imagery tasks [14], and it was considered necessary to preserve this positional information.

### 4.4 Training conditions

**Table 1:** Parameters for training.

| parameters | value |
|---|---|
| learning rate | $1e-4$ |
| batch size | 8 |
| drop rate (InternImage) | 0.0 |
| drop rate (ST-pooling) | 0.1 |
| drop path rate (InternImage) | 0.4 |
| drop path rate (ST-pooling) | 0.0 |

FractalDB [21] was used for the pre-training of InternImage. The rationale behind this choice was that pre-training with FractalDB achieved higher accuracy than pre-training with ImageNet [41] on some image classification datasets such as Places-365 [64] and Omniglot [24]. Additionally, because FractalDB consists of artificial images rather than natural images, it was deemed suitable for

the pre-training of topological maps like those used in this study. The batch size was set to 256, and the learning rate was set at 0.0005 for the pre-training process. Based on the obtained model's accuracy, the three- and four-class tasks were trained for 40 epochs, and the two-class task was trained for 30 epochs.

The training parameters used when training inference models with the targeted dataset are shown in Tab. 1. In our current implementation, the number of InternImage models included in the proposed method equals the number of frames in an input sample. As a result, a small batch size was used considering the limitation of computational resources, even though InternImage-S, one of the smallest implementations of InternImage, with *stage* and $C_1$ in Eq. (1) as 4 and 80, respectively. Adam [22] was used as the optimizer and training was conducted for 50 epochs.

## 5 Results and discussion

This section presents the experimental results and discusses about the effect of the components implemented in the proposed architecture.

### 5.1 Experimental results

Table 2 shows the classification of all tasks by the proposed model, t-CTrans, CNN, EEGNet Fusion, ConTraNet, and HCANN. The classification accuracy achieved by the proposed model reached 88.57%, 80.69%, and 70.20% on two-, three-, and four-class motor imagery tasks, respectively, in cross-individual validation. These values are better than those of all other state-of-the-art methods for this task. HCANN is the state-of-the-art method for BCI Competition IV 2a; however, its accuracy decreased when applied to the PhysioNet Dataset. This decline is likely due to the approximately tenfold increase in the number of subjects involved.

**Table 2:** Classification accuracy (%) in the PhysioNet Dataset in cross-individual classification. Only EEGNet Fusion used data whose length was 4s and HCANN was trained and tested by us using open-source code.

| Models | L/R | L/R/O | L/R/O/F |
|---|---|---|---|
| ours | **88.57** | **80.69** | **70.20** |
| t-CTrans(2022) [54] | 87.80 | 78.98 | 68.54 |
| CNN(2018) [11] | 87.98 | 76.61 | 65.73 |
| EEGNet Fusion(2020) [40] | 83.80 | - | - |
| ConTraNet(2022) [2] | 83.61 | 74.38 | 65.44 |
| HCANN(2024) [19] | 85.43 | 67.66 | 57.25 |

Next, the classification results were analyzed by considering the confusion matrix for the four-class task shown in Fig. 5. The accuracy was 0.74%, 0.71%,

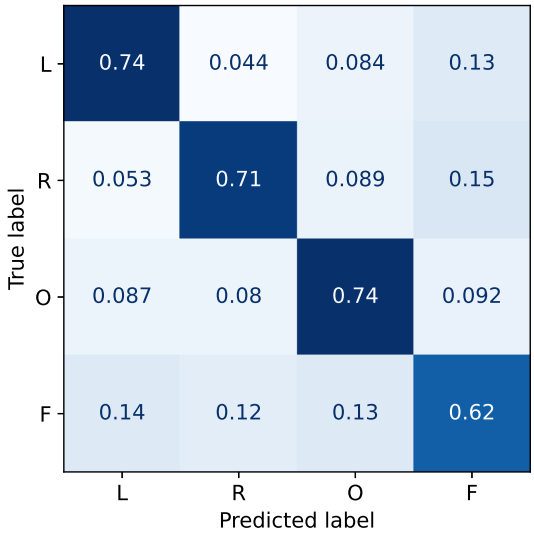



**Fig. 5:** Confusion matrix of four-class classification task.

0.74%, and 0.62% for the estimation of L, R, O, and F, respectively. The accuracy values for L, R, and O were similar, and that for F became was slightly lower. This was attributed to the same cause reported previously [8,14]: EEG signals for the estimation of task F do not have significant differences from those for other tasks, although their intensity increased for tasks O and changed drastically for tasks L and R. To further improve the performance of this task, we must construct a classifier that can extract hidden features in the current representation of input signals.

### 5.2 Preliminary experiments

The proposed model includes several parameters to be defined before the quantitative evaluation. These parameters are determined by considering the results of the preliminary experiments described below.

First, the number of frames $N$ for an input sample whose length was six seconds was determined. Table 3 shows the classification accuracy of the four-class motor imagery task in cross-individual validation. Considering these results, $N$ was set to 60.

Next, we compare the pre-training datasets: FractalDB and ImageNet. Table 4 shows the classification accuracy of the four-class motor imagery task in cross-individual validation. Considering these results, Fractal DB was adopted for pre-training.

**Table 3:** Classification accuracy for several frame sizes.

| N | L/R/O/F |
|---|---|
| 30 | 69.61 |
| 60 | **70.20** |
| 96 | 69.60 |
| 120 | 68.41 |

**Table 4:** Classification accuracy when different datasets were used for pre-training.

| Dataset | L/R/O/F |
|---|---|
| FractalDB-1k | **70.20** |
| ImageNet-1k | 69.16 |

**Table 5:** Classification accuracy when the interpolation method was changed.

| Method | L/R/O/F |
|---|---|
| nearest neighbor | **70.20** |
| Clough-Tocher [1] | 69.29 |

Finally, the interpolation method used in the generation of topological maps was chosen. Table 5 shows the classification accuracy of the four-class motor imagery task in cross-individual validation when the interpolation method was changed. The results showed that simple interpolation based on the nearest neighbor method was better than the sophisticated method widely used in EEG signals classification.

### 5.3 Effect of proposed components

**Table 6:** Classification accuracy when topological maps generation method was changed.

| Method | L/R/O/F |
|---|---|
| parallel projection | 69.30 |
| azimuthal equidistant | 68.71 |
| t-SNE | **70.20** |
| UMAP [36] | 68.42 |

**Table 7:** Classification accuracy when the model structure was changed.

| Method | L/R/O/F |
|---|---|
| ST-pooling | **70.20** |
| Multi-Head Attention | 69.88 |
| PoolFormer [57] | 69.74 |

To analyze the effect of using the proposed method to generate topological maps from EEG signals, the classification accuracy achieved using the different methods is shown in Tab. 6. When using the method based on t-SNE, the accuracy improved by 0.90% and 1.49% compared to those of simple parallel projection and azimuthal equidistant projection, respectively. The results showed that t-SNE is more accurate than UMAP [36], another nonlinear dimensionality reduction method, and that t-SNE is suitable for generating topological maps for motor imagery classification from EEG signals.

To verify the effect of ST-pooling, a quantitative comparison with the multi-head attention used in existing methods and PoolFormer was conducted: the ST-pooling block was replaced with multi-head attention and PoolFormer takes spatiotemporal images after InternImage and the stacking process in the same way as in the proposed method. Table 7 shows the classification accuracy.

Replacing multi-head attention with an ST-pooling layer resulted in an accuracy improvement of 0.32%. The accuracy with ST-pooling was better than PoolFormer by 0.46% even though ST-pooling was inspired by PoolFormer [57]. These results showed that ST-pooling successfully captured the appropriate spatiotemporal correlations concealed in topological maps for an input sample generated from successive EEG signals for this task.

## 6 Conclusion

This study proposed a novel method for motor imagery classification based on EEG signals with three key features for improving accuracy: topological map generation with appropriate coordinate transform, use of InternImage to extract spatial features, and use of ST-pooling to treat spatial and temporal features concealed in a sequence of EEG images. To realize appropriate coordinate transformation, t-SNE, which was originally proposed for nonlinear dimensionality reduction was adopted instead of other simpler methods and other dimensionality reduction methods. ST-pooling was designed by replacing the multi-head attention in PoolFormer to exploit the spatiotemporal information concealed in a sequence of EEG images.

Experimental results using the PhysioNet EEG Motor Movement/Imagery dataset showed that the proposed method improved the classification accuracy to 88.57%, 80.69%, and 70.20% on two-, three-, and four-class motor imagery tasks, respectively, in cross-individual validation; all of these are better than results obtained using existing state-of-the-art methods.

This study showed that the proposed method's functions to treat spatiotemporal features in a sequence of EEG signals represented using topological maps were effective for motor imagery classification. To further demonstrate the capability of the proposed method, we will apply it to other tasks using EEG signals and improve its performance in the future.

## References


1. Alfeld, P.: A trivariate Clough—Tocher scheme for tetrahedral data. Comput. Aided Geom. Des. **1**(2), 169–181 (1984). https://doi.org/https://doi.org/10.1016/0167-8396(84)90029-3
2. Ali, O., Saif-ur Rehman, M., Glasmachers, T., Iossifidis, I., Klaes, C.: ConTraNet: A single end-to-end hybrid network for EEG-based and EMG-based human machine interfaces. arXiv preprint arXiv:2206.10677 (2022). https://doi.org/10.48550/arXiv.2206.10677
3. Altaheri, H., Muhammad, G., Alsulaiman, M., Amin, S.U., Altuwaijri, G.A., Abdul, W., Bencherif, M.A., Faisal, M.: Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: a review. Neural Comput. Appl. **35**(20), 14681–14722 (2023). https://doi.org/10.1007/s00521-021-06352-5
4. Amin, S.U., Alsulaiman, M., Muhammad, G., Bencherif, M.A., Hossain, M.S.: Multilevel Weighted Feature Fusion Using Convolutional Neural Networks for EEG Motor Imagery Classification. IEEE Access **7**, 18940–18950 (2019). https://doi.org/10.1109/ACCESS.2019.2895688
5. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks. In: ICLR. pp. 1–15 (2016)
6. Brunner, C., Leeb, R., Müller-Putz, G., Schlögl, A., Pfurtscheller, G.: BCI Competition 2008–Graz data set A. Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology **16**, 1–6 (2008)
7. Carlson, T., del R. Millan, J.: Brain-Controlled Wheelchairs: A Robotic Architecture. IEEE Robot. Autom. Mag. **20**(1), 65–73 (2013). https://doi.org/10.1109/MRA.2012.2229936
8. Cho, H., Ahn, M., Ahn, S., Kwon, M., Jun, S.C.: EEG datasets for motor imagery brain-computer interface. GigaScience **6**(7), gix034 (2017). https://doi.org/10.1093/gigascience/gix034
9. Craik, A., He, Y., Contreras-Vidal, J.L.: Deep learning for electroencephalogram (EEG) classification tasks: a review. J. Neural Eng. **16**(3), 031001 (2019). https://doi.org/10.1088/1741-2552/ab0ab5
10. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: CVPR. pp. 702–703 (2020)
11. Dose, H., Møller, J.S., Iversen, H.K., Puthusserypady, S.: An end-to-end deep learning approach to MI-EEG signal classification for BCIs. Expert Syst. Appl. **114**, 532–542 (2018). https://doi.org/10.1016/j.eswa.2018.08.031
12. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: ICLR (2021)
13. Du, Y., Xu, Y., Wang, X., Liu, L., Ma, P.: EEG temporal–spatial transformer for person identification. Sci. Rep. **12**(1), 14378 (2022). https://doi.org/10.1038/s41598-022-18502-3
14. Ehrsson, H.H., Geyer, S., Naito, E.: Imagery of Voluntary Movement of Fingers, Toes, and Tongue Activates Corresponding Body-Part-Specific Motor Representations. J. Neurophysiol. **90**(5), 3304–3316 (2003). https://doi.org/10.1152/jn.01113.2002
15. Fadel, W., Kollod, C., Wahdow, M., Ibrahim, Y., Ulbert, I.: Multi-Class Classification of Motor Imagery EEG Signals Using Image-Based Deep Recurrent Con-

volutional Neural Network. In: BCI. pp. 1–4 (2020). https://doi.org/10.1109/BCI48061.2020.9061622
16. Farooq, F., Rashid, N., Farooq, A., Ahmed, M., Zeb, A., Iqbal, J.: Motor Imagery based Multivariate EEG Signal Classification for Brain Controlled Interface Applications. In: ICOM. pp. 1–6. IEEE (2019). https://doi.org/10.1109/ICOM47790.2019.8952008
17. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals. Circulation **101**(23), e215–e220 (2000). https://doi.org/10.1161/01.CIR.101.23.e215
18. Jeong, J.H., Lee, B.H., Lee, D.H., Yun, Y.D., Lee, S.W.: EEG Classification of Forearm Movement Imagery Using a Hierarchical Flow Convolutional Neural Network. IEEE Access **8**, 66941–66950 (2020). https://doi.org/10.1109/ACCESS.2020.2983182
19. Ji, Y., Li, F., Fu, B., Zhou, Y., Wu, H., Li, Y., Li, X., Shi, G.: A novel hybrid decoding neural network for EEG signal representation. Pattern Recognition p. 110726 (2024). https://doi.org/https://doi.org/10.1016/j.patcog.2024.110726
20. Kant, P., Laskar, S.H., Hazarika, J., Mahamune, R.: CWT Based Transfer Learning for Motor Imagery Classification for Brain computer Interfaces. J. Neurosci. Methods **345**, 108886 (2020). https://doi.org/10.1016/j.jneumeth.2020.108886
21. Kataoka, H., Okayasu, K., Matsumoto, A., Yamagata, E., Yamada, R., Inoue, N., Nakamura, A., Satoh, Y.: Pre-training without Natural Images. IJCV **130**(4), 990–1007 (2022). https://doi.org/10.1007/s11263-021-01555-8
22. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization (2015)
23. Kostas, D., Aroca-Ouellette, S., Rudzicz, F.: BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. Front. Hum. Neurosci. **15**, 653659 (2021). https://doi.org/10.3389/fnhum.2021.653659
24. Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B.: Human-level concept learning through probabilistic program induction. Science **350**(6266), 1332–1338 (2015). https://doi.org/10.1126/science.aab3050
25. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. J. Neural Eng. **15**(5), 056013 (2018). https://doi.org/10.1088/1741-2552/aace8c
26. Leeb, R., Brunner, C., Müller-Putz, G., Schlögl, A., Pfurtscheller, G.: BCI Competition 2008–Graz data set B. Graz University of Technology, Austria **16**, 1–6 (2008)
27. Li, M.A., Han, J.F., Duan, L.J.: A Novel MI-EEG Imaging With the Location Information of Electrodes. IEEE Access **8**, 3197–3211 (2019). https://doi.org/10.1109/ACCESS.2019.2962740
28. Li, M.a., Luo, X.y., Yang, J.f.: Extracting the nonlinear features of motor imagery EEG using parametric t-SNE. Neurocomputing **218**, 371–381 (2016). https://doi.org/10.1016/j.neucom.2016.08.083
29. Liu, T., Yang, D.: A Densely Connected Multi-Branch 3D Convolutional Neural Network for Motor Imagery EEG Decoding. Brain Sci. **11**(2), 197 (2021). https://doi.org/10.3390/brainsci11020197

30. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In: ICCV. pp. 9992–10002 (2021). https://doi.org/10.1109/ICCV48922.2021.00986
31. Luo, T.j., Zhou, C.l., Chao, F.: Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network. BMC Bioinf. **19**(1) (2018). https://doi.org/10.1186/s12859-018-2365-1
32. López-Larraz, E., Sarasola-Sanz, A., Irastorza-Landa, N., Birbaumer, N., Ramos-Murguialday, A.: Brain-machine interfaces for rehabilitation in stroke: A review. NeuroRehabilitation **43**(1), 77–97 (2018). https://doi.org/10.3233/nre-172394
33. Ma, Q., Wang, M., Hu, L., Zhang, L., Hua, Z.: A Novel Recurrent Neural Network to Classify EEG Signals for Customers' Decision-Making Behavior Prediction in Brand Extension Scenario. Front. Hum. Neurosci. **15**, 610890 (2021). https://doi.org/10.3389/fnhum.2021.610890
34. Ma, X., Qiu, S., Wei, W., Wang, S., He, H.: Deep Channel-Correlation Network for Motor Imagery Decoding From the Same Limb. IEEE Trans. Neural Syst. Rehabil. Eng. **28**(1), 297–306 (2020). https://doi.org/10.1109/TNSRE.2019.2953121
35. Van der Maaten, L., Hinton, G.: Visualizing Data using t-SNE. JMLR **9**(86), 2579–2605 (2008)
36. McInnes, L., Healy, J., Saul, N., Großberger, L.: UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. **3**(29), 861 (2018). https://doi.org/10.21105/joss.00861
37. Olivas-Padilla, B.E., Chacon-Murguia, M.I.: Classification of multiple motor imagery using deep convolutional neural networks and spatial filters. Appl. Soft Comput. **75**, 461–472 (2019). https://doi.org/10.1016/j.asoc.2018.11.031
38. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python. JMLR **12**, 2825–2830 (2011)
39. Planelles, D., Hortal, E., Costa, Á., Úbeda, A., Iáñez, E., Azorín, J.M.: Evaluating Classifiers to Detect Arm Movement Intention from EEG Signals. Sensors **14**(10), 18172–18186 (2014). https://doi.org/10.3390/s141018172
40. Roots, K., Muhammad, Y., Muhammad, N.: Fusion Convolutional Neural Network for Cross-Subject EEG Motor Imagery Classification. Computers **9**(3), 72 (2020). https://doi.org/10.3390/computers9030072
41. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. IJCV **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
42. Schalk, G., McFarland, D., Hinterberger, T., Birbaumer, N., Wolpaw, J.: BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans. Biomed. Eng. **51**(6), 1034–1043 (2004). https://doi.org/10.1109/TBME.2004.827072
43. Song, Y., Jia, X., Yang, L., Xie, L.: Transformer-based Spatial-Temporal Feature Learning for EEG Decoding. arXiv preprint arXiv:2106.11170 (2021). https://doi.org/10.48550/arXiv.2106.11170
44. Sun, J., Cao, R., Zhou, M., Hussain, W., Wang, B., Xue, J., Xiang, J.: A hybrid deep neural network for classification of schizophrenia using EEG Data. Sci. Rep. **11**(1), 4706 (2021)
45. Tabar, Y.R., Halici, U.: A novel deep learning approach for classification of EEG motor imagery signals. J. Neural Eng **14**(1), 016003 (2016). https://doi.org/10.1088/1741-2560/14/1/016003

46. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is All you Need. In: NeurIPS. vol. 30 (2017)
47. Wairagkar, M., Hayashi, Y., Nasuto, S.J.: Modeling the Ongoing Dynamics of Short and Long-Range Temporal Correlations in Broadband EEG During Movement. Front. Syst. Neurosci. **13**, 66 (2019). https://doi.org/10.3389/fnsys.2019.00066
48. Wairagkar, M., Hayashi, Y., Nasuto, S.J.: Dynamics of Long-Range Temporal Correlations in Broadband EEG During Different Motor Execution and Imagery Tasks. Front. Neurosci. **15** (2021). https://doi.org/10.3389/fnins.2021.660032
49. Wang, P., Jiang, A., Liu, X., Shang, J., Zhang, L.: LSTM-Based EEG Classification in Motor Imagery Tasks. IEEE Trans. Neural Syst. Rehabil. Eng. **26**(11), 2086–2095 (2018). https://doi.org/10.1109/TNSRE.2018.2876129
50. Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D.F., Chao, L.S.: Learning Deep Transformer Models for Machine Translation. In: ACCL (2019). https://doi.org/10.18653/v1/p19-1176
51. Wang, W., Dai, J., Chen, Z., Huang, Z., Li, Z., Zhu, X., Hu, X., Lu, T., Lu, L., Li, H., et al.: InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions. In: CVPR. pp. 14408–14419 (2023). https://doi.org/10.1109/cvpr52729.2023.01385
52. Wang, Z., Cao, L., Zhang, Z., Gong, X., Sun, Y., Wang, H.: Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition. Concurrency Comput. Pract. Exp. **30**(23), e4413 (2018). https://doi.org/10.1002/cpe.4413
53. Wei, Y., Liu, Y., Li, C., Cheng, J., Song, R., Chen, X.: TC-Net: A Transformer Capsule Network for EEG-based emotion recognition. Comput. Biol. Med. **152**, 106463 (2023). https://doi.org/10.1016/j.compbiomed.2022.106463
54. Xie, J., Zhang, J., Sun, J., Ma, Z., Qin, L., Li, G., Zhou, H., Zhan, Y.: A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification. IEEE Trans. Neural Syst. Rehabil. Eng. **30**, 2126–2136 (2022). https://doi.org/10.1109/TNSRE.2022.3194600
55. Xu, B., Zhang, L., Song, A., Wu, C., Li, W., Zhang, D., Xu, G., Li, H., Zeng, H.: Wavelet Transform Time-Frequency Image and Convolutional Network-Based Motor Imagery EEG Classification. IEEE Access **7**, 6084–6093 (2018). https://doi.org/10.1109/ACCESS.2018.2889093
56. Yang, J., Yao, S., Wang, J.: Deep Fusion Feature Learning Network for MI-EEG Classification. IEEE Access **6**, 79050–79059 (2018). https://doi.org/10.1109/ACCESS.2018.2877452
57. Yu, W., Luo, M., Zhou, P., Si, C., Zhou, Y., Wang, X., Feng, J., Yan, S.: MetaFormer is Actually What You Need for Vision. In: CVPR. pp. 10819–10829 (2022). https://doi.org/10.1109/cvpr52688.2022.01055
58. Yu, Y., Zhou, Z., Yin, E., Jiang, J., Tang, J., Liu, Y., Hu, D.: Toward brain-actuated car applications: Self-paced control with a motor imagery-based brain-computer interface. Comput. Biol. Med. **77**, 148–155 (2016). https://doi.org/10.1016/j.compbiomed.2016.08.010
59. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. In: ICCV. pp. 6023–6032 (2019)
60. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond Empirical Risk Minimization. In: ICLR (2017)

61. Zhang, X., Yao, L., Kanhere, S.S., Liu, Y., Gu, T., Chen, K.: MindID: Person Identification from Brain Waves through Attention-based Recurrent Neural Network. IMWUT **2**(3), 1–23 (2018). https://doi.org/10.1145/3264959
62. Zhao, X., Zhao, J., Liu, C., Cai, W.: Deep Neural Network with Joint Distribution Matching for Cross-Subject Motor Imagery Brain-Computer Interfaces. BioMed Res. Int. **2020**, 1–15 (2020). https://doi.org/10.1155/2020/7285057
63. Zhao, X., Zhang, H., Zhu, G., You, F., Kuang, S., Sun, L.: A Multi-Branch 3D Convolutional Neural Network for EEG-Based Motor Imagery Classification. IEEE Trans. Neural Syst. Rehabil. Eng. **27**(10), 2164–2177 (2019). https://doi.org/10.1109/TNSRE.2019.2938295
64. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 Million Image Database for Scene Recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(6), 1452–1464 (2018). https://doi.org/10.1109/TPAMI.2017.2723009
65. Zhu, K., Wang, S., Zheng, D., Dai, M.: Study on the effect of different electrode channel combinations of motor imagery EEG signals on classification accuracy. J. Eng. **2019**(23), 8641–8645 (2019). https://doi.org/10.1049/joe.2018.9073