EasyChair Preprint

№ 6150

ML-Based Classification of Eye Movement Patterns During Reading Using Eye Tracking Data from an Apple iPad Device

Victor Anisimov, Konstantin Ð¡hernozatonsky, Andrey Pikunov, Shedenko Ksenia, Daria Zhigulskaya and Arsen Revazov

July 23, 2021

ML-based classification of eye movement patterns during reading using eye tracking data from an Apple iPad device

Perspective machine learning algorithm needed for reading quality analytics app on an iPad with built-in eye tracking

Anisimov V., Chernozatonsky K., Pikunov A., Shedenko K., Zhigulskaya D., & Revazov Arsen

Oken Technologies Inc.

Russian Federation, Moscow

victor@oken.tech ✉, kc@oken.tech ✉, andrey@oken.tech, ksenia@oken.tech, daria@oken.tech, arsen@oken.tech

***Abstract*— Eye movements in reading can provide important information about readers' perception of texts and, with appropriate algorithms, about the level of their understanding of what is written. Both digital learning and professional training require processing of large volumes of information, mostly in the text format. Management and control of students' or trainees' levels of perception in reading can be accomplished with the use of technology, assessing attention, engagement, understanding, cognitive load and tiredness of readers. These metrics can potentially be inferred from eye tracking data during reading a text. For this task it is important to utilize features of mass-market consumer devices. For example, the latest version of Ipad Pro with a standard iOS operating system has an embedded eye tracker, and thus it provides opportunities for mass adoption in the educational and training settings. Authors of this work built a stable algorithm for detecting saccades and fixations in noisy eye tracking data recorded by iPad Pro 11" and achieved certain progress in applying a machine-learning algorithm for classifying eye movement patterns in reading. The results could be used for creating an interactive reader's assistant in the format of an iOS application.**

Keywords- Eye tracking data, reading, machine learning, text perception.

I. Introduction

Eye tracking data in reading is one of the most informative physiological data channels. While reading, people move their eyes to change the gaze position from word to word and from line to line. Generally the movement pattern consists of two types: rapid eye movements (called saccades) and points of gaze movement pauses (called fixations). It was shown that during the latter, the human brain perceives text meaning [1]. Research has shown that recordings of eye movements and other individual physiological data of students can be used to determine their cognitive state, tiredness [2] and emotions [3] etc. Algorithms on the basis of machine learning (ML) models can analyze complex physiological inputs data and reveal hidden correlations between students' cognitive states and their understanding of the content.

This ambitious task would require a methodology for recognizing patterns in individuals' eye movements. There are two types of approaches that can potentially do this. The first type utilizes direct measurements of certain parameters of eye movements, such as gaze movement speed, dispersion of gaze coordinates, etc. The second type is based on machine learning algorithms. The first type is quite suitable for laboratory environments, where all measurements are under control of the research team, the equipment includes high-frequency professional eye-trackers, and research subjects' heads are placed on rests securing a fixed position. In educational or training environments, however, the quality of data from the frontal camera will always be of inferior quality compared to that in the physiology lab. Therefore for a mass product that we are developing, it makes sense to use the ML algorithm, which is more adaptive and automated.

In the past decade the whole education and professional training industry made a bold transition to online and digital mode. The pandemics of the last year accelerated the transition, making digital the new normal and pushing the development of both the technologies and strategies for distance learning. One of the common trends in education now is *e-learning*. Many articles and the internet resources relate to the opportunities and difficulties of this area. The advantages of e-learning are obvious: 1) courses in e-learning are not so strictly limited, as classical programs in a university are very scrupulously framed and the structure of courses is defined for a long period of time, 2) comfort is a significant advantage, cause many people feel stress among in crowded classrooms, 3) e-learning looks like the initiative position of individuum and potentially is attractive for his employers, 4) in choosing among digital courses it is possible to reveal the most relevant sphere and skills for the person, 5) one of the most important advantages is the decreasing price of education related to optimization of many offline routines [4,5,6]. At the same time new problems occur for this new form of education. For example, it is needed to control the process of such education in conditions when a teacher doesn't communicate with a student in real life. Methods of control in the area of *e-learning* are named *proctoring*. Physiological indicators are good candidates for the role of objective control of student activity.

Nowadays students routinely use their mobile devices, such as iPads, for learning purposes. Since computers and iPads can be equipped with devices for monitoring users' physiological data, we can analyze students' cognitive states and learning behaviors utilizing inputs from these devices. For the most part educational content is still provided in the text format, which justifies prioritization of eye-movement data over other types of physiological data inputs.

Attempts to implement assessment of educational materials perception on the basis of eye movements and other physiological parameters have been made for some time [7, 8, 9], but mostly these were laboratory studies. The objective of their research was establishing some correlations between the materials offered to viewers and their eye-tracking patterns. The analysis of data obtained was either fully performed by experts or was partially automated, mostly with a low degree of automation. Therefore previous studies of eye movements vary greatly in their approaches and quality of data. This and the fact that some results are ambiguous, did not allow us to develop an algorithm for classifying students' cognitive states based on published results. The use of advanced computer methods, such as new mathematical algorithms and data mining for analysis are more appropriate, because it excludes the human factor and enables us to go deeper in the data [10]. Its use for educational and training tasks promises significant progress in providing a flexible and effective solution that will help educators and students evaluate the perception of teaching materials. Innovative digital educational strategies require innovative technologies.

To achieve a reader's cognitive state estimation by physiological parameters, an accurate measuring tool is needed. In actual study authors developed the approach of obtaining eye movement events from an eye tracker built-in an Apple iPad Pro device. The quality of such data is unfortunately quite low yet and can't be used directly for gaze point detection and eye movement events quality.

So we describe a prospective algorithm of eye movement event detection. We also selected texts that differed greatly in their semantic and stylistic features. As discovered by ML algorithms, reading of different texts was associated with different eye movement strategies of our subjects.

II. Methods

A. iPad application and server ML algorithms

The special software for iPad and iPhone (Oken Reader) was developed by present time. The software integrates a convenient interface of text presentation and implements recording of the respondent's eye tracking data using ARKit technology [Fig. 1]. Recorded data are stored on a server, where ML algorithms are applied to it. It allows the use of significant computer power distantly from individual tablet PCs. Quantitative gaze tracking data revealed essential patterns of reading for different types of texts. So based on eye tracking data server algorithms can estimate the current reader's condition (state) based on a database conformity to all other recorded readers. Data set consists of horizontal (vx), vertical (vy), and absolute velocities of the 3D gaze point; averaged angular velocities of the eyeballs and head (ϖEx, ϖHx, ϖEy, ϖHy, ϖEabs, ϖHabs); pairwise differences of the track coordinates (Δijx, Δijy) - parameters recorded by the device.

B. Participants, experimental design

The first study had 78 participants (22_males, 23+5 y.o., min 18 y.o., max 39 y.o.; 56 females, 29+11 y.o., min 18 y.o., max 61 y.o.), Russian speakers, residents of the metropolis. All the participants signed the consent on the personal information processing and the voluntary consent on the participation in the study. Completed reading sessions were monetary rewarded.

Participants were asked to read texts on iPads placed at a comfortable distance of about 50 cm from their eyes. The algorithm ARkit embedded in iPad allows fixing the distance between the screen and the eyes not strictly, but keeping this distance is better for gaze detection. The readers wearing eyeglasses participated in the study as well: the quality of the eye tracking in this case is pretty similar.

For the study we selected 12 texts with different levels of complexity estimated by the Flesch-Kincaid Readability Test, Coleman-Liau Readability Test, SMOG, Dale-Chall readability formula, and Automated Readability Index [11] adopted for Russian language.

There was also pseudo text among others: it was constructed as a linguistically correct but with meaningless structure. The hypothesis was that reading this text is mostly different from reading any normal one.

III. Results

A. Event detection by EMC algorithm

Unfortunately, to date, the signal quality is not sufficient for precise detection of the gaze position in a reading context (Fig. 1). The figure exemplifies gaze track and x, y coordinates over time during the reading of one paragraph. Due to the data noisiness, basic EMC algorithms (for example, based on the velocity threshold) are not applicable.

The study was inspired by the approach proposed in [12], however, instead of the classical Random Forest algorithm we considered the more modern one, Catboost [13]. Moreover, our feature set differs for the one considered by Zemblys et al. and consists of horizontal, vertical, and absolute velocities of the 3D gaze point; averaged angular velocities of the eyeballs and head; density of the gaze trajectory; pairwise differences of the track coordinates. Moreover, one sample was described not only by its own features but also by the features of neighboring samples (from the "future" as well as from the "past" in the track) that helped out EMC algorithm to capture the temporal dependencies better. The resulting size of the feature set was 242.

The training dataset was created as follows. Eye movements during reading were recorded by a 120 Hz iPhone camera with a magnifying lens. Then, the video recording was synchronized with the eye-tracking signals (recorded using OkenReader at 60 Hz). 17 pages read by 9 people were chosen and the corresponding sections of the video recordings were manually marked up by 3 experts who had long experience of work with eye movement data. Experts worked with raw video recording that was synchronized in time with ARkit tracks. It allowed to obtain most of all crucial information of the gaze detection and events validation. Algorithm didn't filter eye tracking data, so it obtained fixations and saccades of the wide duration range (further events could be filtered in terms of ecological validity). The markup consisted of events of two types: fixations and saccades. Such markup was mapped onto the eye-tracking signal. We trained the algorithm using the 3 quarters of the prepared dataset while the rest (1 quarter) was used for the testing. Table 1 reports the main classification metrics.

The detection of such events is crucially important for the work of the algorithm and text categorization on the base of eye movement data directly depends on the quality of such detection.The significant number of events (saccades and fixations) approved by experts allowed to improve the quality of automated detection highly.

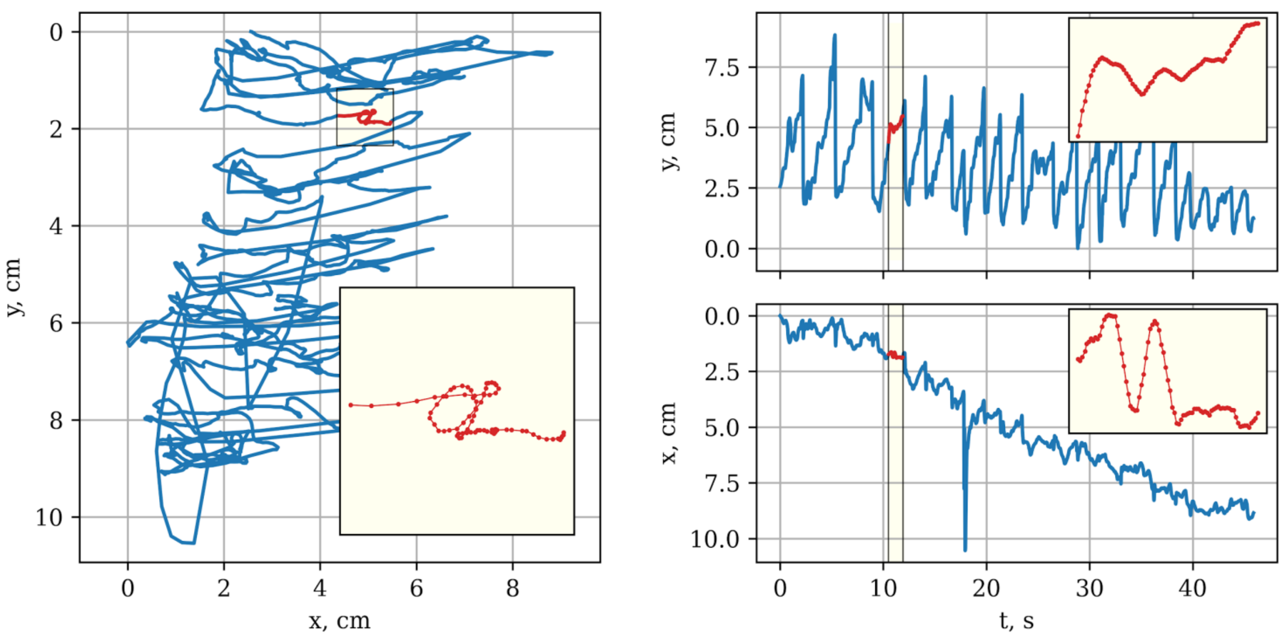

Figure. 1. Example of an eye track during reading (one of the best examples provided). The yellow areas zoom in on a small segment of the track where fixation presumably occurred.

TABLE I. Classification report for the EMC algorithm. Metrics were calculated using the held-out test dataset

	Fixation	Saccade
Precision	0.973	0.661
Recall	0.905	0.881
Fl-score	0.938	0.756
Support	7955	1666

It was still impossible to locate the track accurately on the text, but time parameters became precise after the procedure.

We do not compare our algorithm with others due to the differences in the initial features provided via ARKit. For example, no published algorithms utilize eye and head movements during the reading.

B. Clustering

After improving the quality of temporal features of the eye tracking events an attempt to separate them by relevant mathematical algorithms was made. The works [14, 15] show that reading patterns can be revealed based on eye-tracking data. More specifically, fixations can be assigned to pre-defined Areas of Interest (AOI) after that transition patterns between AOIs (encoded in the transition matrices) can be clusterized.

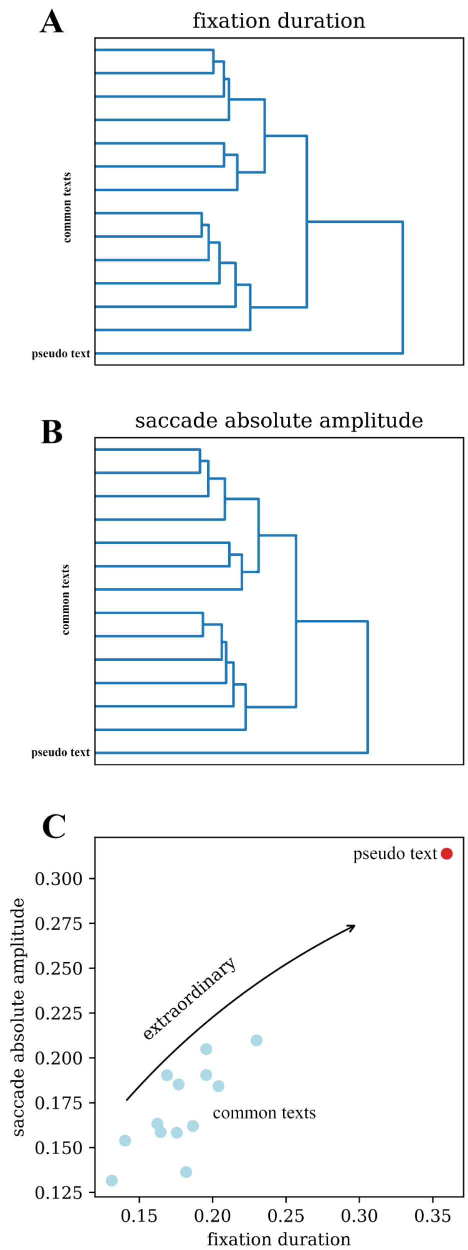

Figure. 2. The results of the agglomerative clustering performed for the distributions of the fixation duration and saccade absolute amplitude shows the difference between reading pseudo text and common texts (see vertical axis for A, B dendrograms). The abscissa is the SJSD calculated for $PDF_{fixation}$ (A) and $PDF_{saccade}$ (B). Pseudo text significantly differs from all common texts. Unlabeled branches correspond to remaining texts. (C) The pairplot shows how texts differ from "the average text" in terms of fixation duration (abscissa) and saccade absolute amplitude (ordinate). Light blue circles denote common texts.

As was stated above, an eye-tracking device does not supply signals that are accurate in terms of spatial coordinates, so it is not able to define whether fixations are located into some AOI. Therefore, a clustering method that uses temporal and local spatial characteristics of eye-movement events neglecting the overall spatial information was implemented. It means that we analyzed timing features and a sequence of fixations but didn't relate to exact positions of the fixations on words.

Using a custom EMC algorithm, distributions of the fixation durations and saccade amplitude (absolute amplitude in degrees) for every text and for all respondents were generated. For further analysis, fixations with durations between 50 and 1000 ms and saccades less than 15 degrees were selected. Such oculomotor events are relevant to normal reading [16].

Then histograms for both types of events (60 equal-sized bins each) and normalized these histograms to obtain the probability density functions ($PDF_{fixation}$ and $PDF_{saccade}$ for fixations and saccades, respectively) were calculated.

Since clustering relies on some similarity metric, squared Jensen-Shannon divergence (SJSD) to clusterize texts that are described by the PDFs was utilized. After calculating the pairwise SJSD agglomerative clustering realized in scikit-learn package [17] was used, namely AgglomerativeClustering with the complete linkage.

The figure (Fig. 2A, 2B) reveals the significant difference between pseudo text and other common types of text. Since dendrograms are built upon only one event type (either fixation or saccade), it's hard to see how these two events work together (Fig. 2A, 2B). To overcome this issue, we plot SJSD calculated using $PDF_{fixation}$ and $PDF_{saccade}$ between every text and "the averaged text" (Fig. 2C). Pseudo text is located in the top-right corner that supports the point that this text differs the most from others in terms of fixation and saccade characteristics. All other (normal) texts are dispersed in the opposite corner (shown as light blue circles).

The result supports that at least the most different pseudo text (that was constructed as a linguistically correct but with meaningless structure) is detected highly significantly. Potentially the algorithm could be used for more slightly differentiated texts.

Compared to the eye picture-based approach of eye-movements classification [18] the approach used in actual study is more robust. First of all, the data from a head-mounted eye tracker and eye tracker from an independent device (for example, phone or tablet PC) are different in the meaning of the degree of freedom for head-movements. Secondly, in addition, the dataset [18] was obtained mainly for working in VR and AR environments. Thirdly, reading static texts from the movable device (2D) includes mainly classical fixations and saccades, and often little or no smooth pursuit events. Fourthly, it also reveals more degree of freedom for device movements. So, we don't use solutions like [19] because it has no task-specific focus and as any ubiquitous tool could limitly solve concrete tasks. At the same time comparison with expert event detection was realized.

IV. Discussion

In the study one of the approaches revealing eye movement strategies of a reader in reading texts with different content was realized. The usage of ML algorithms made it possible to work with a noisy signal recorded by ARkit in iPad Pro and consistently reveal differences in reading patterns for different readers. Such an example was chosen for detection of the case most different from normal reading.

The system recognized the pseudo text with an AUC ROC of 0.93, which is a very good result for registration of eye movements without headrests and utilization of a simplified calibration scheme. If the algorithm is suitable for determining different texts on their parameters, we can suppose that it can be integrated into systems of the reader's cognitive state analysis.

Based on the eye tracking method, reader's state analytics is a promising tool for different fields that deal with information processing from self-learning to publishing house activity. In the first case, a student can choose the most effective strategy of how to prepare for exams using the statistics of his or her attention and mental focus. In the second case, publishers can evaluate readers' attitudes to a book's content for further improvements of the latter. This is extremely important for publishing houses focused on educational and sci-pop literature. In addition, this technology could be useful for school children because of their little learning experience that is insufficient for understanding how the material read was understood.

Despite the high accuracy of the pseudo text identification, the algorithm based on ARKit requires an increase in the accuracy of defining common texts. What features may potentially increase the efficiency of the system? A better determination of the reader's functional state can be achieved by adding other physiological inputs, that is, by integrating data from various devices. In our opinion, the electroencephalogram (EEG) and electrocardiogram (ECG) are promising in this respect. It has been shown that EEG is an effective method for assessing emotional response to presented information [20], as well as its memorization and retrieval from memory [21, 22]. The ECG is a source of data about heart rate variability (HRV), which can be used to assess which of the two types of autonomic nervous system is dominant: sympathetic or parasympathetic [23]. Conclusions about the level of stress and fatigue can be drawn on the basis of HRV [24, 25]. A strong argument in favor of using HRV is the simplicity of ECG registration and the fact that this indicator is quite objective, because the heart rhythm cannot be changed with a person's own conscious efforts. The accuracy of state recognition can also be improved by calculating indicators not for the entire text, but for its parts. This will allow us to identify areas of interest within the text and analyze physiological patterns that have characteristics of quick flashes: for example, a particular EEG rhythms that may be associated with a greater cognitive and attention load.

V. Conclusions

In the current work an algorithm and complex solution based on iPad/iPhone embedded eye tracker was developed. Thus achieving two goals: 1) developed a stable algorithm that can reliably extract eye movement events from noisy eye tracking data, 2) implemented the detection of reading strategy on the model of different texts presentation.

The ML algorithm can analyze big volumes of gaze movement data and reveal features related to the reader's eye movement strategy during perceiving educational and other content. The algorithm recognizes unusual texts (such as pseudo-text and instructed texts) with distinct features that are significantly different from those of other texts. In the future we are planning to discern reading patterns for less idiosyncratic texts. This would open up prospects for using the approach in teaching. For example, when studying for international exams such as TOEFL and IELTS, which contain a number of assignments requiring finding relevant information in texts. The methodology can also be used for educational exercises that describe a particular strategy for finding answers, important details and key information.

With ML methodology, relevant results can be obtained even with the use of common consumer devices, such as iPads. This opens up a wide perspective for the use of integrated eye trackers in school, higher and professional education, targeted at a large number of students. The system can also be incorporated into professional training systems and operator manuals. Currently a new study is being conducted to identify differences between reading English and Russian texts. In the future the use of algorithms for different languages will allow us to develop platforms for other countries, using all capabilities of interactive technologies that accumulate data and bring meaningful results regardless of physical locations of students. The technology can thus provide cognitive feedback to improve the efficiency of the whole e-learning process.

References

[1] Rayner, K. Eye movements in reading and information processing: 20 years of research. Psychological bulletin, 124(3), 1998, p. 372.

[2] Yu, Shaoda, et al. "Support vector machine based detection of drowsiness using minimum EEG features." 2013 International Conference on Social Computing. IEEE, 2013.

[3] Hou, Xiyuan, et al. "CogniMeter: EEG-based emotion, mental workload and stress visual monitoring." 2015 International Conference on Cyberworlds (CW). IEEE, 2015.

[4] Allen, M., Bourhis, J., Burrell, N., & Mabry, E. Comparing student satisfaction with distance education to traditional classrooms in higher education: A meta-analysis. The American Journal of Distance Education, 16(2), 2002, pp. 83-97.

[5] Kasmad, K., Mustakim, M., & Sunarsi, D. Increasing Community School Interest Through Service Quality, Prices and Promotion in Vocational High Schools. Journal of Educational Science and Technology (EST), 6(2), 2020.

[6] Poulin, R., & Straut, T. T. Distance Education Price and Cost Report. WICHE Cooperative for Educational Technologies (WCET), 2017.

[7] Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. Eye-tracking technology in medical education: Asystematic review. Medical teacher, 40(1), 2018, pp. 62-69.

[8] Sun, Y., Li, Q., Zhang, H., & Zou, J. The application of eye tracking in education. In International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Springer, Cham, 2017, pp. 27-33.

[9] Xu, J., & Zhong, B. Review on portable EEG technology in educational research. Computers in Human Behavior, 81, 2018, pp. 340-349.

[10] Maier, M., Elsner, D., Marouane, C., Zehnle, M., & Fuchs, C. DeepFlow: Detecting Optimal User Experience From Physiological Data Using Deep Neural Networks. In AAMAS, 2019, pp. 2108-2110).

[11]Patalay, P., Hayes, D., & Wolpert, M. Assessing the readability of the self-reported Strengths and Difficulties Questionnaire. BJPsych open, 4(2), 2018, pp. 55-57.

[12] Zemblys, R., Niehorster, D. C., Komogortsev, O., & Holmqvist, K. Using machine learning to detect events in eye-tracking data. Behavior research methods, 50(1), 2018, pp. 160-181.

[13] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

[14] Mozaffari, S., Al-Naser, M., Klein, P., Küchemann, S., Kuhn, J., Widmann, T., & Dengel, A. (2020). Classification of Visual Strategies in Physics Vector Field Problem-solving. In ICAART (2), 2020, pp. 257-267.

[15] Kucharský, Š., Visser, I., Truțescu, G. O., Laurence, P. G., Zaharieva, M., & Raijmakers, M. E. Cognitive strategies revealed by clustering eye movement transitions. Journal of Eye Movement Research, 13(1), 2020.

[16] Manor, B. R., & Gordon, E. Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. Journal of neuroscience methods, 128(1-2), 2003, pp. 85-93.

[17] Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2011, pp. 2825–2830.

[18] Fuhl, W., & Kasneci, E. (2021). A Multimodal Eye Movement Dataset and a Multimodal Eye Movement Segmentation Analysis. arXiv preprint arXiv:2101.04318.

[19] Fuhl, W., Kasneci, G., & Kasneci, E. (2021). TEyeD: Over 20 million real-world eye images with Pupil, Eyelid, and Iris 2D and 3D Segmentations, 2D and 3D Landmarks, 3D Eyeball, Gaze Vector, and Eye Movement Types. arXiv preprint arXiv:2102.02115.

[20] Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., ... & Yang, X.. A review of emotion recognition using physiological signals. Sensors, 18(7), 2018, p. 2074.

[21] Numata, T., Kiguchi, M., & Sato, H. Multiple-Time-Scale Analysis of Attention as Revealed by EEG, NIRS, and Pupil Diameter Signals During a Free Recall Task: A Multimodal Measurement Approach. Frontiers in neuroscience, 13, 2019, p. 1307.

[22] Mazzon, G., De Dea, F., Cattaruzza, T., Manganotti, P., Monti, F., & Accardo, A. Memorization test and resting state EEG components in mild and subjective cognitive impairment. Current Alzheimer Research, 15(9), 2018, pp. 809-819.

[23] Khan, A. A., Lip, G. Y., & Shantsila, A. Heart rate variability in atrial fibrillation: The balance between sympathetic and parasympathetic nervous system. European journal of clinical investigation, 49(11), 2019, e13174.

[24] Kim, J. A., & Kang, S. W. (2017). Relationship among sleep quality, heart rate variability, fatigue, depression, and anxiety in adults. Korean Journal of Adult Nursing, 29(1), 2017, pp. 87-97.

[25] Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry investigation, 15(3), 2018, p. 235.