

EEG4Students: An Experimental Design for EEG Data Collection and Machine Learning Analysis

Zheng Zhou*
zhengzhou@brandeis.edu
Brandeis University
Waltham, Massachusetts, USA

Guangyao Dou*
guangyaodou@brandeis.edu
Brandeis University
Waltham, Massachusetts, USA

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). However, during the COVID-19 pandemic, data collection and analysis could be more challenging. The remote experiment during the pandemic yields several challenges, and we discuss the possible solutions. This paper explores machine learning algorithms that can run efficiently on personal computers for BCI classification tasks. The results show that Random Forest and RBF SVM perform well for EEG classification tasks. Furthermore, we investigate how to conduct such BCI experiments using affordable consumer-grade devices to collect EEG-based BCI data. In addition, we have developed the data collection protocol, EEG4Students, that grants non-experts who are interested in a guideline for such data collection. Our code and data can be found at <https://github.com/GuangyaoDou/EEG4Students>.

CCS CONCEPTS

• **Human-centered computing** → **User centered design**.

KEYWORDS

Brain-Machine Interfaces, Data Collection, Data Analysis, EEG, Machine Learning, Interpretable AI

ACM Reference Format:

Zheng Zhou and Guangyao Dou. 2022. EEG4Students: An Experimental Design for EEG Data Collection and Machine Learning Analysis. In *Proceedings of KDD Undergraduate Consortium (KDD-UC'22)*. ACM, New York, NY, USA, 7 pages.

1 INTRODUCTION

Researchers from Computer Science, Neuroscience, and Medical fields have applied EEG-based Brain-Computer Interaction (BCI) techniques in many different ways [2, 13, 17, 19, 21, 31], such as diagnosis of abnormal states, evaluating the effect of the treatments, seizure detection, motor imagery tasks [3–5, 14, 18, 22], and developing BCI-based games [12]. Previous studies have demonstrated the great potential of machine learning, deep learning, and transfer

learning algorithms [1, 6, 16, 20, 23, 24, 37, 41–44] in clinical data analysis.

However, EEG datasets' size is still relatively small compared with peers in Computer Vision and Natural Language Processing [13, 19, 33]. The COVID-19 pandemic makes it harder to conduct BCI experiments in person at the research labs. Also, EEG signals have noise issues, mainly because of the contact of sensors and skin for several current non-invasive consumer-grade devices. The outlier issue is also a concern for the EEG data because of the difficulties subjects have in concentrating on the experimental tasks during the entire session. Although the current debate on whether the clinical or consumer-grade device is more appropriate for research remains, there is no doubt that more affordable products such as the euroSky, Muse, and OpenBCI are showing promising results with validation and repeatability[39]. Therefore, we see this opportunity to utilize the more accessible consumer-grade devices to enlarge the current dataset. Given the simplicity to use, we can also improve the data collection efficiency.

On the other hand, current algorithms and frameworks are more for in-person clinical experiments and less for the possible virtual experiment during the pandemic or with non-expert subjects. Given the need for a larger dataset in the field, our **research questions** are: 1. Can we contribute to the current BCI dataset with data from a large non-expert subject population? 2. Can we design easy-to-use BCI experiments to collect EEG data with consumer-grade devices for students? 3. Can we develop a step-by-step guide for such a data collection and machine learning analysis process so that we can still collect EEG data during the pandemic?

This paper is organized as follows: part two elaborates on the basic experimental design of our research; part three demonstrates a detailed guideline for non-expert subjects' data collection and possible training; part four provides our data analysis workflow; part five presents the experiment result that reflects the experiment success and answers the research questions one and two; parts six and seven are the discussion of limitations along with possible future works and the conclusion, which address the research question three.

2 EXPERIMENT DESIGN

Several EEG experiments focus on college students' high-level cognitive tasks frequently conducted, as mentioned in Table 1.

In this experiment, we are trying to replicate the experiment from the Think-Count-Recall (TCR) paper [30] in both in-person and remote ways via Zoom, which is the classification of five cognitive tasks: Think, Count, Recall, Breathe, and Draw. Scalp-EEG signals are recorded from nineteen subjects. Each subject participates in six sessions, and each session is five minutes long. For each

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD-UC'22, Aug 14–18, 2022, Washington DC, USA

© 2022 Association for Computing Machinery.

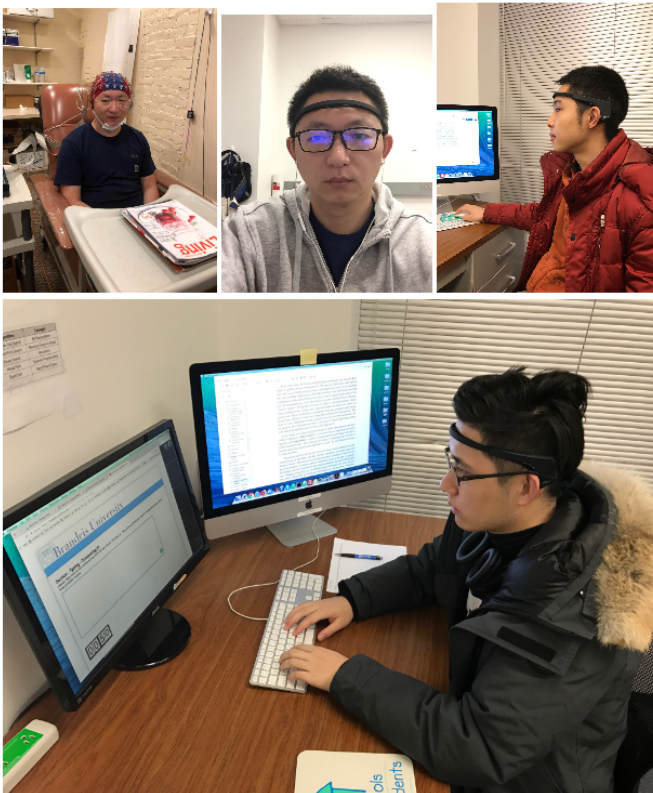


Figure 1: College Students using EEG Devices

session, there are five tasks, and each task is one minute. Tasks are selected by the subjects and researchers based on frequent tasks in study environments for students in their everyday life. Each subject completes six sessions over several weeks. The programs are run on a 2018 Macbook Pro with a 2.2GHz 6-core Intel Core i7 processor and 16 GB of memory. The Python version is 3.8. The scikit-learn version is 0.24.1.

3 DATA COLLECTION

As [10, 15, 27, 28] mentioned, several affordable (less than three hundred dollars) non-invasive consumer-grade EEG headsets are commercially available. As shown in Figure 1, we have pilot-tested several clinical and non-clinical EEG devices with college students. The Top left cell in Figure 1 shows an example of the wearing of a clinical device while others demonstrate consumer-grade devices. Muse Headsets are used as an example for demonstration

3.1 Acquiring Muse Headset Collecting Skills

The data collecting skills are delivered and acquired entirely through in-person settings before the outbreak of COVID-19 and virtual

(E) T	1	2	3	4	5
[28]	Math	Close-eye Relax	Read	Open-eye Relax	None
[33]	Python Passive	Math Passive	Python Active	Math Active	None
[32]	Read	Write Copy	Write Answer	Type Copy	Type Answer
[30]	Think	Count	Recall	Breathe	Draw

Table 1: Tasks (T) in Experiments (E)

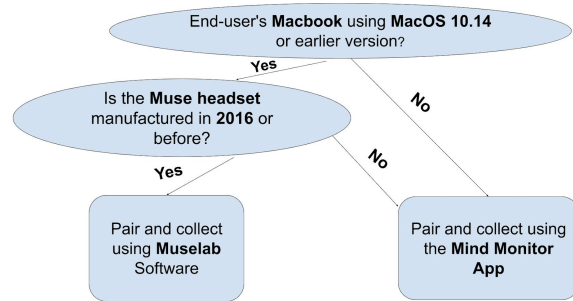


Figure 2: Data Collection App Selection

zoom meetings for virtual participants during the pandemic. The EEG coach demonstrates the headset wearing and posture position on live video to multiple future EEG coaches simultaneously, simulating a virtual classroom setting. As shown in Figure 2, for end-users who have a muse headset manufactured in 2016 or before and have a computer with macOS 10.14 or former version. The MuseLab software is introduced and installed during the virtual session. Via screen-sharing, the EEG coach demonstrates and guides the trainee in step-by-step data acquisition instruction. All applications are installed from Muse Developer Resource as a Muse IO folder, including muse-io, muse-player, Muselab application, and licenses. For Muse headsets manufactured after 2016 or computer systems other than macOS 10.14 and the former version, a different application, "Mind Monitor," is required to collect EEG signals since the latest version headset and the Muselab software are not compatible. Mind Monitor can be directly installed from the ios APP store and paired with the muse headset following the instruction provided by the application.

The Muse headset has four receptive fields, two near the subject's anterior side and two on the lateral sides. The headset wearing is similar to glasses-wearing, but since the headsets are not customized fitting, some participants have trouble recording consistently accurate EEG signals on the lateral receptive fields. In order to resolve the issue, rubber bands are introduced and tightened on the posterior side of the subject's head, securing the headset in place. In turn, the signal consistency is significantly improved.

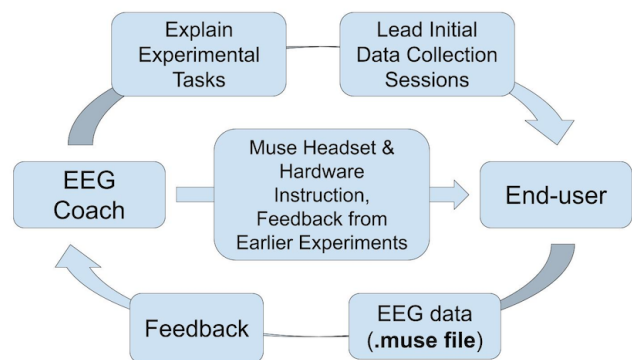


Figure 3: Data Collection Flow Chart, User and Coach

3.2 Data Collection Protocol

Two individuals, the EEG coach and the end-user are usually required to complete the data acquisition of such non-invasive EEG signals via Muse headset and Muse Recording software, As shown in Figure 3. During the pandemic, data collection takes place virtually, meaning the EEG coaches have to communicate with the end-user and provide instruction through a screen-sharing virtual environment throughout the data acquisition.

3.2.1 In-Person Data Collection Sessions. During in-person data collection sessions, an EEG coach and a test subject are located in the same room with two monitor screens. One screen monitors the headset connectivity and signals consistency, and the other displays the Qualtrics protocol to the subject. After the EEG coach successfully establishes the Muse headset connection, the participants put it on and proceed as directed by the EEG coach and the Qualtrics protocol.

3.2.2 Virtual Data Collection Sessions. During the pandemic, all our experiments are forced to move online. Therefore, Muse headsets are mailed to participants or end-users along with instructions to establish a Bluetooth connection. When collecting data, the EEG coach would join the end-user in a zoom meeting and instruct the end-user to connect the headset and walk through the Qualtrics protocol. Zoom meeting allows the end-users to share their screen to help the EEG-coaches to monitor the data collection process and provide a remote control for necessary assistance.

3.2.3 Muse Headset Usage Via MuseLab and Mind Monitor. Muse headset pairing can be initiated by holding the start icon on the headset for five seconds until the light indicator on the headset starts flashing, indicating ready for pairing. The EEG coach instructs the end-user to open the muse.io file on the end user's computer and connect the headset via Bluetooth. Muse.io file will automatically connect to the headset if it is ready for pairing and located in range. When the light indicator stops flashing and remains turned on, and the terminal message on the muse.io file indicates wording such as "connection successful" and displays the headset's battery life, the devices are paired. After the connection is established, the end-user is instructed to open the Muselab application and adjust the OSC port 5000 to receive the EEG signal from the headset. Before putting on the headset, the last step is to instruct the end-user to add a new visualization of EEG signals as a scrolling line graph under the visualizers' drop-down menu.

For end-users who use the MindMonitor app, the data collection protocol is similar. The end-users open the app on their compatible tablets or smartphones and a computer to follow the instructions from the EEG coach and the Qualtrics.

After connecting the Muse headset via Bluetooth, the recording format is changed in the settings to "Muse," and the recording interval is set as "constant." The end-users are then prompted to put on the headset, and the scrolling line graphs from four Muse receptive fields will be visible on the MindMonitor app. The recording process is identical to Muslab, except that the recorded files are saved by default in plain text CSV format and are automatically uploaded to Dropbox.

If the end-users are solely one-time participants, a screen remote control can be done to ensure efficiency. If the end-users ought to

be trained as repeated users or EEG coaches, they will follow the instruction provided by the EEG coach and acquire the experiment protocol during a screen share. As soon as the EEG visualization appears on the application window, the end-user is instructed to put on the headset and open the experiment file on Qualtrics, an online survey system we use to give step-by-step instructions. The end-user is asked to keep the Muselab application, and the Qualtrics file visible on the screen share to ensure the EEG coach can monitor the EEG signal receptive status by the headset while proceeding with the experiment. The end-user will follow the combined instruction by the EEG coach and the Qualtrics file to experiment. Once data collection is finished, the end-user will upload the .muse file onto the shared Google Drive for further data analysis.

4 DATA ANALYSIS

Most EEG recording applications have a toolset to convert the recording files to TXT or CSV files. Afterward, we can pick the subset of data we plan to use for further analysis; we recommend starting with the absolute value of the EEG signals.

As shown in Figure 4, once we have six TXT files for all six sessions, we first execute a Matlab program - preprocess.m - to identify the noises and turn all the TXT files into separate CSV files. Then, we run the cleaning.ipynb to exclude these data in the Pandas Dataframe for further analysis. Next, we execute the TCR.ipynb to train and cross-validate existing machine learning algorithms such as the Random Forest, SVM, and KNN. Lastly, the TCR.ipynb will generate a visualization of task predictions for end-users.

4.1 Experiments

This section describes the subjects, data preprocessing, and validation techniques we use for experiments.

Subjects: Before the pandemic, sixteen healthy subjects participated in the experiment, numbered 1 to 16. Compared to during the pandemic, we successfully collected data from only three participants remotely, and only one subject's data was adapted after denoising, numbered 17 in our experiment. We used the second trial of subject 17. Subjects who lost more than 65% of the total data due to the data cleaning actions would be removed from the

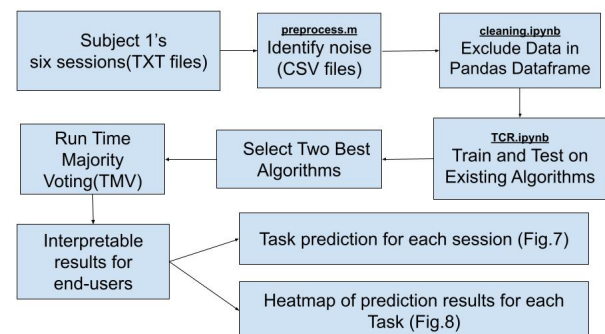


Figure 4: Data Analysis Flow Chart

subsequent analysis. In the end, twelve subjects, including subject 17, were left to continue the analysis. Moreover, for the total of six sessions of each subject, if a session lost more than 65% of the data during the data cleaning phase, then that session would also be excluded for further analysis.

Preprocessing: As mentioned in [30, 32], for each task during each session, the first 30% of the data, which is the first 18 seconds of each task and during the transition phase, will be removed. This has been proved reasonable during the data cleaning phase. During the EEG recording, some electrodes may have temporarily lost contact with the subjects' scalp. The result is that multiple sequential spectral snapshots from one or more electrodes have the same value. In this paper, we decide to remove such anomaly when detected for a consecutive 1.4 seconds. Such a cleaning action causes a different level of data loss for each subject.

Time-wise Cross-Validation: We adopted time-wise cross validation[33, 38]. For each five-minute session, there were five tasks. We divided each task into seven parts, meaning each had six seconds, evenly and continuously. Then, we did seven-fold cross-validation. In each fold, we checked if each fold lost more than 65% of the original data, then that fold was discarded for future analysis. We tested one subset in each fold and trained on the left-out subset.

4.2 Algorithms

We implement several machine learning algorithms commonly used in the field [7–9, 11] from the scikit-learn [26]. For example, Linear Classifiers, Nearest Neighbors, Decision Trees, and Ensemble Methods.

Linear Classifiers: The Linear Discriminant Analysis(LDA), especially Shrinkage Linear Discriminant Analysis (Shrinkage LDA), and Support Vector Machine(SVM), which works well on high dimensional spaces, perform adequately on EEG datasets based on previous research [5, 18]. These algorithms are simple and computationally efficient because they classify data based on linear functions. In this paper, we use 'LinearDiscriminantAnalysis' and set the Shrinkage to 'auto' for Shrinkage LDA. We used SVC with RBF and linear kernel functions from the scikit-learn for SVM.

Nearest Neighbor: Such a classifier implements the K-Nearest Neighbor(KNN). The KNN performs pretty well on EEG datasets [19], compared to most of the other classifiers. We use 'KNeighborsClassifier,' and choose k to be five (as there are five tasks).

Decision Tree: The Decision Trees classifier is easy to understand, implement, and interpret. The computational cost of using the tree is low and can handle both numerical and categorical data. However, it might overfit as trees are too complex. In this paper, we use 'DecisionTreeClassifier.'

Ensemble Methods: We implement Random Forest, Adaboost, and Gradient boosting. Ensemble Machine Learning algorithms combine several weaker learners' predictions and form more robust and accurate predictions. These have been widely used in EEG-based experiments and research [19, 30, 32]. In this paper, we use

Covid/experiment	Scheduled	Female/Male	Actual	After Denoise
Before	16	8/8	16	11
During	6	3/3	3	1

Table 2: Comparison of pre/during pandemic data collection

'RandomForestClassifier', 'AdaBoostClassifier', and 'GradientBoostingClassifier' from the scikit-learn [26].

5 RESULTS

5.1 Remote Data Collection

We scheduled a remote data collection meeting with six end-users during the pandemic, as shown in Table 2. During the zoom meeting, only three participants showed up. Two of them had noise issues. They could not simultaneously get all four electrodes on their skin to generate consistent, stable signals. Only one end-user successfully finished the experiment. We labeled the data from this end-user as data of subject seventeen.

5.2 Classification Results

We report the average accuracy for all subjects and the runtime of each classifier for each participant in table 3. As we can see, Random Forest has the highest accuracy in our experiment, and the SVM with RBF kernel is the second best performing classifier on the dataset. Though Nearest Neighbors does not perform as well as the Random Forest and the RBF SVM, it is one of the fastest algorithms on personal computers. In contrast, although the average accuracy of GradientBoost is 3% higher than that of the KNN, it is the most time-consuming algorithms we experimented with. The Adaboost performs the worst compared to other algorithms. We present the comparison between the average accuracy of EEG4Students and the TCR paper in Table 4.

The individual difference may impact the accuracy of each subject. But we can still recognize a general pattern from Fig 5. We order all twelve subjects by prediction accuracy using Random Forest. Most of the algorithms demonstrate consistent patterns for the different algorithms. Random Forest and SVM with RBF kernel are above most other algorithms. As shown in Fig 6, noise is also a

Algorithm	Average Accuracy	Average run-time (s)
Random Forest	0.63	37.1
RBF SVM	0.61	15.2
GradientBoost	0.58	119.0
Nearest Neighbors	0.55	2.3
Decision Tree	0.5	2.7
Linear SVM	0.48	11.3
Shrinkage LDA	0.48	1.9
LDA	0.48	1.9
Adaboost	0.44	7.1

Table 3: Algorithms with Accuracy and Run-time

Algorithm	Accuracy(EEG4Students)	Accuracy([30])
Random Forest	0.63	0.64
RBF SVM	0.61	0.55
Nearest Neighbors	0.55	0.58
Shrinkage LDA	0.48	0.54
LDA	0.48	0.54
Adaboost	0.44	0.57

Table 4: Accuracy Comparison between EEG4Students and the TCR[30] paper

concern for our data collection and analysis. As we can see, six subjects had more than 50% noise. We keep the threshold of including the subject to 35% of the remaining data, as we believe that when a participant has little data left, the high accuracy contributes little to our research.

Figure 7 and Figure 8 are heatmaps for subject 9 and subject 17 for all five tasks, respectively. Subject 9 yields the highest Random Forest accuracy among in-person participants. Subject 17 is the only participant we experimented with among virtual participants.

As shown in Figure 7, the diagonal is the tasks that have been correctly predicted. The other cells give the team a sense of how one task may be misclassified as other tasks. For subject 9, the Random Forest can predict the tasks Count, Breathe, and Draw with high

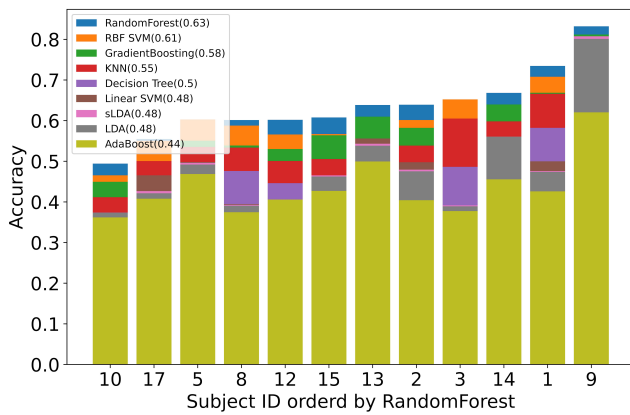


Figure 5: Algorithm performance among different subjects. The X axis is the subject id ordered by the accuracy of the Random Forest classifier, and the y axis is the accuracy. The legend shows the classifiers’ names and average accuracy among all participants.

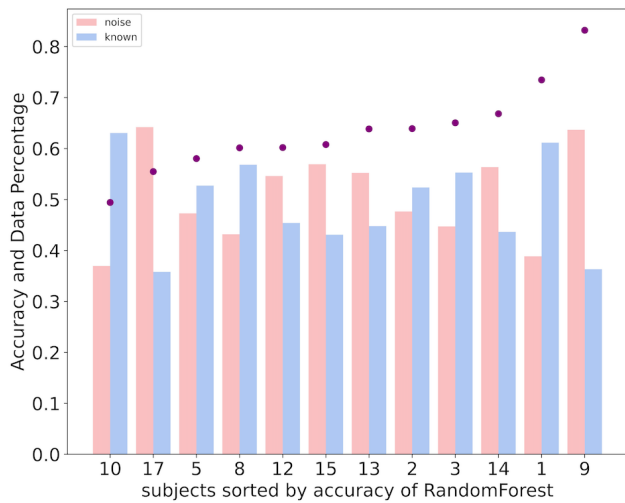


Figure 6: Visualization of data noise and data left to cross-validate for each subject. The X axis is the subject id ordered by the accuracy of the best performing classifier, Random Forest, and the y axis is the percentage. The purple dot represents the corresponding accuracy of the Random Forest.

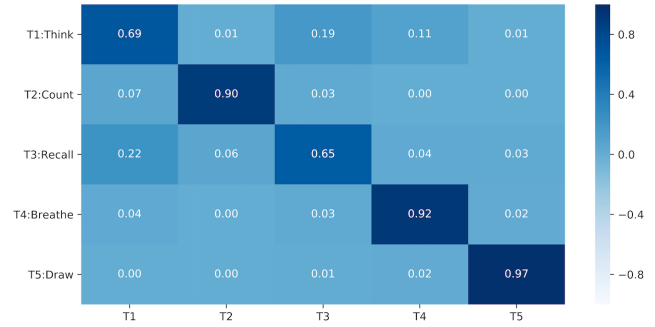


Figure 7: Heatmap for all five tasks of subject 9. The X-Axis is the five designed tasks (Think, Count, Recall, Breathe, and Draw), and the Y-Axis is the predicted tasks.

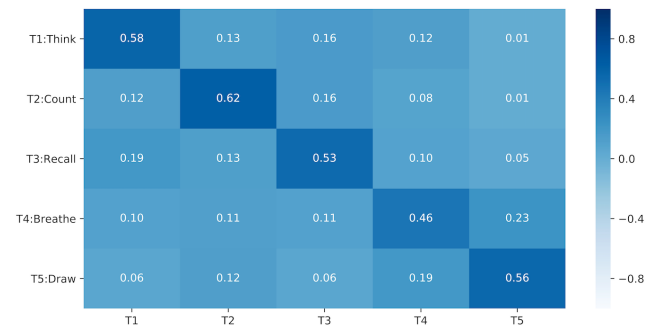


Figure 8: Heatmap for all five tasks of subject 17, the remote participant of our data collection experiment. The X-Axis is the five designed tasks (Think, Count, Recall, Breathe, and Draw), and the Y-Axis is the predicted tasks.

accuracy, while not so good on Think and Recall. However, as we can see from Figure 6, subject 9 has a lot of noise compared to other subjects.

Figure 8 is the classification results for subject 17, the subject we collected completely in virtual settings. The Random Forest performs well, but not so well compared to subject 9. We also note from Figure 6 that subject 17 has much noise that we exclude for training.

Such visual feedback also helped the research team and the end-users better understand what task pairs were easy to be confused with each other.

6 DISCUSSION

6.1 Virtual Coaching Challenges

This paper proposes an approach for non-expert independent researchers to collect EEG-based BCI data with affordable non-clinical devices. When performing test trials with Muse headsets, we provide a general guideline, as shown in figure 2 and 3, which shows promising results in many EEG data collections performed by naive EEG end-users. This general guideline demonstrates a decision tree for non-expert researchers to acquire data collection hardware and software. We also present our data collection flow, which forms a closed loop between the researchers and the experimental subjects. In addition, we elaborate on our data-cleaning analysis procedures.

Significant progress has been made in user training for EEG-based BCI studies, while the EEG4Students serves as a stepping stone for further improved training programs in future research.

However, some limitations are identified along the course of our project. Our project spans from pre-pandemic to post-pandemic time. We find that in-person data collection trials are significantly more efficient than trials that take place virtually during the global pandemic. If an end-user runs into a distinct situation during Equipment connection or data collection due to personal device usage, the problem can hardly be resolved. For instance, about half of the EEG coaches are issued a newer version of the 2018 Muse headset compared to the original ones from 2016; the updated model is so distinct that it is incompatible with the intended Muselab application. In order to resolve this incompatibility, a new EEG recording application, Mind Monitor, is found, and another protocol is formulated specifically for the 2018 Muse headset. Therefore, a different experiment protocol and data collection have to be developed and delivered to the end-users virtually.

During the pandemic, six subjects (three female and three male participants) were recruited, and each subject was issued a Muse headset and instructions for data collection. The entire coaching and data acquisition process was converted into a virtual established environment. Given the distinct individual problematic data collecting process and the communication barrier set up by the pandemic, multiple participants failed to submit EEG data. As shown in Table 2, in the pre-pandemic BCI experiment that took place in person, all sixteen subjects delivered completed data, and eleven datasets were considered meaningful after denoised. By the end of the study, during the pandemic, three out of six participants participated in the experiment. Only one out of the six participants delivered a completed dataset. More BCI user training, as [36] suggested, could help solve such issues. Therefore, We ought to explore more strategies and updated methods that could grant us efficiency when data collection has to be completed in a virtual environment.

Even though as much detail and trial and error experience we managed to include in our guideline, there are chances that individual cases develop distinct issues. Our priority is to return to an in-person setting when allowed for data collection efficiency. As we perfect our data collection protocol, we will recruit more subjects and potential EEG coaches to enlarge the current dataset exponentially. Our future work includes developing a study/work community for college students as end-users and EEG coaches. In this interactive community, they can discuss their experience with non-clinical devices' use of EEG-based data collection and possibly establish solutions to various issues after the encounter.

6.2 Machine Learning Algorithms

Comparatively low accuracies by Machine Learning and Deep Learning algorithms in medical tasks are common [25, 28, 32, 34, 35, 40]. Especially for EEG classification tasks, Machine Learning techniques have struggled to achieve high accuracies due to low signal-to-noise ratio, the outlier issues, and the "mind wandering" states during the experiments [19, 20, 29]. We should note that the results of the EEG4Students are similar to the paper we are trying to replicate, based on the table 4. Due to the small datasets, we did not implement any deep learning models that are popular in the

field of EEG, such as CNN, RNN, and LSTM [13, 37]. We will try to implement these deep learning algorithms as we collect more data with more subjects. In this section, we still provide the guidelines for choosing the Machine Learning classifiers that achieve relatively high accuracies.

Random Forest: Random Forest yields the highest accuracy in predicting the tasks, based on Figure 5. Also, the Random Forest is easy to implement on personal computers from the scikit-learn.

SVM: Support Vector Machine with RBF kernel function should be used instead of the SVM with a linear kernel function, as it performs better for all tested subjects. The average accuracy is close to the Random Forest, and its average accuracy is 13% higher than the linear SVM.

KNN: The K Nearest Neighbors performs adequately. The classification is computed by the majority vote of the nearest neighbors of each data point. The KNN is efficient, as the average runtime of the KNN was 2.3 seconds to perform the timewise cross-validation.

Boosting: Boosting classifiers in the paper provide various levels of prediction accuracy in our experiment. The accuracy for Adaboost is ranked the lowest, but the GradientBoost is ranked the third among classifiers with significant runtime.

We recommend using Random Forest and RBF SVM on personal computers. Also, KNN generates a comparable baseline result with the fastest runtime.

7 CONCLUSION

This paper investigates the data collection for EEG-based BCI to develop larger datasets. We explore the possibility of collecting EEG data from college students with affordable devices. The results demonstrate that the proposed framework could simplify the process and contribute to developing a larger EEG dataset. Here, we present the EEG4Students model. The results show a reasonable prediction accuracy of Random Forest and RBF SVM and interpretable task types for most end-users. In turn could be a building block toward the future of everyone using non-invasive, wireless, and affordable EEG-based BCI systems, whether for college students or other non-experts, to collect BCI data similar to current smartphone usage.

REFERENCES

- [1] Sizhe An, Ganapati Bhat, Suat Gumussoy, and Umit Ogras. 2020. Transfer Learning for Human Activity Recognition using Representational Analysis of Neural Networks. *arXiv preprint arXiv:2012.04479* (2020).
- [2] Aurélien Appriou, Andrzej Cichocki, and Fabien Lotte. 2020. Modern machine-learning algorithms: for classifying cognitive and affective states from electroencephalography signals. *IEEE Systems, Man, and Cybernetics Magazine* 6, 3 (2020), 29–38.
- [3] Pouya Bashivan, Gavin M Bidelman, and Mohammed Yeasin. 2014. Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity. *European Journal of Neuroscience* 40, 12 (2014), 3774–3784.
- [4] Pouya Bashivan, Irina Rish, and Steve Heisig. 2016. Mental state recognition via wearable EEG. *arXiv preprint arXiv:1602.00985* (2016).
- [5] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. 2015. Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448* (2015).
- [6] Jordan J Bird, Luis J Manso, Eduardo P Ribeiro, Aniko Ekart, and Diego R Faria. 2018. A study on mental state classification using EEG-based brain-machine interface. In *2018 International Conference on Intelligent Systems (IS)*. IEEE, 795–800.
- [7] Leo Breiman. 1996. Bagging predictors. *Machine learning* 24, 2 (1996), 123–140.
- [8] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

- [9] Leo Breiman. 2017. *Classification and regression trees*. Routledge.
- [10] Cédric Cannard, Helané Wahbeh, and Arnaud Delorme. 2021. Validating the wearable MUSE headset for EEG spectral analysis and Frontal Alpha Asymmetry. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 3603–3610.
- [11] Jérôme-Alexis Chevalier, Alexandre Gramfort, Joseph Salmon, and Bertrand Thirion. 2020. Statistical control for spatio-temporal MEG/EEG source imaging with desparsified multi-task Lasso. *arXiv preprint arXiv:2009.14310* (2020).
- [12] Damien Coyle, Jose Principe, Fabien Lotte, and Anton Nijholt. 2013. Guest Editorial: Brain/neuronal-Computer game interfaces and interaction. *IEEE Transactions on Computational Intelligence and AI in games* 5, 2 (2013), 77–81.
- [13] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering* 16, 3 (2019), 031001.
- [14] Dieter Devlaminck, Willem Waegeman, Bruno Bauwens, Bart Wyns, Patrick Santens, and Georges Otte. 2010. From circular ordinal regression to multilabel classification. In *Proceedings of the 2010 Workshop on Preference Learning (European Conference on Machine Learning, ECML)*. 15.
- [15] Marcello Ienca, Pim Haselager, and Ezekiel J Emanuel. 2018. Brain leaks and consumer neurotechnology. *Nature biotechnology* 36, 9 (2018), 805–810.
- [16] Murat Kaya, Mustafa Kemal Binli, Erkan Ozbay, Hilmi Yanar, and Yuriy Mishchenko. 2018. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Scientific data* 5, 1 (2018), 1–16.
- [17] Fabien Lotte. 2014. A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces. In *Guide to Brain-Computer Music Interfacing*. Springer, 133–161.
- [18] Fabien Lotte. 2015. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proc. IEEE* 103, 6 (2015), 871–890.
- [19] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. 2018. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering* 15, 3 (2018), 031005.
- [20] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. 2007. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering* 4, 2 (2007), R1.
- [21] Fabien Lotte and Cuntai Guan. 2010. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Transactions on biomedical Engineering* 58, 2 (2010), 355–362.
- [22] Fabien Lotte and Camille Jeunet. 2015. Towards improved BCI based on human learning principles. In *The 3rd International Winter Conference on Brain-Computer Interface*. IEEE, 1–4.
- [23] Fabien Lotte, Camille Jeunet, Jelena Mladenović, Bernard N’Kaoua, and Léa Pillette. 2018. A BCI challenge for the signal processing community: considering the user in the loop.
- [24] Kai J Miller. 2019. A library of human electrocorticographic data and analyses. *Nature human behaviour* 3, 11 (2019), 1225–1235.
- [25] Catherine Ordun, Alexandra N Cha, Edward Raff, Byron Gaskin, Alex Hanson, Mason Rule, Sanjay Purushotham, and James L Gully. 2022. Intelligent Sight and Sound: A Chronic Cancer Pain Dataset. *arXiv preprint arXiv:2204.04214* (2022).
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [27] Roberto Portillo-Lara, Bogachan Tahirbegi, Christopher AR Chapman, Josef A Goding, and Rylie A Green. 2021. Mind the gap: State-of-the-art technologies and applications for EEG-based brain–computer interfaces. *APL bioengineering* 5, 3 (2021), 031507.
- [28] Xiaodong Qu, Mercedes Hall, Yile Sun, Robert Sekuler, and Timothy J Hickey. 2018. A Personalized Reading Coach using Wearable EEG Sensors-A Pilot Study of Brainwave Learning Analytics.. In *CSEdu* (2). 501–507.
- [29] Xiaodong Qu and Timothy J Hickey. 2022. EEG4Home: A Human-In-The-Loop Machine Learning Model for EEG-Based BCI. In *International Conference on Human-Computer Interaction*. Springer, 162–172.
- [30] Xiaodong Qu, Peiyan Liu, Zhaonan Li, and Timothy Hickey. 2020. Multi-class Time Continuity Voting for EEG Classification. In *International Conference on Brain Function Assessment in Learning*. Springer, 24–33.
- [31] Xiaodong Qu, Saran Liukasemsarn, Jingxuan Tu, Amy Higgins, Timothy J Hickey, and Mei-Hua Hall. 2020. Identifying clinically and functionally distinct groups among healthy controls and first episode psychosis patients by clustering on EEG patterns. *Frontiers in psychiatry* (2020), 938.
- [32] Xiaodong Qu, Qingtian Mei, Peiyan Liu, and Timothy Hickey. 2020. Using EEG to Distinguish Between Writing and Typing for the Same Cognitive Task. In *International Conference on Brain Function Assessment in Learning*. Springer, 66–74.
- [33] Xiaodong Qu, Yixin Sun, Robert Sekuler, and Timothy Hickey. 2018. EEG markers of STEM learning. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.
- [34] Salman Razzaki, Adam Baker, Yura Perov, Katherine Middleton, Janie Baxter, Daniel Mullarkey, Davinder Sangar, Michael Taliercio, Mobasher Butt, Azeem Majeed, et al. 2018. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. *arXiv preprint arXiv:1806.10698* (2018).
- [35] Jonathan G Richens, Ciarán M Lee, and Saurabh Johri. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nature communications* 11, 1 (2020), 1–9.
- [36] Aline Roc, Léa Pillette, Jelena Mladenovic, Camille Benaroch, Bernard N’Kaoua, Camille Jeunet, and Fabien Lotte. 2020. A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering* (2020).
- [37] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. 2019. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering* 16, 5 (2019), 051001.
- [38] Sohrab Saeb, Luca Lonini, Arun Jayaraman, David C Mohr, and Konrad P Kording. 2016. Voodoo machine learning for clinical predictions. *Biorxiv* (2016), 059774.
- [39] Phattarapong Sawangjai, Supanida Hompoonsup, Pitshaporn Leelaarporn, Supavit Kongwudhikunakorn, and Theerawat Wilaiprasitporn. 2019. Consumer grade EEG measuring sensors as research tools: A review. *IEEE Sensors Journal* 20, 8 (2019), 3996–4024.
- [40] Hannah L Semigran, David M Levine, Shantanu Nundy, and Ateev Mehrotra. 2016. Comparison of physician and computer diagnostic accuracy. *JAMA internal medicine* 176, 12 (2016), 1860–1861.
- [41] Xiang Zhang, Lina Yao, Xianzhi Wang, Jessica JM Monaghan, David Mcalpine, and Yu Zhang. 2020. A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. *Journal of Neural Engineering* (2020).
- [42] Ziyuan Zhao, Kartik Chopra, Zeng Zeng, and Xiaoli Li. 2020. Sea-Net: Squeeze-And-Excitation Attention Net For Diabetic Retinopathy Grading. In *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2496–2500.
- [43] Ziyuan Zhao, Kaixin Xu, Shumeng Li, Zeng Zeng, and Cuntai Guan. 2021. MT-UDA: Towards Unsupervised Cross-modality Medical Image Segmentation with Limited Source Labels. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 293–303.
- [44] Ziyuan Zhao, Kerui Zhang, Xuejie Hao, Jing Tian, Matthew Chin Heng Chua, Li Chen, and Xin Xu. 2019. Bira-net: Bilinear attention net for diabetic retinopathy grading. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1385–1389.