Multi-Class Classification of Motor Imagery EEG Signals Using Image-Based Deep Recurrent Convolutional Neural Network

Ward Fadel[1, 2], Csaba Kollod[1, 2], Moutz Wahdow[1, 2], Yahya Ibrahim[1], Istvan Ulbert[1, 2]
[1] Faculty of Information Technology and Bionics, Pazmany Peter Catholic University, Budapest, Hungary
[2] Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary
{fadel.ward, kollod.csaba, wahdow.moutz, ibrahim.yahya, ulbert.istvan}@itk.ppke.hu

Abstract*—Classification of EEG signals is a cornerstone of building the motor-imagery (MI) based Brain-computer interface (BCI) systems. EEG signals differ from one subject to another and even for the same subject among different trials, and this is why designing a general classification model is still debated. Deep learning is dominant in so many fields like computer vision and natural language processing but it is still under investigation for EEG signals classification. We followed a new trend in EEG signals classification in which these signals are transformed into images, and so classifying such signals become an image classification problem where Deep learning can work well. The Physionet dataset for EEG motor movement/imagery tasks was used which consists of 109 subjects and the motor imagery EEG signals for three frequency bands (Delta [0.5-4 Hz], Mu [8-13 Hz], and Beta [13-30 Hz]) was transformed into 3-channel images (one channel for each band) using the Azimuthal equidistant projection and Clough-Tocher algorithm for interpolation. These 2-D images represent the input data to our model which consists of Deep Convolutional Neural Network (DCNN) to extract the spatial and frequency features followed by Long Short Term Memory (LSTM) to extract temporal features and then finally to be classified into 5 different classes (4 motor imagery tasks and one rest). Our results were promising (70.64% average accuracy) and 5% better than the results of Support Vector Machine (SVM) method over the same dataset. We noticed that taking Delta band into account increases the classification accuracy by 2.51%.

Keywords*—Brain-Computer Interface (BCI); Classification; Motor Imagery; Convolutional Neural Networks (CNN); Long Short Term Memory (LSTM).

I. Introduction

Brain Computer Interface (BCI) is a communication system that translates brain signals into commands for an interactive application [1]. The most common noninvasive method to record brain activity is electroencephalography (EEG) due to direct measures of neural activities, inexpensiveness, and portability [2]. EEG signals have been used to control assistive devices for disabled people and for rehabilitation purposes [3].

There are many EEG-based BCI paradigms that have been deployed. The sensorimotor rhythms (SMR) paradigm is one of the most popular motor imagery paradigms in which the subject is asked to imagine a kinesthetic movement of large body parts such as hands and feet. This imagined movement modulates the brain activity especially on the motor cortex and causes Event-Related Desynchronization (ERD) and event-Related Synchronization (ERS) in Mu (8-13 Hz) and beta (18-26 Hz) rhythms [4].

In order to use a BCI, firstly, an offline training phase is needed for system collaboration. Secondly, the brain activity is translated into commands during the online phase.

The online BCI system is generally a closed-loop starting with EEG signals (e.g. using motor imagery) produced by the subject. These signals are pre-processed then the features are extracted, after that, appropriate features are selected then classified in order to be translated into commands for an application, finally, feedback is provided to the subjects containing information whether their mental commands were recognized or not [5].

The classification of the EEG features is the main part of BCI systems. The classification of motor imagery activity is a challenging task due to the low signal-to-noise ratio of EEG signals and their non-stationary nature for the same subject and among subjects, the high sensitivity for artifacts, the limited number of training data, and the low reliability of current BCI systems. Therefore, classification algorithms mainly aim to overcome one or more of the previously mentioned challenges. To overcome the low signal-to-noise ratio of EEG signals and the low reliability issues, Riemannian methods [6], [7], tensors and deep learning [8] classification techniques are used in which feature extraction, feature selection, and classification could be merged in a single step.

Neural networks approach has been investigated thoroughly and outperformed many other approaches in computer vision, natural language processing, and robotics, but still not investigated enough for MI-EEG signals classification because the network performance strongly depends on the architecture and metrics, and large input data is needed which is usually

This work was supported by the Hungarian National Research Development and Innovation Office, Thematic Excellence Program 2019, National Brain Research Program, 2017-1.2.1-NKP-2017-00002, National Bionics Program ED_17-1-2017-0009.

978-1-7281-4707-9/20/$31.00 ©2020 IEEE

not the case for EEG data.

The EEG input data for the neural network can be either signal values (e.g. raw data), extracted features (e.g. using Principal Components Analysis (PCA)) or images.

Although neural networks are generally the best choice for image classification, very few researchers deployed transforming MI-EEG signals into images for classification.

The MI-EEG data was preprocessed in [9] and only Mu and Beta bands have been chosen, then Common Spatial Patten (CSP) was used to extract features, after that Fast Fourier Transform Energy Map (FFTEM) was applied to generate energy maps between frequency and channels for feature selection and these images were fed to CNN model for classification. The method was used for single trials and using FFTEM ignores the time dependency between the EEG signal slices.

In [10], C3, C4, and Cz electrodes were used for 2-class MI tasks classification over 9 subjects. Short time Fourier transform (STFT) was applied on 2 seconds trials and then Mu and Beta bands images were extracted from the output spectrum for each of the 3 electrodes before one image was formulated for the 3 electrodes, and thus time, frequency and location information of the EEG signals was preserved. The resulted images formed the input to a CNN for feature extraction followed by Stacked Auto Encoder (SAE) for classification.

In [11], the wavelet transform for signals on C3, C4, and Cz was performed over each trial then the corresponding wavelet time-frequency images for C3 and C4 were combined together after excluding Cz which negatively affected the results, and finally the images were fed to a 2-layers CNN for classification.

A transfer learning method was introduced in [12]. Continuous Wavelet Transform (CWT) was chosen to generate scalogram images which had better time-frequency resolution than the spectrogram images generated by STFT, and then Deep CNN was used based on AlexNet model which was first trained on ImageNet and then the scalogram images were resized to get the same ImageNet images size and then fed to the network.

The goal in [13] was to classify memory/cognitive load. The EEG data gathered from 64 electrodes was transformed into colored images using 3 frequency bands, and these images were fed to a CNN followed by LSTM in order to keep the spatial, frequency and temporal features of the EEG signals. We used a similar approach for MI-EEG classification.

We can notice that choosing only few number out of the 64 electrodes can probably lead to more correlated data especially if the number is less than 4 and this will negatively affect the generalization ability of the neural network and more input data has to be fed to the network unless very precise preprocessing and feature extraction techniques are used. Secondly, the input size in the previous studies is relatively small and it is hard to make data augmentation to increase the input size because flipping or zooming the images is not suitable while adding noise which seems to be more reasonable for augmentation is not an easy mission for such noisy data. Thirdly, because the input size is relatively small, most of the proposed networks are not deep enough to extract the most important features. Finally, most of the studies ignore the time dependency between consecutive input images.

In our approach, we used CNN followed by LSTM and we tried to put into consideration all the above drawbacks by choosing a dataset with 109 subjects taken from 64 electrodes and by transforming the MI-EEG signals into images that preserve the spatial-frequency dimensions of the signals. LSTM was used to extract the temporal information between consecutive images that were classified into 5 different classes.

II. METHODS

A. The dataset

We used the Physionet dataset for EEG motor movement/imagery [14]. The recordings were taken for 109 subjects using the BCI2000 system and obtained from 64 electrodes in the 10-10 system at 160 samples/second. Each subject performed 14 runs, two baseline runs (one with eyes opened and the other with closed eyes) each of which lasts for one minute, followed by 4 different runs repeated for three times, and each run lasts for two minutes. The subject is asked to do one of the following four tasks: real opening and closing the left or right fist depending on whether the target appeared in the left or right side of the screen (task 1), imagined opening and closing the left or right fist (task 2), real opening and closing both fists or both feet (task 3), or imagined opening both fists or both feet (task 4). Each of the 12 runs has 30 tasks, and each task lasts for approximately 4 seconds and annotated as follows: T0 which corresponds to rest, T1 which corresponds to the left fist or both fists movement (real or imagined), and T2 which corresponds to the right fist or both feet (real or imagined). The subject is asked to rest (T0) after performing each task annotated as T1 or T2. We excluded 6 subjects (S043, S088, S089, S092, S100, and S104) because their recordings are not suitable [15], so our analysis is applied over 103 subjects.

Our goal is to classify the five different motor imagery tasks in the dataset (imagined left fist movement, imagined right fist movement, imagined both fists movement, imagined both feet movement, and rest), so we divided the tasks into 9 different classes and then we chose only the tasks corresponding to 5 classes (4 motor imagery movements and one rest).

B. Transforming EEG signals into 2-D images

As mentioned earlier, the motor imagery EEG signals activity is mainly dominant in two bands; Mu [8-13 Hz] and Beta [13-30 Hz], so for the first scenario, the signals for each

subject were bandpass filtered into two bands.
Then, for the second scenario, we also band pass filtered Delta band [0.5-4 Hz] to add another option to the selected bands for feature extraction. Next, we applied Fast Fourier Transform (FFT) on the time series with a window equals to 0.4 second (this means we have 10 different measures per each 4 seconds trial for every electrode), and then the sum of the squared absolute values was calculated for each band, so we got for the 2 bands scenario (64 x 2 x 10 = 1280) different measures for each 4 seconds task duration, and (64 x 3 x 10 = 1920) measures per task for the 3 bands scenario with Delta band. Then the measurements for 64 electrodes were transformed into 2-D images, and thus the problem of EEG signal classification becomes an image classification task..

In order to project the 3-D electrodes space into a 2-D plane, we applied the Azimuthal Equidistant Projection (AEP) in which the positions of the projected electrodes to a plane that is tangent to the head top point are equidistant - relative to that point - compared with the original 3-D space.

We applied Clough-Tocher interpolating algorithm [16] over 32 x 32 mesh and the results are shown in fig. 1, and thus we got the 2-D topology-preserving unicolor images for each band (red refers to Delta, green refers to Mu and blue refers to Beta) which represents the average activity for motor imagery EEG signals over the scalp for 0.4 seconds, then we added the resulting images for each band together to form either 2-channel images for the 2 band scenario or 3-channel images for the 3 bands scenario and so both scenarios have the same number of images. For each subject we have (30 x 10 x 6 = 1800) 2-D images (after excluding the two base runs and the real movement tasks), so we have in total 185400 images for 103 subjects which form the model input.

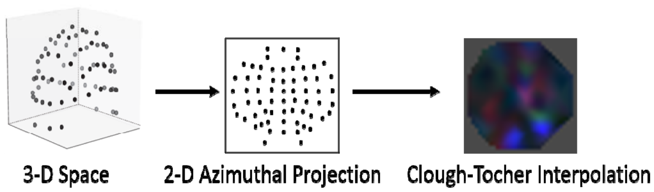

Fig. 1. Electrodes in the 3-D space projected using Azimuthal 2-D projection before interpolation.

C. *The Network Model*

The proposed model is composed of LSTM on the top of CNN. CNN is a multilayer neural network with several convolution and max pooling layers and a fully connected layer in the output. The standard CNN is composed of two parts, feature extraction and classification.

Input images are convolved with several 2-D filters in each convolution layer and then subsampled to a smaller size in the pooling layer. The weights and the filters are learned in the convolution layer by back-propagation to decrease the classification error. Our CNN is inherited by the VGG network architecture [17] which uses multiple convolutional layers stacked together with small receptive field.

We evaluated different configurations of the CNN by changing the following: the number of the stacked convolutional layers, the number of the filters, the size of the filters, the loss function, the drop rate, the learning factor, the batch size and other parameters. We obtained the best results by this network configuration: 3 stacked Convolutional layers (3-64) then max pooling layer then 2 convolutional layers (3-128) then another max pooling layer followed by 2 convolutional layers (3-256) followed by ReLU layer and finally max pooling layer. Stochastic Gradient Descent (SGD) optimization algorithm and cross entropy loss function were used, the batch size was 16 and the number of epochs was 20.

On the other hand, LSTM is a special type of Recurrent Neural Network which is able to remember long term time dependencies. It uses cells, input gate, forget gate, output gate, and cell activation vectors [18]. We used one layer LSTM and examined a different number of cells and best results were obtained for 256 cells. The main purpose of LSTM is to extract time dependent information between consecutive EEG images and these predictions will feed the fully connected layer which is followed then by a softmax layer. Fig.2 shows the proposed approach for motor imagery EEG classification.

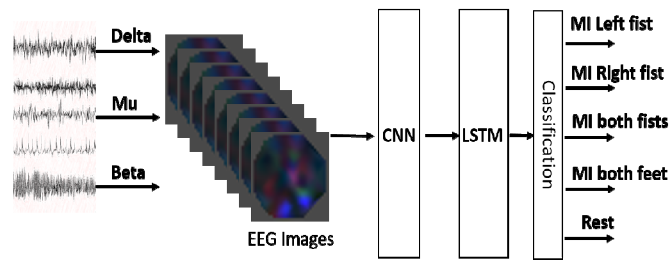

Fig. 2. The proposed approach for MI-EEG signals classification.

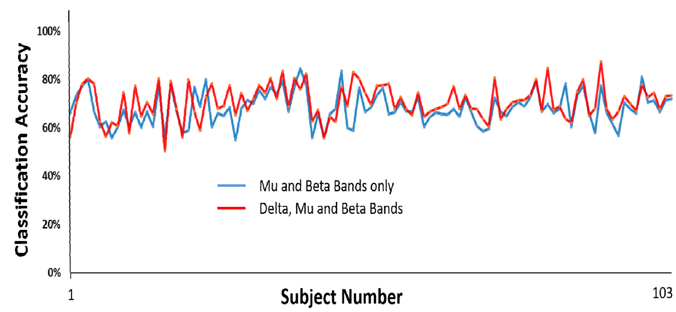

Fig. 3. Classification accuracy with the subject number for the two scenarios.

D. *Results*

We used the Leave-One-Out-Cross-Validation (LOOCV) for evaluating the performance of the model. All the 180 trials for one subject was used for testing and the trials for another randomly chosen subject was used for validation and the rest 101 subjects trials were used as the training set.

The average test accuracy was 70.64% (using 3 bands scenario) and 68.13% (using 2 bands scenario). Fig.3 shows the classification accuracy for both scenarios with the subject number.

The results are promising for 5-class motor imagery EEG classification and can be compared with the state-of-the-art methods. We compared our results with a famous baseline method for EEG classification which is Support Vector Machine (SVM) and our results showed 5% better classification accuracy.

III. Conclusion

Our investigation showed that CNN followed by LSTM can be used as a general classifier for motor imagery EEG-based BCI systems but they are strongly dependent on the architecture and configuration.

We reached 70.64% average classification accuracy by adding Delta band to the image generation process, which is usually ignored in MI-EEG signals classification approaches. Our results were better than SVM by 5%. We expect that by further investigations and applying different network architecture, the accuracy could be highly improved. The results also can be improved by pre-processing the dataset differently or using artifacts reduction techniques.

Choosing carefully a subset of the 64 electrodes which are the most relevant for motor imagery EEG signals classification can reduce the dimensionality of the input and this can increase the model accuracy.

IV. Future Work

We would like to try different CNN architectures like AlexNet, GoogleNet, ResNet, and also to try Gated Recurrent Units (GRU) network instead of LSTM and then test the models for different datasets as the results can be totally different according to the chosen dataset. Additionally, we would like to try different forms of input images.

References

[1] Youngjoo Kim, Jiwoo Ryu, Ko Keun Kim, Clive C. Took, Danilo P. Mandic, and Cheolsoo Park, Motor Imagery Classification Using Mu and Beta Rhythms of EEG with Strong Uncorrelating Transform Based Complex Common Spatial Patterns, Hindawi Publishing Corporation, Computational Intelligence and Neuroscience, 2016.

[2] D. J. McFarland and J. R. Wolpaw, EEG-based brain–computer interfaces, Current Opinion in Biomedical Engineering 2017, 4:194–200 .H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[3] Chuanqi Tan, Fuchun Sun,Big Fan, Tao Kong and Wenchang Zhang,Autoencoder-based transfer learning in brain-computer interface forrehabilitation robot, International Journal Of Advanced Robotic Systems, 2019 .

[4] Reza Abiri et al, A comprehensive review of EEG-based brain–computerinterface paradigms,J. Neural Eng.2019, 16, 011001.

[5] F Lotte1 ,L Bougrain, A Cichocki, M Clerc, M Congedo, A Rakotomamonjy and F Yger, A review of classification algorithms for EEG-basedbrain–computer interfaces: a 10 year update,J. Neural Eng. 2018, 15,031005.

[6] Congedo M, Barachant A and Bhatia R, Riemannian geometry for EEGbased brain–computer interfaces; a primer and a review Brain-Comput. Interfaces 2017, 4, 155–74.

[7] Congedo M, Barachant A and Kharati K, Classification of covariance matrices using a riemannian-based kernel for BCI applications IEEE Trans. Signal Process. 2016 , 65, 2211–20.

[8] Maitreyee Wairagkar, Motor Imagery based Brain Computer Interface (BCI) using Artificial Neural Networks Classifiers, University of Reading, 2016.

[9] W. Abbas and N. Khan, DeepMI : deep learning for multiclass motor imagery classification 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC) pp 219–22,2018.

[10] Y. Tabar and U. Halici, A novel deep learning approach for classification of EEG motor imagery signals J. Neural Eng. 14 016003, 2017.

[11] B. Xu, L. Zhang, A. Song, C. Wu, W. Li, D. Zhang, G. Xu, H. Li, and H. Zeng, Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification, IEEE Access, vol. 7, pp. 6084–6093, 2018.

[12] S. Chaudhary, S. Taran, V. Bajaj, A. Sengur, Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification, IEEE sensors, vol. 19, 2019.

[13] Pouya Bashivan, Irina Rish, Mohammed Yeasin, Noel Codella, Learning Representations from EEG with Deep Recurrent Convolutional Neural Networks, ICLR Conference, 2016.

[14] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R.Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system,", IEEE Transactions on Biomedical Engineering, vol. 51, no. 6,pp. 1034–1043, 2004.

[15] A. Loboda, A. Margineanu, G. Rotariu, and A. M. Lazar, Discrimination of EEG-based motor imagery tasks by means of a simple phase information method, International Journal of Advanced Research in Artificial Intelligence, vol. 3, no. 10, 2014.

[16] Alfeld, Peter. A trivariate cloughtocher scheme for tetrahedral data. Computer Aided GeometricDesign, 1(2):169–181, 1984.

[17] Simonyan, K and Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In ICLR, pp. 1–14, 2015.

[18] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.