**REFEREE REPORT(S):  
Referee: 1  
  
COMMENTS TO THE AUTHOR(S)  
This study uses deep learning for motor imagery classification on EEG-fNIRS data and shows two main findings: 1) the combined use of EEG and fNIRS improved classification performance and 2) the proposed deep learning method enhanced classification performance. For evaluation, two factors (recording modality [EEG, EEG-fNIRS] and classification algorithm [DNN, SVM, LDA]) were investigated and statistical analysis showed that the combined use of EEG and fNIRS achieved the best results when using deep learning algorithm (DNN). Especially, it is very interesting to see that the effect of recording modality was greater compared to the effect of classification algorithm (Fig. 6).  
  
\* major comments:  
  
- The number of training examples was very small (N = 20 per class). Did you investigate the overfitting problem for classification algorithms (e.g., SVM, LDA)? This may be important for comparisons between DNN, SVM, and LDA.**

Although the overall task was repeated 20 times, a classification outcome was provided every second, that, considering the length of the task (20 seconds), gave an overall training set of 200 trials. [non dovremmo spiegare se e come abbiamo valutato l’overfitting?]

**- For comparisons between DNN and other classifier (SVM, LDA), it is also important to report some relevant information (spatial filtering, feature selection, feature extraction, hyperparameter, etc.) for SVM-and LDA-based classification, for example, how to select and extract features and how to optimize hyperparameter (e.g. regularization parameter C), etc. for each data type (single use (EEG) or combined use (EEG-fNIRS) of recording method).**Feature extraction was a simple beta band power within 1 sec of integration time.

**- The authors mentioned that there were statistically significant differences for all possible combinations (comparison pairs) in post-hoc analysis. Did the authors correct the p-values for multiple comparison, e.g., Bonferroni correction? Based on Figure 6, I am not sure whether all combinations were significantly different from each other (e.g., EEG + fNRIS-LDA vs. EEG-SVM, etc.).** Should be corrected  
  
**- Further, the statistical evaluation was not completely clear. The authors mentioned that the analysis was performed on differential accuracies with respect to EEG-LDA. It is not usual for statistical analysis (e.g., ANOVA). The reason for such procedure is also not clear. There were two factors: recording modality (two levels: EEG alone, EEG combined with fNIRS) and classification algorithm (three levels: DNN, SVM, LDA). The data (EEG or EEG-fNRIS) was repeatedly measured (evaluated) depending on classifier algorithms. On the other hand, classification algorithm (DNN, SVM, or LDA) was repeatedly evaluated depending on recording modality. Thus, a two-way repeated measures ANOVA with recording modality (two levels: EEG alone, EEG combined with fNIRS) and classification algorithm (three levels: DNN, SVM, LDA) as two within-subjects factors could be suitable for statistical analysis.**

We should perform rANOVA  
  
**- In discussion, the authors mentioned that CNN was also applied, and the performance was worse compared to DNN. How large were performance differences between CNN and DNN?**

Few point percent.  
  
**- In literature, deep learning algorithms are more effective for subject-to-subject transfer compared to "no transfer". It may be interesting to investigate subject transfer (e.g., LeaveOneSubjectOut).**

We should mention in the paper that we tried the subject transfer with poor results  
  
\* minor comments:  
**- The Figure 1 and Figure 2 are wrongly assigned. Please correct this!**

Fixed  
**- The citation styles in the manuscript are not consistent. Please check this!**

Check the style  
**- Please explain earlier how 200 seconds can be obtained for right or left motor imagery!**

Although the overall task was repeated 20 times, a classification outcome was provided every second, that, considering the length of the task (20 seconds), gave an overall training set of 200 trials.

**- The authors mentioned that the electrode impedances (EEG) were kept below 50 kOhm. Is that maybe typo (5 kOhm)?**

No, it was not a typo. For our EEG system, 50 kΩ is a value that falls within the range recommended by the manufacturer. Indeed, we acquired the EEG signals by using a 128-channel HydroCel Geodesics Sensor Net (Electrical Geodesics, Inc.). Contrarily to traditional EEG systems which usually require sensor-scalp impedances below 5 kΩ, the HydroCel Geodesics Sensor Net provides excellent recordings with impedances up to values in the 50-100 kΩ range thanks to the use of high input impedance amplifiers (see Tucker, 1993; or the Geodesics Sensor Net Technical manual available at http://www.egi.com/manuals-current).

Tucker D.M. (1993), ‘Spatial sampling of head electrical fields: the geodesic sensor net’, *Electroencephalography and Clinical Neurophysiology*, vol. 87, iss. 3, pp. 154-163

**- How many iterations were done during optimization (DNN)? How many training examples were used?**

Report in the text (it is already reported in the figure), 10 fold cross validation means 180 trainings for each iteration  
  
**Referee: 2  
  
COMMENTS TO THE AUTHOR(S)**  
**This study investigated the feasibility of applying a deep learning algorithm to hybrid EEG/NIRS BCI dataset. Deep learning has been attracting growing attention of researchers in BCI fields, and many studies used it to enhance BCI performance in recent. In this context, it is quite interesting to apply deep learning algorithms to hybrid EEG/NIRS BCI data. However, the study is not appropriate for publication with the current from due to mainly the shortcoming of experimental design.   
1.        The number of subjects is only 6, which is too small to prove the hypothesis. I recommend to conduct the experiment with at least more than 15 subjects for proper statistical analysis.**

Few more subjects?  
**2.        About the statistics, the authors used accuracy vales estimated from cross validation, but which is not a correct way to apply statistics. As I mentioned, more subjects should be recruited and one accuracy value that is a mean accuracy obtained from cross-validation should be used for each subject. Also, comparison should be fairly done without comparing single modality results with hybrid modality results.**

We already did as the reviewer suggested.  
**3.        The number of trials is 20 for each hand motor imagery, which is also too small.** **Please check previous related studies for the proper number of trials.**

Although the overall task was repeated 20 times, a classification outcome was provided every second, that, considering the length of the task (20 seconds), gave an overall training set of 200 trials.

**4.        It was mentioned that impedance was kept below 50 k, but which is somewhat high. In general, impedance is kept below 10k for reliably recording EEGs.**

For our EEG system, 50 kΩ is a value that falls within the range recommended by the manufacturer. Indeed, we acquired the EEG signals by using a 128-channel HydroCel Geodesics Sensor Net (Electrical Geodesics, Inc.). Contrarily to traditional EEG systems which usually require sensor-scalp impedances below 5 kΩ, the HydroCel Geodesics Sensor Net provides excellent recordings with impedances up to values in the 50-100 kΩ range thanks to the use of high input impedance amplifiers (see Tucker, 1993; or the Geodesics Sensor Net Technical manual available at http://www.egi.com/manuals-current).

Tucker D.M. (1993), ‘Spatial sampling of head electrical fields: the geodesic sensor net’, *Electroencephalography and Clinical Neurophysiology*, vol. 87, iss. 3, pp. 154-163

5**.        For NIRS, only 8 channels around motor cortex were used, but 123 electrodes from the whole scalp were used for EEG. Shouldn’t a subset of electrodes around motor cortex be used also for EEG as NIRS?**

We will report all electrodes and only 14 around the motor cortex;

6**.        Why only beta band was used for data analysis. In general, a frequency band of 8 – 30 Hz, including mu and beta, is used for decoding motor imagery tasks. Also, a subject-specific frequency band is used.**

Although commonly employed, EEG results were poorer when considering the combined mu and beta band ERD/ERS  
7**.        In figure 4, there is bilateral ERD for left hand motor imagery, not contralateral ERD.**

Although figure nm shows an activation on the ipsilater motor cortex, which was common to many subjects, the image still shows an higher activation in the contralateral cortex.  
8.        **There are no classification results for NIRS. Even though the performance of NIRS is lower than the others (EEG or EEG/NIRS), it would be helpful to provide it.**

Reports fNIRS results (or hide them).

**9.        For Figures 4b) and c), exact channel information should be provided.**

Add channel  
10.      **In Fig. 5, the results obtained from a single subject were shown, but they were meaningless without mean results.**

Average results are reported in figure 6

11.        **The captions of Figures 1 and 2 are flipped.**

Fixed  
**12.        There are also some recent EEG/NRIS hybrid studies.   
- Open Access Dataset for EEG+NIRS Single-Trial Classification," IEEE Transactions on Neural Systems and Rehabilitation Engineering,   
- Evaluation of a Compact Hybrid Brain-Computer Interface System," Biomed Research International.**

Add citations