Convolutional Neural Networks with Reusable Full-Dimension-Long Layers for Feature Selection and Classification of Motor Imagery in EEG Signals

Mikhail Tokovarov(✉)

Lublin University of Technology, Nadbystrzycka 38D, 20-618 Lublin, Poland
m.tokovarov@pollub.pl

Abstract. In the present article the author addresses the task of classification of motor imagery in EEG signals by proposing innovative architecture of neural network. Despite all the successes of deep learning, neural networks of significant depth could not ensure better performance compared to shallow architectures. The approach presented in the article employs this idea, making use of yet shallower, but productive architecture. The main idea of the proposed architecture is based on three points: full-dimension-long 'valid' convolutions, dense connections - combination of layer's input and output and layer reuse. Another aspect addressed in the paper is related to interpretable machine learning. Interpretability is extremely important in medicine, where decisions must be taken on the basis of solid arguments and clear reasons. Being shallow, the architecture could be used for feature selection by interpreting the layers' weights, which allows understanding of the knowledge about the data cumulated in the network's layers. The approach, based on a fuzzy measure, allows using Choquet integral to aggregate the knowledge generated in the layer weights and understanding which features (EEG electrodes) provide the most essential information. The approach allows lowering feature number from 64 to 14 with an insignificant drop of accuracy (less than a percentage point).

Keywords: Motor imagery · Feature selection · Convolutional neural network · Reusable convolutions

1 Introduction

While analyzing EEG signals, researchers discovered an effect appearing when an examined person was imaging hand movements without even actually performing them [7]. The effect was called motor imagery and attracted serious attention of researchers. Further research revealed, that one of the fields, where motor imagery could be used in therapy purposes, was post-stroke rehabilitation, helping patients to recover faster [2, 4].

I. Farkaš et al. (Eds.): ICANN 2020, LNCS 12396, pp. 79–91, 2020.
https://doi.org/10.1007/978-3-030-61609-0_7

Another field, where motor imagery presents interest, is brain computer interface (BCI): it could potentially become a source of control signals for brain computer interfaces, able to help severely disabled people in communication and rehabilitation [14].

The author of the present research proposes a novel model which can solve the task of subject-independent classification with a high level of accuracy.

The structure of the model also allows to interpret the contribution of each EEG electrode, thus performing feature selection and ensuring interpretability of the model. The proposed method of feature selection allows to significantly decrease the number of electrodes: down to 14 from 64 electrodes by the cost of slightly lower accuracy (less than a percentage point). The method of network interpretation based on fuzzy integral calculation is also proposed, so it does not simply compute the weighed sum of all filters, but takes into account their interactions, expressed by the language of set theory.

2 Related Work

Recent years have brought rapid development of machine learning, and, more specifically, its subset - deep learning. Motor imagery classification is no exception and many studies involving deep models have appeared. Various types of deep learning models were tested: deep feed forward networks with dense layers, using separate feature extraction techniques [3], convolutional neural networks, ensuring feature extraction without standalone feature extraction methods [16], recurrent neural network, addressing the sequent nature of EEG signals [17].

Despite all the successes of deep learning models, shallow models have proven to be more effective for EEG signal processing [16], so the proposed model is shallow as well.

One of the advantages, that convolution layers provide lies in the fact, that they do not require feature extraction methods, since feature extraction is performed by the convolutional layers themselves. This approach is called end-to-end processing. Its main advantage is no need of feature extraction, filtering and other kinds of preprocessing [18, 19].

As for feature selection, researchers present various approaches to feature selection in EEG signals: Fisher ranking methods [9], the approaches based on differential evolution [1]. The authors of [12] proposed the method of feature selection based on application of convolutional network filter weights, but no fuzzy measure was introduced, so the weights were simply averaged. Another novelty of the architecture in question is the method of its interpretation and feature selection by means of fuzzy aggregation.

3 Materials and Methods

3.1 Dataset Description

In order to test the performance of the proposed architecture one of the largest publicly available motor imagery signal datasets was used [15]. The dataset contains the samples of EEG signals of 109 volunteers. The recordings were gathered in accordance with 64 electrode 10-10 montage standard. The sampling frequency was equal to 160 Hz.

The present research is focused on developing an architecture allowing to distinguish between two types of motor imagery: imaging left- or right-hand movement. Similarly to the research [5], signals of 4 persons were rejected due to the fact that they contained steady state signal in many motor imagery cues [20]. Each of 105 participants provided 21 observations of each type of motor imagery, which in total gave 4410 observations.

3.2 Proposed Architecture Description

The main idea of the proposed architecture is based on three main points:

- full length 'valid' convolutions,
- reusable layers,
- dense layer connections, i.e. combining input and output of the layer.

Convolutional networks used in image processing have comparatively small filters [6]. As for EEG, this approach proves to be insufficient, so researchers choose longer filters [5, 16, 20]. After preliminary tests it was discovered that introducing longer convolution filters leads to higher accuracy. The present research goes even further and makes use of factorized convolutions analyzing the whole correspondent dimensions.

Figure 1 presents the idea of full length valid convolutions.

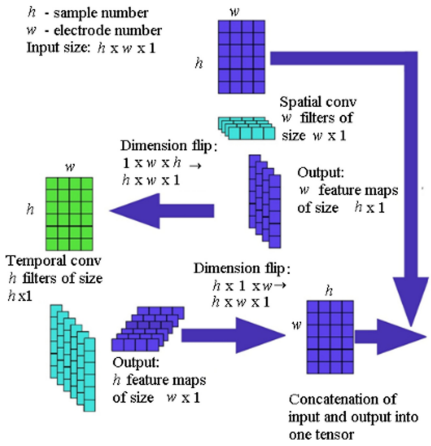

Fig. 1. Full length convolution block with concatenation of input and output

One of the key ideas of the research was to introduce a block consisting of two sequent convolution layers: spatial and temporal convolutions having the same length as the corresponding input's dimension. In order to obtain the output of the same shape as the input the following filters are applied to the h-by-w-by-1 input:

- w filters of shape w-by-1, which are applied in order to extract spatial features,
- h filters of shape h-by-1, being trained to capture the temporal features.

As the result of 'valid' application of the first (spatial) convolution to a *h*-by-*w*-by-1 input, an *h*-by-1-by-*w* tensor is obtained. In order to obtain a tensor with shapes *h*-by-*w*-by-1, a dimension flip is carried out. The output is fed to the temporal convolution layer. The result of the temporal convolution is also flipped back to shapes *h*-by-*w*-by-1. An output tensor of extracted features is concatenated with the input tensor.

Another solution based on repeated use of full-dimension convolution block was introduced. This approach was first presented in [10]. In order to make repeated application of the convolution block possible, a compressing 1-by-1 convolution (denoted as (1) in Fig. 2) is applied.

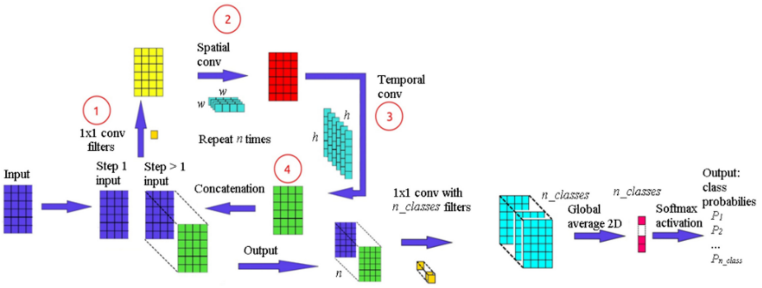

Fig. 2. Complete diagram of the proposed architecture

3.3 Computing Complexity Analysis

The filters are usually separated into temporal and spatial components [5, 16]. This approach employs two subsequent filters of shapes *n*-by-1 and 1-by-*m* instead of one *n*-by-*m* filter.

In the present research the filters are factorized but their length equals to the corresponding dimension. Due to high a number of parameters in every layer it may seem that the computational cost of the proposed architecture is inappropriately high. Actually, this is not the case, due to the fact, that the 'valid' convolution rule is used.

Firstly, a conventional approach will be analyzed. Let the layer input be denoted as I and its size be *h*-by-*w*-by-*d*.

Let the considered convolutional layer have k filters of size *n*-by-1 and 1-by-*m*.

Many packages implementing convolutional models employ cross-correlation function instead of convolution [6]. Discrete cross correlation of multichannel input (I) and convolutional filter (f) for the output pixel (i,j) of the output feature map can be expressed as follows:

$$(I \star f)_{i,j} = \sum_{p=0}^{n-1} \sum_{q=o}^{m-1} \sum_{z=1}^{d} I^*_{i+p,j+q,z} \cdot f_{i+1,j+1,z} + b \tag{1}$$

Where b is the filter's bias and I^* denotes the input padded with zeros in order to implement 'same' rule convolution, i.e. to ensure the output shapes equal to the input shapes.

From formula (1) the number of operations per output pixel (N^{op}_{same}) can be obtained in the following way:

$$N^{op}_{same} = 2 \cdot n \cdot m \cdot d \tag{2}$$

In the case if factorized convolutions are used, (2) can be written as for temporal and spatial convolutions in the following way:

$$N^{op-time}_{same} = 2 \cdot n \cdot d \tag{3}$$

$$N^{op-space}_{same} = 2 \cdot m \cdot k \tag{4}$$

Where k is the number of filters, and thus, the depth of the output of both layers.

Taking into consideration the fact that the output shape is equal to the input shape, one can obtain the total number of operations per one input tensor (N^{total}_{same}):

$$N^{total}_{same} = \left(N^{op-time}_{same} + N^{op-space}_{same}\right) \cdot h \cdot w \cdot k + 2 \cdot N^{act}_{same} \tag{5}$$

$$\begin{aligned} N^{total}_{same} &= \left(N^{op-time}_{same} + N^{op-space}_{same}\right) \cdot h \cdot w \cdot k + 2 \cdot h \cdot w \cdot k \\ &= \left(N^{op-time}_{same} + N^{op-space}_{same} + 2\right) \cdot h \cdot w \cdot k \end{aligned}$$

Where N^{act}_{same} is the number of the activation function application, which is equal to the output shape. The activation is applied twice: for the spatial and temporal convolution.

As for the proposed method, the number of operations per output pixel is expressed in the following way:

$$N^{op-time}_{valid} = 2 \cdot h \tag{6}$$

$$N^{op-space}_{valid} = 2 \cdot w \tag{7}$$

The depth of the input is guaranteed to be equal to 1 by the application of 1-by-1 convolutional layer with 1 filter, so the depth is not presented in (6), (7) and (8). The total number of operations for the proposed method is expressed as follows:

$$N^{total}_{valid} = \left(N^{op-time}_{valid} + N^{op-space}_{valid}\right) \cdot h \cdot w + 2 \cdot N^{act}_{valid}$$

$$\begin{aligned} N^{total}_{valid} &= \left(N^{op-time}_{valid} + N^{op-space}_{valid}\right) \cdot h \cdot w + 2 \cdot h \cdot w \\ &= \left(N^{op-time}_{valid} + N^{op-space}_{valid} + 2\right) \cdot h \cdot w \end{aligned} \tag{8}$$

In order to compare the complexities, divide (5) by (8):

$$\frac{N^{total}_{same}}{N^{total}_{valid}} = \frac{\left(N^{op-time}_{same} + N^{op-space}_{same} + 2\right) \cdot k}{N^{op-time}_{valid} + N^{op-space}_{valid} + 2} = \frac{(2 \cdot n \cdot d + 2 \cdot m \cdot k + 2) \cdot k}{h + w + 2} \tag{9}$$

On the basis of [5] it was assumed that $d = 1$ and the 'valid' rule for the spatial convolution, decreased its output size by w times (due to the fact that the spatial filter has the size of 1-by-w). We thus obtain the following:

$$\frac{N_{same}^{total}}{N_{valid}^{total}} = \frac{\left(2 \cdot n + 1 + \frac{2 \cdot w \cdot k + 1}{w}\right) \cdot k}{2 \cdot h + 2 \cdot w + 2} \tag{10}$$

Using values from [5] in (10) for comparison one can obtain:

$$\frac{N_{same}^{total}}{N_{valid}^{total}} = \frac{\left(2 \cdot 30 + 1 + \frac{2 \cdot 64 \cdot 40 + 1}{64}\right) \cdot 40}{2 \cdot 480 + 2 \cdot 64 + 2} \approx \frac{5640}{1090} \approx 5.17$$

Being approximately 5 times faster (per one repeat), the proposed convolution block can be used repeatedly without causing increase of computational load compared to conventional convolutional layers.

3.4 Regularization and Data Augmentation

The overall number of observations in the dataset is equal to 4410, which is not very high for a convolution network, prone to overfitting in the case of a small dataset [6]. For data augmentation and overcoming overfitting, two measures were taken:

- adding normally distributed noise to the input and intermediate feature maps,
- randomly shifting the start of the analyzed signal segment, presented in Fig. 3. The shift value was generated uniformly from [−m_shift, m_shift].

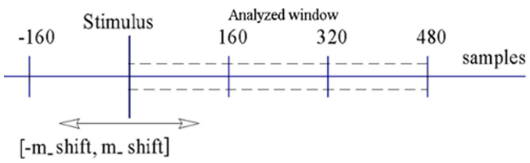

Fig. 3. Data augmentation by means of window random shift

3.5 Interpretation of Feature Maps

Being shallow, the network does not behave as a black box. The fact, that the spatial filters cover the whole electrode set means, that every weight in a spatial filter corresponds to the importance, the specific electrode has in the output of the filter.

Only the feature maps computed for the net input were analyzed, as they contain the information that was obtained directly from the input EEG signal, thus reflecting the electrode importance directly. Figure 4 shows the output of spatial convolution layer. The columns of the output tensor correspond to the output of the spatial convolution filter.

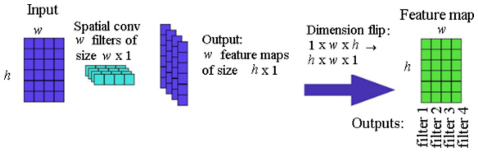

Fig. 4. Output of spatial convolution layer

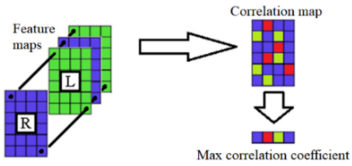

Fig. 5. Obtaining correlation map from feature maps

The coefficients of Pearson correlation between the elements of the feature map and class predictions were chosen to be fuzzy measures of the filters' relevance (Fig. 5). It can be thought of as the measure of consistent behavior of a filter, i.e. that it generates lower values for class 0 and higher values for class 1, or vice versa. Maximal absolute correlation coefficients were considered, since for assessing filters' relevance it was not the sign of correlation coefficient, what was important, but the strength of correlation.

The obtained maximal values of filter correlation coefficients $\{C_n\}$, $n = 1, 2, \ldots$, can be thought of as the values of fuzzy measure over the set of convolutional filters [13].

After obtaining $\{C_n\}$, being the fuzzy measures of the filters' importance, aggregation of the electrode weights can be performed by application of Choquet integral [13]:

$$w(e_j) = \sum_{i=1}^{m} \left[w_i(e_j) - w_{i+1}(e_j) \right] \cdot g(A_i) \tag{11}$$

where:

$w(e_j)$ - aggregated weight of the j-th electrode;

$w_i(e_j)$ - weight of the j-th electrode in the i-th filter, the

weights have to be arranged to produce nonincreasing sequence [13];

$g(A_i)$ - fuzzy measure of subset of first i weights $w_i(e_j)$.

The values of $g(A_i)$ for j-th electrode, can be obtained by the following formula:

$$g(A_i) = \begin{cases} C_1^j, & i = 1 \\ g(A_{i-1}) + C_i^j + \lambda \cdot g(A_{i-1}) \cdot C_i^j, & i > 1 \end{cases} \tag{12}$$

Coefficient λ describing the interaction can be obtained by following formula [13]:

$$1 + \lambda = \prod_{i=1}^{m} (1 + \lambda g_i) \tag{13}$$

Thus, in order to find the value of λ, the roots of the m-degree polynomial have to be found. For the considered case, the degree of polynomial is equal to 64, so direct computing of polynomial roots is too computationally expensive and even not reasonable, as only one root is needed. The root was obtained by Newton-Raphson method.

4 Experiment

4.1 Experiment Procedure

5-fold cross-validation procedure was applied to assess the performance of the model.

Repeat number needed for stable result was found on the basis of partial mean accuracy:

$$Acc_i^{part} = \sum_{j=1}^{i} Acc_j \tag{14}$$

where Acc_j is the cross validation accuracy, obtained in j-th repetition. Number of repetitions was set to 30. Further sections present the effect of selected hyperparameters on accuracy. Differences between separate results presented in the chapter are not high, but fine tuning of hyperparameters allows improving accuracy from 82.33% to 83.26%.

4.2 Activation Function Influence

Several activation functions were examined. The results are presented in Table 1.

Table 1. Accuracies obtained for various activation functions

Activation function	Mean (acc), %
relu	82.33
selu	82.93
tanh	**83.06**

Interestingly, the best performance was obtained with the use of the tanh activation function, so it was used for further experiments.

4.3 Reuse Number Influence

Table 2 contains the values of parameters and accuracies achieved for various reuse numbers.

Table 2. Comparison of accuracy values obtained for various number of repeated convolution block application

Reuse number	Mean (acc), %
1	0.8306
2	**0.8326**
3	0.8315
4	0.8307

The application convolution block repeated twice allows to improve the accuracy by 0.21 percentage point, further increase of repeat number leads to lower accuracy, so the repeat number equal to 2 was chosen for further experiments.

4.4 Dense Connection Influence

In order to investigate the influence of dense connections on the accuracy, the experiment with the network without dense connections was conducted. Table 3 contains the results achieved for the architecture without use of dense connection.

Table 3. Accuracy obtained for the architecture without dense connections.

Dense connection	Mean (acc), %
No	83.08
Yes	**83.26**

4.5 Influence of Multiple Block Application

The tests of deeper architecture were conducted. The deeper architecture was made of sequentially repeated convolutional blocks presented in Sect. 3.2. The results are presented in Table 4.

Table 4. Accuracy obtained for the architectures with various numbers of convolutional blocks.

Conv. block number	Mean (acc), %
1	**83.26**
2	83.05
3	82.95

5 Feature Selection by Analysis of Activation Maps

The method presented in the Sect. 3.5 allows to explain and interpret the weights of spatial convolutions. It also allows to find the most relevant electrodes, thus permitting to perform feature selection. The experiment was conducted in the following way:

1. the network was trained with the training dataset;
2. the procedure of feature selection described in Sect. 3.5 was performed;
3. the training set was modified, only selected electrodes were used;

4. a new model was created and trained on the modified dataset;
5. the performance of the trained model was assessed by applying the test dataset modified in a way similar to the training dataset: only selected electrodes were taken into account.

The procedure was conducted three times (runs): selecting 14, 32 and 48 electrodes. Every run included 30 repetitions. The set of weights, associated with correspondent electrodes was obtained in each repetition. The sets were averaged. Figure 6 presents the heatmaps of the electrode weights obtained with the proposed method. The weights were mapped to 2D structure in accordance with 10-10 EEG montage system.

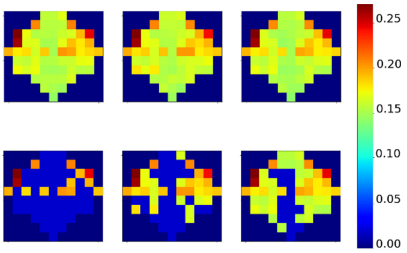

Fig. 6. Heatmaps of the electrode weights obtained in three runs. Top: all electrodes shown. Bottom: only selected electrodes (with first 14, 32, 48 highest weights) shown.

Accuracies obtained for each number of electrodes are presented in Table 5. The proposed procedure of feature selection allows obtaining similarly high results (less than a percentage point difference) for the limited electrode set.

Table 5. Accuracy for various numbers of selected features.

Selected electrode number	Mean (acc), %
14	82.34
32	82.9
48	83.16
64	83.26

6 Discussion and Conclusion

The paper presents the novel architecture of convolutional neural networks, dedicated to classification of motor imagery in EEG signals. The architecture combines the ideas proved to be promising in the areas outside neuroscience, e.g.: convolutional filter reuse and dense connections. The present paper presents the first example of application of

reusable filters for classifying motor imagery. The positive influence of this approach was shown in the Sect. 4.3.

Dense connections proved to be a promising approach as well.

As for making the network deeper, it did not provide any improvement and, on the contrary, led to decrease of classification accuracy, and made computations slower.

Table 6 contains the comparison of the accuracies presented in the published papers. The accuracies were achieved in the same conditions as the method proposed in the present paper, i.e.:

- Physionet benchmark dataset,
- 3-second long signal segments analyzed,
- 105 participants. Many papers present high results obtained for a significantly limited subset of Physionet benchmark dataset. Only the papers, where the analysis of at least a 105-person dataset is presented, are taken into account.

Table 6. Comparison of achieved accuracies.

Paper	Mean (acc), %
[20]	82.43
[8]	80.1
[5]	80.05
[18]	74.71
[11]	70.6
Present research	**83.26**

As it was shown in Sect. 3.3, despite the long filters applied, proposed architecture has low computational complexity, even compared to shallow model, presented in [5].

Another innovation proposed in the present paper was a feature selection method which allowed to find the electrodes, providing the most valuable information, allowing to achieve high classification accuracy even with a limited electrode set. The measure of an electrode's importance was obtained on the basis of Choquet integral, which made possible effective aggregation of the data cumulated in in the network filters, ensuring clear representation of the knowledge regarding the electrode's importance. Constructing fuzzy measures on the basis of correlation coefficients is the approach which can be easily understood and interpreted.

Table 7 presents the classification accuracies achieved by applying a limited set of electrodes in the present paper and in other papers published in recent years.

The results in the table show, that even with a very limited set of electrodes the proposed method of feature selection can ensure sufficient classification accuracy, which is higher compared to the results presented in the literature. It is worth noting that in the proposed method the electrodes are selected on the basis of a clearly understandable measure, which takes into account the "consistency" of the filters' behavior, contrary to

Table 7. Comparison of achieved accuracies for limited electrode set

Paper	Electrode number	Mean (acc), %
[8]	14	80.05
[5]	14	76.66
[20]	19	81.95
	38	81.86
Present research	**14**	**82.34**

approaches presented in other research, such as selecting the electrodes on the basis of empirical observations stating that specific electrodes are related to motor imagery more closely than the others or simple averaging of the electrode weights across all filters of the trained convolutional network. The first approach is too general, as EEG signals are known to be highly subject specific. That is why data driven method presented in the paper allows achieve higher performance. As for the latter approach, not all the filters of the trained network become trained for extracting valuable for classification features to the same degree, due to that reason the approach based on simple averaging will lead to poorer performance of classification model.

The proposed approach of feature selection by network weights interpretation can be used not only for EEG signal processing, but also within any other field where multidimensional data are analyzed.

The author hopes that the present research makes real impact in the field of machine learning and EEG signal processing and, in order to encourage further research in the field, makes the source code publicly available on the github platform upon the paper publication. The code can be found under the following link: https://github.com/mtokovarov/cnn_reuse_full_dim_fs.git.

References

1. Baig, M.Z., Aslam, N., Shum, H.P., Zhang, L.: Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. Expert Syst. Appl. **90**, 184–195 (2017)
2. Belkofer, C.M., Konopka, L.M.: Conducting art therapy research using quantitative EEG measures. Art Ther. **25**(2), 56–63 (2008)
3. Chai, R., Ling, S.H., Hunter, G.P., Nguyen, H.T.: Mental non-motor imagery tasks classifications of brain computer interface for wheelchair commands using genetic algorithm-based neural network. In The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2012)
4. Dijkerman, H.C., Ietswaart, M., Johnston, M., MacWalter, R.S.: Does motor imagery training improve hand function in chronic stroke patients? A pilot study. Clin. Rehabil. **18**(5), 538–549 (2004)
5. Dose, H., Møller, J.S., Puthusserypady, S., Iversen, H.K.: A deep learning MI-EEG classification MODEL for BCIs. 2018 In: 26th European Signal Processing Conference, pp. 1690–1693. IEEE (2018)

6. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
7. Ikeda, A., Lüders, H.O., Burgess, R.C., Shibasaki, H.: Movement-related potentials recorded from supplementary motor area and primary motor area: role of supplementary motor area in voluntary movements. Brain **115**(4), 1017–1043 (1992)
8. Kim, Y., Ryu, J., Kim, K.K., Took, C.C., Mandic, D.P., Park, C.: Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns. Comput. Intell. Neurosci. (2016)
9. Kirar, J.S., Agrawal, R.K.: Relevant feature selection from a combination of spectral-temporal and spatial features for classification of motor imagery EEG. J. Med. Syst. **42**(5), 78 (2018)
10. Köpüklü, O., Babaee, M., Hörmann, S., Rigoll, G. Convolutional Neural networks with layer reuse. In: proceedings of 2019 IEEE International Conference on Image Processing (ICIP), pp. 345–349 (2019)
11. Netzer, E., Frid, A., Feldman, D.: Real-time EEG classification via coresets for BCI applications. Eng. Appl. Artif. Intell. **89**, 103455 (2020)
12. Mzurikwao, D., Ang, C.S., Samuel, O.W., Asogbon, M.G., Li, X., Li, G.: Efficient channel selection approach for motor imaginary classification based on convolutional neural network. In: 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), pp. 418–421 (2018)
13. Pedrycz, W., Gomide, F.: An Introduction to Fuzzy Sets: Analysis and Design. Mit Press, Cambridge (1998)
14. Pfurtscheller, G., et al.: Current trends in Graz brain-computer interface (BCI) research. IEEE Trans. Rehabil. Eng. **8**(2), 216–219 (2000)
15. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans. Biomed. Eng. **51**(6), 1034–1043 (2004)
16. Schirrmeister, R.T., et al.: Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. **38**(11), 5391–5420 (2017)
17. 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)
18. Zhang, D., Chen, K., Jian, D., Yao, L.: Motor imagery classification via temporal attention cues of graph embedded EEG signals. IEEE J. Biomed. Health Inf., 2570–2579 (2020)
19. Zhang, D., Yao, L., Chen, K., Wang, S., Chang, X., Liu, Y.: Making sense of spatio-temporal preserving representations for EEG-based human intention recognition. IEEE Trans. Cybern., 3033–3044 (2019)
20. Wang, X., Hersche, M., Tömekce, B., Kaya, B., Magno, M., Benini, L.: An accurate EEGNet-based motor-imagery brain-computer interface for low-power edge computing. arXiv preprint arXiv:2004.00077 (2020)