

“© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Classification of EEG Motor Imagery Tasks Using Convolution Neural Networks

Sai Ho Ling, *Senior Member, IEEE*, Henry Makgawinata, Fernando Huerta Monsivais, Andre dos Santos Goncalves Lourenco, Juan Lyu, Rifai Chai, *Senior Member, IEEE*,

Abstract— Electroencephalograph (EEG) is a highly nonlinear data and very difficult to be classified. The EEG signal is commonly used in the area of Brain-Computer Interface (BCI). The signal can be used as an operative command for directional movements for a powered wheelchair to assist people with disability in performing the daily activity. In this paper, we aim to classify Electroencephalograph EEG signals extracted from subjects which had been trained to perform four Motoric Imagery (MI) tasks for two classes. The classification will be processed via a Convolutional Neural Network (CNN) utilising all 22 electrodes based on 10-20 system placement. The EEG datasets will be transformed into scaleogram using Continuous Wavelet Transform (CWT) method. We evaluated two different types of image configuration, i.e. layered and stacked input datasets. Our procedure starts from denoising the EEG signals, employing Bump CWT from 8-32 Hz brain wave. Our CNN architecture is based on the Visual Geometry Group (VGG) network. Our results show that layered image dataset yields a high accuracy with an average of 68.33% for two classes classification.

I. INTRODUCTION

A Brain-Computer Interface (BCI) decode EEG signal and transform it into useful messages [1]. One application of such signal translation is for a control system to move a powered wheelchair for people who have lost the motoric ability. For instance, joystick control using the subject's chin movement as a control mechanism can be implemented. However, this method is deemed difficult to use and non-attractive [2]. EEG signal provides hands-free access to control a device for people with severe motor disabilities simply by imagining motor movements of their body limbs.

By imagining limb movement, the brain cortex experiences oscillation of actives which can be captured by non-invasive EEG electrodes placed on the subject's scalp. This overall brain activities can be classified to determine the underlying motor imagery tasks being performed by a pattern recognition method involving a neural network.

Sai Ho Ling, Henry Makgawinata, Fernando Huerta Monsivais, Andre dos Santos Goncalves Lourenco, Juan Lyu are with Faculty of Engineering and Information Technology, University of Technology, Sydney (UTS), Broadway NSW 2007.

Rifai Chai is with Department of Telecommunications, Electrical, Robotics and Biomedical Engineering School of Software and Electrical Engineering Faculty of Science, Engineering and Technology, Swinburne University of Technology, Hawthorn, VIC 3122, Australia, e-mail: Rchai@swin.edu.au.

The versatility of Artificial Neural Network had been utilised in various fields commonly for data modelling, pattern recognition or classification purpose to analyse complex data [3]. Wavelet transform has been widely used to detect correlations in the EEG signals [4] and [5]. The Continuous Wavelet Transform (CWT) can be used to analyse the EEG signal and is revealed to be a good model for EEG signal feature extraction [5]. Deep learning has gained popularity in the computer vision field due to the more powerful graphics processing unit in recent time. There has not been any significant improvement in EEG classification using deep learning methods in recent years [6].

Most MI classification involved two classes which were classified with preselected EEG channels based on 10-20 International system placement. For example, both [7] and [8] classified Left and Right movements with only C4, Cz and C3 electrodes. These electrodes are located in the motor area of the brain. In [7], the EEG signal was processed via Short-time Fourier transform (STFT) to create a 2D image. Meanwhile, CWT was used in [8]. A Convolutional Neural Network (CNN) was then employed to classify their EEG images which were configured in stacked formats.

Using a higher number of electrodes have been observed to provide an advantage in identifying significant oscillation in brain activities. In adults, the use of a higher number of electrodes, i.e. 64-256 electrodes, allow visual brain activity down to the sub-lobar precision. A higher number of electrodes also shows to lower the possibility of spatial aliasing [9].

Hence, in this study we present the approach to use all 22 channels based on 10-20 International system of electrode placement to classify two classes of MI. Our contribution can be summarised that using all 22 channels we evaluated a strategy of splitting the EEG datasets into layered and stacked image configuration. Instead of having the 2D images being stacked and fed into CNN, we propose to generate the EEG images into its electrode image – creating 22 layered images as a data input for the network.

II. METHODOLOGY

A. Dataset

We used EEG raw data 2a four class MI from BCI competition IV [10] by Graz University of Technology, Austria. There were nine subjects participated during the experiment. Each subject performed four MI tasks, namely left hand, right hand, tongue and feet movements. Each MI data is extracted from the 22 electrodes based on 10-20 system

placement. Sampling rate frequency was set at 250 Hz. A notch filter at 50 Hz was applied and a bandpass filter applied between 0.5 and 100 Hz.

The experiment was conducted in two sessions. The EEG extraction process is shown in Fig. 1. At $T = 0$, a monitor in front of the subject displays a fixation cross. At $T = 2$, an image of an arrow pointing to either of the four directions appears, corresponding to the motor imagery tasks. Then, the mental tasks start from $T = 3$ to $T = 6$. At $T > 6$, the fixation arrow disappears. Each session contains nine runs, and in total there were 288 trials collected from each subject.

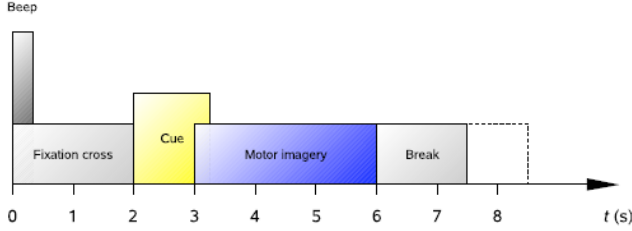


Fig. 1. EEG data extraction model [10]

B. EEG signal processing

The EEG datasets in [10] contained Training and Evaluation. In this study, Evaluation datasets were selected to be analysed. The signal processing and images generation were conducted in Matlab environment. We extracted the data from all 22 channels for every task from each subject and applied the CWT to the EEG motor images data using a Bump wavelet. Then, we generate an image reconstruction of the EEG signal as shown in Fig. 2 with the dimensions of $875 \times 1167 \times 3$, where the y-axis represents frequency in Hertz (Hz), and the x-axis represents time in seconds (s). The parabolic dashed line denotes cone of influence. It shows where the edge effects of the CWT become significant.

Since EEG signals are known for its non-stationary and non-linear characteristic, the CWT can preserve the information in the time-frequency domain by using a windowing technique with variable-sized regions for the analysing function [4]. Long-time intervals are used at low frequencies, and shorter regions are used at higher frequencies. The CWT mathematical formula is given by the expression

$$CWT(\omega, s) = \frac{1}{\sqrt{|s|}} \int f(t) \psi^* \left(\frac{t-\omega}{s} \right) dt \quad (1)$$

where ψ is the wavelet, ω is the scaling and s the shifting parameter [8].

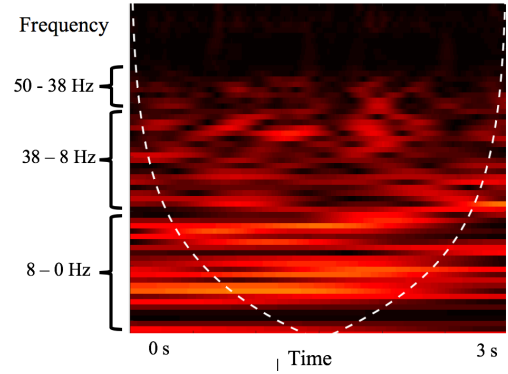


Fig. 2. CWT reconstruction of the EEG signal

C. EEG scaleogram processing

From the image reconstruction obtained illustrated in Fig. 2, the image is then converted from an RGB image to grayscale [0 255], where 0 is black, and 255 is white. Therefore, each image for each channel for a task will be represented by a matrix of 875×1167 , reducing the number of parameters by a factor of 3. Further, we reduce the image size by 80% to allow a faster performance with the CNN. With this, we have a matrix of 175×234 for each channel for each task. A study by [11] suggests that 8-13 Hz and 15-38 Hz band of frequencies are the most relevant for this type of brain activity. An event-related desynchronization (ERD) is the decrease of the power spectral density in the frequency band (8-13 Hz) which occurs during an MI task. In contrast, there is an increase in the frequency band (15-38 Hz) which is called an event-related synchronization (ERS). As a result, the image is cropped so that it contains only range of frequencies 8-38 Hz. Therefore, we have a final image represented in Fig. 3 for one channel for one task with the dimensions of 53×166 . For each subject, we have a 4-dimensional matrix with the size of $288 \times 22 \times 53 \times 166$.



Fig. 3. Final size of the CWT reconstruction of the EEG signal

Two types of EEG image generated from here, i.e. layered and stacked version as shown in Fig. 4. Each version contains the EEG data from the 22 channels. Fig. 4 (a) shows the illustration of the layered images containing 22 images being layered in a 3D alike format. Each image represents each electrode. Fig. 4 (b) is an example of the 2D image of scaleogram of the EEG being stacked as one image. The size of each image in the layered version is 53×166 . The size of the stacked image is 583×83 . These images will be the input dataset for the CNN learning scheme.

III. RESULTS

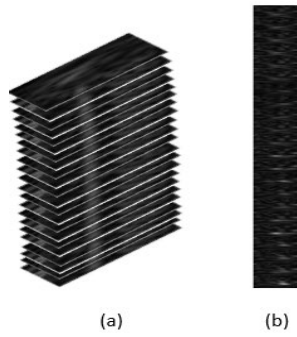


Fig 4. Scaelogram derived from CWT (a) 22 layered image version (b) a stacked image version from 22 channels

D. CNN structure

The use of CNN architecture in this project was based on the Visual Geometry Group (VGG) Network. The two types of image configurations were tested: layered and stacked, creating 3D and 2D datasets, respectively. Our CNN configuration is illustrated in Fig. 5.

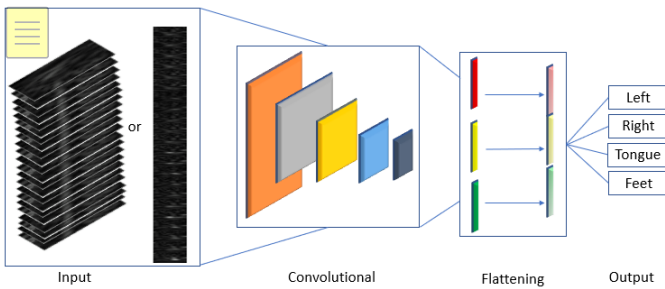


Fig. 5. An illustration of our 1D Convolutional Neural Network

This configuration consists of five hidden layers, each of which contains two convolutional and one pool layers. The filter (or kernel) size was kept at 3×3 . A pooling layer downsamples the input data along spatial dimensions, resulting in an image with a smaller volume. Our pooling layers require two hyperparameters: the spatial extent and stride, both of which were set at 2×2 , which is one of the most common configurations of these layers [12].

The input dataset was divided into 70% for training and 30% for testing. To avoid overfitting, the number of images set for training was augmented by rotating each image 180 degrees and then concatenated it to the original data, which artificially increases the number of samples. The network was trained for 100 epochs each time, and the batch size of the dataset was set to 40. A total of three dropout layers to lessen the number of neurons to a given percentage to avoid their reliance with others, were also implemented with 80%, 50% and 50%, respectively, in our Neural Network structure.

The CNN processing was completed via the University of Technology Sydney ARCLab Supercomputers, using Atlas and Orion clusters. Among their specifications, Atlas consists of an Intel Xeon 2x 3.1 GHz CPU with 2 x 8GB NVIDIA Quadro P4000 GPU. On the other hand, Orion counts with an Intel Xeon 2.9 GHz CPU and from 5 to 8GB NVIDIA Quadro M4000/P4000 GPUs, depending on which node was used.

The process of testing the dataset was repeated three times and the average was calculated. However, the testing was conducted only once if the trend of the result was very obvious such that the network constantly randomly distributed the classification accuracy. This is very noticeable when testing stacked image configuration displaying a very low classification accuracy result.

Table 1 and Table 2 show the result of binary classification between the layered and stacked images using 22 channels. The average accuracy for layered image overall is on average of 68.33%. The highest accuracy is for Left and Tongue classes at 74.32%. The next best accuracy is recorded for Right and Tongue movement at 69.46%. The highest average accuracy is recorded for subject nine with an average of 85.98%. For the same subject, highest accuracy recorded is for Left and Tongue movement with an average of 96.97%. The Left and Right movement for subject nine is also reasonably distinguishable by the network with a classification accuracy of 87.12%.

For comparison purpose, we also tried to classify only using three channels of C3, Cz and C4 with layered image configuration for Left and Right classes for all subjects as shown in Table 2. The overall average accuracy is at 63.43%. The accuracy at subject level varies with subject three being the highest accuracy yield at 87.50% followed by subject nine at 86.57%. The use of 22 channels for Left and Right classes still shows that 22 channels yield a slightly higher accuracy average at 68.32% compared to the preselected three channels described here.

Based on our result, it appears that the accuracy of the movement classification heavily depends on the integrity of the input dataset which corresponds to [13]. At the subject level, it is obvious that some subjects correspond to the MI tasks better than others. This is reflected in the result for subject two, five and six where classification between three preselected channels and 22 channels show low accuracy for these three subjects. The result of the stacked image configuration appears to be randomly distributed, and no pattern can be detected by the network. Stacked images may be more applicable to a low number of channels. This strongly suggests that with the increasing number of electrodes, it is best to employ a layered version of image configuration with CNN scheme to achieve higher classification accuracy.

Classification research by [14] used two types of datasets, and one of the datasets was from the same BCI dataset 2a. Although it shows high classification accuracy by employing the regularized approach for Filter bank Common Spatial Pattern, it appears that their datasets were augmented with another dataset from Wadsworth physiobank data. Hence, the datasets were sourced and combined from two different datasets. Previous studies such as [9], [10] had only been performed with a low number of channels. Thus, to our knowledge, this is the first study ever conducted with 22 channels utilising layered image configuration.

Table 1. Comparison of accuracy (%) and standard deviation of two classes MI classification for all subjects with 22 channels – Layered Images.

Subject	2 Classes Layered Images Accuracy % (mean ± standard deviation)						Subject average
	Left vs Feet	Left vs Right	Left vs Tongue	Right vs Feet	Right vs Tongue	Feet vs Tongue	
1	86.68	68.18	92.04	90.91	96.59	56.81	81.87
2	45.45	63.64	65.91	52.27	54.55	56.81	56.44
3	85.23	82.96	78.41	87.50	87.50	64.78	81.06
4	55.17	64.58	68.06	55.56	69.70	57.57	61.77
5	52.27	47.72	52.27	50.00	50.00	47.72	50.00
6	63.89	52.27	60.60	59.26	59.26	58.34	58.94
7	65.28	69.69	74.07	60.65	64.35	58.80	65.47
8	63.89	78.70	80.55	62.04	75.00	80.30	73.41
9	96.21	87.12	96.97	83.33	68.18	84.09	85.98
Average	68.23	68.32	74.32	66.83	69.46	62.8	68.33
SD	17.33	13.21	14.43	15.89	15.1	11.86	12.75

Table 2. Comparison of accuracy (%) and standard deviation of two classes MI classification for all subjects with 22 channels – Stacked images.

Subject	2 Classes Stacked Images Accuracy % (mean ± standard deviation)						Subject average
	Left vs Feet	Left vs Right	Left vs Tongue	Right vs Feet	Right vs Tongue	Feet vs Tongue	
1	54.54	50.00	52.27	50.00	40.91	52.27	50.00
2	50.00	56.82	52.27	47.73	52.27	47.73	51.14
3	47.73	53.41	52.27	47.73	48.86	57.96	51.32
4	50.76	49.24	50.76	53.03	48.49	50.76	50.50
5	45.45	50.00	46.59	53.41	43.18	46.97	47.60
6	53.41	53.41	53.41	52.27	44.32	47.73	50.76
7	47.73	48.87	53.41	44.32	50.00	46.59	48.48
8	56.82	51.14	54.55	50.00	50.00	59.57	53.68
9	52.27	50.76	51.14	46.59	55.68	46.59	50.15
Average	50.97	51.52	51.85	49.45	48.19	50.68	50.44
SD	3.66	2.57	2.29	3.11	4.65	4.99	1.73

Table 3. Accuracy (%) and standard deviation of MI Classification of Left and Right movements for all subjects

3 Channels (C3, Cz and C4) Left vs Right Classification (mean ± standard deviation)	
Subject	Layered
1	53.24
2	48.15
3	87.50
4	61.11
5	44.91
6	51.39
7	63.89
8	74.07
9	86.57
Average	63.43 ± 16.04

IV. CONCLUSION

Improvement in computer graphics processing unit in recent time allows the opportunity to classify complex EEG data using deep learning scheme. In this study, binary classes classification was conducted for EEG signal extracted for MI movements. This study has shown that the layered image configuration method is appropriate for binary classes

classification with 22 channels. It has also been observed that stacked image configuration takes approximately 400 to 500 seconds to train the network while layered configuration could be half of that time. This presents a potential of using deep learnings scheme with layered image configuration for fast online training adapting it quickly for a new user. Finally, the novelty presented in this study can potentially be used as a stepping stone for four class classification.

REFERENCES

- [1] B. Z. Allison, C. Brunner, V. Kaiser, G. R. Müller-Putz, and C. Neuper, "Toward a hybrid brain-computer interface based on imagined movement and visual attention," *Journal of Neural Engineering*, vol. 7, no. 2, p.026007, March 2010.
- [2] P. B. Taylor and H. T. Nguyen, "Performance of a head-movement interface for wheelchair control," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 1590-1593, 2003.
- [3] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*, MIT Press, Cambridge, MA, 1995.
- [4] S. V. Bozhokin and I. B. Suslova, "Wavelet-based analysis of spectral rearrangements of EEG patterns and of non-stationary correlations," *Physica A: Statistical Mechanics and its Applications*, vol. 21, pp. 151-160, 2015.
- [5] O. Faust, U.R Acharya, H. Adeli and A. Adeli, "Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis," *Seizure: European Journal of Epilepsy*, vol. 26, pp.56-64, 2015.
- [6] F. Lotte, et al., "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p.031005. April 2018.
- [7] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, no. 1, p. 016003. 2017.
- [8] H. K. Lee and Y. S. Choi, "A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequency image", in *International Conference on Information Networking (ICOIN)*, vol. 2018, pp. 906-909, 2018.
- [9] M. Seeck, et al., "The standardized EEG electrode array of the IFCN," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2070-2077. 2017.
- [10] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl and G. Pfurtscheller, "Four class motor imagery," *BNCIHorizon2020*, 2008. [Online]. Available: <http://bnci-horizon-2020.eu/database/data-sets>. [Accessed: 24 -July-2018].
- [11] G. Pfurtscheller and F. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp.1842-1857, 1999.
- [12] A. Karpathi, "CS231n Convolutional Neural Networks for Visual Recognition," Stanford Vision and Learning Lab. Available: <http://cs231n.github.io/convolutional-networks/#conv>. [Accessed: Aug. 11, 2018].
- [13] H. T. Nguyen, L. M. King and G. Knight, "Real-time head movement system and embedded Linux implementation for the control of power wheelchairs," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, San Francisco, CA, pp. 4892 -4895, 2007.
- [14] H. V. Shenoy, A. P. Vinod and C. Guan, "Shrinkage estimator based regularization for EEG motor imagery classification," in *Information, Communications and Signal Processing (ICICSP), 2015 10th International Conference*, Singapore, 2015.