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

Bi-dimensional representation of EEGs for BCI classification using CNN architectures

Edgar Hernández-González[1], Pilar Gómez-Gil[1], Erik Bojorges-Valdez[2], Manuel Ramírez-Cortés[3]

Abstract— **An important challenge when designing Brain Computer Interfaces (BCI) is to create a pipeline (signal conditioning, feature extraction and classification) requiring minimal parameter adjustments for each subject and each run. On the other hand, Convolutional Neural Networks (CNN) have shown outstanding to automatically extract features from images, which may help when distribution of input data is unknown and irregular. To obtain full benefits of a CNN, we propose two meaningful image representations built from multi-channel EEG signals. Images are built from spectrograms and scalograms. We evaluated two kinds of classifiers: one based on a CNN-2D and the other built using a CNN-2D combined with a LSTM. Our experiments showed that this pipeline allows to use the same channels and architectures for all subjects, getting competitive accuracy using different datasets: 71.3 ± 11.9% for BCI IV-2a (four classes); 80.7±11.8 % for BCI IV-2a (two classes); 73.8± 12.1% for BCI IV-2b; 83.6± 1.0% for BCI II-III and 82.10%± 6.9% for a private database based on mental calculation.**

Keywords: 2D signal representation, Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), Short Time Fourier Transform (STFT), Continuous Wavelet Transform (CWT).

I. INTRODUCTION

Brain-computer interfaces (BCI) provide a direct channel for communication and control between human brain and external devices [1], many of them based on electroencefalograms (EEG). The design of a successful BCI pipeline process depends on the wise selection of good strategies for three main steps: signal conditioning, feature extraction and classification. In general, hyper-parameters for these steps must be tuned for each subject when used BCI for several subjects and environments, making difficult to deploy these models. Attempts have been made to design automatic inter-subjects models based on Convolutional Neural Networks (CNN) (for example [2]) because they are able to automatically extract the most discriminatory features from textures in an image [3]. However, in order to explode this capability, it is mandatory to represent meaningful time-space information into each image pixel and into its neighbour relationship. Recently, several works have proposed image representations and architectures using them. For example, Olivas-Padilla and Chacon-Murguia [4] proposed a model called Discriminative Filter Bank Common Spatial Pattern (DFBCSP), which is a modification of the Filter Bank Common Spatial Pattern (FBCSP). This model looks for enhancing those frequency bands that most discriminate among mental activities; features were classified using two CNNs, with monolithic and modular structures respectively. Xu and collaborators [5] presented a model based on scalogram estimation of EEG, using Wavelet transforms at the C3 and C4 channels and combining them in a single image as input for a CNN. Tabar and Halici [6] used *STFT* to extract features from EEG signals classified with a CNN-1D followed by a Stacked Auto Encoder (SAE). However, all of these models required to define the CNN's hyper-parameters for each new set of data subjects being used.

In this paper, we present a simple but robust design of an image representation of EEG segments, using a combination of a CNN layer with a long-short term memory network (LSTM) and a fully connected (FC) network. The main contribution of this design is a pipeline which allows to get meaningful frequencies and organize them in arrangements of rows and columns, which maps the relationship of EEG with respect to time and spatial positions of the electrodes. This topological distribution allows a classification more tolerant to intra-subject deployment than other models, at least in two BCI domains: one based on motor imagery (MI) and the other based on mental calculation activities, using the same parameters for each subject.

II. METHODOLOGY

A. Signal Conditioning

It is common that recordings of EEG signals present noise and/or missing values or artifacts. To replace missing values in EEG we applied a median filter of 255 taps as well as referencing the signal using the common average reference (CAR) method. As usual, our signal conditioning process included a band-pass filter from 8 to 30 Hz, in order to preserve brain waves found in α (8–13 Hz) and β (13–30 Hz) bands, wich are described by several researchers as good indicators of mental activity.

B. Primary feature extraction

Our proposed image representation of EEG was built from data showing a meaningful relationship among EEG frequencies, spatial location of electrodes and time. We evaluated representations obtained from spectrograms, calculated with the Short-time Fourier transform (STFT) and scalograms calculated by a Continous Wavelet Transform (CWT) based on the complex Morlet wavelet; channels were concatenated

*E. Hernández-González thanks to the National Council of Science and Technology (CONACYT) in México.

[1]E. Hernández-González and P. Gómez-Gil are with the computer science department in the National Institute of Astrophysics, Optics and Electronics, Puebla, México.

[2]E. Bojorges-Valdez is with the Engineering for Innovation Department in the Universidad Iberoamericana Ciudad de México, México.

[3]M. Ramírez-Cortés is with the Electronics Department in the National Institute of Astrophysics, Optics and Electronics, Puebla, México.

TABLE I: Parameters of the Short-time Fourier transform (STFT) for the data sets used as benchmarks.

STFT	*BCI IV-2a*	*BCI IV-2b*	*BCI II-III*	*MC*
Window	Hanning	Hanning	Hanning	Hanning
Window size	250	250	128	512
Overlapping	225	225	112	480
FFT size (NFFT)	500	500	256	1024
Image	45×31	45×31	45×17	45×27
Concatenated	990×31	135×31	90×17	1440×27

TABLE II: Size in pixels of the scalograms built using *CWT*.

CWT	*BCI IV-2a*	*BCI IV-2b*	*BCI II-III*	*MC*
Image	45×1000	45×1000	45×384	45×1344
Concatenated	990×1000	135×1000	90×384	1440×1344
Resize	495×500	67×500	90×384	720×672

vertically. The specific parameter values chosen to calculate these transforms and the size of spectograms for each data set used in the experiments are detailed in section III; sizes of final images are showed in the Table I.

Scalogram were calculated using a CWT based on complex Morlet wavelets (cmor3-3) and Mexican hat, being cmor3-3 the one that obtained the best results. Scales were calculated every 0.5 Hz and channels were vertically concatenated; scalograms were resized using an inter-area interpolation, in order to reduce the computational cost during training the CNN. The resulting size of scalograms is showed in the Table II.

C. Classification

Two classifiers were evaluated: a *CNN-2D* and *CNN-2D + LSTM*. The *CNN-2D* automatically extracts features in the convolutional layers and classifies them using a fully connected (FC) layer, while the *CNN-2D + LSTM* uses a combination of a LSTM + FC layers for classification. The idea behind using LSTM is to represent the temporal information embedded in the EEG signals. Table III shows the size of segments and number of channels used for each dataset; images were normalized to [0,1] range. Both (*CNN-2D* and *CNN-2D + LSTM*) have the same convolutional layer structure, which consists of two convolutional layers with activation function ReLU, padding and a max-pooling of 2 $\times$ 2; Fig. 1 shows these architectures.

For the LSTM layer, activation is performed by a hyperbolic tangent function and the last layer of the network is built using Softmax. The dropout used in the hidden layer of the MLP and in the LSTM layer was 0.5. For all networks, Adam optimizer was used with a learning rate of 10^{-4}; the loss function was the categorical cross entropy, batch size was 36 for *BCI IV-2a* and *BCI IV-2b*, 16 for *BCI II-III* and 20 for *mental calculation*. Training was executed using 400 iterations when *STFT* was used for feature extraction and 100 iterations when *CWT* was applied. Hyperparameters of both architectures were selected using a grid search, detailed in (Table IV). Table V and the Fig. 1 show the selected hyperparameters for *CNN-2D* and *CNN-2D + LSTM* networks.

TABLE III: Input sizes for each dataset.

Data set	Signal segment	Number of channels
BCI IV-2a	4 seconds	22
BCI IV-2b	4 seconds	3
BCI II-III	3 seconds	2
mental calculation	2.625 seconds	32

TABLE IV: Grid search values.

Hyperparameters	Possible values
number of filters	2, 4, 8, 16, 32, 64
filters size	(3,3), (15,3)
number of neurons	16, 32, 64, 128, 256
number of units	4, 8, 16, 32

Other training approaches, such as batch normalization, transfer learning and fine-tuning were analyzed, but no improvement in performance was obtained. Also, we analysed the use of an image representing each channel as a frame in a video-style input feeding. However, this results on worse performance than the one-image approach and more computationally expensive. A data augmentation approach was also tested observing an improvement using a few subjects. But, we were not able to test the complete data sets due to training time constraints.

III. EXPERIMENTS AND RESULTS

Our architecture was evaluated using these popular benchmarks: *BCI Competition IV dataset 2a* [7], *BCI Competition IV dataset 2b* [7], *BCI Competition II dataset III* [8] and a private database, based on *mental calculation*, provided to us by its creator [9]. All deep neural networks were developed using Keras and experiments were performed using Google Colab platform. Code is free available at github.

Table VI shows the accuracy of five independent executions using the training and testing sessions provided by the BCI competition. Notice that for *BCI IV-2a* (2 classes: *left hand and right hand*), *BCI IV-2b* and *BCI II-III*, the best results were obtained using a *STFT* representation and the *CNN-2D + LSTM* neural network. Indeed, for *BCI IV-2a* using all 4 classes, the best results were achieved using the *CWT* representation and the *CNN-2D* neural network. Table VI shows average accuracy calculated using ten-cross validation for *BCI IV-2a* (4 classes), *BCI IV-2a* (2 classes: *left hand and right hand*) and *BCI IV-2b* ten-fold cross-validation. Same table includes the results using five-fold cross-validation for *BCI II-III* and *mental calculation*. For data sets *BCI IV-2a* (2 classes), *BCI IV-2b* and *BCI II-III*, the best results were obtained using *STFT* and *CNN-2D*; for *BCI IV-2a* (4 classes) and *mental calculation* data sets the best results were achieved using *CWT* and *CNN-2D + LSTM*.

Table VII shows the results of the proposed architecture along with several state-of-the-art work for each set. An AUC-ROC of 0.89$\pm$0.01 was obtained using five-fold cross-validation, slightly overcoming the work of [9].

In general, the *CNN-2D* and *STFT* were the fastest computational schemes. According to [19] the average training time per subject using *BCI IV-2a* was 2,580 seconds; in [20] the

TABLE V: Hyper-parameters of the *CNN-2D* and *CNN-2D + LSTM* networks.

Dataset	*IV-2a*	*IV-2a*	*IV-2b*	*IV-2b*	*II-III*	*II-III*	*MC*	*MC*
Repre	*STFT*	*CWT*	*STFT*	*CWT*	*STFT*	*CWT*	*STFT*	*CWT*
Input (a)	990x31	495x500	135x31	67x500	90x17	90x384	1440x27	720x672
CNN-2D (b)	16@3x3	16@3x3	4@3x3	8@3x3	64@3x3	8@3x3	2@3x3	8@15x3
Dense (CNN-2D) (d)	128	128	32	16	256	128	16	32
LSTM (c)	8	32	4	8	32	16	4	8
Dense (LSTM) (d)	128	256	32	16	256	128	16	32

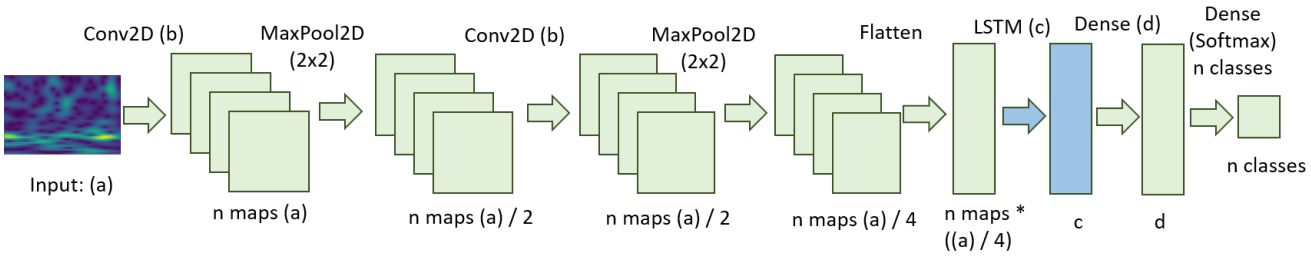

Fig. 1: Proposed architecture, *CNN-2D* (green network) and *CNN-2D + LSTM* (green network plus blue network).

TABLE VI: Average Accuracy for the different data sets using the training and testing sessions provided by the competition calculated using cross validation.

	CNN-2D		*CNN-2D + LSTM*	
Data set	*STFT*(%)	*CWT*(%)	*STFT*(%)	*CWT*(%)
IV-2a (4)	66.4±10.3	**67.0±11.4**	65.9±9.9	66.9±11.0
IV-2a (4) CV	68.4±12.5	69.9±12.1	68.0±12.0	**71.3±11.9**
IV-2a (2)	79.7±12.3	80.6±12.2	**80.7±11.8**	80.3±12.4
IV-2a (2) CV	**79.9±11.2**	79.7±12.4	79.5±11.3	79.9±12.6
IV-2b	73.4±13.2	73.0±12.5	**73.8±12.1**	73.1±12.7
IV-2b CV	**69.9±12.2**	69.6±12.2	69.6±12.5	69.2±11.9
II-III	82.9±0.3	81.4±1.3	**83.6±1.0**	82.9±1.6
II-III CV	**83.6±7.3**	81.8±6.0	83.2±7.2	81.4±3.6
MC CV	76.6±8.4	81.7±7.0	74.7±9.2	**82.1±6.9**

TABLE VII: Accuracy values achieved for our proposal and reported values for the different data sets with the training sessions and tests and with cross validation (CV) (BCI IV-2a (4): 4 classes, BCI IV-2a (2): 2 classes).

Data sets	Related works	Our proposal
IV-2a (4)	**70.6±14.7** [10], 67.7±12.9 [11], 65.5±15.0 [12]	67.0±11.4
IV-2a (4) CV	69.2±14.1 [13]	**71.3±11.9**
IV-2a (2)	**85.0±7.4** [4]	80.7±11.8
IV-2a (2) CV	79.4±13.9 [14]	**79.9±11.2**
IV-2b	**84.2±11.3** [15], 80.0±13.0 [11]	73.8±12.1
IV-2b CV	**77.5±7.7** [6], 76.1±8.4 [16]	69.9±12.2
II-III	**90.1±1.0** [17], 90.0 [6], 89.2 [8]	83.6±1.0
II-III CV	**91.4** [18], 89.5 [5]	83.6±7.3

reported time was 3,300 seconds; in [21] the average training time per subject was 1.328 seconds and the time to predict an example was 0.05 ms. However, in our proposal, for *BCI IV-2a* using *CWT* and *CNN-2D* (the combination that generates the best results while not the fastest), the training time is 93 seconds; in this way, state of the art works is surpassed with respect to training time. With respect to prediction time, our proposal requires 4 ms, which may be considered reasonable. With respect to *BCI IV-2b*, the work of [6] reported a training time of 1,157 seconds and a prediction time of 400 ms; in [16] a training using only 60 samples took 18 seconds. In the other hand, our proposal took 42 seconds for training using 400 samples and 1.09 ms for a prediction. For more details on the results of this research, see [22].

IV. DISCUSSION

Our proposal has a higher classification performance when the data set has many channels as in *BCI IV-2a* (22 channels) and *mental calculation* (32 channels). Indeed, the best results are obtained when the signal is longer in time (except for *BCI II-III*); most of the published works showed the opposite, due to high dimension in data. With respect to the two proposed representations, *STFT* was better for few channels (*BCI IV-2b* and *BCI II-III*) while *CWT* was better for many channels (*BCI IV-2a* and *mental calculation*).

The neural networks *CNN-2D* and *CNN-2D + LSTM* may be considered shallow, since they only have two convolutional layers with max-pooling and one FC layer. Besides, they are very similar to each other, their only difference is that *CNN-2D + LSTM* network contains an LSTM layer before the FC layer; notice that actually none of them outperforms the other one with respect to performance.

It is important to point out that several works do not report results using cross-validation, but just accuracy for each subject. Even BCI paradigms has been used long time ago, there is no standard about how to correctly report their performance and this issue is specially troubling when test are done over static datasets and no-online implementations. In this work the mean and standard deviation of five independent folds were reported, which gives a better information about the behavior of the model during other executions.

V. CONCLUSIONS

We present a pipeline for BCI that is able of classifying EEG signals from *motor imagery* or *mental calculation*, using DL techniques that uses CNNs alone or combined with LSTM networks. In this work, the same preprocessing,

number of channels and network architecture were used for all subjects of each data set. This gives some insight that our framework is more versatile than others. In addition, our proposed obtained reliable results both in the classification of *motor imagery* and *mental calculation*. Our experiments showed that CNN architectures are able to automatically extract features from the *STFT* and *CWT* representations to build robust 2D images. Indeed, the training time of the proposed model overcame the time reported in other works. This was because our network contains only two convolutional layers. Considering that the prediction time is fast (from 1.09 ms to 1.69 seconds), our proposal is useful to implement pre-processing, feature extraction and classification phases of a BCI system. As future work, we propose to analyze other pre-processing and statistical techniques over the generated 2D image in order to enhance the differences among them with respect to each BCI class keeping inter-subject invariance.

REFERENCES

[1] Jonathan R Wolpaw, Niels Birbaumer, William J Heetderks, Dennis J McFarland, P Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A Quatrano, Charles J Robinson, and Theresa M Vaughan. Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering*, 8(2):164–173, 2000.

[2] M. Jiménez-Guarneros and P. Gómez-Gil. Cross-subject classification of cognitive loads using a recurrent-residual deep network. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–7, 2017.

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[4] Brenda E Olivas-Padilla and Mario I Chacon-Murguia. Classification of multiple motor imagery using deep convolutional neural networks and spatial filters. *Applied Soft Computing*, 75:461–472, 2019.

[5] Baoguo Xu, Linlin Zhang, Aiguo Song, Changcheng Wu, Wenlong Li, Dalin Zhang, Guozheng Xu, Huijun Li, and Hong Zeng. Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification. *IEEE Access*, 7:6084–6093, 2018.

[6] Yousef Rezaei Tabar and Ugur Halici. A novel deep learning approach for classification of EEG motor imagery signals. *Journal of neural engineering*, 14(1):016003, 2016.

[7] Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot Mueller-Putz, et al. Review of the BCI competition IV. *Frontiers in neuroscience*, 6:55, 2012.

[8] Steven Lemm, Christin Schafer, and Gabriel Curio. BCI competition 2003-data set III: probabilistic modeling of sensorimotor/spl mu/rhythms for classification of imaginary hand movements. *IEEE Transactions on Biomedical Engineering*, 51(6):1077–1080, 2004.

[9] Erik Bojorges-Valdez, Juan C Echeverría, and Oscar Yanez-Suarez. Evaluation of the continuous detection of mental calculation episodes as a BCI control input. *Computers in biology and medicine*, 64:155–162, 2015.

[10] Siavash Sakhavi, Cuntai Guan, and Shuicheng Yan. Parallel convolutional-linear neural network for motor imagery classification. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 2736–2740. IEEE, 2015.

[11] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in neuroscience*, 6:39, 2012.

[12] Jing Luo, Jie Wang, Rong Xu, and Kailiang Xu. Class discrepancy-guided sub-band filter-based common spatial pattern for motor imagery classification. *Journal of neuroscience methods*, 323:98–107, 2019.

[13] Huijuan Yang, Siavash Sakhavi, Kai Keng Ang, and Cuntai Guan. On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2620–2623. IEEE, 2015.

[14] Shan Guan, Kai Zhao, and Shuning Yang. Motor imagery EEG classification based on decision tree framework and riemannian geometry. *Computational intelligence and neuroscience*, 2019, 2019.

[15] Na Lu, Tengfei Li, Xiaodong Ren, and Hongyu Miao. A deep learning scheme for motor imagery classification based on restricted boltzmann machines. *IEEE transactions on neural systems and rehabilitation engineering*, 25(6):566–576, 2016.

[16] Sumit Soman et al. High performance EEG signal classification using classifiability and the twin SVM. *Applied Soft Computing*, 30:305–318, 2015.

[17] Zhiwen Zhang, Feng Duan, Jordi Solé-Casals, Josep Dinarès-Ferran, Andrzej Cichocki, Zhenglu Yang, and Zhe Sun. A novel deep learning approach with data augmentation to classify motor imagery signals. *IEEE Access*, 7:15945–15954, 2019.

[18] Jie Zhou, Ming Meng, Yunyuan Gao, Yuliang Ma, and Qizhong Zhang. Classification of motor imagery EEG using wavelet envelope analysis and LSTM networks. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 5600–5605. IEEE, 2018.

[19] Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed Amine Mekhtiche, and M Shamim Hossain. Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Generation Computer Systems*, 101:542–554, 2019.

[20] Syed Umar Amin, Mansour Alsulaiman, Ghulam Muhammad, Mohamed A Bencherif, and M Shamim Hossain. Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification. *IEEE Access*, 7:18940–18950, 2019.

[21] Sahar Selim, Manal Mohsen Tantawi, Howida A Shedeed, and Amr Badr. A CSP\AM-BA-SVM approach for motor imagery BCI system. *IEEE Access*, 6:49192–49208, 2018.

[22] Edgar Hernández-González. Clasificación de señales EEG basada en representaciones bidimensionales y redes neuronales convolucionales. Master's thesis, INAOE, 2020.