

Evaluation on the Effect of EEG Pre-processing and Hyper parameters Tuning to the Performance of Convolutional Neural Network Motor Execution Classification

Mohamed Ragab Mahmoud Farghaly¹, Kim Chuan Lim², Yin Fen Low³, Feng Duan⁴ and Yew Guan Soo⁵

^{1,2,3,4,5}Machine Learning and Signal Processing, Centre for Telecommunication Research & Innovation, Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka

⁴Nankai University, College of Artificial Intelligence, Tianjin, P. R. China, 300350

Article Info

Article history:

Received Sep 20, 2021

Revised Nov 13, 2021

Accepted Nov 22, 2021

Keywords:

Convolutional Neural Network
Electroencephalogram
Hyperparameters Tuning
Pre-processing

ABSTRACT

Electroencephalogram (EEG) based classification has achieved a promising performance using deep learning models like Convolutional Neural Network. Various pre-processing strategies such as smoothing the EEG data or filtering are commonly used to pre-process the captured EEG signal before the subsequent feature extraction and classification while hyperparameters tuning might help to improve the classification performance. As well, the number of layers used in the CNN can affect the performance of the classification. In this paper, the number of layers needed for the CNN to classify the EEG data correctly, the effect of apply smoothing to pre-process the EEG signal for modern end-to-end CNN and the effect of enabling hyperparameters tuning during the training phase of CNN is investigated and analyzed. Two CNN models, namely Deep CNN with 5 layers and Shallow CNN with 1 layer, with convincing classification accuracy on motor execution classification as reported in the literature were chosen for this study. Both the CNN models are trained on EEG motor execution dataset with different training strategies and dataset pre-processing. Based on the obtained training and test classification accuracy, Shallow CNN trained with enabling hyper parameters tuning and without smoothing the EEG data achieved the best classification accuracy with average training accuracy of 99.9% and test accuracy of 96.87%. This indicates that CNN does not need to have many layers to correctly classify the motor execution data and the EEG data does not require smoothing.

Copyright © 2021 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Kim Chuan Lim,
Machine Learning and Signal Processing,
Centre for Telecommunication Research & Innovation,
Faculty of Electronics and Computer Engineering,
Universiti Teknikal Malaysia Melaka,
Jalan Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
Email: kimchuan@utem.edu.my

1. INTRODUCTION

Brain-computer interfaces (BCI) or also known as brain machine interface allows interacting with surrounding devices with brain signal without the interfering of peripheral nerves or muscle [1]. BCI fully depends on the studies of brain function that reflect its activity with electroencephalogram (EEG) signal which it is collected over the scalp without the need for surgery. Brain activity can then be analyzed by attaching a non-invasive electrodes cap to the scalp and the subject or patient doesn't need to be stationary which considered as an advantage to it unlike the functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) [2]. Motor movement is one of EEG studies that deal with analyzing and

understanding of the brain activity during doing the movement. Two types of movements are studied using EEG which are the imagination of the movement, called motor imagery, and the actual executing of the movement known as motor execution [3]. The difference between executing and imagining the movement is that imagining the movement doesn't involve muscles activity.

With the aid of available accelerated computing devices (e.g. GPU, FPGA) neural network based deep learning shows a great achievement in computer vision field especially in recognition task with a lots of labelled data set like image, videos, speech and text recognition and many other domains such as drug discovery and genomics [4]. The highly flexible function represented by the neural network and all parameters fitting enabled computer to perform end-to-end learning. For example, learning the required features in the signal to help the subsequence classification given a sequence of recorded raw EEG signal and the respective label. It's known that EEG signal have low noise-to-signal ratio and pre-processing techniques is commonly applied to help reduce the influence of the noise and the size of available public EEG data set is limited but recent research shows the ability of applying such an algorithm like CNN to classify EEG signal [5]–[10]. For instance, CNN shows an outstanding performance in decoding and classifying EEG motor execution/imagery with accuracy ranged between 85%~96% [7], [8], [11]–[15] and EEG emotion with accuracy close to 98% [16], [17].

The usage of deep learning models to classify EEG data is not strict to only CNN, in fact, some works used recurrent neural network (RNN) as a method of classifying EEG data [18]–[22]. But in term of the performance and the model usage, CNN had been used the most across most of the available works [23]. Before using the dataset for model's training, the dataset needs to be pre-processed. While filters are commonly used, some works use additional techniques to pre-process the data. Table 1 summarize some recent work in term of dataset used, pre-processing techniques, number of CNN layers, and lastly the achieved classification accuracy.

Schirmeister et al. [8] introduced two CNN, namely Shallow CNN that consist of 1 layer of convolution and Deep CNN which have 4 layers of convulsions. Both CNNs were used to investigate their capability of decoding and classifying movement-related tasks. Two types of datasets were used. The first dataset was a motor imagery dataset provided by BCI Competition IV referred to as BCI Competition IV dataset 2a (BCIC IV 2a for short) [24]. The other dataset was motor execution dataset called high-gamma dataset (HGD for short) provided by [8]. additional for using the band-pass filter, electrode-wise exponential moving standardization was used to smooth the EEG signal. the EEG data were cleaned by removing the trials that have a least one electrode had a value outside $\pm 800 \mu\text{V}$. for the HGD dataset, only 44 electrodes out of 128 were used and the data was resampled from 512Hz to 250Hz. And only 4.5 seconds were used from each dataset (from -0.5s to 4s).

V. J. Lawhern et al. [11] introduces EEGNet that able to decode and classify different type of EEG paradigm like movement-related tasks. EEGNet consists of one layer of depth-wise convolution followed up by a layer of depth-wise separable convolution. the movement-related task datasets used are as following, a motor imagery dataset provided by BCI Competition IV referred to as BCI Competition IV dataset 2a (the authors referred to it as SMR but here in this paper it's referred to as BCIC IV 2a) [24] and a motor execution dataset collected by [25] referred to as MRCP. the preprocessing used for MRCP is that a band pass filter between 0.1-40 Hz. 64 out of 256 electrodes were used. The data was resampled from 1024Hz to 128Hz and only 1.5 seconds were used (from -0.5 to 1s). For BCI IV 2a dataset, the data was resampled from 250Hz to 125Hz and only 2 seconds of the data were used (from 0.5 to 2.5s).

Tang et al. [7] introduced their own CNN model based on [5], a motor imagery dataset used by them were recorded. Before pre-processing the dataset from the CNN training, an Event-Related Desynchronization / Event-Related Synchronization (ERD/ERS) analysis was conducted to identify the important frequency range for each subject than a band-pass filter was used. The trials were segmented with a duration of 4 seconds.

Compared to the previous discussed work, Nurse et al. [13] and Tabar and Halici [15] used images instead of EEG signal to train their own CNN. Nurse et al. [13] introduced a CNN model that consists of two layers of convolutions and one fully connected layer. The dataset used is a recorded motor execution dataset by [26]. The dataset was band-pass filtered between 0.1-100Hz, and the trial were segmented into 400ms (-100 to 300ms). Then the data were converted into images to be used for training the CNN.

Tabar and Halici [15] introduced two models, the first one is a CNN that consist of a layer of convolution followed up by a fully connected layer. And the second one is like the first with addition of stacked autoencoder after the fully connected (CNN+SAE). Two motor imagery datasets used in this work one is BCI Competition IV dataset 2b [27] and the other is BCI Competition II dataset 3 [28]. Both datasets were converted into time-frequency maps images, and the images is used to train their proposed model.

As presented in Table 1 for the discussed related work, most of the work need to filter the data before using it to train their own CNN model. In term of the other pre-processing steps each work used their own preferences. the design of the CNN also varied in term of number of layers as the number of layers varied between one to five layers. It is shown that the performance varied between the works. Only [8] achieved the best accuracy classification accuracy for classifying motor execution EEG data.

Table 1. Summary of related work

Author	EEG detection Problem	Dataset	Pre-processing				Additional steps	No. of layers	Accuracy
			No. of electrodes	Filter type and range (Hz)	Sampling rate (Hz)	Trial length (ms)			
Nurse et al. [13]	Motor execution [image based]	Recorded	64	Bandpass 0.1-100	250	400	Convert the trials into images	4	81%
Tabar and Halici [15]	Motor imagery [image based]	BCIC IV dataset 2B [27]	3	Bandpass 0.5-100	250	2000	Convert the time-frequency maps to images	CNN: 2 CNN+SAE: 2+8	CNN = 89.3%
		BCIC II dataset 3 [28]	3	Bandpass 1-50	250	2000			CNN = 77.6%
Tang et al. [7]	Motor Imagery	Recorded	28	Band-pass Depend on the subjects	1000	4000	Used ERD/ERS to determine the frequency range for the filter	3	86.41 ± 0.77%
Schirmeister et al. [8]	Motor Imagery	BCIC IV dataset 2A [24]	22	Bandpass 4-38	250	4500	Apply smoothing to the data	Deep : 4 Shallow : 1	Deep = 70.9%
	Motor execution	High-gamma [8]	44	Highpass 4	250	250			Shallow = 73.7%
V. J. Lawhern et al. [11]	Motor Imagery	BCIC IV dataset 2A [27]	22	Bandpass 0.5-100	125	2000	-	2	EEGNet-4 = 64%
	Motor execution	MRCP [25]	64	Bandpass 1-50		1500	-		EEGNet-8 = 68%
									EEGNet-4 = 78%
									EEGNet-8 = 81%

In this paper, the influence of number of layers chosen for the CNN with the effect of applying smoothing to pre-process the EEG motor execution data and enabling hyperparameter tuning with convolutional neural network (CNN) as joint feature extraction and classification is analyzed and investigated. The structure of this paper is as follows, the pre-processing technique as well as the architecture of the CNN (shallow and deep CNN) used for motor execution dataset are discussed in Section 2. The results obtained for the motor execution classification are statistically analyzed and the obtained results for each pre-processing, architecture of CNNs, as well as enabling hyperparameters tuning are presented in Section 3. Lastly, the obtained results are concluded, and future works are discussed at the end of the paper.

2. RESEARCH METHOD

In this section, the method used to carry out the investigation is discussed. The dataset, the CNN models and the pre-processing steps used is inspired by [8] with some modification to it. For fair comparison, the results of [8] had been reproduced and compared to the proposed work. Figure 1 shows an overview of the research method.

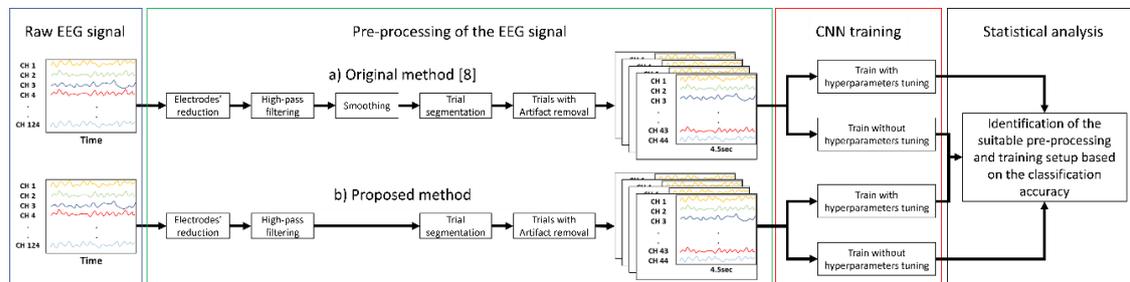


Figure 1. overview of the investigation carried out

2.1. Dataset

High-gamma dataset, which had been recorded and provide by [8], is used for this analysis. The dataset contains a recording of motor execution task. The dataset is recorded by [8] from 14 healthy subjects using 128 electrodes with approximately 1000 trials with 4s of executed movements and 3-4 s inter-trials divided into 13 runs per subject. The four classes of movements are movements of either the left hand (LH),

the right hand (RH), both feet (F), and rest (R). the dataset was collected in an environment that suitable for non-invasive detection of high frequency movement-related EEG components as it's proven that an increase in the gamma frequency components is observed during the execution of the movement [29]–[35].

2.2. Dataset Pre-processing

Subject 1 to Subject 13 recording were used in the analysis. Subject 14 data was excluded as the data suffered from some data loss during the recording as reported by [8] in the dataset repository. Following the pre-processing steps used by [8], only 44 electrodes, out of the 128 electrodes, that cover the motor cortex are used for the investigation. Figure 2 shows the location of these 44 electrodes. The datasets were filtered from 4 Hz by using a butterworth high-pass filter with order of 3 as these frequencies hold the important data for motor execution. The EEG data was smoothed using electrode-wise exponential running standardization (ERS) [8] with a decay factor of 0.999. The expositional mean and variance for each electrode is computed to standardize the continuous data as following the same steps provided by [8]:

$$x'_t = \frac{(x_t - \mu_t)}{\sqrt{\sigma_t^2}} \tag{1}$$

$$\mu_t = 0.001x_t + 0.999\mu_{t-1} \tag{2}$$

$$\sigma_t^2 = 0.001(x_t - \mu_t) + 0.999\sigma_{t-1}^2 \tag{3}$$

Where x'_t : is standardized signal, x_t is original signal, μ_t is mean, and σ_t^2 is variance. The first 1000 means values of μ_t and the first 1000 variance values of σ_t^2 were set to the mean and variance of the 1000 samples window of the dataset. Figure 3 shows the effect of applying smoothing during the pre-processing of the EEG signal Versus the non-smoothed signal. A copy of the dataset is kept without applying electrode-wise exponential running standardization to study the effect of smoothing on the performance of the CNN.

2.3. Convolutional Neural Network

The two CNN architectures used in this study is based on the best accuracy reported by [8]. The first CNN architecture is Deep CNN which consists of four layers of convolution filters with max-pool and one output layer. The first layer is split into two convolutional blocks where the EEG signal of each electrode will be firstly convolute across time and followed with convolution across the EEG electrodes. Exponential liner unit (ELU) is the activation function between each layer and SoftMax is used for the output classification layer. Figure 4-a illustrate Deep CNN

The second architecture is Shallow CNN which consist of one convolution layer with mean-pool and one output layer. The first layer is identical to the Deep CNN but the activation function between the convolution and mean-pool is square function as presented in Figure 4-b.

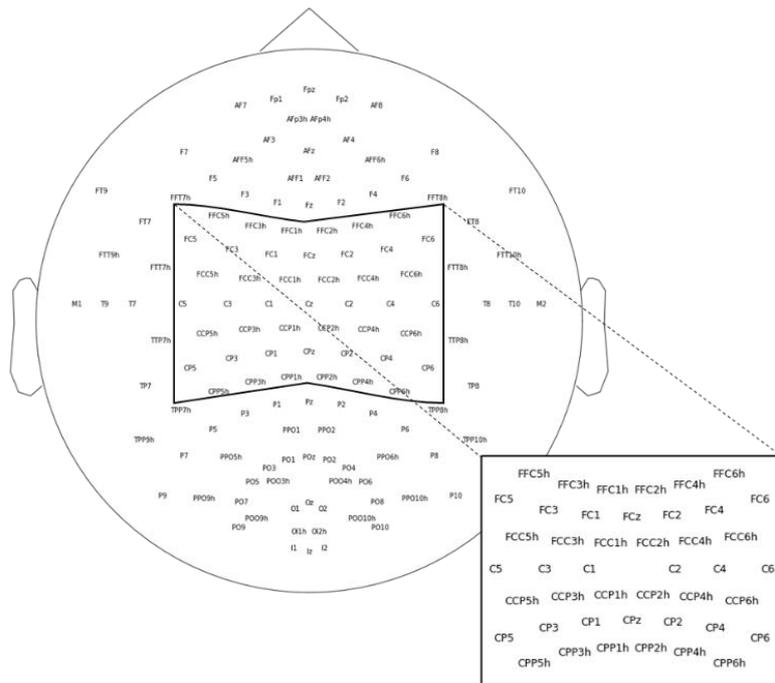


Figure 2. Location of the 44 electrodes.

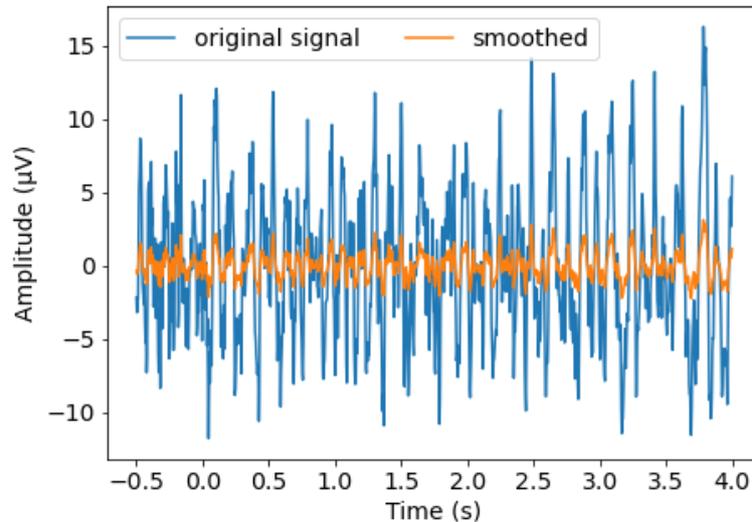


Figure 3. Comparison between before and after applying smoothing for a single electrode

2.4. Training Strategy

Using the recorded EEG trials and the respective motor execution (ME) label, the CNN is then train with and without enabling hyperparameters tuning (H, XH). The hyperparameters tuning are batch normalizations and dropout. Augmenting the training dataset is commonly applied when the available number of training sample is little. The goal to increase total number of training dataset can be achieved by applying sliding window sampling within each trial. A total of two seconds EEG signal samples are treated as the input of the CNN similar to [8]. The optimization function used is Adam optimization algorithm [36] with learning rate of 0.01. As well, early stopping technique is used. The training set is split into training and validation set and training of the architecture stop when the validation accuracy is not improving for specific number of epochs. the training for each method were run 30 times.

2.5 Statistical analysis

Wilcoxon rank-sum test is used to measure whether there is any significant difference between the accuracies when i) which CNN model would generally outperform the other in term of number of layers ii) whether hyperparameters tuning is needed during the training process or not and lastly iii) when smoothing is applied during the pre-processing, The confidence level used in this test is 95%.

3. RESULTS AND DISCUSSION

Using the proposed training methodology (the numbers of layer decided for each CNN model, EEG pre-processing with and without smoothing; the usage of hyperparameters tuning during the neural network training) as discussed in Section 2, the result of training both Shallow and Deep CNN is summarized as shown in Table 2. It is found that Shallow CNN always reports the best accuracy as compared to Deep CNN except for subject #3 which Deep CNN has achieved the highest accuracy of 98.98% without smoothing applied during the pre-processing and with hyperparameters tuning enabled as compared to 98.79% when trained with Shallow CNN with smoothing applied during the pre-processing and with hyperparameters tuning enabled. Ten subjects show that hyperparameters tuning being enabled helped in improving the accuracy while the other three subjects (subject #10, #11, and #12) showed that Shallow CNN can achieve high accuracy with the disable of hyperparameters tuning. As for the smoothing, four subjects (subject #1, #4, #7, and #8) needed smoothing (see Shallow H in Table 2 under smoothing) as compared to the other nine subjects. While subject #5 showed that it has a stable accuracy whether smoothing is applied to the EEG signal or not.

To validate whether the trained Shallow CNN classifier is significantly more accurate as compared to Deep CNN i) with or without smoothing applied during the pre-processing, ii) with hyperparameters being enabled (H) or disabled (XH), rank-sum statistical test is firstly applied to the result of classification on the test dataset by the trained CNN and as shown in Table 3. Based on the p-values presented in Table 3, Shallow CNN outperformed Deep CNN in all aspect no matter smoothing is applied during the pre-processing and enable or disable of the hyperparameters tuning during the training. This indicates that the number of layers used in CNN can affects the performance of the classification.

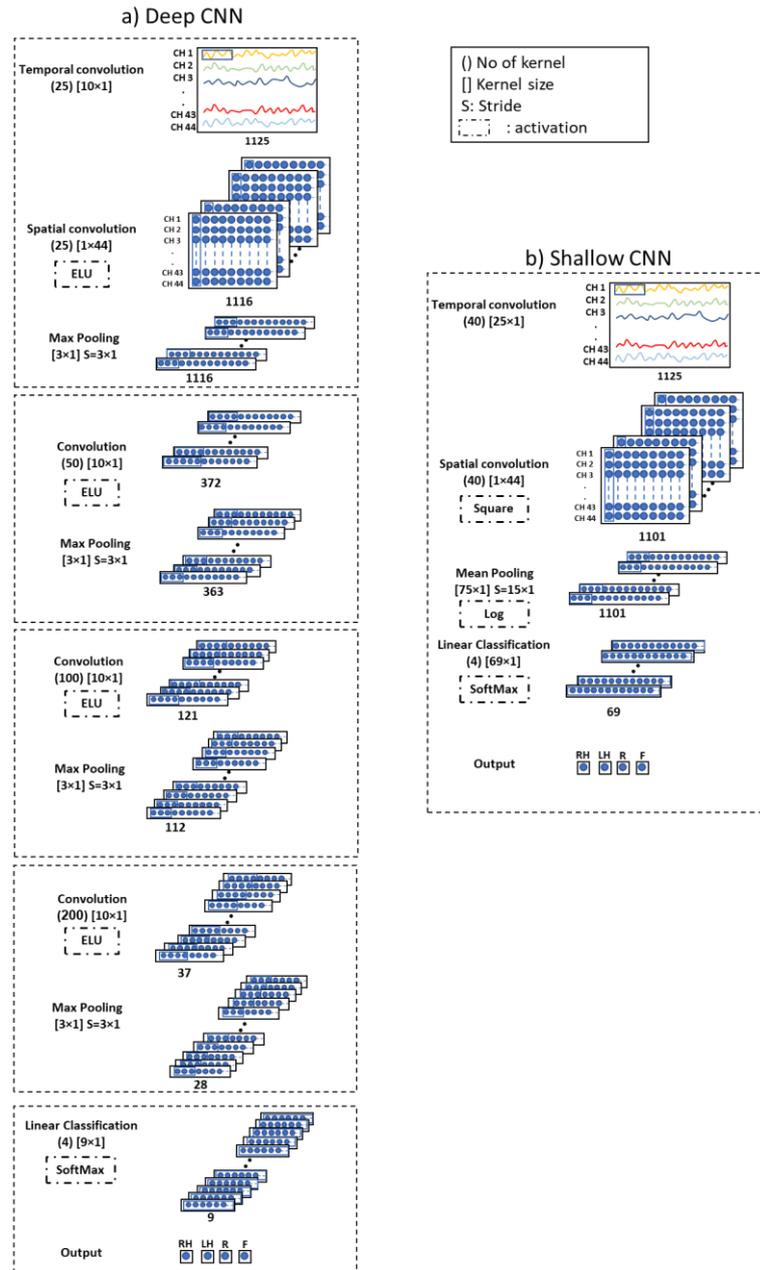


Figure 4. Deep and Shallow CNN based on [8].

Table 2. Test accuracy (%) averaged across 30 runs les

Pre-Processing	Model Method	Smoothing				Without-Smoothing			
		Deep		Shallow		Deep		Shallow	
		XH	H	XH	H	XH	H	XH	H
Subjects	1	81.4	61.73	91.35	95.67	82.9	66.42	93.9	93.9
	2	86.1	88.35	89.5	89.92	87.25	86.23	90	90.44
	3	95.67	98.02	98.21	98.79	93.12	98.98	96.37	98.17
	4	98.92	98.77	99.96	100	96.48	99.21	99.96	99.96
	5	97.83	98.19	99.58	100	96.19	98.19	99.92	100
	6	93.31	92.42	96.37	96.88	86.62	94.23	97.54	98.08
	7	85.97	90.17	93.92	97.61	84	92.12	90.36	95.09
	8	88.27	93.54	97.5	98.17	87.23	93.75	97.81	96.88
	9	96.12	93.96	98.54	98.58	95.29	96.17	98.54	99.29
	10	90.56	91.52	92.79	92.5	91.31	93.73	94.96	93.08
	11	94.08	91.12	98.08	96.92	94.15	94.58	98.35	97.5
	12	96.31	97.23	96.17	97	94.6	98.15	98.75	97.63
	13	93.08	94.17	96.92	96.9	93.04	92.83	98.55	99.35
	Avg	92.12	91.48	96.07	96.84	90.94	92.66	96.54	96.87

Table 3. Rank-sum average results for Shallow CNN vs Deep CNN test accuracy

Method	Model	H			XH		
		Deep (%)	Shallow (%)	P-value	Deep (%)	Shallow (%)	P-value
Pre-Processing	Smoothing	91.48	96.84	7.41e-32	92.12	96.07	3.35e-29
	Without smoothing	92.66	96.87	2.42e-19	90.94	96.54	4.01e-49

Then, four version of trained Shallow CNN are statistically compared in term of i) whether the EEG data need to be smoothed and ii) the need of hyperparameters tuning during the training process and summarized into Table 4. Based on the result of statistical test as presented in Table 4, Shallow CNN trained without applying smoothing and with the hyperparameters tuning disabled (XH) is statistically more accurate as compared to when smoothing is applied. However, no significant difference is observed when Shallow CNN is trained with the enabled of hyperparameters tuning on input EEG signal both with or without pre-processing applied to the input EEG signal. From the perspective of whether to smooth or not to smooth the input EEG signal during the pre-processing stage, the rank sum statistical test is conducted. Shallow CNN trained with smoothing applied during the pre-processing couple with hyperparameters tuning enabled is able to classify more accurately with average accuracy of 96.84 % (p-value 0.000551). However, no significant difference is found when both Shallow CNN, average accuracy of 96.54% and 96.87%, are trained without smoothing applied during the pre-processing stage with hyperparameters tuning being disabled or enabled. Three out of four Shallow CNN with average test accuracy of 96.87%, 96.84%, and 96.54% is further analyzed.

Table 4. Rank-sum average results for smoothing vs without smoothing test accuracy for Shallow ConvNet

Pre-Processing	XH			H		
	Smoothing (%)	Without-Smoothing (%)	P-value	Smoothing	Without-Smoothing	P-value
	96.07	96.54	0.0012	96.84	96.87	0.10

To identify which training method (hyperparameters enabled or disable) and pre-processing should be used on Shallow CNN among the top 3 accurate tested Shallow CNNs, the respective training accuracy is observed as shown in Table 5. Then rank sum test is used to compare between the highest achieved average training accuracy (99.999%) with the other two training accuracy which presented in Table 6. From Table 5 and Table 6, Shallow CNN trained with hyperparameters tuning enabled and without applying smoothing during the pre-processing outperformed the other two methods during the training as this method has two subjects (subject #8 and #10) that did not reach 100% accuracy averaged across the 30 runs. Out of the 30 runs, subject #8 had only two runs that the training accuracy did not reach 100% and subject #10 had only one run did not reach 100%.

The results presented in this section indicates the following, Smoothing the data would affect the performance of CNN classification due to the reduction in the amplitude of the EEG signal as presented in Figure 2. in term of the number of layers, Shallow CNN outperformed Deep CNN and that explains that the model used to classify the data does not require many layers, like Deep CNN, to perform well. Lastly, enabling hyperparameters tuning managed to improve the training accuracy as only subject #8 did not perform well for both run No. 1 and run No. 2 and subject #10 did not perform well for run No. 1 (see Table 5 for average training accuracy). However, in terms of test accuracy, there is no statistical difference between using Shallow ConvNet that trained with hyperparameters tuning being enabled with and without applying smoothing.

Table 5. Training accuracy (%) averaged across 30 runs

Pre-Processing Method	Shallow			
	Smoothing H	Without-Smoothing XH	Without-Smoothing H	
Subjects	1	100	100	100
	2	100	98.126	100
	3	99.894	99.894	100
	4	100	100	100
	5	100	100	100
	6	100	99.989	100
	7	100	100	100
	8	100	99.837	99.99
	9	100	99.989	100
	10	99.861	100	99.996
	11	99.992	100	100
	12	100	100	100
	13	100	99.946	100
	Avg	99.981	99.829	99.999
No of subjects < 100	3	6	2	

Table 6. Rank-sum average results for shallow convnet's training accuracy

Pre-Processing	Without smoothing (%)		P-value	Smoothing (%)		P-value
Method	H	XH		XH	H	
	99.999	99.829	0.00054	99.999	99.981	0.00045

4. CONCLUSION

The study on these CNN architectures introduced by [8] indicates that the CNN does not require many layers so it can be able to classify the EEG trials correctly. Smoothing the data would lead in reduction in the performance due to the reduction in the signal amplitude. And enabling hyperparameters tuning managed to boost the classification accuracy. This shows that such advanced deep learning algorithm can extract features and classify the EEG signal. It will help in designing applications that would benefit the mankind (artificial limbs, ability to control devices with consumer EEG cap) in the future. More studies will be carried out on different types of datasets like using emotions or motor imagery EEG datasets. Recording a new dataset with different subjects and modifying the current architecture to achieve better accuracy will be investigated as well. As well introducing different type of deep learning architectures (Like recurrent convolution neural network) might help as well in giving better results.

ACKNOWLEDGMENTS

The authors would like to thank Universiti Teknikal Malaysia Melaka for the financial support of this work.

REFERENCES

- [1] M. M. Fouad, K. M. Amin, N. El-Bendary, and A. E. Hassanien, "Brain computer interface: A review," *Intell. Syst. Ref. Libr.*, vol. 74, pp. 3–30, 2015, doi: 10.1007/978-3-319-10978-7_1.
- [2] L. Bozhkov, "Overview of Deep Learning Architectures for Classifying Brain Signals," vol. IX, no. 1, pp. 54–59, 2016.
- [3] N. Mizuguchi, H. Nakata, Y. Uchida, and K. Kanosue, "Motor imagery and sport performance," *J Phys Fit. Sport. Med.*, vol. 1, no. 1, pp. 103–111, 2012, doi: 10.7600/jpfs.1.103.
- [4] Y. LeCun, Y. Bengio, G. Hinton, L. Y., B. Y., and H. G., "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [5] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, 2011, doi: 10.1109/TPAMI.2010.125.
- [6] R. Manor and A. B. Geva, "Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI," *Front Comput Neurosci.*, vol. 9, p. 146, 2015, doi: 10.3389/fncom.2015.00146.
- [7] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Opt. - Int. J. Light Electron Opt.*, vol. 130, pp. 11–18, 2017, doi: 10.1016/j.ijleo.2016.10.117.
- [8] R. T. Schirrmester *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, 2017, doi: 10.1002/hbm.23730.
- [9] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2017, doi: 10.1109/TNSRE.2017.2721116.
- [10] S. Mousavi, F. Afghah, and U. Rajendra Acharya, "Sleeppeegnet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PLoS One*, vol. 14, no. 5, pp. 1–15, 2019, doi: 10.1371/JOURNAL.PONE.0216456.
- [11] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, doi: 10.1088/1741-2552/aace8c.
- [12] D. Wang, D. Miao, and G. Blohm, "Multi-class motor imagery EEG decoding for brain-computer interfaces," *Front. Neurosci.*, no. OCT, 2012, doi: 10.3389/fnins.2012.00151.
- [13] E. Nurse, B. S. Mashford, A. J. Yepes, I. Kiral-Kornek, S. Harrer, and D. R. Freestone, "Decoding EEG and LFP signals using deep learning: Heading truenorth," 2016, doi: 10.1145/2903150.2903159.
- [14] I. Sturm, S. Lopuschkin, W. Samek, and K. R. Müller, "Interpretable deep neural networks for single-trial EEG classification," *J. Neurosci. Methods*, vol. 274, pp. 141–145, 2016, doi: 10.1016/j.jneumeth.2016.10.008.
- [15] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, 2017, doi: 10.1088/1741-2560/14/1/016003.
- [16] H. Xu and K. N. Plataniotis, "Affective states classification using EEG and semi-supervised deep learning approaches," *2016 IEEE 18th Int. Work. Multimed. Signal Process. MMSP 2016*, 2017, doi: 10.1109/MMSP.2016.7813351.
- [17] X. Chai, Q. Wang, Y. Zhao, X. Liu, O. Bai, and Y. Li, "Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition," *Comput. Biol. Med.*, 2016, doi: 10.1016/j.combiomed.2016.10.019.
- [18] Y. Wang and D. Wu, "Deep Learning for Sleep Stage Classification," *Proc. 2018 Chinese Autom. Congr. CAC 2018*, pp. 3833–3838, 2019, doi: 10.1109/CAC.2018.8623637.

- [19] E. Bresch, U. Großekathöfer, and G. Garcia-Molina, "Recurrent deep neural networks for real-time sleep stage classification from single channel EEG," *Front. Comput. Neurosci.*, vol. 12, no. October, pp. 1–12, 2018, doi: 10.3389/fncom.2018.00085.
- [20] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, "EEG-based Intention Recognition from Spatio-Temporal Representations via Cascade and Parallel Convolutional Recurrent Neural Networks," 2017, [Online]. Available: <http://arxiv.org/abs/1708.06578>.
- [21] X. Zhang, W. Kou, E. I. C. Chang, H. Gao, Y. Fan, and Y. Xu, "Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device," *Comput. Biol. Med.*, 2018, doi: 10.1016/j.combiomed.2018.10.010.
- [22] 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. Institute of Physics Publishing, Apr. 16, 2018, doi: 10.1088/1741-2552/aab2f2.
- [23] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, 2019, doi: 10.1088/1741-2552/ab260c.
- [24] C. Brunner and R. Leeb, "BCI Competition 2008 – Graz data set A," *Knowl. Creat. Diffus. Util.*, pp. 1–6, 2008, doi: 10.1109/TBME.2004.827081.
- [25] S. M. Gordon, V. Lawhern, A. D. Passaro, and K. McDowell, "Informed decomposition of electroencephalographic data," *J. Neurosci. Methods*, vol. 256, pp. 41–55, 2015, doi: 10.1016/j.jneumeth.2015.08.019.
- [26] E. S. Nurse, P. J. Karoly, D. B. Grayden, and D. R. Freestone, "A generalizable brain-computer interface (BCI) using machine learning for feature discovery," *PLoS One*, vol. 10, no. 6, pp. 1–22, 2015, doi: 10.1371/journal.pone.0131328.
- [27] C. Brunner and R. Leeb, "BCI Competition 2008 – Graz data set B," *Knowl. Creat. Diffus. Util.*, pp. 1–6, 2008, doi: 10.1109/TBME.2004.827081.
- [28] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473–482, Dec. 2007, doi: 10.1109/TNSRE.2007.906956.
- [29] T. Ball *et al.*, "Movement related activity in the high gamma range of the human EEG," *Neuroimage*, vol. 41, no. 2, pp. 302–310, 2008, doi: 10.1016/j.neuroimage.2008.02.032.
- [30] C. A. Usanos, L. Boquete, L. de Santiago, R. B. Navarro, and C. Cavaliere, "Induced gamma-band activity during actual and imaginary movements: EEG analysis," *Sensors (Switzerland)*, vol. 20, no. 6, 2020, doi: 10.3390/s20061545.
- [31] G. Pfurtscheller, "Event-Related Desynchronization (ERD) and 40-Hz Oscillations in a Simple Movement Task," in *Oscillatory Event-Related Brain Dynamics*, 1994.
- [32] H. Tan *et al.*, "Complementary roles of different oscillatory activities in the subthalamic nucleus in coding motor effort in Parkinsonism," *Exp. Neurol.*, vol. 248, pp. 187–195, 2013, doi: 10.1016/j.expneurol.2013.06.010.
- [33] N. E. Crone, D. L. Miglioretti, B. Gordon, and R. P. Lesser, "Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band," *Brain*, vol. 121 (Pt 12), no. 12, pp. 2301–2315, 1998, doi: 10.1093/BRAIN/121.12.2301.
- [34] J. Hammer *et al.*, "Predominance of Movement Speed Over Direction in Neuronal Population Signals of Motor Cortex: Intracranial EEG Data and A Simple Explanatory Model," *Cereb. Cortex (New York, NY)*, vol. 26, no. 6, p. 2863, Jun. 2016, doi: 10.1093/CERCOR/BHW033.
- [35] F. Quandt, C. Reichert, H. Hinrichs, H. J. Heinze, R. T. Knight, and J. W. Rieger, "Single trial discrimination of individual finger movements on one hand: a combined MEG and EEG study," *Neuroimage*, vol. 59, no. 4, pp. 3316–3324, Feb. 2012, doi: 10.1016/J.NEUROIMAGE.2011.11.053.
- [36] D. P. Kingma and J. L. Ba, "Adam: a Method for Stochastic Optimization," *Int. Conf. Learn. Represent. 2015*, pp. 1–15, 2015, doi: <http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503>.

BIOGRAPHY OF AUTHORS



Mohamed Ragab Mahmoud Farghaly received his B.Eng (2017) in Computer Engineering from Universiti Teknikal Malaysia Melaka, Malaysia and currently pursuing his master's degree in machine learning and signal processing.



Kim Chuan Lim received his B.Eng (2003), MSc. (2005), and PhD. (2009) degrees from the Faculty of Arts, Computing, Engineering and Sciences, Sheffield Hallam University., UK. From 2005 he was a research assistant in the Mobile Machines and Vision Laboratory (MMVL) at Sheffield Hallam University. He is currently an Assoc. Prof. at the Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka, Malaysia.



Yin Fen Low received her bachelor's degree in electrical engineering from University of Technology Malaysia (UTM) in 2001 and master's degree (Engineering Science) in Medical image compression from Multimedia University Malaysia (MMU) in 2003. Later on, she completed her doctoral degree specializing in EEG signal processing from Saarland University, Germany in 2011. She worked as Research Officer at MMU from 2001 – 2003. Since Oct 2003, she is a faculty member of the Department of Computer Engineering, Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM). Yin Fen involves actively in the research of EEG signal processing, machine learning, as well as EEG applications in different domains including brain computer interface, tinnitus, person authentication



Feng Duan received the B.E. and M.E. degrees in mechanical engineering from Tianjin University, China, in 2002 and 2004, respectively, and the M.S. and Ph.D. degrees in precision engineering from The University of Tokyo, Japan, in 2006 and 2009, respectively. He is currently a Professor with Nankai University, China. His research interests include cellular manufacture systems, rehabilitation robots, and brain machine interfaces



Yew guan Soo received B.Eng (2001) and M.Eng in Electrical Engineering from Universiti Teknologi Malaysia, and PhD. (2010) in Bio Signal Processing from The University of Tokyo, Japan. Currently his is an Assoc. Prof. in Universiti Teknikal Malaysia Melaka (UTeM). His research interests include Embedded System and Robotics, particularly on the myo-electric robotic prosthesis and sensory substitution for amputees.