EEG4Home: A Human-In-The-Loop Machine Learning Model for EEG-Based BCI

Xiaodong Qu[1,2][0000−0001−7610−6475] and Timothy J Hickey[1]

[1] Brandeis University, Waltham, MA, 02453, USA
[2] Swarthmore College, Swarthmore, PA, 19081, USA
{xiqu,tjhickey}@brandeis.edu

Abstract. Using Machine Learning and Deep Learning to predict cognitive tasks from electroencephalography (EEG) signals has been a fast-developing area in Brain-Computer Interfaces (BCI). Yet, one fundamental challenge is that EEG signals are vulnerable to various noises. This paper identifies two types of noise: external noise and internal noise. External noises are caused by subjects' movement or sensors' instability, and internal noises result from the subjects' random mental activities due to the subjects' mind wandering during the experiment. When the participants conduct other mental activities researchers cannot infer, it will result in data corresponding to' unknown' tasks. We pioneer a Human-In-The-Loop (HITL) machine learning model, EEG4Home, to handle both types of noise and increase the accuracy of predicting known tasks. We introduce a plateau threshold to remove external noise and an unknown threshold set to detect unknown tasks to remove internal noise. Both unsupervised (such as K-Means) and supervised (Such as Random Forests, CNN, and RNN) learning algorithms are implemented in this HITL approach. We use the Thinking1 BCI experiments dataset with sixty subjects[3]. The average prediction accuracy of known tasks has increased from 56.8 percent to 65.1 percent. Overall, this EEG4Home model enables researchers or end-users to gain higher prediction accuracy and more interpretable results.

Keywords: Brain Machine Interface · Machine Learning · Interpretability · BCI for Everyone · Human-Centered Computing

1 Introduction

Researchers from both clinical and non-clinical fields have applied EEG-based Brain Computer Interaction (BCI) techniques in many different ways [13,4], such as diagnosis of abnormal states, evaluating the effect of the treatments, seizure detection, motor imagery tasks [5], and developing BCI-based games [3]. Previous studies have reviewed existing machine learning and deep learning algorithms [27,14,30,24,15,11] for classifying EEG-based BCI tasks.

[3] available to academic researchers by request

However, EEG-based BCI remains an expensive and time-consuming approach, more for research or clinical settings than for non-expert users to use at home. Also, EEG signals have a relatively low signal-to-noise ratio, mainly because of the contact of sensors and skin for several current consumer-grade devices. The outlier issue is also a concern for the EEG data. Some participants have had difficulties generating stable EEG signals for various reasons, such as head shape, hairstyle, and the EEG headsets electrodes. Current EEG-based BCI algorithms and frameworks are more for clinical-style devices and less for an affordable instrument with at-home settings.

Can we develop an easy-to-use machine learning EEG framework for non-expert end-users to classify their everyday learning tasks better? This paper proposed a framework, EEG4Home, to eliminate these disturbing artifacts by utilizing a Human-In-The-Loop (HITP) machine learning approach. We evaluated this approach on four EEG data sets and presented the main results with increased prediction accuracy and more interpretable feedback to both researchers and end-users.

2 Related Work

There are several existing frameworks for analyzing EEG data, such as BCI2000 [27]; OpenVibe [23]; USC Brainstorm [28]; Harvard HAPPE [6]; Stanford EEG-BIDS [18]; and JHU BCI2000web [16]. These frameworks are more for EEG data with clinical devices, usually 128 or 64 electrodes.

Since 2015, several consumer-grade EEG-based BCI headsets have been released, which are more affordable for researchers and non-expert end-users, as shown in Figure 1.

After comparing different non-invasive, easy-to-use options [10,26], We selected Muse and (Muse 2) headsets to collect EEG data for our experiments [22,19,21,20].

Device Name	Manufacturer	Price*	Channels	Main Functions advertised
Muse 2	InteraXon(Canada)	$220	4	Meditation, self-assessment
Muse S	InteraXon(Canada)	$300	4	Meditation, self-assessment, Sleep
Insight	Emotiv Systems	$300	5	Self-assessment, device control
Epoc+	Emotiv Systems	$700	14	Self-assessment, device control
MindWave Mobile 2	Neurosky	$100	1	Meditation,Self-assessment, Gaming, device control
'Mark IV'	Open BCI	$500 to 600	8 to 16	(unassembled), self-assessment open-development,

* price, as of Jun 20, 2020. From the Manufacturers' official websites.

Fig. 1. EEG Head Sets

Exp (Task)	(1)	(2)	(3)	(4)	(5)
[19]	Math	Close-eye Relax	Read	Open-eye Relax	None
[22]	Python Passive	Math Passive	Python Active	Math Active	None
[21]	Read	Write Copy	Write Answer	Type Copy	Type Answer
[20]	Think	Count	Recall	Breathe	Draw

Table 1. Tasks in Experiments

After getting approval from the IRB, we invited more than a hundred college students to participate in four EEG recording experiments, mainly asking them to conduct learning-related cognitive tasks, such as reading, writing, and typing. Table 1 is a summary of the activities in these experiments. Each row is an experiment mentioned in a paper. Each of the first two experiments has four tasks. So the fifth task in the table is labeled as 'None'. Each of the last two experiments has five tasks. The researchers and end-users co-developed these tasks. The goal is to make the experimental tasks as similar as possible to the frequently conducted tasks at home.

3 Methods

Researchers continue to develop new algorithms for EEG-based BCI, such as [9,12,1,17,8,7,25,29,2]. However, the quality of the results heavily rely on the quality of the datasets. We proposed a human-centered framework to collect data faster, with less noise, and provide participants with more interpretable results.

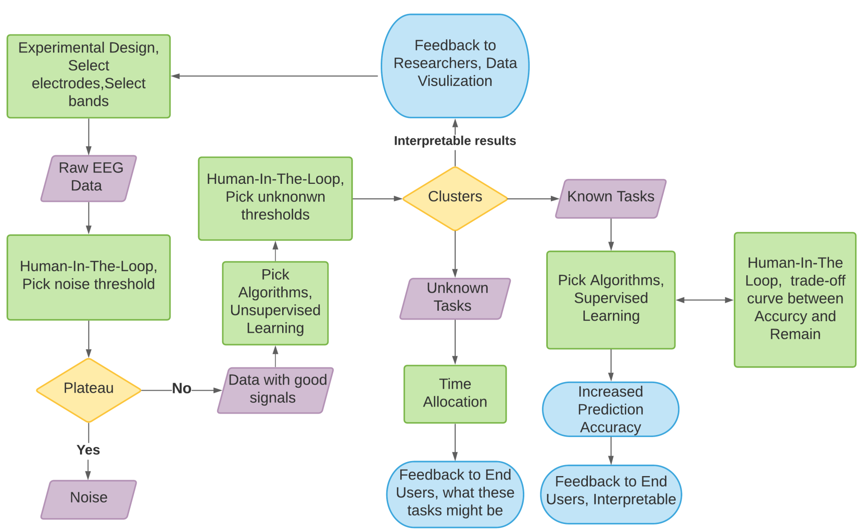

Fig. 2. Experiment and Data analysis Flow Chart.

3.1 Framework Figure 2 demonstrated our closed-loop framework, EEG4Home. The starting point is the experimental design on the top left. Here the researchers from computer science and neuroscience discussed the selection of electrodes. We talked about what EEG devices we use and which electrodes we care about the most. For example, we chose the four-electrode Muse headsets in the four experiments above. And we pay more attention to the AF7 and AF8 electrodes.

Then we collected the raw EEG data. For Muse 2016, we used the research software provided by the Muse manufacturer, InteraXon. For Muse 2018 and later, because InteraXon no longer supports the research software, we switched to Mind Monitor, a third-party application. We used Matlab and Python 3 for data analysis.

Next, we implemented the Human-In-The-Loop approach. Figure 3 is an example of a five-minute session from the Read-Write-Type (RWT) experiment. The plateaus of the EEG signal the noise (5.4%), and researchers decided how long a meaningful noise threshold should be for this experiment. For example, we chose 1.4 seconds as the noise threshold in these four experiments. If a plateau lasts more than that, we marked this time range of signals as the noise.

After removing the noise, we picked algorithms to identify the tasks. First, we used unsupervised learning to figure out the natural clusters of the tasks. We tested unsupervised clustering algorithms, such as DBSCAN, K-Means and GMM. K-Means provided the most interpretable results for our experiments. When we presented the clustering results from k-means to the end-users, they

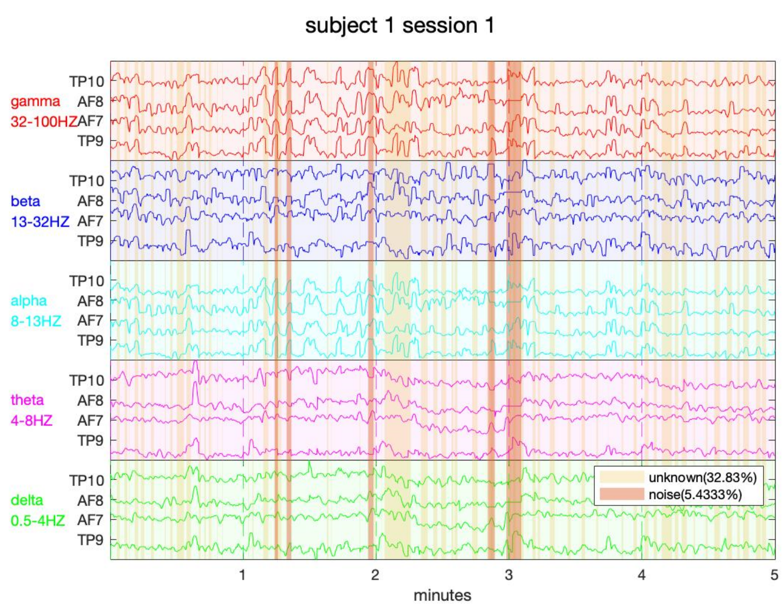

Fig. 3. Noise, unknown, and known tasks, experiment RWT: subject one, session one.

could recall the difference between known and unknown tasks for more than 80% of the experiment time. Typically, they described a 'mind wandering' state that was not the task in the experimental design.

For the unknown tasks, we marked it on the feedback images to the end-user and started an investigation about what it could be. The 'mind wandering' state is the leading reason for such unknown tasks.

For the known tasks, we then picked the supervised learning algorithms. The Random Forest and LSTM beat most other machine learning and deep learning algorithms, and combining them using the Time-Continuity-Voting algorithm outperformed others. After removing the noise and unknown tasks, the prediction accuracy increased substantially, with increased interpretability to the end-users.

Then the researchers and the end-users reviewed the noise, the unknown tasks, and known tasks together. Based on this discussion, the researchers and end-users developed the next version of the experiment. This closed-loop approach generated less noise and unknown tasks during each iteration.

3.2 Algorithms Figure 4 is an overview of how the pair of cluster number K in K-means and unknown threshold co-determined the prediction accuracy and the percentage of data remains for the known tasks. We recognized the pattern that, both indicators contribute to the results, and there is an intersection line of the two surfaces of Accuracy, and the Percentage of the Data remains. The researchers from computer science and neuroscience then discussed the trade-off between Accuracy and the Percentage of the Data that remains.

One crucial innovation here is the Human-In-The-Loop approach. The pattern is when the percentage of data remains decreased. We can get a higher prediction accuracy for the known tasks using this approach. But how to decide the best

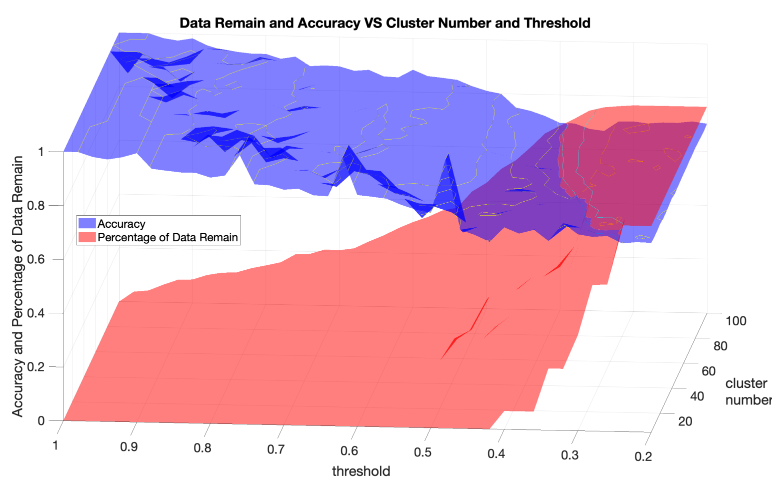

Fig. 4. Accuracy and Remain, 3D

threshold? Is the intersection line necessarily the best? After several rounds of discussion from computer scientists, neuroscientists, and end-users, we agree that interpretability is also essential in such an approach. We eventually selected the threshold to link the end-users reflection best to the machine learning clustering and prediction results.

4 Results

We found the following machine learning results with this framework.

Figure 5 demonstrated the trade-off between prediction accuracy and the percentage of data remains. The X-Axis ranks threshold pairs (cluster number K in K-means and the unknown threshold). We included the top 150 pairs, corresponding to 0.87 to 0.74 for the known tasks. Compared with the random of 0.2 (because we predict a task from five possible tasks), this accuracy provides meaningful feedback to the end-users.

The Y-Axis is the prediction accuracy for the blue line. It also shows the percentage of data-remaining using the orange line. As we can observe from this figure, when the accuracy decreased from 0.87 to 0.74, the data-remaining increased from 0.37 to 0.94.

During our Human-In-The-Loop discussions, the neuroscientists preferred to keep more data remaining as it represents more experiment time. Computer scientists prefer high accuracy. The end-users prefer a model that can classify

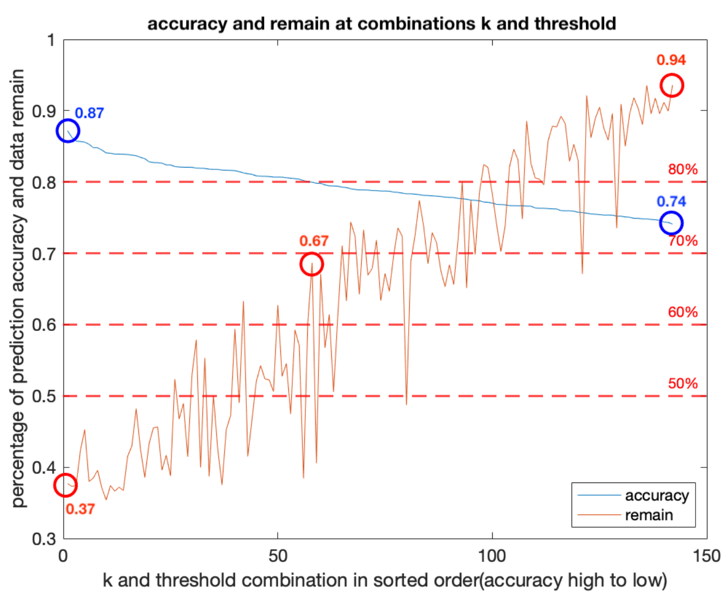

Fig. 5. Accuracy and Remain, 2D

```
noiseRemoved_data = readcsv("subjID.csv");

for threshold = 0.3:0.02:0.6
    % removed unknown using kmeans
    unknownRemoved_data = Kmeans(noiseRemoved_data,threshold);
    % calulate REMAIN
    REMAIN = size(unknownRemoved_data)/size(original_data);
    % use randomforest to predict tasks
    prediction = RandomForest.fit(unknownRemoved_data);
    % calulate ACCURACY
    ACCURACY = compare(prediction,label);
end

plot2D(ACCURACY,REMAIN,"sort","descend ACCURACY ");
```

Fig. 6. Pseudo Code

their experience – the noise, unknown tasks, and known tasks. The model is expected to match the end-users reflection about the experiment. Ideally, when the results show the end-user was doing an unknown task from 3 minutes and 5 seconds to 3 minutes and 15 seconds, the end-user might remember and agree with that type of classification.

Although the individual difference is significant, the trade-off pattern between accuracy and data-remaining is the same for more than 90% of the end-users in these experiments. Figure 6 is the pseudo code to generate the results in Figure 5 for each end-users. Figure 7 is the results of the noise, unknown tasks, and known tasks for different participants in the same experiment.

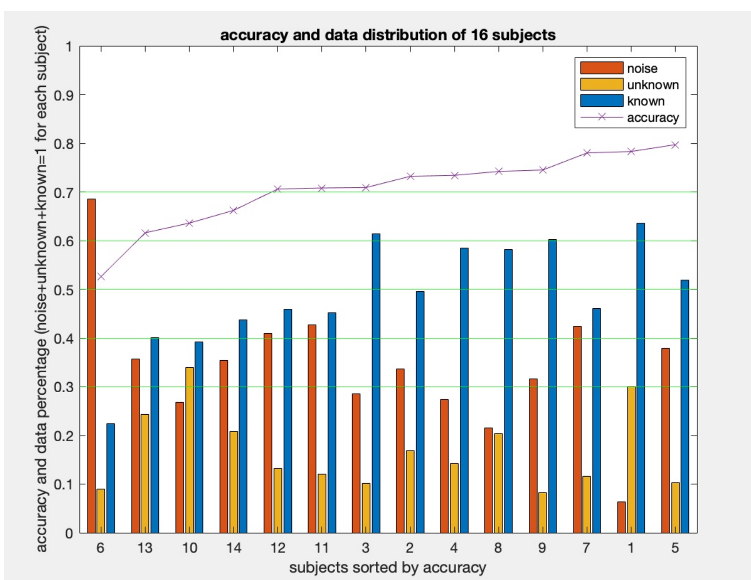

Fig. 7. Experiment RWT: noise, unknown, and known tasks percentage.

The X-axis in Figure 7 is the subject ID of the experiment, ordered by prediction accuracy, from low to high. The Y-axis is the prediction accuracy for that accuracy line. And the percentage of the data for the colored bars. As we can see, subject 5 has about 35% of noise, 10% of unknown tasks, and 55% of the known tasks. Using this approach, the prediction accuracy for the known task for this subject 1 is close to 80%. While subject 6 has almost 70% of noise, that implies the improvement for the recording of this Subject's EEG could be more focused on how to avoid such a high noise level.

Figure 8 is the summary of prediction accuracy and data-remaining for different end-users. The X-axis is the subject ID ordered by the prediction accuracy of the known task. The Y-axis is the percentage. The blue bar for the prediction accuracy. After removing the noise and unknown tasks, the orange bar for the data-remaining is the known tasks. The Yellow bar and purple bar on the right are the averages of all the participants in this experiment.

Figure 9 is an example of the visual feedback to each end-user. The X-axis is the predicted task in this experiment. The Y-axis is the actual task. The diagonal cells represent the accuracy of successfully predicting a certain task. For example, Task 5, as T5 in this figure, is classified with 89% accuracy and is not confused with other tasks. While task 2 is 69% successfully predicted, it has an 11% chance to be misclassified as Task 1 or Task 4.

5 Discussion

This paper presents an EEG4Home framework, which aims to be a fast and more interpretable solution for non-expert BCI end-users at home. The short-term goal

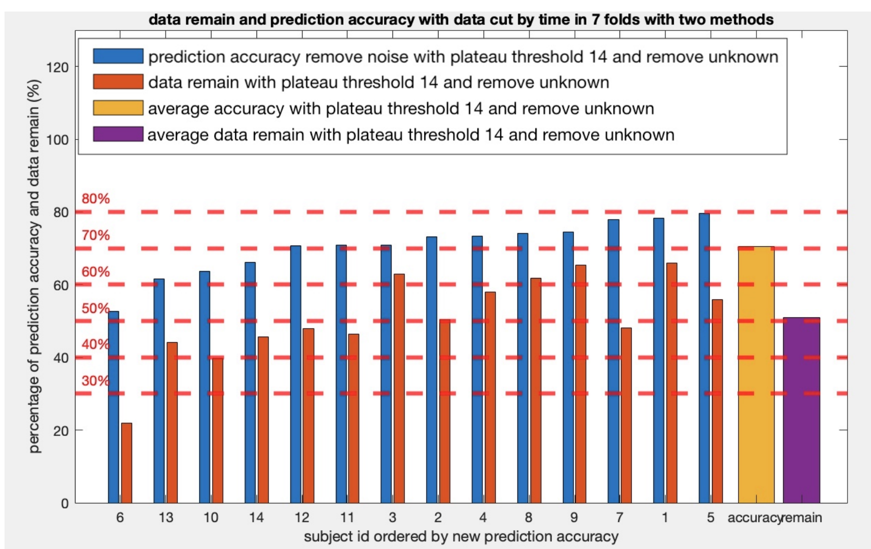

Fig. 8. Experiment RWT: prediction accuracy.

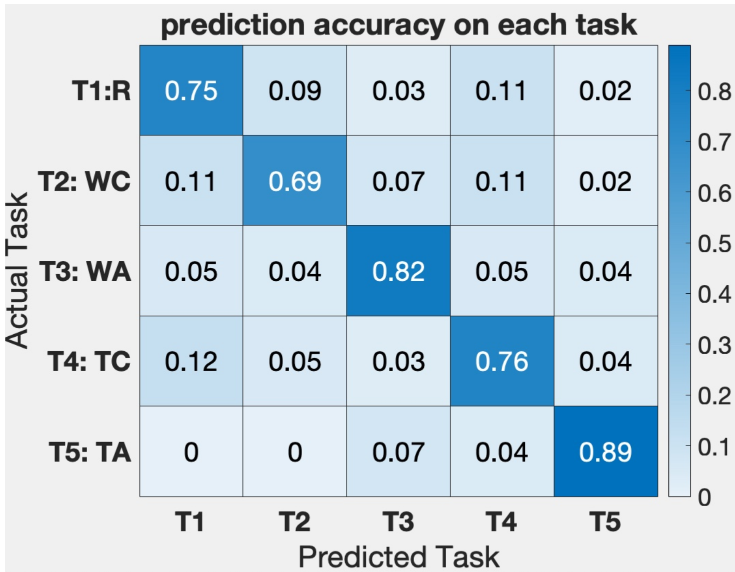

Fig. 9. Feedback to End-users

of this EEG4Home algorithm development is to inspire new machine learning and deep learning approaches for decoding such cognitive tasks from the Human Brain. The long-term goal of this research is to get more end-users involved in using EEG-based BCI. Eventually, using EEG-based BCI could be as common and simple as using a smartphone nowadays.

Based on the research we have reviewed in this paper, we identified some guidelines on whether to keep the noise and unknown tasks before classifying the known tasks. We also recognized several open research questions that deserve to be answered to make EEG-based BCI more user-friendly to non-expert end-users.

The following guidelines could help end-users conduct similar experiments. The current non-invasive EEG devices still generate a large amount of noise, but user training can help control this noise. An EEG coach role could be constructive in this situation. Mind-wandering is an unknown everyday activity in our experiments. When designing tasks, think about the mind-wandering activities, and assume at least one more task as unknown or mind-wandering could appear in the classification results.

Providing visual feedback to end-users and gathering user experience are the best practices to improve the experimental design. Non-expert end-users can usually record meaningful EEG signals with two to three training sessions. But some individuals struggled to get reliable data for the experiments. So this EEG-based BCI still has limitations before most smartphone users can easily use it every day. One exciting finding is that many end-users are motivated to use EEG-based BCI more for fun, game-like activities. That could be a direction worth more investigation.

6 Conclusion

In this paper, we presented the EEG4Home model. Four existing experiments were used in this research. Human-In-The-Loop Machine learning and deep learning approaches have been implemented in data analysis. The results showed a reasonable high prediction accuracy and interpretable task types for most end-users. This framework could be a building block towards the future of everyone using non-invasive, wireless, and affordable EEG-based BCI systems every day at home, similar to current smartphone usage.

References

1. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448 (2015)
2. Chevalier, J.A., Gramfort, A., Salmon, J., Thirion, B.: Statistical control for spatio-temporal meg/eeg source imaging with desparsified multi-task lasso. arXiv preprint arXiv:2009.14310 (2020)
3. Coyle, D., Principe, J., Lotte, F., Nijholt, A.: Guest editorial: Brain/neuronal-computer game interfaces and interaction. IEEE Transactions on Computational Intelligence and AI in games **5**(2), 77–81 (2013)
4. Craik, A., He, Y., Contreras-Vidal, J.L.: Deep learning for electroencephalogram (eeg) classification tasks: a review. Journal of neural engineering **16**(3), 031001 (2019)
5. Devlaminck, D., Waegeman, W., Bauwens, B., Wyns, B., Santens, P., Otte, G.: From circular ordinal regression to multilabel classification. In: Proceedings of the 2010 Workshop on Preference Learning (European Conference on Machine Learning, ECML). p. 15 (2010)
6. Gabard-Durnam, L.J., Mendez Leal, A.S., Wilkinson, C.L., Levin, A.R.: The harvard automated processing pipeline for electroencephalography (happe): standardized processing software for developmental and high-artifact data. Frontiers in neuroscience **12**, 97 (2018)
7. Hersche, M., Millán, J.d.R., Benini, L., Rahimi, A.: Exploring embedding methods in binary hyperdimensional computing: A case study for motor-imagery based brain-computer interfaces. arXiv preprint arXiv:1812.05705 (2018)
8. Hersche, M., Rellstab, T., Schiavone, P.D., Cavigelli, L., Benini, L., Rahimi, A.: Fast and accurate multiclass inference for mi-bcis using large multiscale temporal and spectral features. In: 2018 26th European Signal Processing Conference (EUSIPCO). pp. 1690–1694. IEEE (2018)
9. Hosseini, M.P., Hosseini, A., Ahi, K.: A review on machine learning for eeg signal processing in bioengineering. IEEE reviews in biomedical engineering **14**, 204–218 (2020)
10. Ienca, M., Haselager, P., Emanuel, E.J.: Brain leaks and consumer neurotechnology. Nature biotechnology **36**(9), 805–810 (2018)
11. Kaya, M., Binli, M.K., Ozbay, E., Yanar, H., Mishchenko, Y.: A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. Scientific data **5**(1), 1–16 (2018)

12. Li, G., Lee, C.H., Jung, J.J., Youn, Y.C., Camacho, D.: Deep learning for eeg data analytics: A survey. Concurrency and Computation: Practice and Experience **32**(18), e5199 (2020)
13. Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. Journal of neural engineering **15**(3), 031005 (2018)
14. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain–computer interfaces. Journal of neural engineering **4**(2), R1 (2007)
15. Miller, K.J.: A library of human electrocorticographic data and analyses. Nature human behaviour **3**(11), 1225–1235 (2019)
16. Milsap, G., Collard, M., Coogan, C., Crone, N.E.: Bci2000web and webfm: browser-based tools for brain computer interfaces and functional brain mapping. Frontiers in Neuroscience p. 1030 (2019)
17. Orsborn, A.L., Moorman, H.G., Overduin, S.A., Shanechi, M.M., Dimitrov, D.F., Carmena, J.M.: Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. Neuron **82**(6), 1380–1393 (2014)
18. Pernet, C.R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., Oostenveld, R.: Eeg-bids, an extension to the brain imaging data structure for electroencephalography. Scientific data **6**(1), 1–5 (2019)
19. Qu, X., Hall, M., Sun, Y., Sekuler, R., Hickey, T.J.: A personalized reading coach using wearable EEG sensors-a pilot study of brainwave learning analytics. In: CSEDU (2). pp. 501–507 (2018)
20. Qu, X., Liu, P., Li, Z., Hickey, T.: Multi-class time continuity voting for eeg classification. In: International Conference on Brain Function Assessment in Learning. pp. 24–33. Springer (2020)
21. Qu, X., Mei, Q., Liu, P., Hickey, T.: Using eeg to distinguish between writing and typing for the same cognitive task. In: International Conference on Brain Function Assessment in Learning. pp. 66–74. Springer (2020)
22. Qu, X., Sun, Y., Sekuler, R., Hickey, T.: EEG markers of stem learning. In: 2018 IEEE Frontiers in Education Conference (FIE). pp. 1–9. IEEE (2018)
23. Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., Lécuyer, A.: Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. Presence **19**(1), 35–53 (2010)
24. Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J.: Deep learning-based electroencephalography analysis: a systematic review. Journal of neural engineering **16**(5), 051001 (2019)
25. Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., Engemann, D.A.: Manifold-regression to predict from meg/eeg brain signals without source modeling. arXiv preprint arXiv:1906.02687 (2019)
26. Sawangjai, P., Hompoonsup, S., Leelaarporn, P., Kongwudhikunakorn, S., Wilaiprasitporn, T.: Consumer grade eeg measuring sensors as research tools: A review. IEEE Sensors Journal **20**(8), 3996–4024 (2019)
27. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: Bci2000: a general-purpose brain-computer interface (bci) system. IEEE Transactions on biomedical engineering **51**(6), 1034–1043 (2004)
28. Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M.: Brainstorm: a user-friendly application for meg/eeg analysis. Computational intelligence and neuroscience **2011** (2011)

29. Tu, T., Paisley, J., Haufe, S., Sajda, P.: A state-space model for inferring effective connectivity of latent neural dynamics from simultaneous eeg/fmri. Advances in Neural Information Processing Systems **32**, 4662–4671 (2019)
30. Zhang, X., Yao, L., Wang, X., Monaghan, J.J., Mcalpine, D., Zhang, Y.: A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. Journal of Neural Engineering (2020)