**A Brute-force CNN Model Selection for Accurate Classification of Sensorimotor Rhythms in BCIs**

Berdakh Abibullaev, Irina Dolzhikova, Amin Zollanvari

We would like to express our gratitude to the reviewers for their feedback and constructive comments. The manuscript has been revised to incorporate and address the reviewers’ comments and suggestions where appropriate while keeping it concise. In an effort to address the points, we have made the following corrections and additions to the manuscript (reviewer comments in *italic*). All the changes made to the revised version are marked in blue in the manuscript. We eagerly await your feedback on the revised manuscript.

Recommendation: Accept (minor edits)  
  
Comments:

*1. The discussion section needs some improvement.  The authors need to provide a rational why their method performs well comparable with the benchmark method. [Amin]*

Deep learning has offered unprecedented opportunities to construct remarkably accurate classifiers by integrating the process of feature extraction into the classifier training. However, this integration process comes at the price of a large number of (algorithmic and structural) hyperparameters to be tuned. This state of affairs has partially led many studies to rely on existing well-known architectures such as AlexNet or ResNet, use the domain knowledge to construct the final architecture, or have an unclear *ad hoc* strategy deployed for model selection. This raises the question as to whether training accurate deep learning models using a principled model selection is possible, or, alternatively, experience and developing high intuition regarding the collected BCI data is the most prudent way to go about tuning the hyperparameters.

To address this question, we sought to compare the performance of a standard CNN architecture trained using systematic model selection (with the most straightforward one being the brute-force search) with a well-known architecture, namely, EEGNet [1], which has been partially inspired by domain knowledge and has shown the state-of-the-art performance in various EEG classification tasks. Although the EEGNet is also based on convolutional layers, these layers have been combined in a way to resemble the Filter Bank Common Spatial Patterns [2]. Our results show the possibility of using standard CNN within a brute-force model selection to achieve comparable classification accuracy as EEGNet. In other words, our study is a comparison between the use of prior knowledge versus data in the context of model selection. Naturally, if a “good” prior knowledge about the nature of data and a mechanism for encoding this knowledge into the structure of a classification rule are available, we may expect training predictive models with high accuracy; however, in the absence of such prior knowledge, we may look into conducting a data-driven brute-force model selection as a viable option. Nevertheless, the performance of the selected structure, which is the outcome of this brute-force model selection, depends on the pre-specified space of hyperparameters. Here, we showed that in so far as classification of sensorimotor rhythms that arise during motor imagery tasks is concerned, a pre-specified space that was restricted by our computing power and defined based on some common values of hyperparameters can lead to accurate classifiers.

To better highlight these points in addressing this comment:

1. we have added the following statement at the outset of Section VI (Discussion):

“Deep learning technologies has offered unprecedented opportunities to construct remarkably accurate classifiers by integrating the process of feature extraction into the classifier training. However, this integration process comes at the price of tuning a large number of (algorithmic and structural) hyperparameters. This has partially led many studies to rely on existing well-known architectures such as AlexNet or ResNet, use the domain knowledge to construct the final architecture, or have an unclear *ad hoc* strategy deployed for model selection. This raises the question as to whether training accurate deep learning models using a principled model selection is possible, or, alternatively, experience and developing high intuition regarding the collected BCI data is the most prudent way to go about tuning the hyperparameters.”;

2. we have added the following statement as the fourth paragraph in Section VI (Discussion)

“In other words, our study is a comparison between the use of prior knowledge versus data in the context of model selection. Naturally, if “good” prior knowledge about the nature of data and a mechanism for encoding this knowledge into the structure of a classification rule are available, we may expect training highly accurate predictive models; however, in the absence of such prior knowledge, we may look into conducting a data-driven brute-force model selection as a viable option. Nevertheless, the performance of the selected structure, which is the outcome of this brute-force model selection, depends on the pre-specified space of hyperparameters. Here, we showed that in so far as classification of sensorimotor rhythms is concerned, a pre-specified space that was restricted by our computing power and defined based on common values of hyperparameters can lead to accurate classifiers.”

[1] 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,” Journal of neural engineering, vol. 15, no. 5, 2018.

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

**2. It would be of interest to include the duration of time in seconds it took for the simulations to run on the PC with the specifications they have provided.  [Berdakh]**

Figure below provide the details of the training time (in seconds) of the selected ConvNet Architecture {C[128, 64, 32, 16, 8]\_K(3 x 8)} below on four pooled datasets (Physionet, Weibo2014, BCI-DataSet2A, and BCI-DataSet2B).

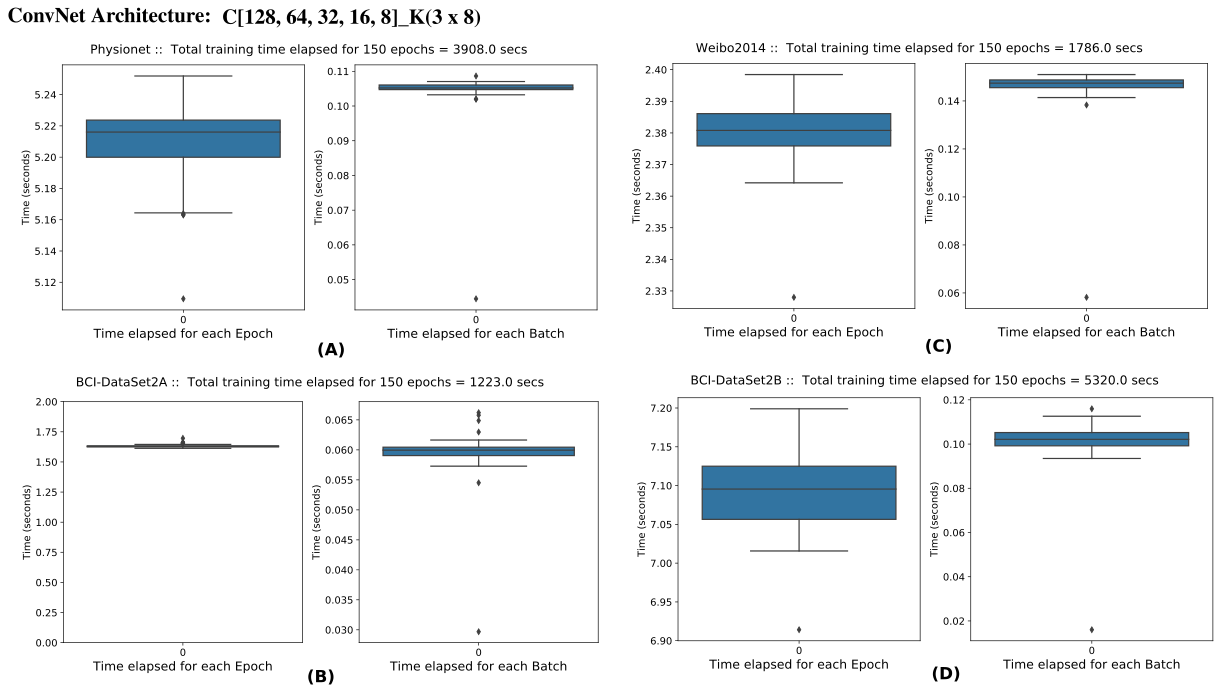
Specifically, we measured the average time taken for each epoch and for each batch as well as the total training time elapsed of the ConvNet model.

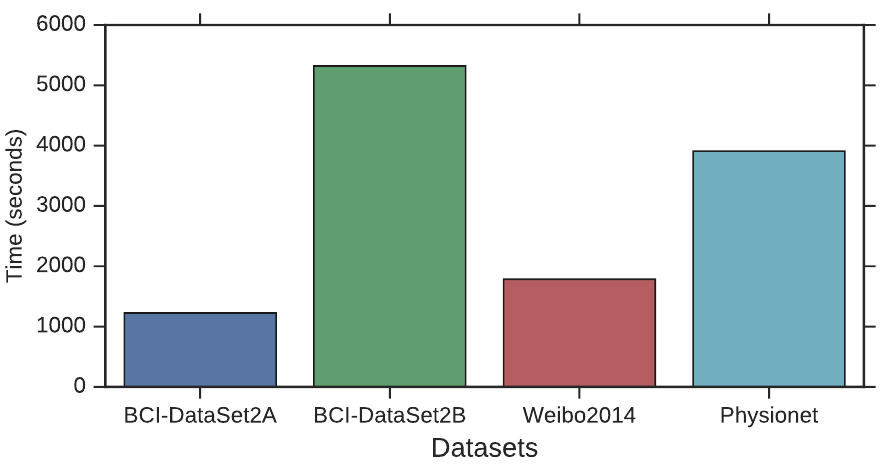
All this process was conducted using a Linux workstation with Intel Core i9-9900K, (3.6 GHz) processor, 32 GB of RAM, and Nvidia GeForce RTX 2080 Ti (RAM = 11GB, CUDA Cores: 4352).

As mentioned in the paper, we use the following hyperparameters:

* the maximum number of examined epochs for all models was set to 150;
* the mini-batch gradient descent (batch size of 64) with Adam optimizer with a learning rate of 0.001 and a decay of 0.0001 ;

the loss function was cross-entropy





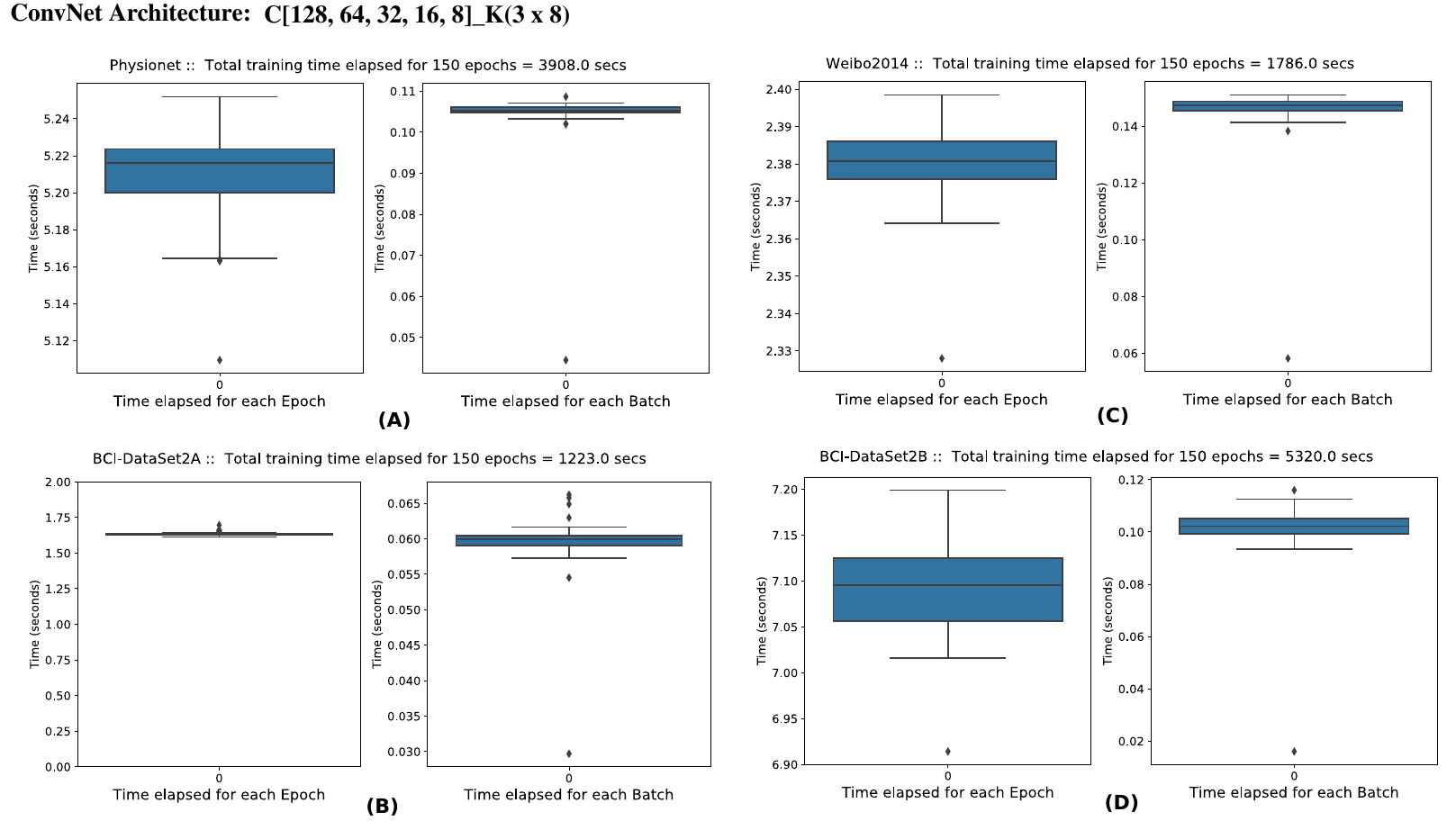
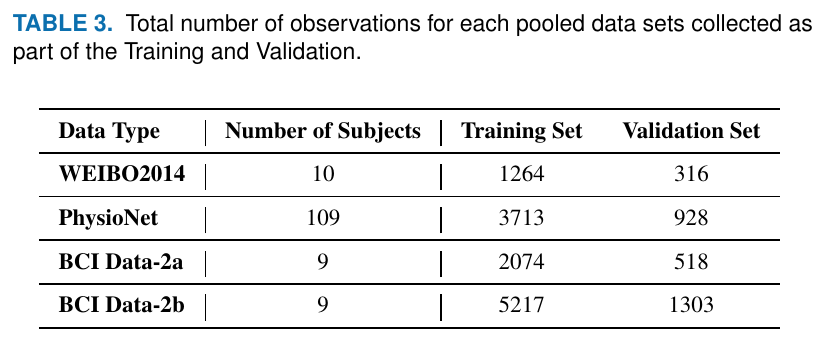


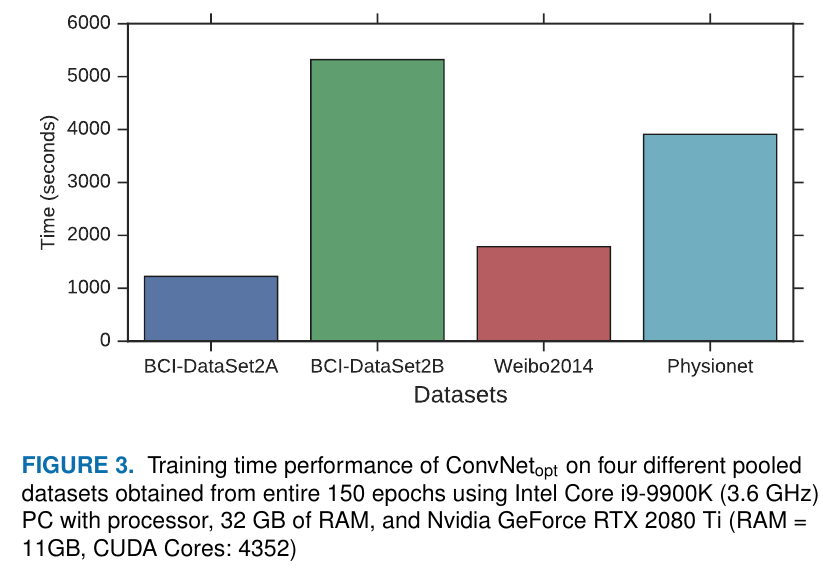
Figure : The training time of the final selected ConvNet Architecture on four pooled datasets (Physionet, Weibo2014, and BCI-DataSet2A, and BCI-DataSet2B).

We note that the longer training time were associated with the size of the dataset (see Table 3). For instance, the BCI-DataSet2B took 5352 seconds, which is the longest duration among other datasets, elapsed to complete the training of the model.



In the results Section V, we added the information about the training-time performance highlighed in blue, and the Figure 3

shown below:



3. Also, couple of sentences defining the standard CNN and defining the state of the art deep learning architectures. What is the difference between the two. Why the CNN performs well compared to just a deep learning neural network? *[Irina and Amin; Irina, please write a summary of EEGNet structure here. Then I will go through that and highlight the difference with ours.]*

The introductory paragraph of Section III was paraphrased to make it easier to understand what is defined as standard CNN and state-of-the-art deep learning architecture. The building blocks of standard CNN are considered in Section III. A and the state-of-the-art deep learning architecture that is regarded as a benchmark for comparison, namely EEGNET structure, is described in section III.B.  The main point that we wanted to highlight in this article on the difference of what we do and what is demonstrated in the EEGNet paper is that the final structure of our CNN model comes from a purely data-driven brute-force model selection, while EEGNet is inspired by domain knowledge (i.e., FBCSP). This inspiration is manifested in the simulations, from which it could be noticed that the useof depthwise convolutions allow EEGNET to learn band-specific spatial filters in a way similar to FBCSP.

Regarding the performance of CNN in comparison to other deep learning neural networks, “several studies have already demonstrated the efficiency of CNNs in terms of training time (as compared to some of the popular recurrent neural networks) and effectiveness to capture latent features from raw EEG data [19].”

In order to better clarify the above mentioned points in the manuscript, the following changes were implemented in Section III:

**“**The key question in this study is whether using standard convolutional neural networks within a systematic model selection would possibly lead to a comparable classification accuracy for motor imagery tasks as the current state-of-the-art deep learning architectures, which are partially inspired by domain knowledge. Therefore, we first briefly describe the main building block standard CNNs, which is the convolutional operation, and we then briefly describe the EEGNet architecture [37], which serves as the benchmark for comparison as it has shown the state-of-the-art performance in various EEG classification tasks.”

**Additional Questions: [The following comments are basically what the reviewer is asking above so no need to rewrite things, we just refer to above comments]**

**Does the paper contribute to the body of knowledge?:** *The paper does not contribute to knowledge, however the paper provides a good survey of works in this area.  The authors have presented results of CNN in various configurations and compared it to a EEGnet benchmark.*  
**Is the paper technically sound?:** *Yes, the paper is technically sound.  It has provided some exhaustive simulations and work.*  
  
**Is the subject matter presented in a comprehensive manner?:** *Yes, the subject matter is presented with exhaustive results, however the discussion section needs some improvement.  The authors need to provide a rational why their method performs well comparable with the benchmark method. It would be of interest to include the duration of time in seconds it took for the simulations to run on the PC with the specifications they have provided.*  
  
**Are the references provided applicable and sufficient?:** *Yes, references are sufficient.*

**Reviewer 2:**

Recommendation: Reject (update and resubmit encouraged)

Comments:  
*1. The contribution of this paper is not explained clearly. Through reading the introduction of this paper, it gives the readers a feeling that the authors just apply standard convolutional neural networks on EEG datasets and optimize hyper-parameters by an exhaustive search. The superiority of the method proposed should be clarified more clearly in the revised version of the paper. [Amin]*

Please note that we are not concluding that our constructed classifier is superior or inferior to the well-known EEGNet architecture. As a matter of fact, as mentioned at the end of Section V (Results), we write: “a two-sided Wilcoxon rank sum test does not reject the null hypothesis of having difference between these two sets of accuracies (p = 0.51). In other words, the classification accuracies achieved by ConvNetopt and EEGNet are comparable on BCI-DataSet2B”. Nevertheless, this observation indeed justifies the underlying hypothesis of our investigation; that is to say, conducting a purely data-driven brute-force model selection within a restricted space of hyperparameters leads to classification accuracies that are comparable to that of EEGNet, which has been inspired by Filter Bank Common Spatial Patterns (i.e., domain knowledge). To better highlight the contribution and conclusion of the paper, please note that:

1. the following statement is now added at the beginning of Section VI to highlight the underlying question behind our study:

“Deep learning technologies has offered unprecedented opportunities to construct remarkably accurate classifiers by integrating the process of feature extraction into the classifier training. However, this integration process comes at the price of tuning a large number of (algorithmic and structural) hyperparameters. This has partially led many studies to rely on existing well-known architectures such as AlexNet or ResNet, use the domain knowledge to construct the final architecture, or have an unclear *ad hoc* strategy deployed for model selection. This raises the question as to whether training accurate deep learning models using a principled model selection is possible, or, alternatively, experience and developing high intuition regarding the collected BCI data is the most prudent way to go about tuning the hyperparameters.”;

1. in the third paragraph of Section VI (Discussion), we write:

“Going back to the main question of this study, this observation shows the possibility of using standard CNNs within a systematic brute- force model selection to achieve comparable classification accuracy as the state-of-the-art deep learning architectures used in the classification of motor imagery tasks.”;

1. the following paragraph is now added in the Discussion section to better highlight the main conclusion of the paper from a general perspective; that is to say, from the general standpoint of machine learning:

“In other words, our study is a comparison between the use of prior knowledge versus data in the context of model selection. Naturally, if “good” prior knowledge about the nature of data and a mechanism for encoding this knowledge into the structure of a classification rule are available, we may expect training highly accurate predictive models; however, in the absence of such prior knowledge, we may look into conducting a data-driven brute-force model selection as a viable option. Nevertheless, the performance of the selected structure, which is the outcome of this brute-force model selection, depends on the pre-specified space of hyperparameters. Here, we showed that in so far as classification of sensorimotor rhythms is concerned, a pre-specified space that was restricted by our computing power and defined based on common values of hyperparameters can lead to accurate classifiers.”

*2. In "INDEX TERMS", the term of "motor imagery" appears twice.*

**Reply:** The duplication was removed.

*3. Generally, three kinds of EEG signals are utilized in constructing BCI systems,  which are steady-state evoked potential (SSVEP),  motor imagery, and event-related potential (ERP). A comparison of them should be added to make the paper more comprehensive.*

We thank the reviewer for the useful comments. The suggested information about different EEG signals that could be utilized to construct BCI systems was added to manuscript.

in the second paragraph of Section I (Introduction), we write:

The design of a high-performance EEG-based BCI system is an open research problem and requires accurate decoding of the EEG signals generated by neuro-electrical activities in the brain [9]–[11]. EEG-based BCI systems are classified as exogenous and endogenous, where the former requires an external stimulus to excite certain responses in the brain [12]. Depending on the type of stimulation, the exogenous BCIs use steady state visual evoked potentials (SSVEPs) [13], or on event-related potentials (ERPs), brain responses elicited in response to cognitive or sensory events [12]. The advantages of exogenous BCIs are related to high information transmission rate with little user training requirement [12]. However, an exogenous BCI system contrain a user to focus on the visual stimuli and, as a result, its usefulness may be limited, especially for severely motor-impaired people [19].

This paper focuses on the endogenous category of BCI systems which utilize sensory-motor rythms (SMRs) for control of external devices, independent of any stimuli. The SMRs represent the modulations of oscillatory activity in EEG induced by motor imagery of limb movement as input features [15], [16]. These changes in the oscillatory activity are known as the event-related desynchronization and synchronization (ERD/ERS) and are commonly used for mental state classification [17], [18].

[9] M. Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, and et al., “The brain-computer interface cycle,” J. Neural. Eng., vol. 6, no. 4, pp. 1–9, 2009.

[10] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi et al., “A review of classification algorithms for EEG-based brain–computer interfaces,” Journal of neural engineering, vol. 4, Jun. 2007.

[11] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, “A review of classification algorithms for EEGbased brain–computer interfaces: a 10 year update,” Journal of neural engineering, vol. 15, no. 3, p. 031005, 2018.

[12] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, “Eeg-based braincomputer interfaces using motor-imagery: Techniques and challenges,” Sensors, vol. 19, no. 6, p. 1423, 2019.

[13] S. P. Kelly, E. C. Lalor, R. B. Reilly, and J. J. Foxe, “Visual spatial attention tracking using high-density ssvep data for independent braincomputer communication,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 13, no. 2, pp. 172–178, 2005.

[14] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” Sensors, vol. 12, no. 2, pp. 1211–1279, 2012.

[15] G. Pfurtscheller, “Spatiotemporal erd/ers patterns during voluntary movement and motor imagery,” in Supplements to Clinical neurophysiology. Elsevier, 2000, vol. 53, pp. 196–198.

[16] G. Pfurtscheller and C. Neuper, “Future prospects of erd/ers in the context of brain–computer interface (bci) developments,” Progress in brain research, vol. 159, pp. 433–437, 2006.

[17] G. Pfurtscheller, “Eeg event-related desynchronization (erd) and synchronization (ers),” Electroencephalography and Clinical Neurophysiology, vol. 1, no. 103, p. 26, 1997.

[18] H. Cho, M. Ahn, M. Kwon, and S. C. Jun, “A step-by-step tutorial for a motor imagery–based bci,” in Brain–Computer Interfaces Handbook. CRC Press, 2018, pp. 445–460.

[19] J. Xie, G. Xu, J. Wang, M. Li, C. Han, and Y. Jia, “Effects of mental load and fatigue on steady-state evoked potential based brain computer interface tasks: a comparison of periodic flickering and motion-reversal based visual attention,” PloS one, vol. 11, no. 9, 2016.

*4. The paper deals with a two-class classification problem and achieves the highest average accuracy of 74.62% (C[128, 64, 32, 16, 8]\_K(3 × 8)) with the trial length of 4 s. In such a BCI system, the information transfer rate is not high, especially compared with SSVEP-based BCIs, which achieve accuracy beyond 90%  in a 40-class classification problem with the trial length of 1s (Masaki Nakanishi et al. 2018). Therefore,  when could we use such a BCI in practice? [Irina; We will check your response]*

The suggested paper demonstrates a method for high-speed BCI system with average accuracy of 90%. However, the method is related to detection of steady state visual evoked responses for BCI speller, while our manuscript concerns different type of BCI paradigm, MI-based BCI (spontaneous BCI). Although, the performance parameters related to trial length and accuracy in the aforementioned article are higher than those presented in our paper, the types of the signals that are used for evoked and spontaneous BCI are fundamentally different; that is to say, the P300 and SSVEP based BCIs belong to the family of dependent-BCI paradigms. These BCI paradigms constraint the users to focus on visual stimuli, for example, LCD monitor.  As a result, despite the fact that the P300 and SSVEP-based BCIs provide high accuracy and information transfer rate, the usefulness of such BCI system is limited because the user needs to focus on the screen all the time. This results in the mental fatigue of the user [3]; moreover, a long-term use of this system may cause severe eye strainbecause of the flashing lights. In contrast, motor imagery-based BCIs may provide more natural self-paced (independent of any external stimuli) control of the external environment, which will be useful and practical to most users, especially severely paralyzed people who are bed prone, even though they are lower in terms of accuracy, and speed.

Regarding information transfer rate, literature demonstrates that P300 and SSVEP show a relatively high information transfer in comparison with other BCI paradigms [4]. Thus, it might be expected that information transfer rate of the BCI system that is considered in our manuscript might be lower than in SSVEP-based BCI. Moreover, the presented results are based on the publicly available datasets and not on the self-conducted experiments where we could have affected the trial length.

To answer the last question about the use of proposed BCI in practice we added the following paragraph in Section I (Introduction):

The methods proposed in our manuscript could be applied to develop motor-imagery based BCI speller, which is a typical BCI-based application that helps people with disabilities to express thoughts [42]. Another possible area for application is robot control [43].

[3] J. Xie, G. Xu, J. Wang, M. Li, C. Han, and Y. Jia, “Effects of mental load and fatigue on steady-state evoked potential based brain computer interface tasks: a comparison of periodic flickering and motion-reversal based visual attention,” PloS one, vol. 11, no. 9, 2016.

[4] H. J. Baek, M. H. Chang, J. Heo, and K. S. Park, “Enhancing the usability of brain-computer interface systems,” Computational intelligence and neuroscience, vol. 2019, 2019.

[42] L. Cao, B. Xia, O. Maysam, J. Li, H. Xie, and N. Birbaumer, “A synchronous motor imagery based neural physiological paradigm for brain computer interface speller,” Frontiers in human neuroscience, vol. 11, p. 74, 2017.

[43] M. Aljalal, R. Djemal, and S. Ibrahim, “Robot navigation using a brain computer interface based on motor imagery,” Journal of Medical and Biological Engineering, vol. 39, no. 4, pp. 508–522, 2019.

*5. In the datasets used in this paper, data were recorded with different numbers of electrodes. I am confused about how to apply the method to these data with different channels. I hope that the authors could make it more clear in the revised version. [Berdakh, do you have these info to add?]*  
  
**Additional Questions:**  
**Does the paper contribute to the body of knowledge?:** Yes, but the authors need to explain their contributions more clearly.  
  
**Is the paper technically sound?:** Yes.  
  
**Is the subject matter presented in a comprehensive manner?:** Yes.  
  
**Are the references provided applicable and sufficient?:** Yes.