Multi-Class Time Continuity Voting for EEG Classification

Xiaodong Qu[1][0000−0001−7610−6475], Peiyan Liu[1], Zhaonan Li[1], and Timothy Hickey[1]

Brandeis University, Waltham, MA 02453, USA
{xiqu,peiyanliu,zli,tjhickey}@brandeis.edu

Abstract. In this study we propose a new machine learning classification method to distinguish brain activity patterns for healthy subjects. We used ElectroEncephaloGraphic (EEG) data associated with five user-defined mental tasks. We collected a data set using the Muse headband with four EEG electrodes (TP9, AF7, AF8, and TP10). Sixteen healthy subjects participated in this six-session experiment. In each session, we instructed them to conduct five different one-minute tasks, we abbreviated the tasks as think, count, recall, breathe and draw. After dealing with noise and outliers, we first fairly compared the performance of existing classifiers, including linear classifiers, non-linear Bayesian classifiers, nearest-neighbor classifiers, ensemble methods, and deep learning, with the same settings. Among these, Random Forest (RF), and Long Short-Term Memory (LSTM) outperform others. We then introduced a new ensemble classifier called Time Continuity Voting (TCV), combining these top two. The timewise cross-validation results showed that TCV could correctly classify the five tasks (20% by chance) with an accuracy of 70% which is at least 6% higher than the top individual classifiers.

Keywords: Time Continuity Voting · Machine Learning · Deep Learning · Classification · ElectroEncephaloGraphy · EEG

1 Introduction

Brain Computer Interfaces (BCI), have been widely used for clinical and non clinical applications, such as diagnosis of abnormal states, evaluation the effect of the treatments, helping patients with motor disabilities to move a mouse or to control a motorized wheelchair, mental workload, seizure detection, motor imagery tasks (left hand, right hand, foot and tongue) [9], BCI based games [6] and passive BCI [26]. Most data are collected in a clinical or research setting[7, 16]. For other domains, such as ImageNet for image classification, or MNIST for handwritten digit recognition, more data of the same kind can be generated directly from the general end users, and more general patterns could be recognized based on such large scale data. To narrow this gap, we designed a pilot study towards building such a large scale EEG data set, for multi-class classification of user-centered tasks, generated by general end users. This is a simple experiment design with our new algorithm, Time Continuity Voting (TCV), to collect and analyze more data.

Think (T) For this task, subjects were asked to think of a certain number of random objects, then in following steps they were asked to recall and draw such objects. We asked the participants to think of different objects each time, and different numbers for each session, in session one and two, six objects; three and four, seven objects; five and six, eight objects. Within one minute, they typed these objects down in a text entry box on the screen.

Count (C) Subjects counted numbers aloud, from 200 towards 0, each time subtracting by seven, e.g. 200, 193, 186, 179 ..., the subjects were asked to count as slowly and clearly as possible, with eyes open, and minimize the movements of the body, especially the eyes and jaws.

Recall (R) Each subject recalled the objects they had just typed in the Think (T) task, in the correct order, if possible. The results were again entered in a text entry box through the keyboard.

Breathe (B) Subjects were instructed to breathe deeply with eyes open. They were asked NOT to think about the objects they just thought or recalled in earlier task(s), ideally, they just focused on the breathing task itself.

Draw (D) Each subject was asked to draw the objects they thought about in the earlier task Think (T), with a pen, on a blank A4 paper. The objects were listed on the screen in front of them so they did not need to recall, just focus on drawing.

Fig. 1. Tasks in this experiment

Previous studies[4, 7, 16, 20] demonstrated that EEG signals could successfully be used to distinguish several kinds of cognitive tasks. In this study, we designed these five user-centered tasks, abbreviated them as Think(T), count(C), recall(R), breathe(B) and draw(D). The order of the five tasks was randomly shuffled for each of the six sessions, as shown in Figure 2.

Our result is that the Time Continuity Voting (TCV) algorithm can recognize the different tasks at a 70% level, compared with 20% by chance.

Beyond the clinical devices with 128 or 64 electrodes, commercial wearable devices for monitoring brain activity (EEG) are now widely available on the consumer market at an affordable price [11, 18], including the Neurosky Mindwave, the Emotiv Epoc, the Open BCI Mark IV headset, and the Muse Headband etc. In this experiment, EEG data were collected using Figure 3)Muse headsets developed by Interaxon.

Such EEG data collection devices could be used by the general end users in order to support better understanding of cognitive tasks, emotion states, health monitoring, and diagnosis of abnormal states. Also it can be used in the classroom to facilitate teaching and learning [8, 19, 25].

Our long term goal is to 1) develop a personalized EEG-based biofeedback system to help more users better understand their unique individual brain signal patterns; 2) make it easier for general end users to collect, analysis and share

S/T	1	2	3	4	5
1	T	C	R	B	D
2	B	T	C	R	D
3	T	B	R	D	C
4	T	C	R	B	D
5	C	T	R	D	B
6	T	C	B	R	D

Fig. 2. Session (S) with Task (T) order shuffled

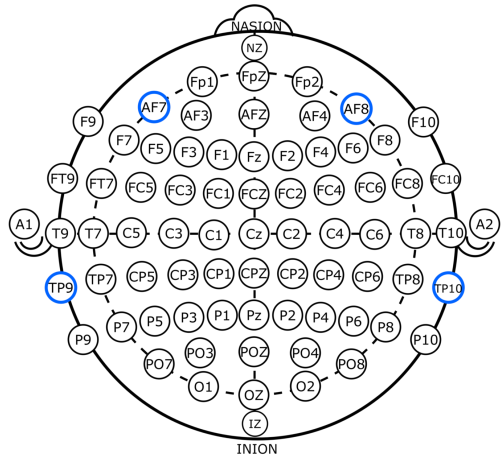

Fig. 3. 10-20 System, four electrodes used on Muse Headset were highlighted

more EEG data in daily life, and eventually discover the general patterns for most people based on large scale data.

2 Algorithms

We reviewed and implemented several classifiers commonly adopted in the field, as summarized in [16], [17] and [7], such as Linear classifiers, ensemble methods, and deep learning. The code discussed in this paper is available online (the Github link is hidden for the double blind review).

2.1 Existing Algorithms

Linear classifiers: Both Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) are discriminative classifiers. Because they use linear functions to classify, the computational requirement is usually low, which make them good for personal computers and smartphones. Previous research has shown the shrinkage LDA (sLDA) and SVM perform from adequately to reasonably well with little data in EEG research. [3, 15].

Nearest Neighbour: Such classifiers look for the nearest neighbours; and we used the k-Nearest Neighbour(kNN) classifier. It is very sensitive to the curse-of-dimensionality, especially with a typical 64 or 128 electrodes EEG headset; however, it works well with low-dimennsional feature vectors [16, 17]. We suspected that since our headset only has 4 electrodes, it would perform adequately.

Ensemble Classifiers: Boosting [10], bagging and Random Forest [5] are types of Ensemble Machine Learning Algorithms which use a group of weaker learners (e.g. random decision trees) to make a stronger classification. These have been the gold standard for several EEG-based classification experiments, here we implemented Random Forest (RF), Adaboost and RusBoost.

Deep Learning: Deep learning, especially Convolutional Neural Network(CNN) performed well in several previous EEG band power (feature) based research [7, 13]. We implemented it with two layers with maximum pooling layers. We also adopted Long Short Term Memory (LSTM) based on our task type [2, 3, 7, 24]. As motioned in previous research, we expect it to perform well with two recurrent layers followed by two fully-connected layers.

2.2 Our Algorithm

In this paper, we propose a new voting approach based on temporally nearest neighbors. Voting is a popular way of classifier combination, [12]. In EEG-based research, the following approaches have performed well: majority voting, weighted majority voting, and voting based on classification entropy [1, 28].

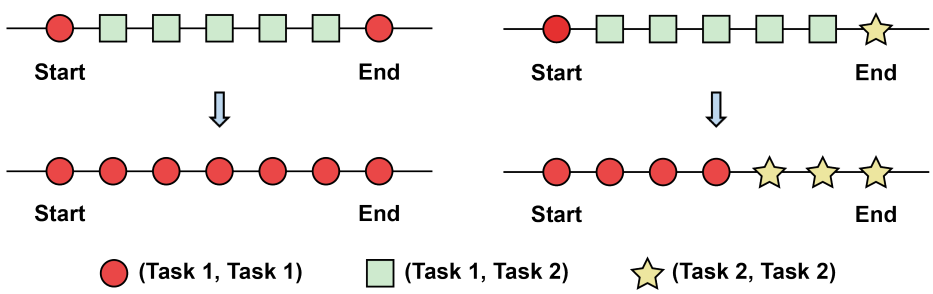

Fig. 4. Time Continuity Voting (TCV)

Time Continuity Voting As shown in Figure 4, for each data point, if classifier A and classifier B have the same prediction on a task, e.g. a data point is predicted as task 1, then the prediction result is 1, and it is marked as 'start'. If the two predictions are different. e.g. classifier A predicts this data point as task 1, and classifier B predicts the same data point as task 2, the algorithm then move on to the next data point until find the nearest neighbor, where the two predictions from the two classifiers are the same, and mark it as the 'end', within the task time frame. Because of the time continuity effect [21, 22] of EEG signals, we assume the task and its nearest neighbor have a high probability to have the same label. Thus the algorithm is trying to find the nearest neighbor where both of the classifiers have the same prediction. If there are two such nearest neighbors but they have different classifications at these two different data points, then the algorithm divides the data points in between evenly, first half as the 'start' data point, and second half as the 'end' data point. If the number of data points between 'start' and 'end' is odd, then the middle data point is randomly assigned the same label as either 'start' or 'end' data point.

3 Experiment

In this experiment, scalp-EEG signals were recorded from sixteen subjects. Each one was tested in six sessions, each session is five minutes long, with five tasks, each task is one minute. Tasks were selected by the subjects together with the researchers, based on frequent tasks in study environments for students in their everyday life. Each subject completed six sessions over several weeks.

Each subject first signed the informed consent form. Then, they put on the Muse headbands and test the recording. Subjects then completed an entrance survey. After these preliminaries, Official EEG recording began. Subjects were directed by an online data collection system, which kept track of time and alerted the subjects to change their tasks after every 60 seconds. After subjects completed all the five tasks, the EEG recording stopped, subjects then completed a short exit survey.

Data cleaning: When collecting EEG data, one or more electrodes may have momentarily lost contact with the subjects' scalp. The result was that multiple sequential spectral snapshots from one or more electrodes had exactly the same 32 bit value. When we detected this anomaly, we set that entire spectral snapshot of 20 values to 0, while keeping the time-stamped value, even if the anomaly was only detected on one electrode. Such cleaning action resulted in a loss of 43% of the entire data. This result echoed with other researches facing the same challenge of low signal-to-noise ratio.

Subjects: Sixteen healthy subjects finished the experiment. Data from four subjects have less than 35 percent data points left after removing noises. Thus these four subjects were excluded from subsequent analysis.

Six males and six females are included in the final data set. Ten of the twelve retained subjects were undergraduate students, the other two were graduate students. Seven subjects were computer science majors; the remaining five were math, biology or psychology majors, or had not yet decided on a field of concentration. The average age of the subjects was 20.2.

All twelve subjects completed the six sessions, producing a data set comprising 360 minutes of EEG recordings (12 subjects x 6 sessions per subject x 5 minutes per session).

Feature Extraction and Feature Selection: We used the Band Powers (BP) features, the absolute band power for a given frequency range (for instance, alpha, 9-13 Hz) is the logarithm of the power spectral density of EEG signals summed over that frequency range. [16]. The Muse headsets, as shown in Figure 3, are equipped with seven dry electrodes that make contact with the subjects' scalp, three of them are reference, the other four are input. The four input electrode locations corresponded to sites TP9, AF7, AF8, and T10 [23]. The Muse EEG recording application automatically filtered out muscle artifacts, such as eye blinking and jaw movements. The EEG system down-sampled sensor signals from 12k Hz to 220 Hz, with 2uV (RMS) noise. Spectral analysis was performed on-board the Muse device and then transmitted wirelessly at 10 Hz to the researcher's workstation. Each of these spectral snapshots consists of 20 numeric values – five spectral values for each of the four electrodes. This procedure generated a total of 3,000 spectral snapshots per subject per session (10 snapshots/second * 300 seconds).

Timewise cross validation Samples, if randomly selected, could be near to each other chronologically in both the training set and the testing set. This may cause over-fitting because EEG signals change slowly. To lessen the possible over-fitting effect, we adopted the timewise cross validation from ([21, 22]). For each five minute session there were five tasks, we divided each task into 10 parts, evenly and contiguously, each part had 10% of the task based on time.

Then we did a 10 fold cross validation first, realized the first 20% and last 10% data are demonstrating low accuracy due to the task transition effect. We then cut off these transition time and only use the rest 70% of the data. In each fold, We trained on the six subsets and tested on the left-out subset. Figure 5 showed a graphical representation of this seven-fold timewise cross validation approach. We also did a session-wise cross validation, but the accuracy is average 18% to 25% lower than this within-session cross validation, such results were expected, probably because of the relatively small number of sessions.

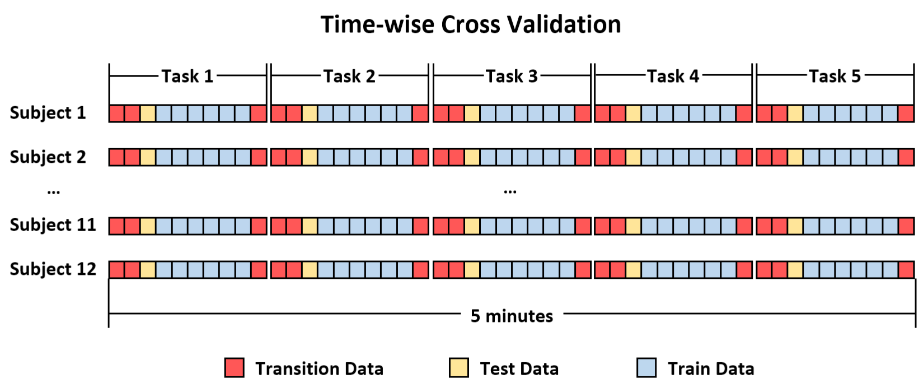

Fig. 5. Timewise Cross Validation

4 Results

Two main results were observed from the EEG classification. First, the machine learning/ deep learning algorithms can distinguish these five tasks fairly well, especially our Time Continuity Voting (TCV). (See Fig. 6). Second, although there are individual differences, (see Fig. 7), the user defined tasks demonstrated general patterns (See Fig. 8) with just six-minute of data from twelve subjects, TCV 0.7, RF 0.64, LSTM 0.64 showed similar trends (See Fig. 6), here we just include the RF figures, the rest figures are available on Github).

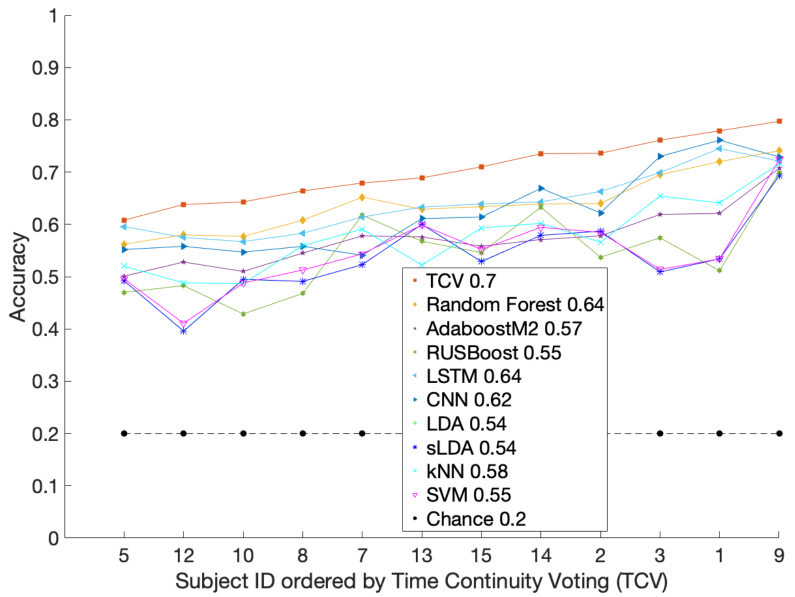

Fig. 6. Compare Algorithms

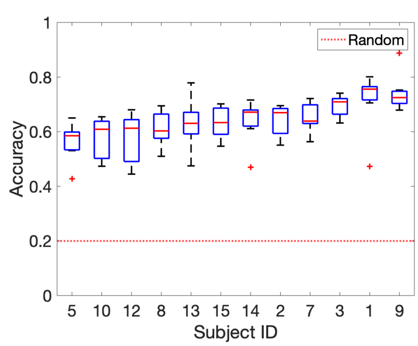

Fig. 7. Subject Difference

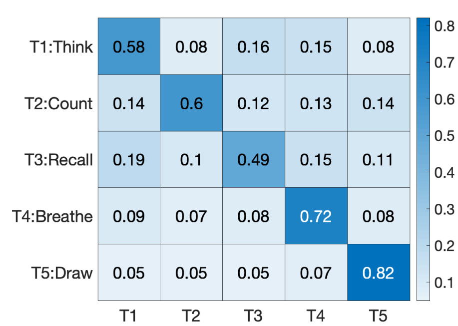

Fig. 8. Task Prediction Accuracy, average of all twelve subjects.

5 Conclusion and Discussion

The main contributions of this paper are the demonstration that Time Continuity Voting (TCV) algorithm can improve the accuracy of EEG classification of such user-centered tasks. TCV outperformed all other widely adopted classifying approaches in the offline classification task(See Fig. 6). It is voting based on the output of the other individual learners, this version is using the top two: RF and LSTM, it captured the different patterns recognized by these different individual machine learning and deep learning learners, and achieved a six percent increase of classification accuracy.

Fig. 6 showed that Linear methods like LDA and SVM still performed adequately. Among all the ensemble methods, Random Forest (Bag) still the best. The two deep learning options, CNN and LSTM, are still very strong predictors even with such a relatively small size of data, but the training time jumped from minutes to hours compared with the traditional machine learning techniques.

The separability of these five different tasks (See Fig. 8, echoed with previous research with clear and distinct EEG patterns from five or more different classes [14, 27], implied that the multi-class task detection could be classified from adequately to good. Although there are still individual difference (see Fig. 7), the similar trends can be observed with these twelve subjects.

Interesting general patterns could be observed from these results(See Fig. 8), such as task Draw and task Breathe are more accurately predicted, possible reasons including muscle related sub-tasks such as inhale, exhale or draw, Also the different objects in each session actually showing the framework is sensitive to the different content subjects are thinking, the low accuracy of task Think and task Recall probably comes from each session the subjects were asked to come up with different objects. The individual difference (see Fig. 7) in task Count is a little surprise to the researchers, we expected it to be as clearly classified as task Breathe, but it was not, probably because the difference of calculating such numbers in mind, between subjects and within subject between sessions, some subjects in some session count it so fast, reaching zero while there were still several second left, while others may just stopped at around a hundred when the time went up, also some subjects count it wrong during the process. These all could lead to the individual difference reflected here.

Such experiment and analysis framework is also sensitive enough to catch the task transition, and early task finishing, as shown in Figure 5, some subjects finish the task Think, task Recall, task Draw earlier than the designed one minute time, thus the last 10% of time is hard to predict, as subjects could be daydreaming or thinking about something else. Also the transition effect is pretty clear, although we observe the subject switching to the next task pretty fast physically between tasks, it is still an obvious general pattern that the EEG signals changed slowly, it took the first 20% of time, which is twelve second to actually enter the stable states of the next task.

Overall, our Time Continuity Voting (TCV) framework outperformed benchmark individual classifiers in this multi-class task classification experiment. Thus TCV could be a useful option for future research of such EEG classification.

Bibliography

1. Ahangi, A., Karamnejad, M., Mohammadi, N., Ebrahimpour, R., Bagheri, N.: Multiple classifier system for eeg signal classification with application to brain–computer interfaces. Neural Computing and Applications **23**(5), 1319–1327 (2013)
2. Arifoglu, D., Bouchachia, A.: Activity recognition and abnormal behaviour detection with recurrent neural networks. Procedia Computer Science **110**, 86–93 (2017)
3. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448 (2015)
4. Bird, J.J., Manso, L.J., Ribeiro, E.P., Ekart, 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)
5. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
6. 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)
7. 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)
8. Deeb, F.A., DiLillo, A., Hickey, T.: Using spinoza log data to enhance cs1 pedagogy. In: International Conference on Computer Supported Education. pp. 14–36. Springer (2018)
9. 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)
10. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences **55**(1), 119–139 (1997)
11. Gang, P., Hui, J., Stirenko, S., Gordienko, Y., Shemsedinov, T., Alienin, O., Kochura, Y., Gordienko, N., Rojbi, A., Benito, J.L., et al.: User-driven intelligent interface on the basis of multimodal augmented reality and brain-computer interaction for people with functional disabilities. In: Future of Information and Communication Conference. pp. 612–631. Springer (2018)
12. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: A review. IEEE Transactions on pattern analysis and machine intelligence **22**(1), 4–37 (2000)
13. Kwak, N.S., Müller, K.R., Lee, S.W.: A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. PloS one **12**(2), e0172578 (2017)

14. Lindig-León, C., Bougrain, L.: Comparison of sensorimotor rhythms in eeg signals during simple and combined motor imageries over the contra and ipsilateral hemispheres. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 3953–3956. IEEE (2015)
15. Lotte, F.: Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. Proceedings of the IEEE **103**(6), 871–890 (2015)
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. Mihajlović, V., Grundlehner, B., Vullers, R., Penders, J.: Wearable, wireless EEG solutions in daily life applications: what are we missing? IEEE journal of biomedical and health informatics **19**(1), 6–21 (2014)
19. Poulsen, A.T., Kamronn, S., Dmochowski, J., Parra, L.C., Hansen, L.K.: Eeg in the classroom: Synchronised neural recordings during video presentation. Scientific reports **7**, 43916 (2017)
20. 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)
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. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C., Kording, K.P.: Voodoo machine learning for clinical predictions. Biorxiv p. 059774 (2016)
23. Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., He, B., Beniczky, S.: The standardized EEG electrode array of the IFCN. Clinical neurophysiology **128**(10), 2070–2077 (2017)
24. Sha, L., Hong, P.: Neural knowledge tracing. In: International conference on brain function assessment in learning. pp. 108–117. Springer (2017)
25. Tarimo, W.T., Deeb, F.A., Hickey, T.J.: Early detection of at-risk students in cs1 using teachback/spinoza. Journal of Computing Sciences in Colleges **31**(6), 105–111 (2016)
26. Zander, T.O., Kothe, C.: Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. Journal of neural engineering **8**(2), 025005 (2011)
27. Zeyl, T., Yin, E., Keightley, M., Chau, T.: Partially supervised p300 speller adaptation for eventual stimulus timing optimization: target confidence is superior to error-related potential score as an uncertain label. Journal of neural engineering **13**(2), 026008 (2016)
28. Zhang, J., Wu, Y., Bai, J., Chen, F.: Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers. Transactions of the Institute of Measurement and Control **38**(4), 435–451 (2016)