2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)
Bari, Italy. October 6-9, 2019

EEG Representation in Deep Convolutional Neural Networks for Classification of Motor Imagery

Neethu Robinson, *Member, IEEE,* Seong-Whan Lee, *Fellow, IEEE,* Cuntai Guan, *Fellow, IEEE*

***Abstract*— With deep learning emerging as a powerful machine learning tool to build Brain Computer Interface (BCI) systems, researchers are investigating the use of different type of networks architectures and representations of brain activity to attain superior classification accuracy compared to state-of-the-art machine learning approaches, that rely on processed signal and optimally extracted features. This paper presents a deep learning driven electroencephalography (EEG) -BCI system to perform decoding of hand motor imagery using deep convolution neural network architecture, with spectrally localized time-domain representation of multi-channel EEG as input. A significant increase in decoding performance in terms of accuracy of $+6.47\%$ is obtained compared to a wideband EEG representation. We further illustrate the movement class specific feature patterns for both the architectures and demonstrate that higher difference between classes is observed using the proposed architecture. We conclude that the network trained by taking into account the dynamic spatial interactions in distinct frequency bands of EEG, can offer better decoding performance and aid in better interpretation of learned features.**

I. Introduction

Electroencephalography (EEG) is the widely used as the non-invasive brain data acquisition modality in Brain Computer Interface (BCI) research. BCI functions by decoding the neural activity and translating the brain states directly to output commands that communicate and control an interfaced external devices [1, 2]. Hence, to ensure a robust and reliable performance of BCI, a variety of decoding tools have been proposed aiming higher accuracy and lesser training time. One of the widely investigated BCI paradigms is motor-imagery (MI) in which participants perform mental rehearsal of movement which forms characteristic Sensorimotor rhythm (SMR) modulations in EEG [3, 4]. SMR-BCI studying activations in EEG has identified event related desynchronization/ synchronization (ERD/ERS) as the fundamental indicator of MI [4]. These are the localized neural synchrony variations in distinct contralateral and ipsilateral sensory motor regions. EEG-BCI research have reported many decoding techniques to classify right and left hand motor imagery including linear and non-linear classifiers, nearest neighbor classifiers, neural networks, adaptive classifiers, matrix and tensor classifiers, transfer learning and deep learning [5].

Deep Learning has emerged as a powerful machine learning tool with superior performance in speech recognition and computer vision [6]. There is a growing interest among neuroengineering researchers to use deep learning methods in building BCI systems to attain superior decoding performance. The potential of deep learning tools to learn optimal features and decoding models, voids the need to perform signal processing and machine learning based feature engineering in conventional BCI. Researchers have presented wide range of deep network architectures which are validated on various EEG datasets [7-13]. The recent literature indicates that deep learning methods outperform the state-of-the-art machine learning approaches using processed EEG. Even though the results look promising, the choice of brain signal representation, type of network and its hyperparameters to be used, largely vary among these reported methods and datasets used, and hence a conclusion regarding the optimal deep neural network (NN) models are not yet available [5, 13].

In EEG decoding using deep learning, one of the critical and initial steps is to determine how to represent time series data recorded from each sensors, considering that it can cause the most information loss and computational cost. Researchers have reported the use of raw, wide and narrow band and filter-bank filtered and time-frequency spectrum of EEG as input to deep neural networks [7-13] to classify it. In [7, 8] envelope representation of EEG using Hilbert transformation and passing it through a convolutional NN (CNN) was proposed. In [9], a sequence of topology-preserving multi-spectral images were obtained from EEG and a recurrent-convolutional NN was used to decode cognitive load. In [9], a 5-layer CNN was proposed built on spatiotemporal characteristics of EEG to classify MI. CNN-based framework to study neural patterns underlying attention and consciousness are reported in [10] and [11] respectively. In [12], short time Fourier transform was used to convert EEG into 2D images, followed by 1D convolution along time axis. In [13], the EEG was represented as a 2D array with the number of time steps as the width and the number of electrodes as the height, and a deep ConvNet architecture was proposed.

In the research reported in this paper, our objective is to propose an EEG signal representation in temporal, spectral and spatial domain. The goal is to allow the network to process and localize the class-specific information in these three domains thereby better capturing the movement task-specific dynamics of EEG. The first convolution block performs filtering over time followed by spatial domain. The spatial filtering is carried out in each band following which the information is pooled together.

Research is supported by Agency for Science, Technology and Research, Singapore.

N. Robinson and C. Guan are with School of Computer Science and Engineering, Nanyang Technological University, Singapore, 50 Nanyang Avenue, S639798 (e-mail: {nrobinson, ctguan}@ ntu.edu.sg.

S. -W. Lee is with the Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, South Korea (e-mail: sw.lee@korea.ac.kr).

978-1-7281-4569-3/19/$31.00 ©2019 IEEE

The deep convolutional NN architecture proposed to decode this signal uses deep ConvNet [13] as a baseline. Further, we also studied how the EEG signal is transformed into feature maps at different levels in the network. The output from the proposed input convolution-max pooling block is obtained for both classes of data and we illustrated how adding spectral information offers more distinction between classes. To the best of our knowledge, a similar architecture validated on EEG dataset and comparison of the effect of different input signal representations in deep learning methods has not been reported in literature.

The rest of this paper is organized as follows: Section II explains the methodology, dataset and data processing steps. Section III reports the results and discussion of the research followed by conclusions in Section IV.

I. METHODOLGOY

The objective of this research is to define a deep convolution neural network architecture to decode binary right hand and left hand motor imagery EEG data. We propose an EEG input representation by preserving temporal, spectral and spatial information of the signal. The input convolution-max pooling layer is then designed to filter the signal in two stages, detecting the information along time domain, and along channels, separately in each frequency band. This is followed by other convolution-max pooling layers and softmax layers. The details of the architecture, analysis and the dataset used for validation is given in the following sub-sections.

A. Dataset

The EEG dataset used in this research is recorded at the Department of Brain and Cognitive Engineering, Korea University. The data from fifty-four healthy subjects (ages 24-35, 25 females) performing binary class motor imagery (MI) were recorded using BrainAmp (Brain Products; Munich, Germany). EEG signals were recorded with a sampling rate of 1000 Hz and collected with 62 Ag/AgCl electrodes. The MI paradigm used was the well-established protocol as per [14]. For all blocks, the rest 3s of each trial began with a black fixation cross that appeared at the center of the monitor to prepare subjects for the MI task. Afterwards, the subject performed the task for 4s when the right or left arrow appeared as a visual cue. After each task, the screen remained blank for 6s (±1.5). For more details on data and experiment protocol, please see [15].

Each subject participated in two data recording sessions and each session had an offline training phase to record data and construct classifier and a test phase that provided visual feedback to the subject by decoding data using the classifier. For the research reported in this paper, we used EEG data from first session that consists of 200 trials of data, with equal number of trials in left (class-0) and right hand (class-1) MI.

The EEG data from 4 seconds of MI task, for every trial was segmented from the continuous data. The data is then downsampled by a factor of 2. For further processing, data from 34 channels {F9, F7, F3, Fz, F4, F8, F10 FC5, FC3, FC1, FC2, FC4, FC6, C5,C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4,CP6, P7, P3, P1, Pz, P2, P4, P8} were used.

B. Proposed EEG representation and CNN architecture

In neuroengineering studies, the widely followed approach for processing multi-channel EEG data is to decompose the signal into frequency bands. The power modulation in specific bands originating from distinct areas of the brain are proven to be the source of neural activity correlated to motor task [3, 4]. This is the basis of various feature extraction algorithms used in EEG-BCI. The filter bank decomposition of EEG in mu and beta ranges, followed by spatial filtering of data using common spatial pattern (FBCSP) to provide features is the state-of-the-art decoding approach using shallow classifiers [16]. Similar approach has been employed in CNN architectures as well, that uses 2D image representation of EEG, including time, frequency and location information, followed by convolution in time axis [12]. The proposed architecture in this paper, is build based on the deep CNN architecture proposed in [13] and available in the open source toolbox, Braindecode. The authors have reported extensive analysis of this architecture and compared its performance against machine learning methods, making this a strong basis for our study.

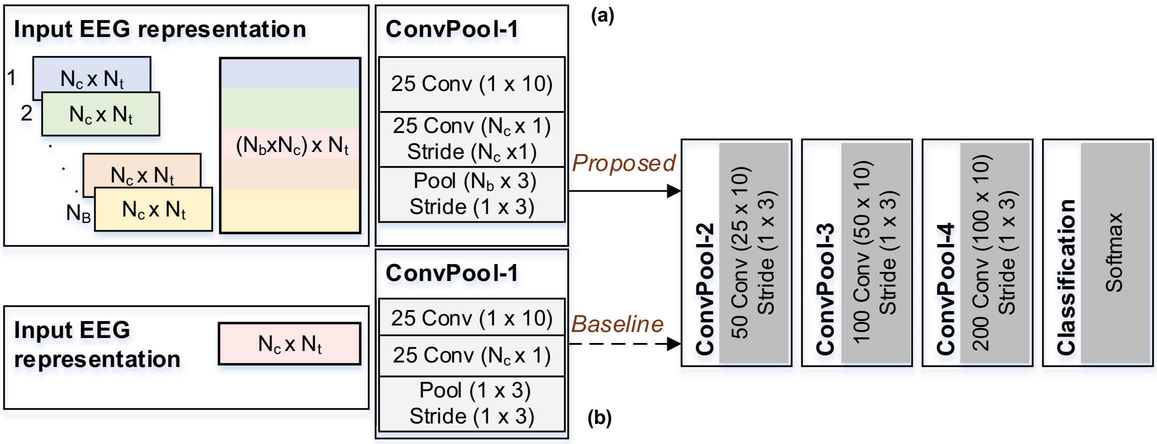

Fig. 1. Deep CNN architectures. (a) proposed and (b) baseline input representation and convolution-max pooling block 1, followed by the standard convolution-max pooling blocks and classification layer.

In this paper, we propose a deep CNN architecture as indicated in Fig. 1. As indicated in Fig. 1(a), a multiband, multi-channel EEG input to the network, represented as a 2D image. The epoched data for each trial denoted by $X \in \mathbb{R}^{N_c \times N_t}$, is decomposed to N_b overlapping frequency bands of 2 Hz bandwidth, spanning from 8 to 30 Hz. The data is concatenated as indicated in Fig.1 and is fed into the network. The input convolution block in the network, provides a 2 stage convolution – first along time, followed by convolution along channel, in individual bands. This step creates a spatial filter for each spectrally localized EEG data, and thus offers a generic spectro-spatial filtering of EEG. The goal of this approach is to preserve and localize spatial information in each frequency band which is significant in discriminating 2 classes of data, thereby aiding in decoding. We believe that a data representation in this manner, allows the network to learn the temporal evolution of interactions between channels in each frequency band, which is otherwise lost in case of a wideband data. In the interest of interpreting the features generated and propagated through the network, we also study how data from each movement class is modulated by each CNN architecture.

The proposed architecture is compared against a wideband EEG representation indicated in Fig. 1(b). We use band pass filtered EEG from 8 to 30 Hz as input to the network. The input block here performs a temporal convolution followed by a spatial convolution over all channels. The rest of the architecture follows the same approach as Deep4Net presented in [13], and its pipeline indicated with the in Fig. 1. After the first convolution-max-pooling block, the network features three convolution-max-pooling blocks, using batch normalization and dropout, followed by a dense softmax classification layer. All layers use exponential linear units (ELUs) as nonlinearities. The research in [13] reported used full bandwidth and high-pass filtered (>4 Hz) data as input to the network.

C. Evaluation and analysis of network

The decoding performance of the proposed architecture is computed in terms of classification accuracy for each subjects. For each subject, we perform the performance evaluations using EEG data from first session. The data undergoes a train-test split, and test classification accuracy is computed and reported. Since the optimization parameters are not computed for this data, we train the network multiple times and statistical results are reported.

In this research, we also investigate how the proposed EEG representation and input layer of network represents the data and its correlates to movement class label. For this, to determine the input from each movement class, trial-averages, over all the samples. Once the network is trained, we study how it translates the input representing each movement class as features. The intermediate output (features) obtained after the input convolution-max pooling block is then studied to identify how the discrimination between each movement class is preserved in the network.

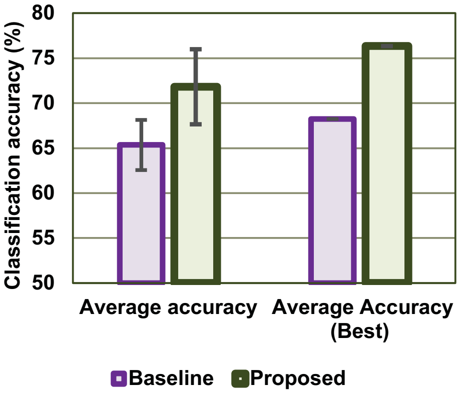

Fig. 3. Decoding performance, average accuracy over all subjects.

II. Results and Discussion

The results of classification and analysis using the proposed architecture is reported in this section. The results are compared with baseline architecture. Further, the input-level features are illustrated to study how the network transforms the EEG data.

A. Classification accuracy

The results of the performance evaluation indicated in Section II-C is presented in Fig. 2. The accuracy obtained using the baseline and proposed methods for each subject is shown. It can be observed that, for majority of the subjects, accuracy improved using the proposed method. The average values over all subjects are indicated in Fig. 3. As mentioned earlier, since the network optimization parameters are not computed for the data, the model is trained multiple ($N = 10$) times. The first set of bars in Fig 3(a) reports average test performance over these N times. The accuracy achieved are $71.82\ \%\ (\pm 4.2\%)$ and $65.4\ \%\ (\pm 2.8\ \%)$ for proposed and baseline methods respectively. The second set of bars indicate the average

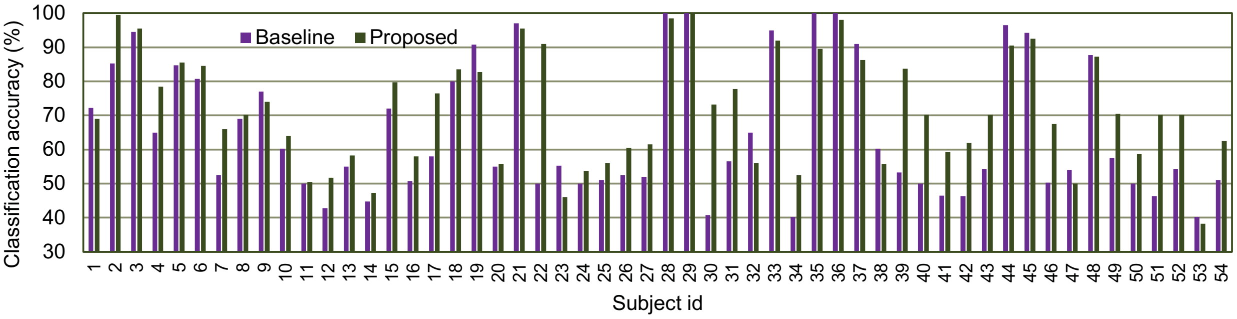

Fig. 2. Average accuracy in the classification of binary right hand v/s left hand MI.

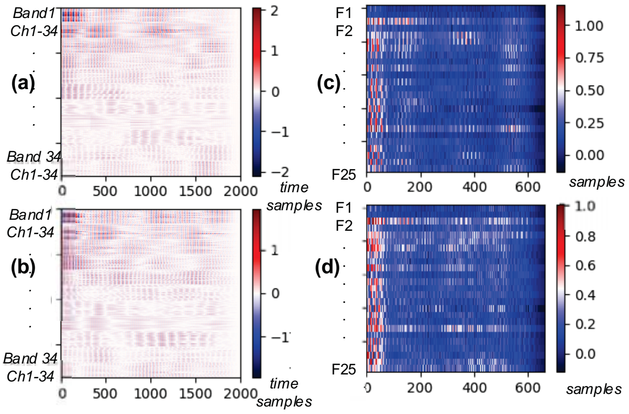

Fig. 4. Proposed architecture input-feature signal illustration. (a) $(N_b x N_c)$ x N_t data for MI class - 0 (b) $(N_b x N_c)$ x N_t data for MI class - 1 (c) Output from convolution-max pooling block 1 for MI class - 0 (d) Output from convolution-max pooling block 1 for MI class - 0. F1-25 indicates output from each of 25 convolution filters.

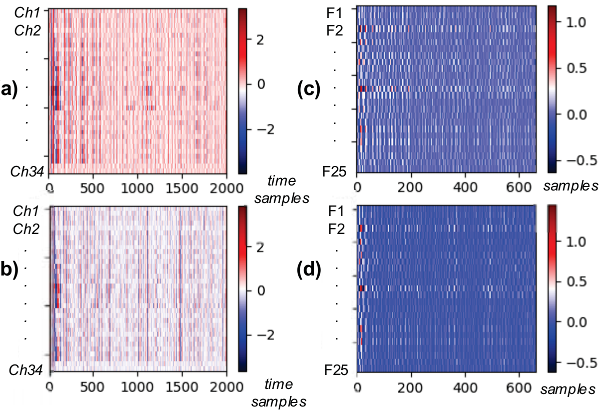

Fig. 5. Baseline architecture input-feature signal illustration. (a) N_c x N_t data for MI class – 0 (b) N_c x N_t data for MI class – 1 (c) Output from convolution-max pooling block 1 for MI class - 0, (d) Output from convolution-max pooling block 1 for MI class - 1. F1-25 indicates output from each of 25 convolution filters.

accuracy, by considering only three (out of ten) models with the best test accuracy. The accuracy obtained are 76.36% and 68.24% for proposed and baseline architectures respectively. This result indicates that there is scope of improvement in decoding performance by identifying an optimized (cross-validated) model for each subject. The increase in accuracy for the proposed method, $+6.47\%$ and $+8.12\ \%$ in both sets are statistically significant $(p < 0.001)$ as well.

The decoding performance for the same data reported in [15] using the state-of-the-art filter band common spatial pattern [16] method is 68.8%, and it can be noted that the proposed deep learning method outperforms this method.

B. *Input data and feature representations*

A gap in current BCI research using deep learning is the interpretation of trained networks and the generated features. It can be presumed that input representations that preserve most distinct class-specific information and network architecture that carries this information forward can offer higher decoding performance. Since the key addition in the proposed architecture is the transformation of input to feature after the first convolution-max pooling block, we illustrate the data at these levels in Fig. 4. We use the data from subject S2 to present the approach (selected since the data obtained good performance in both architectures). The proposed EEG input representation for class-0 and its output after the first convolution max-pooling block is given in Fig. 4 (a) and (c). Similarly the input and output for class-1 is given in Fig. 4(b) and (d). Differences can be observed between the classes and the output from this network level in Figs. 4(c, d) seem to have condensed all the information spread-out in temporal, spatial and spectral domains in Figs. 4(a, b). In comparison, the input-output signal using the baseline architecture is given in Fig. 5. Here, 5(a) and 5(c) are for class-0 and 5(b) and 5(d) are for class-1, with the first column giving input EEG and second column giving output of the first convolution-max pooling block. Differences can be observed in these figures as well.

To visualize the differences between classes, we simply take the difference of values from class-0 and class-1. This difference, denoted as δ, is illustrated in Fig. 6. $\delta_{proposed}$ gives the difference between Fig. 4 (c) and 4 (d) and is shown in Fig. 6 (a), whereas Fig. 6 (b) gives difference between 5 (c) and 5(d) as $\delta_{baseline}$. The x-label of the figures represent time samples after convolution and pooling, and y-label indicates the output from each convolution filter. It can be observed that, certain localized patterns can be observed in Fig. 6 (a), which are not quite stronger in Fig. 6 (b). Fig. 6 (c) illustrates the δ for both architectures averaged over all the output channels. The values are smoothed with a moving average filter (window-size=5, Hanning window), for better visualization. Each plot represents the difference between both movement classes as identified by the input convolution- max pooling block of the network. It can be observed that difference between both movement classes are more distinct in the proposed architecture, which used the added information in frequency domain. Similar analysis can be conducted for all the subjects to obtain the various feature representations. Further, we presume that the patterns observed in Fig. 6(c) for the proposed approach, will have correlation to ERD/ERS patterns that form the basis of MI neural activity in EEG, and thus has potential for further investigation.

C. *Discussion*

In this paper, we present an EEG representation in time, frequency and space domains to be applied as input to a deep CNN architecture. The decoding performance reported in Figs. 2 and 3 indicates the superior performance that can be attained by incorporating the spectral characteristics of data to the network. This indicates that in the proposed architecture, the network learns the time-varying spatial interactions in each frequency band, which effectively captures the dynamic characteristics of EEG.

To investigate how the EEG representations can influence the network, we also looked into the data input and output features from the convolution-max pooling block from the first layer of network. The results are presented in Figs.4-6 and as explained

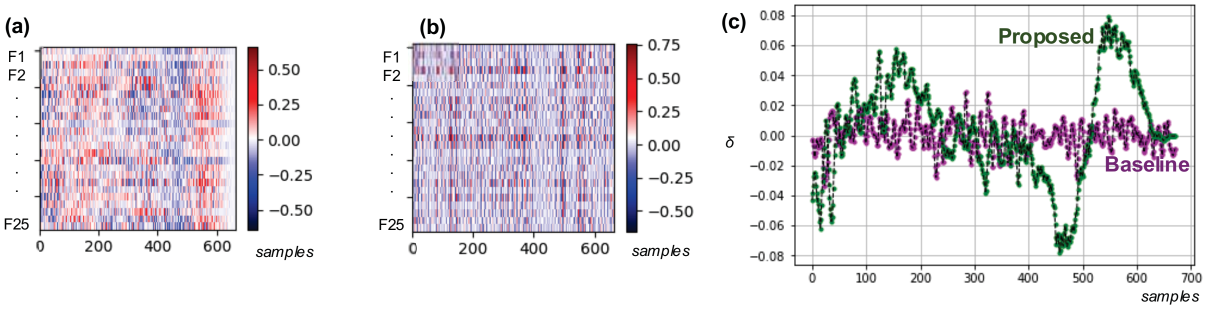

Fig. 6. Difference between features of class – 0 and class – 1, from convolution-max pooling block 1. (a) $\delta_{proposed}$ (b) $\delta_{baseline}$ (c) δ averaged over all the filters.

it offers a simple visualization of how the different networks transform the class-specific information represented in two different ways. Based on the illustration in Fig. 6, we conclude that the between-class difference is more distinct in the proposed architecture. We believe that this difference has contributed to the increase in classification accuracy for the proposed approach. Further research is required to define quantitative approaches to understand the signal propagation through network and statistical measures to characterize and differentiate class-specific activations. The correlation of such feature representation to the standard sensorimotor rhythm patterns will also be of interest for more interpretability of results.

Since adding more spectral information effectively increases the number of parameters to be trained by the network, the deep learning approaches demand more data samples to train the model without overfitting. To facilitate this, future research will also build and evaluate the network in a subject-independent leave-one subject-out approach, by pooling data from all subjects. Furthermore, the correlation of ERD/ERS activations between the input and output data will be studied to better interpret and explain the deep networks.

III. Conclusion

The paper presented a deep CNN architecture for classification of binary hand motor imagery EEG data. An input EEG representation preserving temporal, spatial and spectral information is proposed, followed by input convolution-max pooling block that creates a set of spatial filters for each band of data. The proposed architecture offers a significant increase of +6.47 % in average classification accuracy. The paper also illustrated the distinct between class differences in features created by the proposed architecture that may have aided the increase in performance.

Acknowledgment

This work was supported by the Agency for Science, Technology and Research, Singapore (No. IAF311022), and the Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (No. 2017-0-00451).

References

[1] D. J. McFarland and J. R. Wolpaw, “Brain-computer interfaces for communication and control,” *Communications of ACM*, vol.54(5), pp. 60-66, 2011.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller and T. M. Vaughan, “Brain–computer interfaces for communication and control,” *Clinical neurophysiology*, vol. 113(6), pp.767-791, 2002.

[3] C. Neuper, M. Wörtz, and G. Pfurtscheller, “ERD/ERS patterns reflecting sensorimotor activation and deactivation,” *Progress in brain research,* vol. 159, pp. 211-222, 2006.

[4] B. He, B. Baxter, B. J. Edelman, C. C. Cline and W. Y. Wenjing, “Noninvasive brain-computer interfaces based on sensorimotor rhythms,” *Proceedings of the IEEE*, vol. 103(6), pp.907-925, 2015.

[5] 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,” *Journal of neural engineering*, vol. 15(3), pp. 031005, 2018.

[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521(7553), pp.436, 2015.

[7] S. Sakhavi, C. Guan and S. Yan, “Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 9(99), pp. 1-11, 2018.

[8] S. Sakhavi and C. Guan, “Convolutional Neural Network-based Transfer Learning and Knowledge Distillation using Multi-Subject Data in Motor Imagery BCI”, *8th International IEEE EMBS Conference on Neural Engineering (NER)*, May, 2017.

[9] P. Bashivan, I. Rish, M. Yeasin and N. Codella, “Learning representations from EEG with deep recurrent-convolutional neural networks,” *arXiv*, pp. 1511.06448, 2015.

[10] F. Fahimi, Z. Zhang, W. B. Goh, T. S Lee, K. K. Ang, C. Guan, “Inter-subject transfer learning with end-to-end deep convolutional neural network for EEG-based BCI”, *Journal of Neural Engineering*, vol. 16(9), 026007, 2019.

[11] M. Lee, S.-K. Yeom, B. Baird, O. Gosseries, J. O. Nieminen, G. Tononi and S.-W. Lee, “Spatio-temporal analysis of EEG signal during consciousness using convolutional neural network,” in *6th International Conference on Brain-Computer Interface*, pp. 1-3, 2018.

[12] Y. R. Tabar and U. Halici, “A novel deep learning approach for classification of EEG motor imagery signals,” *Journal of neural engineering*, vol. 14, pp. 016003, 2017.

[13] 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, pp. 5391-420, 2017.

[14] G. Pfurtscheller and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proceedings of the IEEE*, vol. 89(7):1123–1134, 2001.

[15] M. H. Lee, O. Kwon, Y. J. Kim, H. K. Kim, Y. E. Lee, J. Williamson, S. Fazli and S. W. Lee, “EEG Dataset and OpenBMI Toolbox for Three BCI Paradigms: An Investigation into BCI Illiteracy,” *GigaScience*, pp. 1-15, 2019.

[16] K. K. Ang, Z. Y. Chin, H. Zhang and C. Guan, “Filter bank common spatial pattern (FBCSP) in brain-computer interface,” *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390-2397, 2008.