MIPRO 2021, September 27 - October 01, 2021, Opatija, Croatia

Classification of Cognitive Load based on Oculometric Features

Mate Gambiraža, Ivan Kesedžić, Marko Šarlija, Siniša Popović, Krešimir Ćosić
Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia
mate.gambiraza@fer.hr

***Abstract* - Cognitive load is related to the amount of working memory resources used in the execution of various mental tasks. Different multimodal features extracted from peripheral physiology, brain activity, and oculometric reactions have been used as non-intrusive, reliable, and objective measures of cognitive load. In this paper, we use data from 38 participants performing a four-level difficulty n-back task (0-, 1-, 2-, and 3-back task), with their oculometric reactions simultaneously recorded. Based on the neuroanatomic structure and function of the visual system, 26 oculometric features are extracted and organized into 3 groups related to: pupil dilation (PD), blinking, and fixation. The discriminative power of each group of features was evaluated in four-level cognitive load classification using a support vector machine (SVM) model and feature selection, and the achieved classification accuracies were: 33.33% using only pupil dilation features, 30.90% using only blink-related features, 30.21% using only fixations-related features. Finally, a 36.11% classification accuracy was achieved using a combination of all extracted oculometric features. The presented results show that various groups of oculometric features provide complementary information about the subject's cognitive load. The comparison of the extracted groups of features is given, and the most important features in terms of classification performance are discussed.**

Keywords - cognitive load; classification; eye tracking; pupillometry; SVM

I. INTRODUCTION

Cognitive load is related to the amount of working memory resources used in the execution of various mental tasks and is generally considered a multidimensional construct [1]. Measurable dimensions of cognitive load are mental load, mental effort, and performance [1]. Mental load is considered an indication of the expected cognitive capacity demands and a priori estimate of cognitive load [2]. The second aspect of cognitive load, mental effort, refers to the cognitive resources that are allocated to execute the given task and is considered to reflect the actual cognitive load. Performance, as the third aspect of cognitive load, is defined in terms of person's achievements, e.g. number of correct test items, number of errors, and execution time, and can be determined during the task or thereafter. Cognitive load can be increased by a variety of factors, such as the difficulty of a task, environment (distracting vs. quiet), person's expertise, etc. [2]. High cognitive load induces stress, resulting in decreased task performance, while a manageable level of the cognitive load has a lower effect on the performance. Therefore, automated assessment of cognitive load can provide insights to maximize user's performance by easing the cognitive load to avoid stress, frustration, and errors [3].

Four primary methods to cognitive load assessment are: subjective (self-report) measures, performance measures, behavioural measures, and physiological measures [4]. Physiological measures are based on the assumption that changes in cognitive functioning induce changes in physiological variables [2] related to the heart, brain, skin, and eye activity and are often used in cognitive load assessment [5, 6]. Oculometric, i.e. eye activity, measures have several advantages over the other methods, with non-intrusiveness of the measurement being the main advantage. There are various techniques used for eye activity tracking such as electrooculographic, galvanometric, or corneal reflective techniques [7], while eye-tracking setups vary: some are head-mounted, some function remotely and automatically track the head during motion, etc. Depending on the chosen technique, setup, environment, and research hypotheses, the extracted oculometric features vary, but mostly include features related to pupil dilation, blinking, gaze fixations, and saccades.

Pupillometry is proven to be a reliable tool for studying cognitive and emotional processes [8]. Pupil diameter (PD) reflects the activity of the locus coeruleus nucleus (LC) [7, 9-13], a brain structure that promotes noradrenergic signalling in the brain during the stress response and regulates arousal and autonomic activity. Pupil dilation occurs as a consequence of mental effort exertion and numerous sources of the accompanying psychological stress [15-16]. Blinking, precisely spontaneous eyeblink rate, correlates with dopamine levels in the central nervous system and is associated with learning processes and certain aspects of cognitive flexibility [9, 17]. Studies have shown that cognitive load influences the rate and duration of eyeblinks [18-21].

Superior colliculus (SC), a brain structure located in the midbrain, controls motor planning and execution of saccades [22]. It has been shown that saccadic peak velocity correlates with measures of SC activity [23, 24] and decreases towards the end of the task [25, 26]. Furthermore, a longer saccades length was found to indicate higher cognitive load [27], with a higher saccade count being associated with higher visual complexity [28]. However, the saccadic measures highly depend on the performed cognitive task. Ocular fixation is actively controlled by the same brain structures involved in the control of saccadic eye movements (e.g. SC, cerebellum, and reticular formation) [29] and longer fixation duration indicates higher demands on working memory [27].

The classification of cognitive load is a widely investigated research topic [30-37]. Assessing cognitive

IK has been supported through the funding from the Croatian Science Foundation.

load from oculometric features is effective and convenient, and the data can be obtained in real-time. Real-time classification, through non-intrusive measurement, is desirable in various situations such as driving, piloting, and other environments where direct physiological measurements are inconvenient or may limit the subject's task performance.

One of the most widely used tasks in cognitive load estimation is the n-back task [30]. Several studies investigated cognitive load classification induced by different levels of the n-back task from eye-related measures including pupil dilation, blink duration, number of blinks, eye movements, etc. [31-34]. Some studies used a combination of eye-related measures with other physiological modules such as heart activity, electrodermal activity, electroencephalography (EEG), etc. [33-35]. The number of classes in classification studies based on oculometric features on n-back task vary: from pairwise classification (0-back vs. 2-back, 1-back vs. 2- back) [31] to classifying 3 levels of task difficulties [32].

Besides the n-back task, cognitive load classification based on oculometric features was investigated in the context of other cognitively demanding tasks, such as Stroop task [36], flight simulator [37], driving simulator [16], and real test drive [38]. Most of the studies focused on one task only [16, 31, 37], while some of them investigated a combination of various tasks [32, 38]. Used classification models included convolutional neural networks (CNNs) [16, 32], Extra-Trees [31], Gaussian Mixture Models (GMMs) [34], and other models such as support vector machines (SVMs), k-nearest neighbours (k-NN), and linear discriminant analysis (LDA) [39].

The studies briefly reviewed above have generally shown the discriminative power of oculometric features during increased cognitive demand and their potential in cognitive load assessment. However, none of the mentioned studies investigated cognitive load classification of 4 levels of n-back task solely from oculometric features organized into different functional groups of features, which is the topic of this paper.

II. METHODS

A. Participants

The participants in this study were 38 students from the University of Zagreb (34 male and 4 female; mean age = 23.97; SD = 1.77 years; age range = 22-31). All participants signed informed consent according to the guidelines of the University of Zagreb. The study was approved by the Ethics Committee of the University of Zagreb Faculty of Electrical Engineering and Computing.

B. Experimental protocol and setup

The participants underwent a series of n-back tasks. The participants were seated in front of a screen where the tasks were displayed and were instructed to respond with a left arrow on the keyboard in front of them whenever they were presented the same stimulus as the one presented n trials previously, and with a right arrow whenever the stimulus was not the same as the one presented n trials previously, where n is a prespecified integer (n = 0, 1, 2, or 3), similar to the experiment design in [6]. At the bottom of the screen there were always shown words "yes" on the left side of the screen representing the positive response, and "no" on the right side representing that the stimulus is not the same as the one presented n trials previously, where a white box around the selected answer would appear, once answered. The participants' reaction time and accuracy were measured, along with their oculometric responses during the task. The stimuli in the n-back task were uppercase consonant letters. Each letter was presented on the screen for 0.5 s, with a period of 2.5 s for the participants' response. The experimental paradigm consisted of 17 blocks, which included eight blocks of tasks, and nine blocks of resting periods in duration of 45 s, where blocks of tasks and resting blocks are presented in alternating order. During the resting periods, the participants were asked to relax and continue looking at the screen in front of them. The blocks of tasks were randomly shuffled for each participant. The subjects underwent a total of 120 trials of the n-back task that were evenly distributed into four difficulty levels. The total duration of the protocol was 765 s.

The Gazepoint GP3 HD remote eye tracking system was used to record the dynamics of the oculomotor system, PD, and eye blinking (Fig. 1). The device is non-invasive and allows head movement. Data is recorded at a frequency of 150 Hz using markers obtained by the reflection of infrared light on the pupil of the eye, and the compensation of head movements is calculated using a depth camera. The characteristics of the device used to monitor the gaze are following:

- 0.5-1 degree of visual angle accuracy
- 150 Hz sampling frequency
- 9-point calibration
- 35 cm (horizontal) x 22 cm (vertical) movement
- ±15 cm range of depth movement

C. Data Processing and Feature Extraction

Oculometric data were collected during all 17 segments of the n-back task, which included PD for both pupils, blinking events, and gaze coordinates (Fig. 2). The data were divided into 3 functional groups related to different eye activities (pupillary response, blinking events, and fixations) for further feature extraction. Data from two participants were removed due to recording errors.

1. **Pupillary response:** Pupillary response is calculated as PD change compared to the baseline value. During blink time eye-tracking device holds the last PD recorded and these values need to be removed from the dataset. From processed pupillary response data, the following oculometric features were calculated for each task segment:
    a. Mean pupil diameter (PD_{mean})
    b. Median PD (PD_{median})
    c. Standard deviation of PD (PD_{std})
    d. Minimum PD (PD_{min})
    e. Maximum PD (PD_{max})
    f. Slope of PD (PD_{slope})
    g. Standard deviation of PD slope (PD_{std_slope})

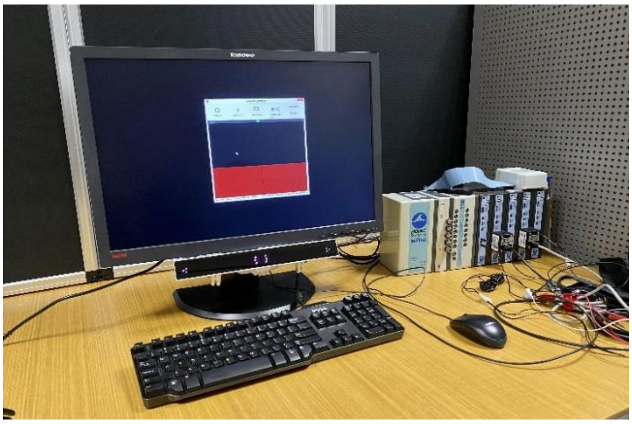

Figure 1. The equipment used in the experiment: Gazepoint GP3 HD remote eye tracking system

2. **Blinking events:** Further data processing consisted of detecting starting and ending points of blink events. Following oculometric features were calculated for each task segment:
    a. Number of blinks (NoB)
    b. Total time eyes were closed during the segment (BD_{total})
    c. Mean blink duration (BD_{mean})
    d. Standard deviation of blink duration (BD_{std})
    e. Median blink duration (BD_{median})
    f. Mean time between blinks (TBB_{mean})
    g. Standard deviation of time between blinks (TBB_{std})
    h. Median time between blinks (TBB_{median})
3. **Fixations:** From the gaze coordinates data, the following oculometric features were extracted:
    a. Number of fixations (NoF)
    b. Total gaze distance, i.e. measured trajectory length divided by Euclidian distance between gaze coordinates (TGD)
    c. Mean fixation duration (FD_{mean})
    d. Median FD (FD_{median})
    e. Standard deviation of FD (FD_{std})
    f. Maximum FD (FD_{max})
    g. Minimum FD (FD_{min})
    h. Distance per fixation, i.e. total gaze distance divided by the number of fixations (DpF)

Besides the features computed using the oculometric data during a specific task (T), as described above, the following ratio-based features were calculated for all feature groups: task feature value divided by the preceding rest feature value (T/B), task feature value divided by the following rest feature value (T/A), and feature value during the preceding rest divided by the following rest feature value (B/A). Total number of calculated features was 92, i.e. 23 initial features and 3 ratio-based versions of each feature, as described above. Participant-based normalization was done, i.e. for each participant mean feature values were subtracted and divided by its standard deviation. After normalization, the ANOVA test was applied as a preliminary feature elimination step.

D. Classification Models

The classification was performed using the SVM learning model. All data extraction and processing were performed in Matlab R2020b. Four classifiers were trained to discriminate four classes of tasks difficulty. Single feature groups as well as the combination of all three feature groups were evaluated in the models. The training process for all models was carried out using a nested leave-one-subject-out (LOSO) cross-validation, which incorporates hyperparameter tuning and model evaluation [5, 6]. The data were first divided into 36 folds, where each fold comprised instances from a single participant. Data from one participant, i.e. one fold, was always held as the testing set, and the remaining data, i.e. 35 folds, were used for training. This process was repeated 36 times, testing on a different participant in each iteration. Classification accuracy for each testing set was calculated by dividing the number of correctly predicted labels by the total number of testing samples from each subject (i.e. correct/total). Mean classification accuracy and standard deviation (SD) for all models were calculated from the obtained 36 classification

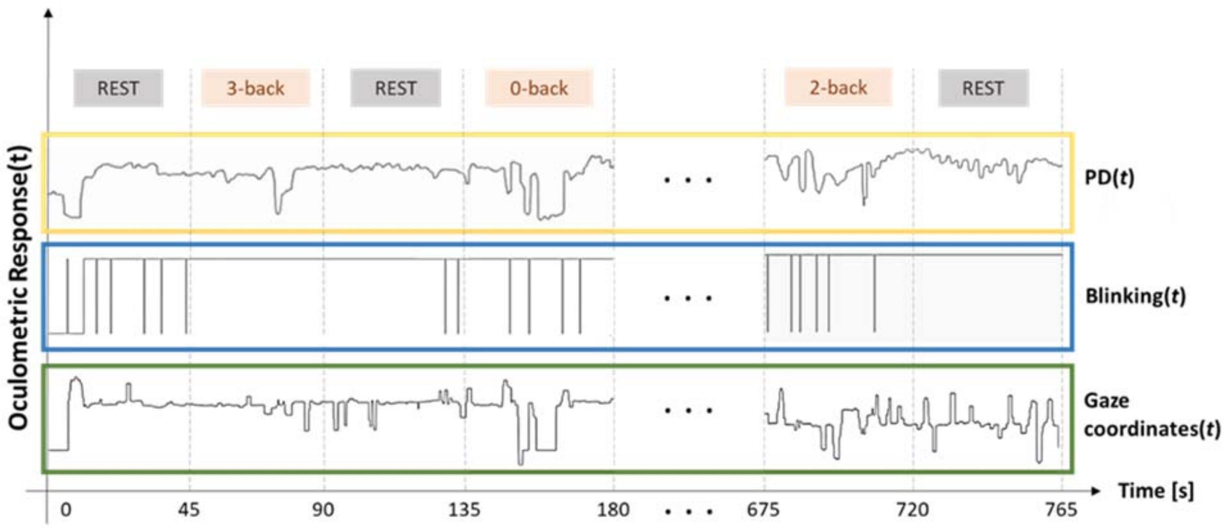

Figure 2. A time-domain plot of the recorded oculometric signals for a specific participant. Blocks of the paradigm are denoted at the top of the plot

accuracies on each of the test folds. The kernel used for the SVM model was the radial basis function kernel, and the hyperparameters tuned for the SVM classifier, using Bayesian optimization, included gamma (the kernel coefficient, search range from 0.01 to 100) and C (the penalty for the error, search range from 0.01 to 100).

III. RESULTS AND DISCUSSION

In this study, the cognitive load classification based on oculometric features was investigated for 4 different levels of the n-back task (0-, 1-, 2-, and 3-back task). As a preliminary feature elimination step, we then applied the ANOVA test ($p<0.1$) and 14 features from PD group, 4 features from Blink group, and 8 features from Fixation group were selected for classifiers training, making a total of 26 features for the final classifier based on all groups of features. The remaining features were eliminated due to their weak discriminative power on a between-task level. Selected features were:

1. **PD group:** $PD_{mean\text{-}T}$, $PD_{median\text{-}T}$, $PD_{min\text{-}T}$, $PD_{std\_slope\text{-}T/B}$, $PD_{mean\text{-}T/A}$, $PD_{median\text{-}T/A}$, $PD_{min\text{-}T/A}$, $PD_{max\text{-}T/A}$, $PD_{mean\text{-}B/A}$, $PD_{median\text{-}B/A}$, $PD_{std\text{-}B/A}$, $PD_{min\text{-}B/A}$, $PD_{max\text{-}B/A}$, PDstd_slope-B/A
2. **Blink group:** NoBT, $BD_{std\text{-}T/A}$, $FD_{mean\text{-}T/A}$, $FD_{median\text{-}T/A}$
3. **Fixation group:** TGDT, $FD_{median\text{-}T/B}$, $TGDT_{T/A}$, $FD_{mean\text{-}T/A}$, $NoF_{T/A}$, $TGD_{B/A}$, $FD_{mean\text{-}B/A}$, $NoF_{B/A}$

To distinguish between 4 different levels of cognitive load represented by the n-back levels, four different SVM classifiers were trained and evaluated for cognitive load assessment. One classifier was trained for each of the three different feature groups and one classifier was trained using all feature groups combined. The evaluation scores of the trained classifiers are presented in Table I and Table II. The mean classification accuracies vary between 30.21% and 33.33% when using only one feature group, while the mean classification accuracy based on all feature groups is 36.11%. The mean classification accuracy achieved for the PD group is 33.33%. One of the 14 features in the PD group of features, median pupil diameter, is already reported in [31, 40, 41] as an important feature in cognitive load estimation. The mean accuracy achieved for the blink group is 30.90%. One of the oculometric features from this group, number of blinks, is also reported in several previous studies [31, 42]. The mean classification accuracy achieved for the fixation group is 30.21% and is the lowest accuracy for a single feature group classification. Finally, we showed that the classification based on all features achieved the highest mean classification accuracy of 36.11%, and that 4-level cognitive load can be assessed using only oculometric features with a better-than-chance level classification, i.e. better than 25%.

As expected, classification based only on PD features achieved the highest accuracy among classifiers using only one feature group, which is already reported in [31] where the authors did pairwise classification of 3 levels of n-back task and used only two feature groups: PD group and blink-related group. Still, small changes in PD for 0- and 1-back task, as shown in a previous study [44], might have constrained the classification accuracy. Furthermore, the addition of blink feature group and fixation feature group

TABLE I. MEAN CLASSIFICATION ACCURACIES (IN %) OF TRAINED SVM MODELS FOR 4-LEVEL CLASSIFICATION

Feature set	*Mean accuracy ± SD*
PD	33.33 ± 21.96
Blink	30.90 ± 12.50
Fixation	30.21 ± 15.92
PD + blink + fixation	36.11 ± 25.14

TABLE II. RECALL AND PRECISION OF TRAINED SVM MODEL ON ALL GROUPS OF FEATURES FOR 4-LEVEL CLASSIFICATION

N-back	*Recall*	*Precision*
0-back	0.7561	0.5962
1-back	0.4118	0.5676
2-back	0.3585	0.4043
3-back	0.7143	0.5882

increased mean classification accuracy of the final model, showing that various groups of oculometric features provide complementary information about the subject's cognitive load.

Table II shows comparably higher recalls for classification of 0- and 3-back tasks vs. 1- and 2- back tasks. This indicates more pronounced differences in the oculometric features during 0- and 3-back task segments, relative to 1- and 2-back task segments.

In addition to the classification accuracy, we provided information on the error distance, i.e. the absolute distance between the labelled difficulty level and the level estimated by the classifier, for all trained classifiers, as shown in Fig. 3. Most of the classification errors are between neighbouring classes, i.e. the classes that differ in absolute values by one (0- vs. 1-back task; 1- vs. 2-back task, and 2- vs. 3-back task). Such classification error is expected since feature values are more similar between these neighbouring classes. For random guessing, error distribution would be 25% for distance 0, 37.5% for distance 1, 25% for distance 2, and 12.5% for distance 3, based on the ratio of relevant outcomes vs. all possible outcomes on 4x4 confusion matrix, i.e. 4/16, 6/16, 4/16, 2/16. Furthermore, the sum of the first two columns, i.e. error distance 0 and 1, for random guessing would be 62.5% and it is shown in Fig. 3 that the sum of those 2 columns is higher for all classifiers ranging from 71.87% to 75% which indicates neighbouring class similarity.

In addition to 4-level classification, we did the 3-level classification for all combination of classes, as well as the pairwise classification. The achieved accuracies are presented in Table III, while acknowledging relatively large SDs which might be due to inter-individual differences in eye tracking features as well as several potential sources of error. When compared to [31], a study which is most similar to the present study, we achieved lower mean accuracies for 0- vs. 2-back and 1- vs. 2-back classification (76.8% for 0-vs. 2-back and 71.5% for 1- vs. 2-back is achieved in [31]). However, slightly higher accuracies were achieved for 0- vs. 1-back and 0- vs. 1- vs. 2-back classification (e.g., 54.0% for 0- vs. 1-back, and 46.7% for 0- vs. 1- vs. 2-back is achieved in [31]).

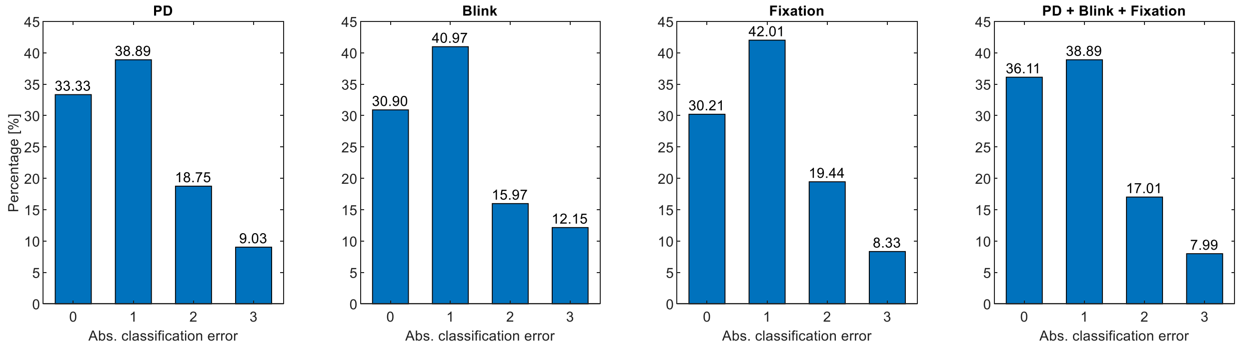

Figure 3. Error Distance: The absolute distance between the labeled difficulty level and the level estimated by the classifier, for all trained 4-level task difficulties classifiers

These findings are likely a consequence of the methodological differences between the study by Appel et al. [31] and the present study. Most notably, [31] implements a cross-subject classification method where the test subject is matched with n subject-specific classifiers and votes from each classifier are weighted according to a similarity measure between the test subject and the train set subjects. On the other hand, authors used less features to train while we used a more comprehensive set of features, including fixation-related features. In any case, this paper represents in-progress research on eye tracking measures for cognitive load estimation, which will in future research require investigating various potential sources of classification error, like:

- the quality and robustness of eye tracking system (Gazepoint GP3 HD system in our work vs. 250 Hertz SMI in reference [31]);
- no chin rest in our work, which is used in reference [31] to fix head position and minimize head movements;
- data sampling approach: our work had 8 samples per subject, i.e. 2 samples per cognitive load level, while reference [31] had 100 samples per cognitive load level for each subject, which were computed on 1-through-5-second sequences randomly extracted from eye tracking recordings on 0-, 1-, and 2-back tasks;
- as well as feature computation and classification modelling approach, whose differences with regard to [31] have already been outlined.

TABLE III. MEAN CLASSIFICATION ACCURACIES (IN %) OF TRAINED SVM MODELS FOR 2- AND 3-LEVEL

Levels	*Mean accuracy ± SD*
0 vs. 1	61.11 ± 24.96
0 vs. 2	66.67 ± 24.64
0 vs. 3	68.06 ± 27.78
1 vs. 2	62.55 ± 27.06
1 vs. 3	63.89 ± 28.31
2 vs. 3	53.47 ± 19.96
0 vs. 1 vs. 2	48.15 ± 23.50
0 vs. 1 vs. 3	50.00 ± 23.90
0 vs. 2 vs. 3	43.06 ± 17.98
1 vs. 2 vs. 3	46.30 ± 23.27

IV. CONCLUSION

In this paper, four difficulties of n-back task were classified using solely oculometric data. Previous studies did cognitive load classification of n-back task either based on a combination of oculometric and other physiological measures or on three or less levels of difficulty [31, 33, 34]. The main aim of this work was to further investigate the potential of a comprehensive set of oculometric features in the task of unimodal 4-level cognitive load classification. We used LOSO validation to evaluate the effectiveness of our models. We achieved better than chance classification accuracies and the comparison of the different oculometric groups of features is given.

One of the main strengths of this work is the randomized protocol design which eliminates the chances of overestimating classifier's performance. Due to randomly shuffled blocks of tasks for each participant, in the future work other classification models, besides the SVM, could be used in cognitive load classification such as more complex sequence models, like recurrent neural networks or long short-term memory networks.

REFERENCES

[1] F. G. Paas and J. J. Van Merrienboer, "Instructional control of cognitive load in the training of complex cognitive tasks," Educ. Psychol. Rev., vol. 6, no. 4, pp. 351-371, Dec. 1994.

[2] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," Educ. Psychol., vol. 38, no. 1, pp. 63-71, Mar. 2003.

[3] F. Chen, J. Zhou, Y. Wang, K. Yu, and S. Z. Arshad, "Multimodal Behavioral and Physiological Signals as Indicators of Cognitive Load," Springer International Publishing, The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition, vol 2, pp. 287–329, Oct. 2018.

[4] F. Chen, J. Zhou, Y. Wang, K. Yu, S. Z. Arshad, A. Khawaji, and D. Conway, "Robust Multimodal Cognitive Load Measurement," Springer International Publishing, pp. 13-32, 2016.

[5] I. Mijić, M. Šarlija, and D. Petrinović, "MMOD-COG: A database for multimodal cognitive load classification," In 2019 11th Int. Symp. Image Signal Process. Anal. ISPA, pp. 15-20, Sep. 2019.

[6] I. Kesedžić, M. Šarlija, J. Božek, S. Popović, and K. Ćosić, "Classification of Cognitive Load based on Neurophysiological Features from Functional Near-Infrared Spectroscopy and

Electrocardiography Signals on n-back Task," IEEE Sens. J., vol. 21, no. 13, pp. 14131-14140, Jul. 2021.

[7] D. B. Carr and P. Grover, "The Role of Eye Tracking Technology in Assessing Older Driver Safety", Geriatrics (Basel), vol. 5, no. 2, no. 36, Jun. 2020.

[8] E. E. Granholm and S. R. Steinhauer, "Pupillometric measures of cognitive and emotional processes," Int. J. Psychophysiol., vol. 52, no. 1, Mar. 2004.

[9] M. K. Eckstein, B. Guerra-Carrillo, A. T. Miller Singley, and S. A. Bunge, "Beyond eye gaze: What else can eye tracking reveal about cognition and cognitive development?" Dev. Cogn. Neurosci., vol. 25, pp. 69–91, Jun. 2017.

[10] D. H. McDougal, and P. D. Gamlin, "Autonomic control of the eye," Compr. Physiol., vol. 5, no. 1, pp. 439-473, Jan. 2011.

[11] J. F. Hejtmancik, P. Cabrera, Y. Chen, O. M'Hamdi, and J. M. Nickerson, "Vision," Conn's Translational Neuroscience, pp. 399-438, 2017.

[12] S. Joshi, Y. Li, R. M. Kalwani, and J. I. Gold, ''Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex,'' Neuron, vol. 89, no. 1, pp. 221–234, Jan. 2016.

[13] S. Mathôt, "Pupillometry: Psychology, physiology, and function," J. Cogn., vol. 1, no. 1, pp. 16, Feb. 2018.

[14] M. L. H. Vo, A. M. Jacobs, L. Kuchinke, M. Hofmann, M. Conrad, A. Schacht, and F. Hutzler, F., "The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect," Psychophysiology, vol. 45, pp. 130–140, 2008.

[15] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," Psychophysiology, vol. 45, no. 4, pp. 602-607, Jul. 2008.

[16] M. Pedrotti, M. A. Mirzaei, A. Tedescho, J. R. Chardonnet, F. Merienne, et al., "Automatic Stress Classification With Pupil Diameter Analysis," Int. J. Hum. -Comput. Int., Taylor & Francis, vol. 30, no. 3, pp. 220-236, Jan. 2014.

[17] H. Ledger, "The effect cognitive load has on eye blinking," The Plymouth Student Scientist, vol. 6, no. 1, pp. 206-223, 2013.

[18] S. Benedetto, M. Pedrotti, L. Minin, T. Baccino, A. Re, and R. Montanari, "Driver workload and eye blink duration," Transp. Res. Part F Traffic. Psychol. Behav., vol. 14, no. 3, pp. 199-208, May 2011.

[19] K. F. Van Orden, W. Limbert, S. Makeig, and T.P. Jung, "Eye activity correlates of workload during a visuospatial memory task," Hum. Factors, vol. 43, no. 1, pp. 111-121, Feb. 2001.

[20] Y. Liang, and J. Lee, "A hybrid Bayesian Network approach to detect driver cognitive distraction," Transp. Res. Part C Emerg. Technol., vol. 38, pp. 146–155, Jan. 2014.

[21] A. F. Kramer, "Physiological metrics of mental workload: A review of recent progress," Multiple-task performance, pp. 279-328, 2020.

[22] D. L. Sparks, "Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus," Physiol. Rev., vol. 66, no. 1, pp. 118-171, Jan. 1986.

[23] J. A. Seideman, "A dynamic, imperturbable link between midbrain activity and saccade velocity," J. Neurophysiol., vol. 123, no.2, pp. 451-453, Jan. 2020.

[24] I. Smalianchuk, U.K. Jagadisan, and N.J. Gandhi, "Instantaneous midbrain control of saccade velocity," J. Neurosci., vol. 38, no. 47, pp. 10156-10167, Nov. 2018.

[25] R. Z. Marandi, P. Madeleine, Ø. Omland, N. Vuillerme, and A. Samani, "Eye movement characteristics reflected fatigue development in both young and elderly individuals," Sci. Rep., vol. 8, no. 1, pp. 1-10, Sep. 2018.

[26] E. Abdulin and O. Komogortsev, ''User eye fatigue detection via eye movement behaviour,'' in Proc. 33rd Annu. ACM Conf. Extended Abstracts Hum. Factors Comput. Syst., pp. 1265–1270, Apr. 2015.

[27] J. Zagermann, U. Pfeil, and H. Reiterer, "Measuring Cognitive Load using Eye Tracking Technology in Visual Computing," Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16). Association for Computing Machinery, New York, NY, USA, pp. 78–85. 2016.

[28] S. Lallé, C. Conati, and G. Carenini, "Predicting Confusion in Information Visualization from Eye Tracking and Interaction Data," In IJCAI, pp. 2529-2535, Jan. 2016.

[29] R. J. Krauzlis, L. Goffart and Z. M. Hafed, "Neuronal control of fixation and fixational eye movements," Phil. Trans. R. Soc. B372, Feb. 2017.

[30] W. K. Kirchner, "Age differences in short-term retention of rapidly changing information," J. Exp. Psychol., vol. 55, no. 4, pp. 352, Apr. 1958.

[31] T. Appel, C. Scharinger, P. Gerjets, & E. Kasneci, "Cross-subject workload classification using pupil-related measures," In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, pp. 1-8, Jun. 2018.

[32] L. Fridman, B. Reimer, B. Mehler, and W. T. Freeman, "Cognitive Load Estimation in the Wild," Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Association for Computing Machinery, New York, NY, USA, Paper 652, pp. 1–9, Apr. 2018.

[33] P. Vanneste, A. Raes, J. Morton, et al., "Towards measuring cognitive load through multimodal physiological data," Cogn. Tech. Work., pp. 1-19, Apr. 2020.

[34] F. S. Chen, J. A. Minson, M. Schöne, and M. Heinrichs, "In the eye of the beholder: Eye contact increases resistance to persuasion," Psychol. Sci., vol. 24, no. 11, pp. 2254-2261, Nov. 2013.

[35] T. M. Gable, A. L. Kun, B. N. Walker, and R. J. Winton, "Comparing heart rate and pupil size as objective measures of workload in the driving context: initial look," Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15), Association for Computing Machinery, New York, NY, USA, pp. 20–25, Sep. 2015.

[36] B. Laeng, M. Ørbo, T. Holmlund, and M. Miozzo, "Pupillary stroop effects," Cogn. Process, vol. 12, no. 1, pp. 13-21, Sep. 2010.

[37] M. D. Babu, D. V. Jeevithashree, G. Prabhakar, K. P. S. Saluja, A. Pashilkar, and P. Biswas, "Estimating pilots' cognitive load from ocular parameters through simulation and in-flight studies," J. Eye Mov. Res., vol. 12, no. 3, Sep. 2019.

[38] S. Yang, J. Kuo, and M. G. Lenné, "Analysis of Gaze Behavior to Measure Cognitive Distraction in Real-World Driving," Proc. Hum. Factors Ergon. Soc. Annu. Meet., vol. 62, no. 1, pp. 1944–1948, Sep. 2018.

[39] H. Rahman, M. U. Ahmed, S. Barua, P. Funk, and S. Begum, "Driver's Cognitive Load Classification based on Eye Movement through Facial Image using Machine Learning," 2021.

[40] P. Weber, F. Rupprecht, S. Wiesen, B. Hamann, and A. Ebert, "Assessing Cognitive Load via Pupillometry,", 2020.

[41] P. Van der Wel, and H. Van Steenbergen, "Pupil dilation as an index of effort in cognitive control tasks: A review," Psychon Bull Rev, vol. 25, no. 6, pp. 2005–2015, Dec. 2018.

[42] K. Fukuda, J. A. Stern, T. B. Brown, and M. B. Russo, "Cognition, blinks, eye-movements, and pupillary movements during performance of a running memory task," Aviat. Space. Envir. Md., vol. 76, no. 7, pp. C75-C85, Jul. 2005.

[43] L. Perkhofer, O. Lehner, "Using Gaze Behavior to Measure Cognitive Load," Davis F., Riedl R., vom Brocke J., Léger PM., Randolph A. (eds) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol 29. Springer, Cham., 2019.

[44] C. Scharinger, Y. Kammerer, and P. Gerjets, "Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading," PloS one, vol. 10, no. 6, e0130608, Jun. 2015.