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

Filter Bank Sinc-ShallowNet with EMD-based Mixed Noise Adding Data Augmentation for Motor Imagery Classification

Jiaming Chen[1], Weibo Yi[2] and Dan Wang[1*]

Abstract— **Motor imagery-based brain computer interface (MI-BCI) is a representative active BCI paradigm which is widely employed in the rehabilitation field. In MI-BCI, a classification model is built to identify the target limb from MI-based EEG signals, but the performance of models cannot meet the demand for practical use. Lightweight neural networks in deep learning methods are used to build high performance models in MI-BCI. Small sample sizes and the lack of multi-scale information extraction in frequency domain limit the performance improvement of lightweight neural networks. To solve these problems, the Filter Bank Sinc-ShallowNet (FB-Sinc-ShallowNet) algorithm combined with the mixed noise adding method based on empirical mode decomposition (EMD) was proposed. The FB-Sinc-ShallowNet algorithm improves a lightweight neural network Sinc-ShallowNet with a filter bank structure corresponding to four sensory motor rhythms. The mixed noise adding method employs the EMD method to improve the quality of generated data. The proposed method was evaluated on the BCI competition IV IIa dataset and can achieve highest average accuracy of 77.2%, about 6.34% higher than state-of-the-art method Sinc-ShallowNet. This work implies the effectiveness of filter bank structure in lightweight neural networks and provides a novel option for data augmentation and classification of MI-based EEG signals, which can be applied in the rehabilitation field for decoding MI-EEG with few samples.**

I. INTRODUCTION

Brain computer interface (BCI) is a communication system between the human brain and external devices without the peripheral nerves and muscles [1]. MI-BCI is the representative paradigm of BCI and can output the movement intention without external stimulation [2]. The classification of MI-based EEG (MI-EEG) signals is the main topic in MI-BCI studies. A classification model that can detect the target limb of the subject is trained by features extracted from the MI-EEG data. However, the classification accuracy of existing methods is not high enough for practical use. Building a high-performance model is still the challenge of MI-BCI studies.

Machine learning methods and deep learning methods are frequently used for feature extraction and classification in MI-BCI. In machine learning methods, Common Spatial Patterns (CSP) [3], Filter Bank Common Spatial Patterns (FBCSP) [4] and Riemannian geometry-based methods [5] are frequently employed because of their stability. In recent years, researchers have found that deep learning methods can get stronger classification performance than machine learning methods [6], and feature engineering is not required so raw EEG signals can be used as the input [7]. For intra-subject studies in MI-BCI with deep learning methods, lightweight neural networks [8] have high decoding performance because they have simple structure, which can alleviate the overfitting problem. For example, the ShallowConvNet [8] and the Sinc-ShallowNet [9] algorithm are lightweight neural networks and can achieve 71% and 72.8% mean accuracy respectively on the BCI Competition IV IIa dataset [10] with four target limbs (including left hand, right hand, both feet and tongue). However, few samples and the lack of multi-scale information extraction in frequency domain limit the performance of lightweight neural networks.

Few samples (or insufficient sample sizes) can cause the overfitting problem that limits the performance of lightweight neural networks. To solve this problem, data augmentation methods commonly used in deep learning can help increase the sample size of MI-EEG data to meet the large data demand of deep learning methods [11]. Window slicing, noise adding and generative adversarial networks (GAN) are commonly used data augmentation methods in MI-BCI. Window slicing is suitable for processing raw EEG signals [8], but the effect is not significant when accompanied with lightweight neural networks [8]. Noise adding is frequently used with lightweight neural networks and can augment raw EEG [12]. GAN is often used with the time-frequency representation of EEG signals [13]. Besides, the EMD method [14] is also used for data augmentation of time-frequency representation of MI-EEG signals by switching the intrinsic mode function (IMF) components acquired from EMD [15]. According to aforementioned data augmentation methods, the noise adding method is suitable for data augmentation in lightweight neural networks that process raw EEG signals. However, directly adding noise to MI-EEG signals that have low SNR will generate trials with lower SNR, which could limit the effect of data augmentation.

Additionally, the input of lightweight neural networks is usually MI-EEG signals filtered by a bandpass filter with one frequency band that covers multiple motor rhythms. The lack of the capability of extracting information from different rhythms could limit the decoding performance because the specific frequency band could be different for different types of movement imagination. The filter bank structure that corresponding to different motor rhythms can help solve this problem by extracting useful features from different tasks.

Therefore, in this work, we proposed a lightweight neural

This work was supported by the National Natural Science Foundation of China under Grand 62006014.

* indicates the corresponding author. wangdan@bjut.edu.cn
[1]Faculty of Information Technology, Beijing University of Technology, Beijing, China
[2]Beijing Machine and Equipment Institute, Beijing, China

network FB-Sinc-ShallowNet to improve Sinc-ShallowNet with a filter bank corresponding to different rhythms to increase the decoding accuracy. Accompanied with FB-Sinc-ShallowNet, we designed the EMD-based mixed noise adding data augmentation method to solve the few samples problem. The mean accuracy of the proposed method on the BCI Competition IV IIa dataset is 77.2%, about 6.34% higher than that of the state-of-the-art method Sinc-ShallowNet. Moreover, the accuracy of FB-Sinc-ShallowNet combined with the proposed data augmentation method is 1.84% higher than that of FB-Sinc-ShallowNet with other data augmentation methods. These indicate that proposed methods can improve the decoding performance of the MI-EEG classification model. The structure of this paper is as follows: Section 2 describes the proposed method in detail. The results are presented and discussed in Section 3. Section 4 is the conclusion of this paper.

II. METHOD

In this section, we introduce the Euclidean alignment (EA) preprocessing method, EMD-based mixed noise adding data augmentation (DA) method and the FB-Sinc-ShallowNet algorithm. The structure of the three methods is shown in Fig. 1. The proposed methods were evaluated on the BCI Competition IV IIa dataset [10] that has two sessions. Raw EEG signals in each session were filtered by a third-order Butterworth band-pass filter in the 4-40hz frequency band and standardized with an exponential moving average window with a decay factor of 0.999. Trials were extracted from 0.5 s to 2.5 s after the cue. Trials and labels are denoted as a matrix $\boldsymbol{X}$ with dimension $N \times E \times T$ and a vector $\boldsymbol{y}$ with length N respectively, where $N(N = 288)$ is the number of trials, $E(E = 22)$ is the number of channels, and $T(T = 500)$ is the number of sampling points. In session 1, 80% and 20% of trials were used as the training set and validation set respectively. All trials in session 2 were used as the test set.

A. Euclidean Alignment

The Euclidean alignment method is used to preprocess the MI-EEG signal to improve the decoding performance by transforming the mean covariance matrix of MI-EEG signals to an identity matrix [16]. Given the trials $\boldsymbol{X}$ of a subject, the arithmetic mean of the covariance matrix of N trials are computed to construct a transformation matrix shown in (1).

$$\begin{aligned} \bar{\boldsymbol{X}} &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i\boldsymbol{X}_i^T \\ \boldsymbol{P}_i &= \bar{\boldsymbol{X}}^{-\frac{1}{2}}\boldsymbol{X}_i \end{aligned} \tag{1}$$

For intra subject experiments in this work, MI-EEG data are collected from the same subject but at different time. Euclidean alignment can help reduce the difference on features of data from the two sessions.

B. EMD-based Mixed Noise Adding Data Augmentation

To improve the quality of generated trials in the noise adding method, we proposed a mixed noise adding data augmentation method based on EMD [14]. The method includes the following four steps.

1) EMD and IMF Components Selecting: The original signal is decomposed by EMD to get the principal components (called IMF components) of the signal. The signal $X_{i,e}$ of trial i on channel e, denoted as S, can be decomposed into J IMF components and the residue r. Then, the correlation coefficients of signal S and every IMF component are computed. The IMF components with correlation coefficients less than 0.1 are dropped. The number of retained IMF components is J'.

2) Estimating the Energy of IMF Components: The energy of the first IMF component is estimated by (2).

$$\hat{E}_1 = \sum_{t=1}^{T} IMF^2(t) \tag{2}$$

Then, the energy of other IMF components are estimated by (3), where H is the Hurst index, and β_H and ρ_H are parameters that vary with H. In practice, $H = 0.5$, $\beta_H = 0.719$, and $\rho_H = 0.201$ [17].

$$\hat{E_{j'}} = \frac{\hat{E}_1}{\beta_H}\rho_H^{-2(1-H)j'}, j' \in \{2, 3, \cdots, J'\} \tag{3}$$

3) IMF Components Filtering with Adaptive Thresholds: After estimating the energy of IMF components, each IMF component is filtered by two steps. First, the IMF component j' is divided into P segments arranged in ascending order according to their extrema. The extremum of a segment is denoted as $e_p^{(j')}\,(p \in \{1, 2, \ldots, P\})$ whose energy is $E_p^{(j')} = (e_p^{(j')})^2$. Next, the cumulative sum of energy is computed in ascending order until (4) is satisfied.

$$\sum_{p}^{q-1} E_p^{(j')} \le \hat{E_{j'}} < \sum_{p}^{q} E_p^{(j')} \tag{4}$$

The absolute value of the extremum of segment q is taken as the adaptive threshold $T_{j'} = \left|e_q^{(j')}\right|$. Then, according to the rule shown in (5), the signal in each segment is filtered where $IMF_{j',p}$ is segment p of IMF component j', and $e_{j',p}$ is the extremum of segment p.

$$IMF_{j',p} = \begin{cases} 0, & e_{j',p} < T_{j'} \\ IMF_{j',p}, & e_{j',p} \ge T_{j'} \end{cases} \tag{5}$$

After filtering all J' IMF components based on the adaptive threshold, the sum of all IMF components are computed to reconstruct signal S'.

4) Mixed Noise Generating: After extracting the main information S' of the original signal, a white noise sequence with a specific SNR is generated based on S'. Let $SNR = s$, and the energy of white noise E_{noise} can be computed by (6).

$$E_{noise} = \frac{E_{signal}}{10^{s/10}} = \frac{\frac{1}{T}\sum_{t=1}^{T} S'(t)^2}{10^{s/10}} \tag{6}$$

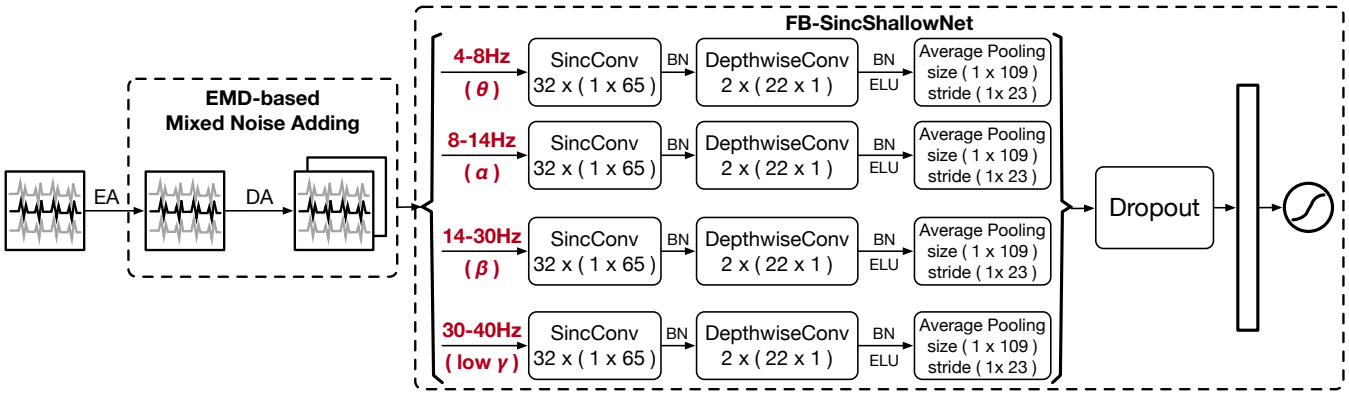

Fig. 1. Framework of employed methods

Then, a white noise sequence $\boldsymbol{a} = [a_0, a_1, \ldots, a_T]$ with length T, zero mean and standard deviation σ is multiplied by E_{noise} and then added to the signal S' to get the mixed noise sequence $\boldsymbol{a_{gen}}$. Note that the SNR s and the standard deviation σ need to be adjusted. Here, they are set to $\sigma = 0.02, s = 1dB$. The mixed noise is added to the original signal after generated for data augmentation. After training on the larger dataset, the decoding performance of the model can be improved.

C. FB-Sinc-ShallowNet

In order to extract features from different rhythms to further improve the decoding performance, we designed the FB-Sinc-ShallowNet algorithm to improve Sinc-ShallowNet. The structure of FB-Sinc-ShallowNet is shown in Fig. 1. Four groups of feature maps are extracted from four branches, concatenated along the channel axis, regularized by a dropout layer [18] with dropout rate set to 0.5, and finally classified by a SoftMax function with the maximum norm constraint of 0.5 after flattened. Layers in each branch is described below.

1) SincConv layer: A sinc convolutional (SincConv) layer performs a temporal convolution between the input signal and a Finite Impulse Response (FIR) filter to extract more meaningful features [19]. Concretely, a SincConv layer is defined as (7),

$$y[T] = x[T] * h(T, f_l, f_h) \tag{7}$$

where f_l (the lower frequency bound) and f_h (the upper frequency bound) are trainable parameters, $x[T]$ is a signal with length T, $y[T]$ is a filtered signal with length T, and $h(T, f_l, f_h) = 2f_h sinc(2\pi f_h T) - 2f_l sinc(2\pi f_l T)$ is the convolutional kernel where $sinc(\cdot)$ denotes the sinc function $sinc(x) = sin(x)/x$. The Hamming window function is introduced into the kernel function to produce a better band-pass filter, as shown in (8), where L is the length of the convolution kernel in time domain.

$$h[T, f_l, f_h] := h[T, f_l, f_h] \cdot [0.54 - 0.46cos(\frac{2\pi T}{L-1})] \tag{8}$$

The initialization of a SincConv layer determines the learned frequency band of interest. The frequency band of interest in Sinc-ShallowNet is [4,40) which covers four type of MI-EEG rhythms. We believe that extracting temporal information from different rhythms separately can help extract multi-scale features in time-frequency domain. To this end, we split the frequency band of interest from [4, 40) to [4,8), [8,14), [14,30) and [30,40) (corresponding to θ rhythm, α rhythm, β rhythm and low γ rhythm respectively). Four SincConv layers are initialized by these frequency bands of interest with four groups of signals used as the input of each layer respectively. Then, a batch normalization layer [20] is placed after the SincConv layer, and its parameters are set to $m = 0.99, \epsilon = 1e-3$ to improve the stability.

2) Depthwise convolutional layer and Pooling layer: A depthwise convolutional layer [21] (kernel size $E \times 1$) with a maximum norm constraint of 1 is used to extract the spatial information from feature maps. Then, a batch normalization layer followed by an exponential linear unit (ELU) activation function (with the parameter α set to 1) is employed. Next, an average pooling layer (pooling size 1×109, stride 23) is employed to compress the time-domain information of the feature map.

D. Training Procedure

1) Parameter Setting and Initialization: The cross-entropy was chosen as the loss function. Adam [22] was used as the optimizer with learning rate set to 0.001 and other parameters set to default. The batch size is 64. The parameters of SincConv layers were initialized by the method described in Section II-C. Other parameters were initialized by the Xavier uniform initialization [23].

2) Training and Testing: There are two phases in the training process with 800 maximum number of epochs for each phase. In the first phase, when the validation loss reaches the minimum, early stopping was performed to prevent overfitting. In the second phase, with validation set merged into the training set, the training is terminated when the validation loss is less than the training loss in the first phase.

III. RESULTS

The proposed method and existing methods were implemented by a deep learning framework TensorFlow 2 with Keras API. The performance of the proposed method was

verified in comparative experiments, and the necessity of the two proposed methods were evaluated in the ablation study. To verify the significance, one sided Wilcoxon signed-rank test [24] with false discovery rate (FDR) [25] correction (α = 0.05) was performed to detect the significance of the difference between the proposed method and other methods. This section describes the experiment process and results in detail.

A. Comparison with State of the Art

The performance of the proposed method was compared with state-of-the-art methods including DeepConvNet [8], ShallowConvNet [8], EEGNet [7] and Sinc-ShallowNet [9]. All compared methods have openly (or partially) available codes and were implemented using the intra-subject version. If their parameters are different from the settings described in Section II-D, adjustments were performed according to the design of the original work to reproduce similar results with the same experiment setup.

The result is listed in Table I where the highest accuracy or the lowest standard deviation (denoted as Std) are shown in bold, and DeepConvNet, ShallowConvNet and Sinc-ShallowNet are denoted as Deep, Shallow and Sinc respectively. As listed in Table I, the average accuracy of the proposed method is about 6.3% higher than the optimal compared method Sinc-ShallowNet. The accuracy of our proposed method on all subjects are significantly higher than compared methods with all p-values of Wilcoxon signed-rank test less than 0.05. This indicates the significantly superior performance of the proposed method. In addition, our method has the lowest standard deviation, which shows that the proposed method has the highest stability.

TABLE I
COMPARISON RESULT

Subject	Deep	Shallow	EEGNet	Sinc	proposed
1	0.667	0.806	0.771	0.819	**0.844**
2	0.34	0.493	0.448	0.497	**0.590**
3	0.705	0.847	0.816	0.84	**0.889**
4	0.253	0.726	0.569	0.74	**0.750**
5	0.243	0.583	0.625	0.559	**0.667**
6	0.257	0.545	0.503	0.538	**0.597**
7	0.684	0.795	0.663	0.844	**0.872**
8	0.649	0.809	0.74	0.847	**0.875**
9	0.736	0.792	0.767	0.854	**0.868**
Average	0.504	0.711	0.656	0.726	**0.772***
Std	0.209	0.126	0.122	0.142	**0.117**

B. Ablation Study

To verify the necessity of two proposed methods, an ablation study was performed. The following methods were used: (1) the proposed method without the proposed data augmentation method, denoted as "proposed (w/o DA)"; (2) the proposed method without using the filter bank structure, i.e., the Sinc-ShallowNet method with the proposed data augmentation method, denoted as "proposed (w/o FB)". Results are shown in Table II. The highest accuracy or the lowest standard deviation (denoted as Std) are shown in bold. Table II shows that the proposed data augmentation method and the filter bank structure of FB-Sinc-ShallowNet can enhance the accuracy by 1.85% and 2.52% respectively (with all p-values less than 0.05), and the accuracy of proposed method on 7 subjects are higher than other methods. Additionally, the standard deviation of the proposed method accuracy is the lowest, which shows that the proposed data augmentation method combined with FB-Sinc-ShallowNet algorithm has higher stability.

TABLE II
ABLATION STUDY OF PROPOSED METHODS

Subject	proposed(w/o DA)	proposed(w/o FB)	proposed
1	0.861	0.844	0.844
2	0.545	0.576	**0.590**
3	0.878	0.882	**0.889**
4	0.729	0.729	**0.750**
5	0.660	0.639	**0.667**
6	0.552	0.535	**0.597**
7	0.868	0.861	**0.872**
8	0.854	0.844	**0.875**
9	0.872	0.865	0.868
Average	0.758	0.753	**0.772***
Std	0.132	0.129	**0.117**

C. Comparison of Data Augmentation Methods

Aiming at verifying the effect of the proposed data augmentation method, window slicing, white noise adding and the proposed EMD-based mixed noise adding methods were compared, with the FB-Sinc-ShallowNet employed as the classification method. The window slicing method was implemented according to the design of Borra et al.[9]. The window length is 2s, overlap is 0.5s, and the data from 0.5s to 4s after cue were extracted. In the white noise adding method, the mean, standard deviation and SNR were set to the same value as the EMD-based mixed noise adding method.

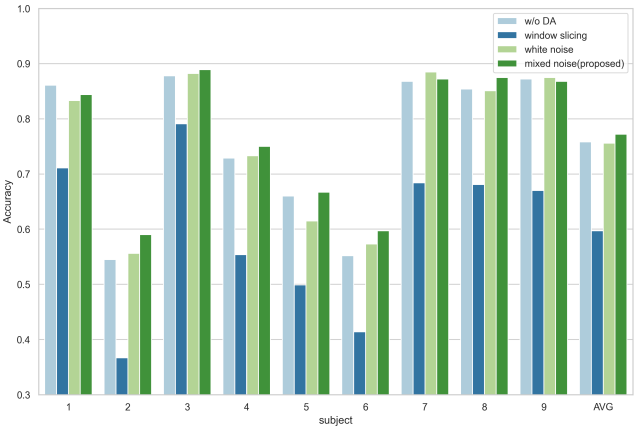

Fig. 2. Comparison of data augmentation methods

As depicted in Fig. 2, the proposed method achieves higher accuracy for most subjects, and the average accuracy of the proposed method (77.2%) is higher than that of the white

noise adding method (75.8%) with statistical significance. Additionally, the proposed method yields superior accuracy on all subjects compared with the window slicing method. In this work, the proposed FB-Sinc-ShallowNet is a lightweight neural network that processes raw EEG signals. Therefore, noise adding methods have better effect than the window slicing method, which is consistent with the conclusion of existing research [9]. Moreover, in noise adding methods, the proposed EMD-based mixed noise method is better. However, the computational cost of EMD limits the speed of the proposed method. Optimizing the computational efficiency of EMD is expected in future work.

IV. CONCLUSION

In this work, we proposed FB-Sinc-ShallowNet, a lightweight neural network that leverages the filter bank structure to improve Sinc-ShallowNet, accompanied with the EMD-based mixed noise adding method that improves the noise adding data augmentation method with EMD. A comparison study was performed between the proposed method and state-of-the-art methods. Superior performance (the average accuracy is 77.2%) and stability (the standard deviation is 0.117) of proposed methods were verified. Therefore, we believe that FB-Sinc-ShallowNet combined with the EMD-based mixed noise adding method can improve the decoding performance of MI-EEG classification. Future work includes: (1) speeding up the EMD process; (2) exploring the effect of proposed methods in inter subject experiments.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: a review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, 2000.

[2] W. Yi, S. Qiu, K. Wang, H. Qi, X. Zhao, F. He, P. Zhou, J. Yang, and D. Ming, "Enhancing performance of a motor imagery based brain–computer interface by incorporating electrical stimulation-induced SSSEP," *J. Neural Eng.*, vol. 14, Jan. 2017. Art. no. 026002.

[3] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng*, vol. 8, no. 4, pp. 441–446, 2000.

[4] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition iv datasets 2a and 2b," *Front. Neurosci.*, vol. 6, p. 39, 2012.

[5] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: a review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, 2016.

[6] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018. Art. no. 031005.

[7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, July 2018. Art. no. 056013.

[8] 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," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[9] D. Borra, S. Fantozzi, and E. Magosso, "Interpretable and lightweight convolutional neural network for EEG decoding: Application to movement execution and imagination," *Neural Netw.*, vol. 129, pp. 55 – 74, 2020.

[10] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI Competition IV," *Front. Neurosci.*, vol. 6, 2012. Art. no. 00055.

[11] E. Lashgari and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *J. Neurosci. Methods*, vol. 346, no. February, 2020. Art. no. 108885.

[12] M. Parvan, A. R. Ghiasi, T. Y. Rezaii, and A. Farzamnia, "Transfer learning based motor imagery classification using convolutional neural networks," in *Iranian Conf. Elect. Eng. (ICEE)*, pp. 1825–1828, 2019.

[13] X. Zhang, Z. Wang, D. Liu, Q. Lin, and Q. Ling, "Deep adversarial data augmentation for extremely low data regimes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 15–28, 2021.

[14] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A: Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[15] Z. Zhang, F. Duan, J. Solé-Casals, J. Dinarès-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[16] H. He and D. Wu, "Transfer learning for brain–computer interfaces: A euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, 2020.

[17] P. Flandrin, P. Gonçalves, and G. Rilling, "Emd equivalent filter banks, from interpretation to applications," in *Hilbert-Huang transform and its applications*, pp. 57–74, World Scientific, 2005.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[19] M. Ravanelli and Y. Bengio, "Speaker recognition from raw waveform with sincnet," in *IEEE Spoken Lang. Technol. Workshop (SLT)*, pp. 1021–1028, 2018.

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

[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, July 2017.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, pp. 249–256, 2010.

[24] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, no. 1, pp. 1–30, 2006.

[25] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc.: Ser. B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.