EEG daydreaming, a machine learning approach to detect daydreaming activities

Ruyang Wang[1], Xiaodong Qu[2]

[1] Brandeis University, Waltham MA 02453, USA
ruyangwang@brandeis.edu
[2] Swarthmore College, Swarthmore PA 19081, USA
xqu1@swarthmore.edu

Abstract. In this paper, we propose a new method to detect noise hindrances in Electroencephalographic (EEG) signals caused by mental distractions, which we named "daydreaming signals." Our approach is based on sliding windows and aims to detect and locate these daydreaming signals to specific points in time. We expect to get cleaner data and, therefore, higher prediction accuracy in current available EEG datasets by removing these daydreaming signals. Beyond these improvements to existing data, this approach also has the potential to improve the quality of future data collection, as researchers can discover the pattern of daydreaming signals in trial rounds and deal with these signals accordingly.

Keywords: Design methods and techniques · Machine learning · Supervised learning · Electroencephalography(EEG) · Sliding windows · EEG signal classification.

1 Introduction

Electroencephalography (EEG) is widely used in Brain-Machine Interfaces research [15]. The classification of cognitive tasks using EEG signals has been the focus of discussion in the past few decades [16]. The low signal-to-noise ratio is a common hindrance in EEG signal classification. While multiple types of machine learning and deep learning algorithms have been used for cognitive task classification [4, 12, 23], the accuracy of EEG signal classification can hit a bottleneck without a proper separation of noise. Noise in EEG data could come from a variety of sources, which can be identified as two main types: i) noise from the outside, including factors like environmental noise, noise caused by experimental settings, and noise caused by static electricity; and ii) noise from the human body, including physical activity such as blinking and breathing, and mental activity such as distracting thoughts [25]. Although large quantities of research have been conducted and shown success in removing the external noise [14], the problem of detecting and removing internal noise remains an area in need of more exploration. This paper focuses on the latter source of noise, noise from mental activity, and aims to design an algorithm to detect and remove noise caused by mental distractions.

In experiments, subjects sometimes lose focus on the task assigned to them. Instead of performing the designated tasks, they may engage in other activities intentionally or unintentionally. Such distractions can lead to noise in EEG signals that is different from expected signals of expected activities. We designed an algorithm to detect and remove this kind of noise, which we have named "daydreaming signals." Our approach first removed the environmental noise with plateau threshold [19], then analyzed the cleaned data by two rounds of baseline classification analysis combined with sliding windows. We implemented this algorithm with the publicly available Thinking1 BCI experiments dataset.

Since the Thinking1 BCI experiments dataset is a relatively small dataset, we focused on classifiers that work well with EEG data and small datasets. We chose three classifiers for baseline classification: Random Forest (RF), Support Vector Machines (SVM), and Long Short-Term Memory (LSTM). Random Forest is a classifier that combines numerous randomized decision trees [6]. Random Forest has shown outstanding performance when there is a larger number of variables compared with the number of observations [3]. It was also proved in EEG dataset that the Random Forest classifier outperforms other classifiers when the sample set is small [16]. Support Vector Machines is a linear classifier that implements a hyperplane to identify classes [8]. General advantages of using SVM include good generalization properties, as well as their resistance to overtraining and the curse of dimensionality [17]. Support Vector Machines are commonly employed in human-computer interaction to identify physiological patterns, and have shown high performance in prediction accuracy of EEG data [22]. Long Short-Term Memory is a deep learning architecture that uses an artificial Recurrent Neural Network (RNN) [10]. Long Short-Term Memory has been widely used on EEG signal classification and showed high accuracy when combined with other classification techniques [9, 2]. Neural Networks usually work well on large datasets, yet Long Short-Term Memory can show high accuracy with small datasets [7, 13, 20, 24].

As for results, Table 1 shows the old (without sliding windows analysis) and new (with sliding windows analysis) prediction accuracy of three baseline methods. Compared with traditional Random Forest, Support Vector Machines, and Long Short-Term Memory classifications, this approach increased the average prediction accuracy from 55.0 percent to 66.1 percent, from 41.2 percent to 65.5 percent, and from 46.0 percent to 56.3 percent, respectively.

Table 1. Average Prediction Accuracy for Baseline Methods

	RF	SVM	LSTM
Prediction Accuracy (old)	55.0	41.2	46.0
Prediction Accuracy (new)	**66.1**	**65.5**	**56.3**

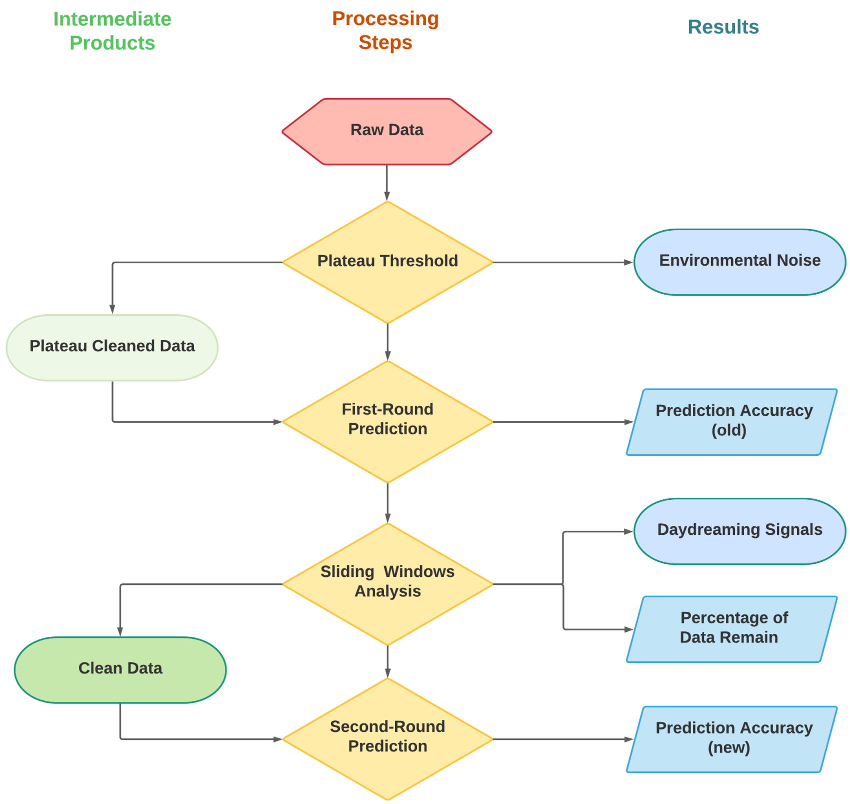

Fig. 1. Workflow chart for the proposed method that detects and removes daydreaming signals. After removing the environmental noise with plateau threshold, the cleaned data is analyzed by two rounds of baseline method analysis combined with sliding windows. Three types of EEG signals are identified: environmental noise (e), daydreaming signals (d), and clean data (c); data remaining percentage, old and new prediction accuracy are recorded.

.

2 Proposed Method

Figure 1 shows the workflow of our algorithm with four significant steps contained: Plateau Threshold, First-Round Prediction, Sliding Windows Analysis, and Second-Round Prediction. The detailed method is described in this section.

2.1 Data Source and Preprocess The Thinking1 BCI experiments dataset contains 16 health non-expert subjects. Signals are captured using Muse neuromonitoring headset, which are suggested for meditation evaluations and non-medical usages [11]. Despite having a smaller number of available electrodes

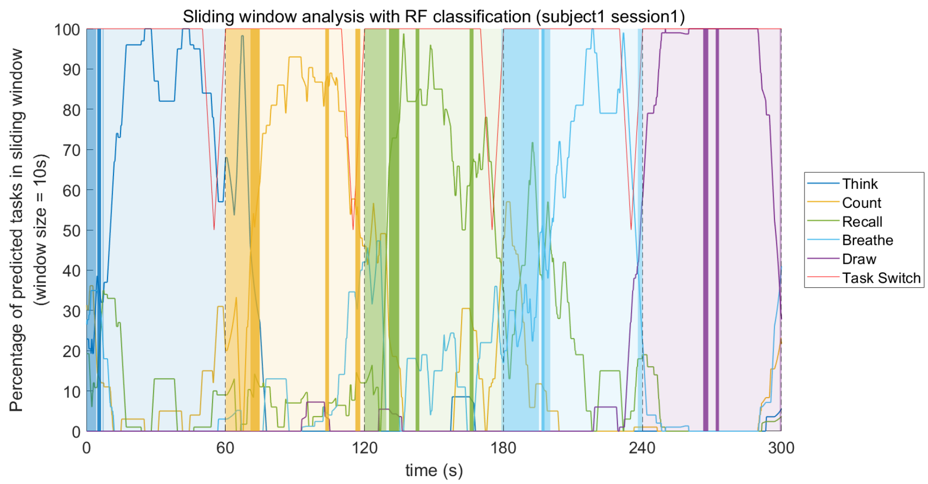

Fig. 2. Result of sliding window analysis on subject 1 session 1 using Random Forest as baseline method. Each task is represented by one color. Environmental noise, day-dreaming signals, and clean data are marked with different degrees of background color saturation within each task. Environmental noise has the largest degree of saturation, daydreaming signals have a middle degree of saturation, and clean data has the smallest degree of saturation.

.

compared with medical-use devices (4 electrodes vs. 16 electrodes), Muse neuromonitoring headset can collect accurate data at a lower cost [5, 1]. In Thinking1 BCI experiments dataset, each subject conducted a 30-minute experiment that contained six sessions on five tasks (Think [T], Count [C], Recall [R], Breathe [B], Draw [D]) in a random sequence. Before conducting our proposed algorithm, all EEG data for analysis were pre-processed to remove external noise. In the analysis of this experiment, environmental noise (e) caused by poor contact was removed using the "plateau threshold method." For more details, please refer to the previous work [19].

2.2 Baseline Prediction For the baseline prediction, We divided each session into ten folds chronologically. Fold 1 represents the first one-tenth of recorded data in the time sequence, fold 2 represents the second one-tenth of recorded data in the time sequence, and so forth. We conducted 10-fold cross-validation with the baseline prediction method (RF, SVM, or LSTM) and recorded the classification results for each EEG data point as $label_X$. According to the experiment setup, there are five possible labels for each data point, representing five cognitive tasks: $label_T$, $label_C$, $label_R$, $label_B$ are $label_D$. Prediction accuracy and the percent of data remaining for each subject were also recorded for future comparison.

2.3 Create Sliding Windows Sliding windows were created to calculate the distribution of predicted labels in a given time period. In this approach, we created a sliding window with a size of 100 data points (10s) with 99 data points (9.9s) overlap in the time sequence. Within each sliding window, we calculated the percentage of each predicted label. For the $label_x$ (x is either T, C, R, B, or D) that has the highest percentage in one sliding window, We would consider the 100 data points in this sliding window to have more notable features to the corresponding $task_x$. That is to say, the subject may have a brain activity closer to $task_x$.

2.4 Detect and Remove Daydreaming Signals After creating the sliding windows and calculating the percentage of labels in sliding windows, the next step was to mark the daydreaming signals. To find the daydreaming signals, we looked into each sliding window to find the $label_x$ that had the highest percentage among the 100 labels; each window is then marked as $label_x$ as a whole. We took the designed $task_x$ in each session as the ground truth to compare with $label_x$. If $label_x$ were different from the ground truth, we would mark the first data point in this sliding window as a "daydreaming signal (d)" and remove this data point. The data remaining after this step is identified as "clean data (c)." Figure 2 shows an example of sliding window percentage and distribution of daydreaming signals in subject 1 session 1.

2.5 Second-Round Prediction After removing the daydreaming signals, we repeated step 2 (10-fold cross-validation with the same baseline classification method) for the rest of the clean data and calculated the final prediction accuracy and percent of data remaining. The result was compared with prediction accuracy and the percentage of data remaining in the first-round prediction.

3 Results

3.1 Prediction Accuracy and Data Remaining

In Figure 3, we show the prediction accuracy of baseline methods and the improved prediction accuracy after removing daydreaming signals. Three baseline classification algorithms are marked in three different colors. The baseline accuracy is represented in dashed lines, and the new prediction accuracy by implementing our algorithm is represented in solid lines. Compared with traditional Random Forest classification (red dashed line), our algorithm (red solid line) improved the average accuracy from 55.0 percent to 66.1 percent. Compared with traditional Support Vector Machines classification (green dashed line), our algorithm (green solid line) improved the average accuracy from 41.2 percent to 65.5 percent. And compared with traditional Long Short-Term Memory classification (blue dashed line), our algorithm (blue solid line) improved the average accuracy from 46.0 percent to 56.3 percent. Our algorithm has significantly increased prediction accuracy compared with all three baseline classification algorithms.

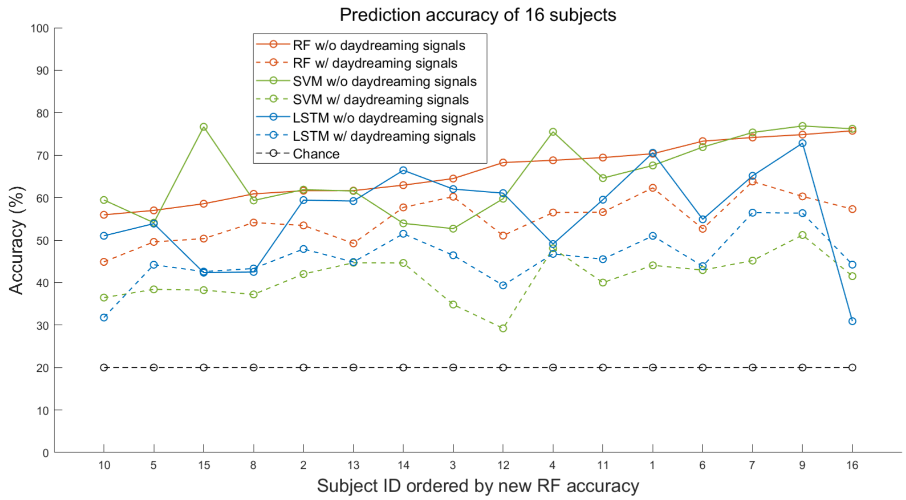

Fig. 3. Prediction accuracy of 16 subjects: baseline classification algorithms (dashed lines) vs. our improved algorithm (solid lines)

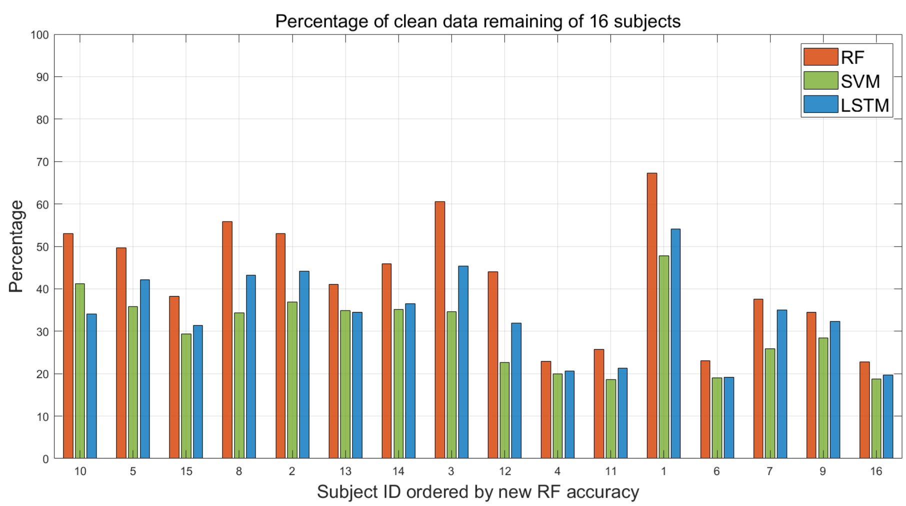

Fig. 4. Percentage of clean data remaining of 16 subjects after removing environmental noise and daydreaming signals for three baseline methods: Random Forest (red bars), Support Vector Machines (green bars), Long Short-Term Memory (blue bars)

We also provide the data remaining condition for three baseline algorithms after removing daydreaming signals in Figure 4. Compared with Support Vector Machines (with an average of data remaining of 30.2 percent) and Long Short-

Term Memory (with an average of data remaining of 34.1 percent), RF classifier has a higher average data remaining percent (42.2 percent). On average, using the RF classifier as a baseline algorithm leaves more data remaining than the condition of using the Support Vector Machines classifier and the Long Short-Term Memory classifier. The Long Short-Term Memory classifier leaves slightly more data remaining than the Support Vector Machines classifier.

3.2 Analysis of Daydreaming Signals

Choice of Baseline Method for Daydreaming Signals Analyses Our algorithm includes different steps to remove different types of noises. In the step of plateau threshold detection, environmental noise is removed, and our algorithm detects and removes daydreaming signals; the remaining data is classified as clean data. Since the environmental noises are detected using the same plateau threshold, the original data remaining is the same among all three baseline methods.

However, the daydreaming signal is detected based on the result of the first-round classification. Although the percentage of daydreaming signals and its distribution condition is quite similar among the three baseline methods, slight differences exist among different baseline conditions. According to the result of our experiment, the Random Forest classifier has the highest prediction accuracy, and the highest data remain. Therefore, Random Forest is considered the most accurate detection algorithm for daydreaming signals among the three baseline algorithms, and we choose Random Forest to present the results in this section. Detailed discussion on the choice of baseline method will be provided in the Discussion and Future Work section. The percentage of different types of data in the condition of using the other two baseline classifiers is provided in the supplement.

Distribution Pattern of Daydreaming Signals Focusing on the condition of using Random Forest classifier as baseline method, we further analyzed the distribution pattern of the daydreaming signals.

In Figure 5, we provide a bar graph of the percentage of each type of signal (environmental noise, daydreaming signals, and clean data) of all sessions in total for each subject. Daydreaming signals are detected in all 16 subjects, and the percentage of daydreaming signals can be as high as 34.3 percent or as low as 7.2 percent, showing significant differences among subjects.

We then utilize a box and whisker plot to determine daydreaming signal distribution in different tasks (Figure 6). By looking at the distribution pattern of daydreaming signals in different tasks, we can see that: i) a considerable variance exists between different subjects. As the average percent of detected daydreaming signals is between 10 to 20 percent for all five tasks (Think [T], Count [C], Recall [R], Breathe [B], Draw [D]), the upper extreme value could be as high as over 50 percent. This variance shows that the individual differences in daydreaming signals are substantial. A few subjects have a high proportion

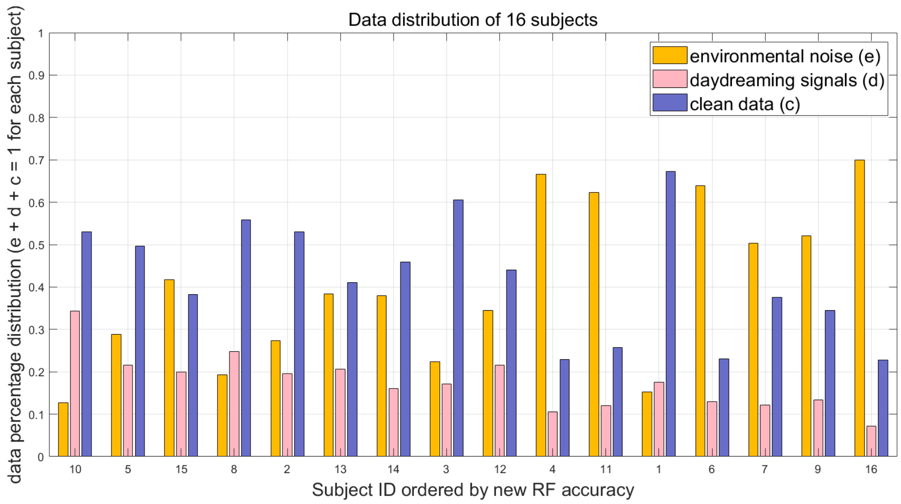

Fig. 5. The percentage of environmental noise (e), daydreaming signals (d) and clean data (c) for each subject using Random Forest as baseline classification method. Note that e + d + c = 100%.

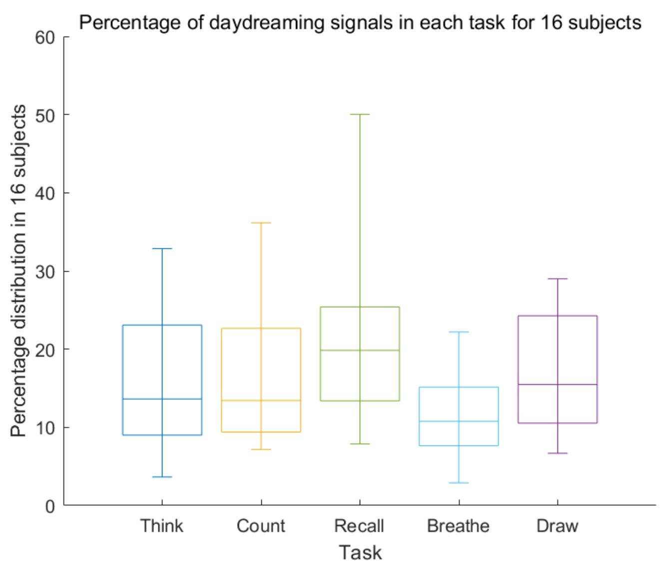

Fig. 6. Task-wise distribution of daydreaming signals in 16 subjects.

of daydreaming signals, perhaps because these subjects have difficulty concentrating during the experiment, and ii) daydreaming signals tend to have higher proportions in the task of Recall. Compared with the other four tasks, the percentage of daydreaming signals is the highest in the task of Recall in extremes values, quartile values, and the median, indicating a general trend that subjects fall into the state of "daydreaming" more easily when doing the task of Recall. We believe that this phenomenon corresponds to the design and setup of the experiments. When performing the Recall task, the subjects were asked to recall the last task they did. At this time, the state of EEG activity of the subject may be closer to the state of the previous task rather than the behavior of Recall itself, leading to the noise of unclear EEG patterns. However, this is only a simple assumption based on the experimental design; further analysis might be needed to make valid conclusions.

4 Discussion and Future Work

Our proposed algorithm is proven effective for detecting internal noise (daydreaming signals) and increasing prediction accuracy in cognitive task classification. Since subjects may get distracted unconsciously when performing the designated tasks, our algorithm can help to locate the specific time when the subjects are not focused on assigned tasks, either consciously or unconsciously; this algorithm can accordingly help remove noise and increase prediction accuracy. This approach to detecting "daydreaming signals" can also be helpful in adjusting experiment setups.

This paper tested our algorithm with three baseline algorithms for accuracy: the Random Forest classifier, the Support Vector Machines classifier, and the Long Short-Term Memory classifier. We proved that accuracy is increased by removing daydreaming signals compared with all three baseline methods. However, the majority of our daydreaming signal distribution analysis is conducted with Random Forest as the baseline classifier. Random Forest has proven to be a valuable tool for small training data sets and is less affected by the curse-of-dimensionality [16], it is the ideal baseline algorithm to use for Thinking1 BCI experiments dataset and personalized uses. Considering Random Forest is not 100-percent accurate and daydreaming signal is a newly identified class of signals, further analysis on the distribution pattern of daydreaming signals in the case of different baseline methods will be helpful to improve the analysis on daydreaming signals.

Sliding windows is another critical approach used in this paper. Although sliding windows have become a popular approach of processing EEG signals, windows size and overlap choice vary among researchers, and the consequence of using different window sizes and overlaps are still not clear [23, 18]. In this paper, we chose a window size of 10s with 99 percent overlap based on the result of the plateau threshold. For future work, we plan to test our algorithm with various sizes of sliding windows and overlaps based on the responsive time window of neurons and to further explore the distribution pattern of daydreaming signals.

For dataset selection, we chose the Thinking1 BCI experiments dataset. Although small EEG datasets can be useful for personalizing analysis [18, 21], the small size of the dataset leads to hindrance in training and more specific examination on variables. Further investigation on the performance of our algorithm on larger datasets is warranted.

5 Conclusion

The accuracy of EEG signal classification on cognitive tasks can be affected by multiple noise sources. This paper focused on "daydreaming signals," the noise caused by mental distractions. We designed a new algorithm to detect and remove these "daydreaming signals" based on sliding windows. We compared the prediction accuracy between the traditional algorithms and our improved algorithm. As a result, the average prediction accuracy is increased for all three baseline algorithms: 55.0 percent to 66.1 percent for Random Forest classifier, 41.2 percent to 65.5 percent for Support Vector Machines classifier, and 46.0 percent to 56.3 percent for Long Short-Term Memory classifier. Our approach can be helpful for increasing prediction accuracy, for designing and adjusting experimental setups, and for personalizing uses.

References

1. Bashivan, P., Rish, I., Heisig, S.: Mental state recognition via wearable eeg. arXiv preprint arXiv:1602.00985 (2016)
2. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from eeg with deep recurrent-convolutional neural networks (2016)
3. Biau, G., Scornet, E.: A random forest guided tour. Test **25**(2), 197–227 (2016)
4. Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.M., Robbins, K.A.: The prep pipeline: standardized preprocessing for large-scale eeg analysis. Frontiers in neuroinformatics **9**, 16 (2015)
5. Bird, J.J., Manso, L.J., Ribeiro, E.P., Ekárt, A., Faria, D.R.: A study on mental state classification using eeg-based brain-machine interface. In: 2018 international conference on intelligent systems (IS). pp. 795–800. IEEE (2018)
6. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
7. Chen, Y., Zhang, D.: Well log generation via ensemble long short-term memory (enlstm) network. Geophysical Research Letters **47**(23), e2020GL087685 (2020)
8. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297 (1995)
9. 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)
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
11. Ienca, M., Haselager, P., Emanuel, E.J.: Brain leaks and consumer neurotechnology. Nature biotechnology **36**(9), 805–810 (2018)
12. Jas, M., Engemann, D.A., Bekhti, Y., Raimondo, F., Gramfort, A.: Autoreject: Automated artifact rejection for meg and eeg data. NeuroImage **159**, 417–429 (2017)
13. Lei, J., Liu, C., Jiang, D.: Fault diagnosis of wind turbine based on long short-term memory networks. Renewable energy **133**, 422–432 (2019)

14. Leske, S., Dalal, S.S.: Reducing power line noise in eeg and meg data via spectrum interpolation. Neuroimage **189**, 763–776 (2019)
15. 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)
16. 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)
17. 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)
18. 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)
19. 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)
20. 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)
21. 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)
22. Quitadamo, L., Cavrini, F., Sbernini, L., Riillo, F., Bianchi, L., Seri, S., Saggio, G.: Support vector machines to detect physiological patterns for eeg and emg-based human–computer interaction: a review. Journal of neural engineering **14**(1), 011001 (2017)
23. 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)
24. Wang, Y.B., You, Z.H., Yang, S., Yi, H.C., Chen, Z.H., Zheng, K.: A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. BMC medical informatics and decision making **20**(2), 1–9 (2020)
25. Zhang, X., Yao, L., Wang, X., Monaghan, J., Mcalpine, D., Zhang, Y.: A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. Journal of Neural Engineering **18**(3), 031002 (2021)