

BrainActivity1: A Framework of EEG Data Collection and Machine Learning Analysis for College Students

Zheng Zhou¹[0000-0001-9313-2106], Guangyao Dou¹[0000-0001-8011-9658], and Xiaodong Qu^{1,2}[0000-0001-7610-6475]

¹ Brandeis University, Waltham MA 02453, USA
{zhengzhou, guangyaodou}@brandeis.edu

² Swarthmore College, Swarthmore PA 19081, USA
xqu1@swarthmore.edu

1 Introduction

Previous research in Computer Science, Neuroscience, and Medical fields has implemented EEG-based Brain-Computer Interfaces (BCI) in several ways, [1, 13, 14, 19, 21, 23, 25, 33, 34, 35], such as diagnosis of Alzheimer’s, emotion recognition, mental workload, motor imagery tasks [3, 4, 5, 15, 22, 26]. Machine learning, deep learning, and transfer learning algorithms have demonstrated the great potential in such biomarker data analysis [2, 6, 7, 16, 17, 20, 24, 27, 28, 31, 37, 38, 39, 40, 41].

However, EEG datasets’ size is still relatively small compared with peers in Computer Vision and Natural Language Processing [13, 23, 36].

Our **research questions** are: 1. Can we develop a larger dataset with data from a larger audience? for example, college students? 2. Can we design easy-to-use BCI experiments to collect EEG data with consumer-grade devices for college students? 3. Can we develop a step-by-step guide for such a data collection and machine learning analysis process?

2 Methods

As [11, 18, 30, 32] 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 showed an example of the wearing of a clinical device while others demonstrated the consumer-grade devices. Muse Headset were used as an example for demonstration, followed by more details in the 6-page full-length page.

As shown in Figure 2, first, we installed the data collection software or application. The following method section was formulated based on the example of using Muse Headset. We have used this device since 2016. First, the end user’s OS version must match the version of the EEG recording application. If the end-user had a newer Macbook or the latest version of Muse headset, they ought

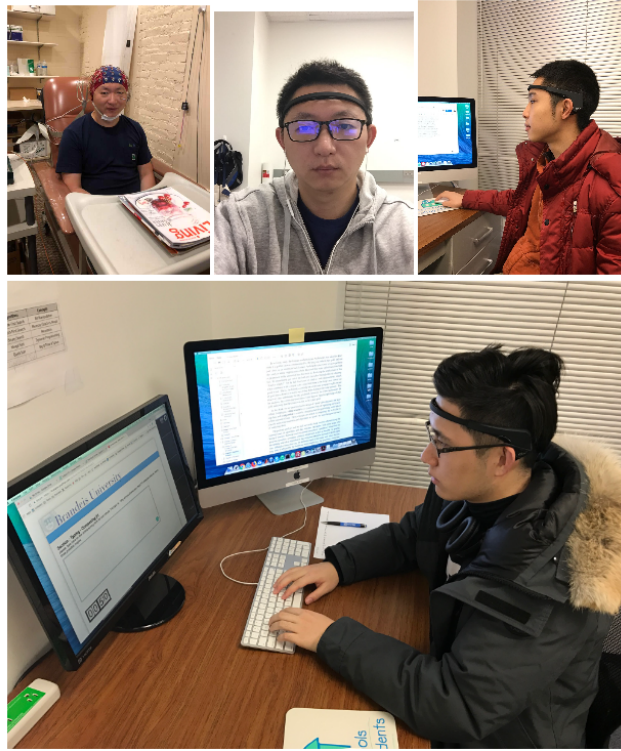


Fig. 1. College Students using EEG Devices

to use the Mind Monitor Application to record EEG signals. Otherwise, the end-user may still use the Muselab application for EEG recording.

Two individuals, the EEG coach, and the end-user, were 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. First, the EEG Coach introduced the Muse Headset and Muse recording application to the end-user. Then the EEG coach explained the details of a specific experiment. Next, the EEG coach and the end-user started the experiment to collect the data. The data was saved as a MUSE file. The end-user then summarized the experimental feedback to the EEG coach. Together with the research team, the EEG coach generated visual feedback to the End-user. Such feedback may contribute to the development of future experiments.

If the end-user was interested, they could learn the data analysis themselves, which took two hours on average for students who major in computer science.

Most EEG recording applications came with a toolset to convert the recording files to TXT or CSV files. Afterward, we could pick the subset of data we planned to use for further analysis; we recommended starting with the absolute value of the EEG signals.

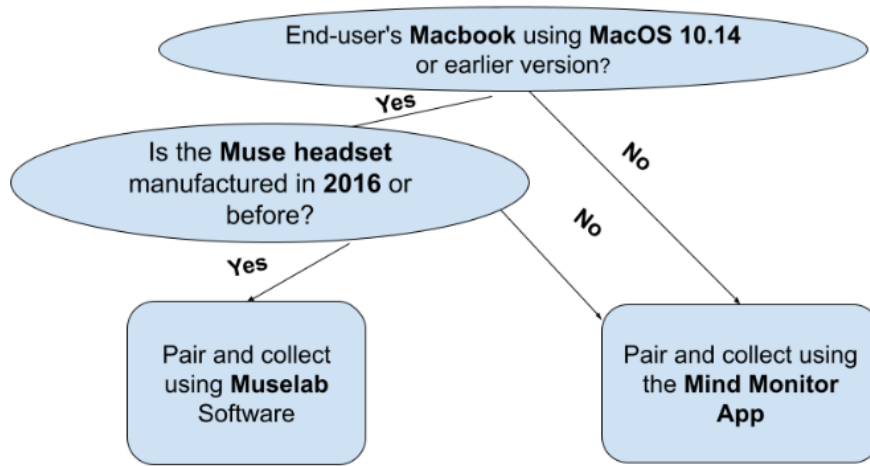


Fig. 2. Data Collection App Selection

We implemented several machine learning algorithms commonly used in the field [8, 9, 10, 12] from the scikit-learn [29]. For example, Linear Classifiers, Nearest Neighbors, Decision Tree, and Ensemble Methods.

As shown in the Figure 4, once we had six TXT files for all six sessions of data, we first executed a Matlab program - preprocess.m - to identify the noises

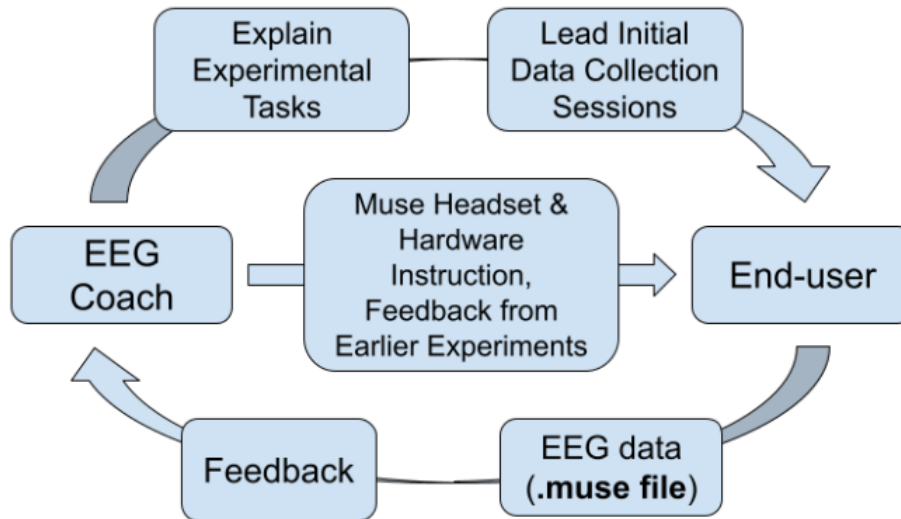


Fig. 3. Data Collection Flow Chart, User and Coach

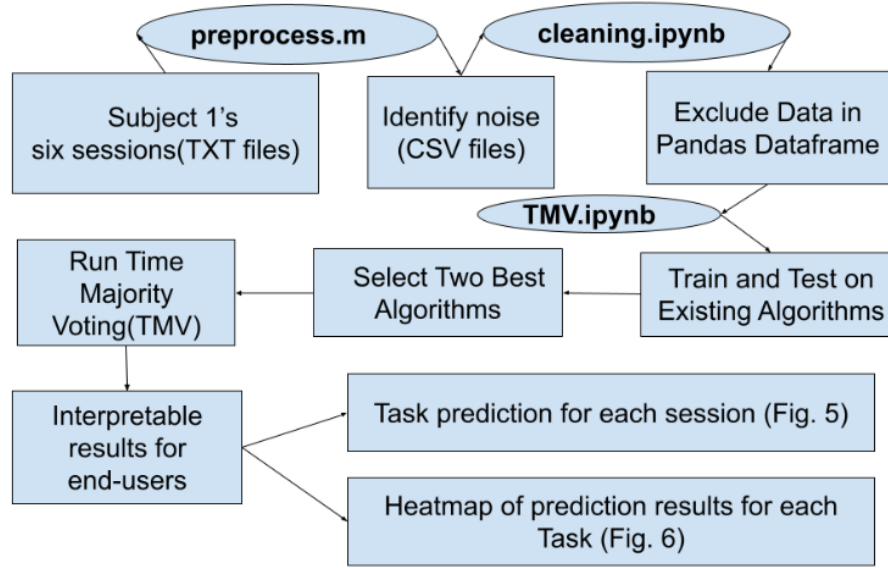


Fig. 4. Data Analysis Flow Chart

and turn all the TXT files into separate CSV files. Then, we ran the Clean.ipynb to exclude these data in the Pandas Dataframe for further analysis. Next, we executed the TMV.ipynb to train and cross-validate existing machine learning algorithms such as the Random Forest, SVM, and KNN. Then we selected the top two best-performing algorithms and used these two algorithms to perform Time Majority Voting. In our case, these two algorithms were Random Forest and RBF SVM. Lastly, the TMV.ipynb would generate a visualization of task predictions for end-users. Figure 5 and a heatmap for all six sessions (Fig 6). These two figures are examples of task one's prediction results for all six sessions of subject 1 in the TCR Experiment.

3 Results

As shown in Figure 5, sessions designed task were clearly recognized. All six sessions of subject 1's task 2 showed consistent patterns. Such visual feedback was provided to the EEG coach and end-user who collected this set of data. From the experiment notes, we ran into a signal issue behind the right ear of the end-user throughout the session, then we recorded this case and potential solutions to improve future experiments.

As shown in Figure 6, the X-Axis is the designed tasks, the Y-Axis is the predicted tasks. The diagonal means the designed tasks matched the prediction. 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.

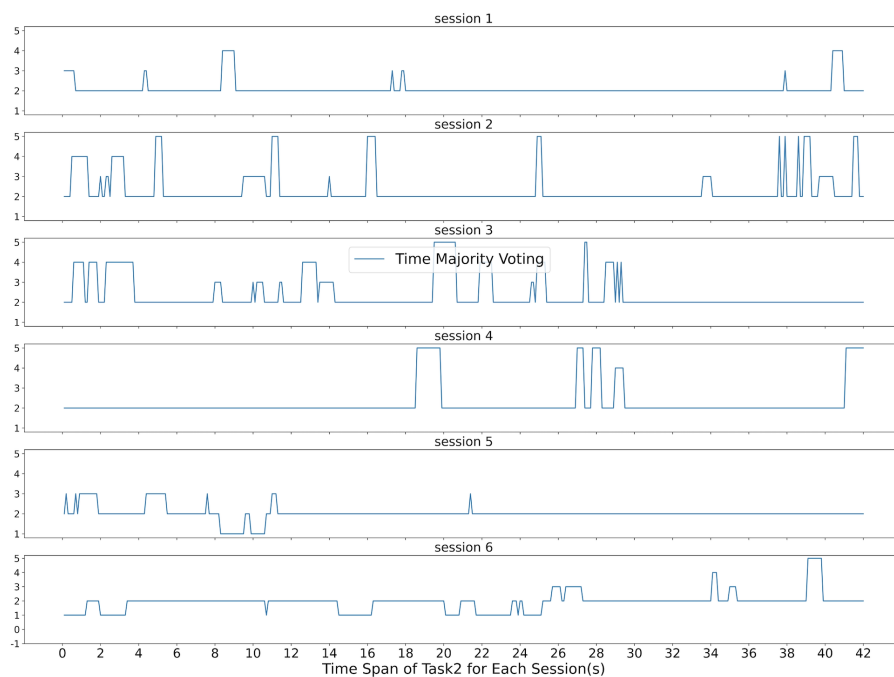


Fig. 5. Subject 1 Task 2 Prediction for all Six Sessions

4 Discussion

This paper proposed 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 provided a general guideline, as shown in figure 2 and 3, which showed promising result in many EEG data collections performed by naive EEG end-users. This general guideline demonstrated a decision tree for non-expert researchers to acquire data collection hardware and software. We also presented our data collection flow, which formed a closed loop between the researchers and the experimental subjects. In addition, we elaborated our data-cleaning analysis procedures.

Significant progress has been made in user-training for EEG-based BCI studies, while the framework proposed in this paper serves as a stepping stone for further improved training programs in future research. However, some limitations were identified along the course of our project. Our project spanned from pre-pandemic to post-pandemic time. We found that in-person data collection trials were significantly more efficient than trials that took place virtually during the global pandemic. We ought to explore more strategies and updated methods that could grant us the efficiency when data collection has to be completed in a virtual environment.

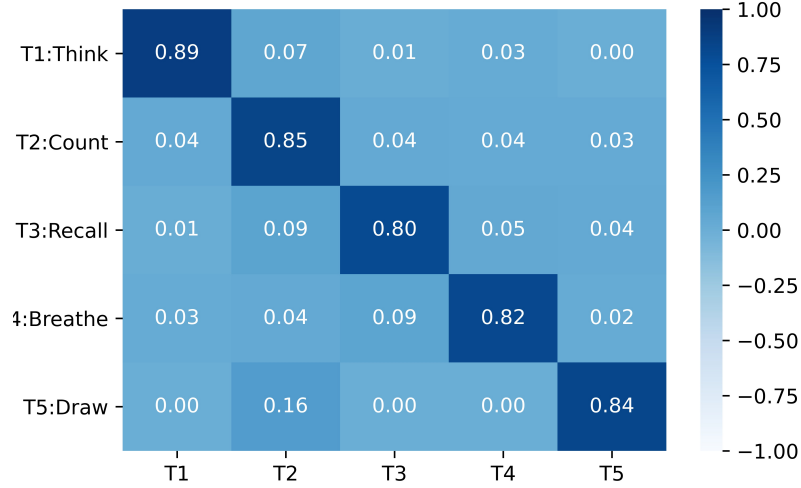


Fig. 6. Subject 1's TMV Heatmap for all Six Sessions

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 future work includes research/EEG coach-based student study/work community, in which they can learn and discuss their experience with non-clinical devices collecting EEG-based data and possibly establish solutions to various issues after the encounter.

5 Conclusion

This paper investigated the data collection for EEG-based BCI to develop larger datasets. We explored the possibility of collecting EEG data from college students with affordable devices. The results demonstrated that the proposed framework could simplify the process and contribute to developing a larger EEG dataset.

Bibliography

1. Appriou, A., Cichocki, A., Lotte, F.: Modern machine-learning algorithms: for classifying cognitive and affective states from electroencephalography signals. *IEEE Systems, Man, and Cybernetics Magazine* **6**(3), 29–38 (2020)
2. Basaklar, T., Tuncel, Y., An, S., Ogras, U.: Wearable devices and low-power design for smart health applications: challenges and opportunities. In: 2021 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED). pp. 1–1. IEEE (2021)
3. Bashivan, P., Bidelman, G.M., Yeasin, M.: Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity. *European Journal of Neuroscience* **40**(12), 3774–3784 (2014)
4. Bashivan, P., Rish, I., Heisig, S.: Mental state recognition via wearable eeg. arXiv preprint arXiv:1602.00985 (2016)
5. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448 (2015)
6. Bhat, G., Tuncel, Y., An, S., Lee, H.G., Ogras, U.Y.: An ultra-low energy human activity recognition accelerator for wearable health applications. *ACM Transactions on Embedded Computing Systems (TECS)* **18**(5s), 1–22 (2019)
7. 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)
8. Breiman, L.: Bagging predictors. *Machine learning* **24**(2), 123–140 (1996)
9. Breiman, L.: Random forests. *Machine learning* **45**(1), 5–32 (2001)
10. Breiman, L.: Classification and regression trees. Routledge (2017)
11. Cannard, C., Wahbeh, H., Delorme, A.: Validating the wearable muse headset for eeg spectral analysis and frontal alpha asymmetry. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 3603–3610. IEEE (2021)
12. 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)
13. 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)
14. Darvishi, A., Khosravi, H., Sadiq, S., Weber, B.: Neurophysiological measurements in higher education: A systematic literature review. *International Journal of Artificial Intelligence in Education* pp. 1–41 (2021)
15. 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)
16. Dongare, S., Padole, D.: Categorization of eeg using hybrid features and voting classifier for motor imagination. In: 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). pp. 217–220. IEEE (2021)
 17. Gu, J., Zhao, Z., Zeng, Z., Wang, Y., Qiu, Z., Veeravalli, B., Goh, B.K.P., Bonney, G.K., Madhavan, K., Ying, C.W., et al.: Multi-phase cross-modal learning for noninvasive gene mutation prediction in hepatocellular carcinoma. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 5814–5817. IEEE (2020)
 18. Ienca, M., Haselager, P., Emanuel, E.J.: Brain leaks and consumer neurotechnology. *Nature biotechnology* **36**(9), 805–810 (2018)
 19. Jamil, N., Belkacem, A.N., Ouhbi, S., Guger, C.: Cognitive and affective brain–computer interfaces for improving learning strategies and enhancing student capabilities: A systematic literature review. *IEEE Access* (2021)
 20. 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)
 21. Lotte, F.: A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces. In: *Guide to Brain-Computer Music Interfacing*, pp. 133–161. Springer (2014)
 22. 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)
 23. 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)
 24. 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)
 25. Lotte, F., Guan, C.: Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering* **58**(2), 355–362 (2010)
 26. Lotte, F., Jeunet, C.: Towards improved bci based on human learning principles. In: *The 3rd International Winter Conference on Brain-Computer Interface*. pp. 1–4. IEEE (2015)
 27. Lotte, F., Jeunet, C., Mladenović, J., N’Kaoua, B., Pilette, L.: A bci challenge for the signal processing community: considering the user in the loop (2018)
 28. Miller, K.J.: A library of human electrocorticographic data and analyses. *Nature human behaviour* **3**(11), 1225–1235 (2019)
 29. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)

30. Portillo-Lara, R., Tahirbegi, B., Chapman, C.A., Goding, J.A., Green, R.A.: Mind the gap: State-of-the-art technologies and applications for eeg-based brain-computer interfaces. *APL bioengineering* **5**(3), 031507 (2021)
31. Qian, P., Zhao, Z., Chen, C., Zeng, Z., Li, X.: Two eyes are better than one: Exploiting binocular correlation for diabetic retinopathy severity grading. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 2115–2118. IEEE (2021)
32. 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)
33. 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)
34. Qu, X., Liukasemsarn, S., Tu, J., Higgins, A., Hickey, T.J., Hall, M.H.: Identifying clinically and functionally distinct groups among healthy controls and first episode psychosis patients by clustering on eeg patterns. *Frontiers in psychiatry* p. 938 (2020)
35. 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)
36. 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)
37. 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)
38. Xu, K., Zhao, Z., Gu, J., Zeng, Z., Ying, C.W., Choon, L.K., Hua, T.C., Chow, P.K.: Multi-instance multi-label learning for gene mutation prediction in hepatocellular carcinoma. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 6095–6098. IEEE (2020)
39. 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)
40. Zhao, Z., Chopra, K., Zeng, Z., Li, X.: Sea-net: Squeeze-and-excitation attention net for diabetic retinopathy grading. In: 2020 IEEE International Conference on Image Processing (ICIP). pp. 2496–2500. IEEE (2020)
41. Zhao, Z., Xu, K., Li, S., Zeng, Z., Guan, C.: Mt-uda: Towards unsupervised cross-modality medical image segmentation with limited source labels. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 293–303. Springer (2021)