Modeling Cognitive Load in Mobile Human Computer Interaction Using Eye Tracking Metrics

Antony William Joseph[1(✉)], J. Sharmila Vaiz[2], and Ramaswami Murugesh[2]

[1] IT-Integrated Design, National Institute of Design, Bengaluru, Karnataka, India
willi@nid.edu

[2] Department of Computer Application, Madurai Kamaraj University, Madurai, India

Abstract. Modeling cognitive load of user interaction based on ocular parameters have become a dominant method for exploring usability evaluation of interfaces for systems and applications. Growing importance of Artificial Intelligence in Human Computer Interaction (HCI) has proposed many approaches to understand users' need and enhance human centric method for interface design. In particular, machine learning-based cognitive modeling, using eye tracking parameters have received more attention in the context of smart devices and applications. In this context, this paper aims to model the estimated cognitive load values for each user into different levels of cognition like very high, high, moderate, low, very low etc., while performing different tasks on a smart phone. The study focuses on the use behavioural measures, ocular parameters along with eight traditional machine learning classification algorithms like Decision Tree, Linear Discriminant Analysis, Random Forest, Support Vector Machine, Naïve Bayes, Neural Network, Fuzzy Rules with Weight Factor and K-Nearest Neighbor to model different levels of estimated cognitive load for each participant. The data set for modeling consisted of 250 records, 11 ocular parameters as prediction variables including age and type of task; and three types of classes (2-class, 3-class, 5-class) for classifying the estimated cognitive load for each participant. We noted that, Age, Fixation Count, Saccade Count, Saccade Rate, Average Pupil Dilation are the most important parameters contributing to modeling the estimated cognitive load levels. Further, we observed that, the Decision Tree algorithm achieved highest accuracy for classifying estimated cognitive load values into 2-class (86.8%), 3-class (74%) and 5-class (62.8%) respectively. Finally, from our study, it may be noted that, machine learning is an effective method for predicting 2-class-based (Low and High) cognitive load levels using ocular parameters. The outcome of the study also provides the fact that ageing affects users' cognitive workload while performing tasks on smartphone.

Keywords: Ocular parameters · Modeling cognitive load · Machine learning · Classification · Eye tracking metrics · Cognitive load levels · Human-computer interaction

T. Z. Ahram et al. (Eds.): AHFE 2021, LNNS 271, pp. 99–106, 2021.
https://doi.org/10.1007/978-3-030-80624-8_13

1 Introduction

The concept of mental workload is coordinated with the demand imposed by tasks on the human's limited mental resources [1] in which the demand is less than the capacity available or the demand exceeds the capacity [2]. It is not only influenced by the demand/resource but also factors like time pressure, scenario complexity, individual experience and ability [3]. If higher amount of information is delivered to a person at once, it is likely that the person will not retain it for longer duration. Therefore, it is essential to manage the cognitive load of each individual efficiently, for effective learning. Modeling cognitive load of user interaction has become a dominant method for exploring usability evaluation of interfaces of system and applications. Moreover, it can be a useful technique in development cycle of designing systems and benefits HCI [4].

Growing importance of Artificial Intelligence (AI) in the society has proposed many approaches to interpret inferences from machine learning models [5]. Lately, machine learning based cognitive modeling using eye tracking data has received more attention in the context of various smart technologies and applications. Machine learning (ML) is an area that focuses on creation of algorithms which learn on their own, based on data and experience. Since ML models learn to perform tasks by generalizing from instances, they are considered cost effective [6]. In eye tracking research, ML and classification methods were employed to study and analyze eye movement data in order to perform a classification and to discover the right category of information. Thus, this study aims to model the estimated cognitive load values for each user, into different levels of cognition using traditional ML classifiers. A set of eye tracking parameters were extracted along with user interaction behavioural measures [7] for each participant while performing five different tasks on an android mobile phone.

Next section focuses on related work that have shown different approaches followed in ML modeling using ocular parameters. Section 3 explains about research methodology, particpularly participants, materials and design of experiment. We present the experiment results in Sect. 4. research methodology we used in our study in Sect. 4. We discussed the results in Sect. 5. Finally, we draw conclusions and future work in Sect. 6.

2 Related Work

Researchers applied supervised machine learning to classify cognitive load levels [8] from early 1990s. However, most work has been focused on binary classification; and only a limited number of studies classified cognitive load into three levels predominantly in the area of emotion recognition [9] and attention. There are number of studies explored using ML algorithms and eye metrics to measure mental effort and users' attention.

Klami et al. [10] presented preliminary results on inferring the relevance of images based on implicit feedback about users' attention measured using eye tracking features with two level classifications. They collected eye data from 27 participants using Linear Discriminant Analysis (LDA) classifier and achieved performance accuracy of 91.2%. The outcome of their study suggests that it is possible to predict relevant images accurately with a simple classifier and ocular parameters. Eivazi and Bednarik [11] explored modeling problem-solving behaviour and performance levels, from visual attention data with 14 participants. They employed Support-Vector Machine (SVM) on a set of ocular parameters with 3-class and achieved predictive accuracy of 87%. Their findings confirmed that eye tracking features carries important information in predicting problem-solving behaviour and performance levels from visual attention. Li et al. [12] on the other hand used SVM to predict spatial visualization problem' difficulty level from eye tracking data. They collected data from 88 participants and achieved an accuracy of 87.60%.

Behroozi and Parnin [13] conducted a study to predict stressful technical interview settings through eye tracking saccade metric. They used Naive Bayes, Random Forest, Multi-Layer Perceptron, KNN, Logistic Regression and Decision Tree with two classes-Low and High respectively. They collected data from 11 participants and noted Random Forest (RF) algorithm as the best binary classificaiton technique to model stress and cognitive load with 92.24% accuracy.Appel et al. [14] explored predicting cognitive load in an emergency simulation based on behavioral and physiological measure with 47 right-handed participants. They modeled cognitive load using RF algorithm with 2-class (low and high respectively) and achieved performance accuracy of 72%. The outcome of their study indicated that RF prediction model is reliable for prediction of cognitive load level.

In summary, we observed that many studies employed ML algorithms to classify eye tracking data predominantly with two and three classes respectively. We noted that, existing research lacked correlation between eye parameters and ageing for estimating cognitive load levels. There was not enough attention paid to model different levels of cognitive load while performing tasks on mobile phones with smaller sample size confined to a particular age group. Thus, our study includes a wide range of participants aged between 20–60+ years, performing five different tasks on a smartphone. We modeled and analised the cognitive level classification for 2-class, 3-class and 5-classes respectively.

3 Proposed Approach

This study was designed to model different cognitive load levels for participants of different age groups while performing different tasks on a mobile phone. The study consisted of 50 participants aged between 20 years to 60+ years. Each participant was asked to perform five different tasks on an android mobile phone using Tobii Pro Glass 2eye tracking glasses. More details on participants and their selection criteria, task description, experiment design and experimental flow can be found in a paper by Joseph et al. [15].

We recorded ocular parameters for each participant and extracted relvant features from raw eye gaze data using an extraction algorithm [15]. Ocular parameters like pupil dialation for left and right eyes were used to estimate the cognitive load for each participant while performing each task using low pass filter (LPF) [15]. Thus, in total we documented 250 records with cognitive load values which were used to classify estimated cognitive load values.

In this study, we used eight traditional machine learning classifiers such as Decision Tree (DT) [16], Linear Discriminant Analysis (LDA) [17], Random Forest (RF) [18], Support Vector Machine (SVM) [19], Naïve Bayes (NB) [20], Neural Network (NN) [21], Fuzzy Rules with Weight Factor (FRWF) [22] and K-Nearest Neighbours (KNN) [23] that are commonly used in eye tracking research [24–27] to classify the estimated cognitive load values. We modelled the collected cognitive load values into two types of classification namely binary and multiclass classification. Binary classification consisted of two classes (2-class) such as Low and High. Multiclass classification, on the other hand, consisted of 3-class as Low, Moderate and High; and 5-class as Very-Low, Low, Moderate, High, Very-High respectively.

In recent studies [28, 29], behavioural measures were employed for measuring cognitive load levels. Similarly, in our study, we considered various behavioural measures such as time taken on a task, number of steps taken to complete task, task completion or failure, gaze interaction behaviour; along with ocular parameters like Fixation Count (FC), Fixation Rate (FR), Saccade Count (SC), Saccade Rate (SR), Average Fixation Duration (AFD), Standard Deviation Fixation Duration (SDFD), Maximum Fixation Duration (MFD), Average Pupil Dilation (APD) and Standard Deviation of Pupil Dilations (SDPD) [15, 30] captured for each participant to classify estimated cognitive load values into different classes as mentioned above.

4 Results

In this section, we presented a detailed report on results obtained for classifying the estimated cognitive load values into 2-class, 3-class and 5-class respectively; performance [25] of each ML algorithm to classify the estimated cognitive values; accuracy [14] of each cognitive value being reported as classified rather than being unpredicted; and features contributing to modelling the proposed classification scheme.

It may be noted from figure (Fig. 1) that, 2-class classification is reported to have less number of average unpredicted values with 51.25 when compared to 3-class with 82.5 and 5-class with 139.38 across all ML algorithms. In a 2-class classifier the average number cognitive values classified as Low were 145.75 and high were 53 respectively. From this we can conclude that, estimated cognitive load values are best classified as High or Low respectively rather than very-low, moderate and very-high. Additionally, it may be noted that, out of eight ML algorithms used, DT was considered to perform better with an average unpredicted value of 63.67 (2-class = 33; 3-class = 65; 5-class = 93). *Age* was considered as the root node to construct the decision tree for 2-class, 3-class and 5-class classification.

Further, we investigated accuracy of the defined binary and mulitclass classifiers to classify each estimated cognitive load value using each ML algorithm. Results report that, 2-class classifer achieved highest accuracy of 86.8% for correctly predicted values using DT algorithm (Table 1). Overall average percentage of accuracy for correctly predicted values for 2-class (79.59%) was found to be greater than 3-class and 5-class. DT outperformed with an average percentage of accuracy (74.53%) for correct prediction of cognitive load values (Table 1) across binary and multiclass classification.

Finally, we employed the SVM-Recursive Feature Elimination (RFE) method [24] for identifying and selecting important features that can contribute to modelling the proposed classification scheme. We noted that, for 2-class, 3-class and 5-class classifiers,

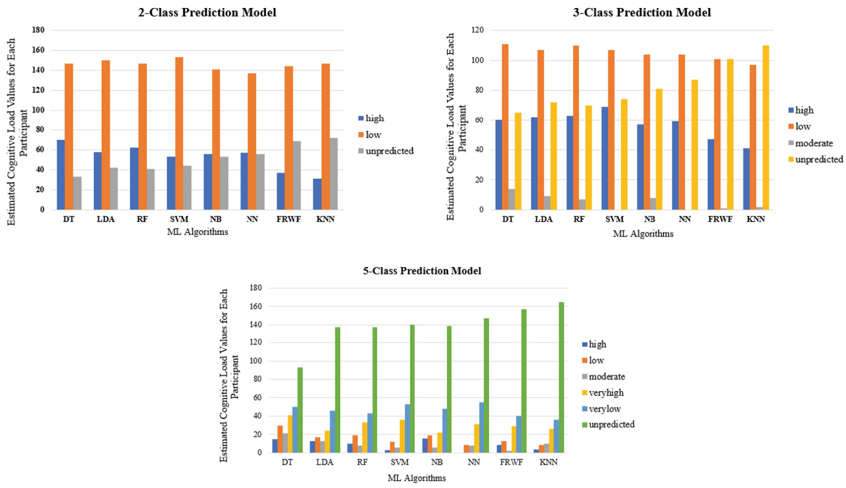

Fig. 1. Results for modelling prediction of three classes using ML algorithms.

Table 1. Accuracy of 2-class, 3-class and 5-class using different ML algorithms.

Algorithm	Accuracy		
	2-class	3-class	5-class
Decision Tree	86.8%	74%	62.8%
Linear Discriminant Analysis	84%	72.4%	45.2%
Random Forest	83.6%	71.6%	45.5%
Support Vector Machine	82.4%	71.2%	43.5%
Naive Bayes	78.8%	67.6%	43.6%
Neural Network	77.9%	65.2%	42.3%
Fuzzy Rules with Weight Factor	72%	62%	35.5%
K-Nearest Neighbours	71.2%	58.8%	33.9%

Age, FC, SC and APD were the parameters considered to contribute high accuracy for cognitive load value classification. Additionally, SR and Task were also considered as parameters that contributed to high accuracy for 3-class and 5-class respectively.

5 Discussion

Study of cognitive workload in HCI domain has gained greater importance in the recent past and considered one of the valuable sources of information for user experience today. The result of cognitive modeling in our study shows that we achieved moderately high classification performance using eye tracking metrics. Results of 8 cognitive load classification models in the present eye tracking experiment achieved moderately good classification performance in estimating cognitive load levels compared with the classification accuracy achieved using psycho-physiological methods in earlier studies. We observed a strong relationship between estimated labels and observations gained from task performance.

We achieved highest classification results in all cases using DT classification technique in the 2-class 87%, in the 3-class 74% and in the 5-class 63% cognitive load measurement. Further, the cognitive load classification results suggest that Age, FC, SC, APD, SR are the most important ocular parameters in predicting user's cognitive load level of different age groups while performing different tasks on mobile phone. It is observed that age remains as a substantial factor along with other ocular parameters influencing cognitive load. Among 8 classifiers used in our study, we noted that 4 classifiers namely, DT, LDA, RF and SVM predicted performance accuracy of 81% and above for 2-class, 71% and above for 3-class and 43% and above for 5-class prediction respectively. This supports the view that ocular parameters can determine the level of cognitive load in relation to age and type of task performed. This has proved that irrespective of type of task (simple or complex) performed by users, ageing plays an important role in determining cognitive load using ocular parameters [15, 30].

Results of this study represent real world situation like different types of tasks performed on a mobile phone, complexity of each task, luminance condition for each task, sequence of task performance and performance of tasks on different days. As discussed in the literature review, eye tracking study conducted in the past to estimate cognitive load and predict performance, had only limited number of samples. However, our study focused on a large sample size, consisting of varied age groups ranging from 20 years to 60+ years, performed five different tasks in real time on a mobile phone. Using real time task with eye movement features, establishes a new direction for eye tracking for cognitive load estimation and prediction research study.

6 Conclusion

This paper aims to model the estimated cognitive load values for each user into different levels of cognition as 2-class, 3-class, and 5-class classifiers. Our study suggests that eye tracking technology along with ML algorithms can be employed to predict users cognitive load levels with regard to age and type of tasks. The study began with the aim to measure cognitive load and investigate the relationship between mental workload and eye tracking variables. We selected eight ML algorithms to model different levels of cognition and found Decision Tree to outperform with an average accuracy of 74.53% for both binary and multiclass classifier. Additionally, we noted that a 2-class classifier is most suitable to classify the estimated cognitive load values with lesser values being unpredicted. The outcome of our prediction result shows a strong relationship between estimated cognitive load levels and observations gained from task performed on mobile phone. Dominant features contributing to prediction accuracy were Age, FC, SC and APD. Age remains as an important factor for increased cognitive load along with ocular parameters. This proves that irrespective of the type of task performed by user, age plays a substantial role in contributing to cognitive load. This is a new achievement in estimating cognitive load in the field of HCI.

Acknowledgment. The authors are thankful to the Department of Computer Application, Madurai Kamaraj University, Madurai, India and National Institute of Design, Bengaluru Campus, India for their encouragement, motivation, and relentless support in carrying out our study.

References

1. Moray, N.: Mental Workload: Its Theory and Measurement. Plenum, New York (1979)
2. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance, 3rd edn. Prentice Hall, Upper Saddle River (2000)
3. Galy, E., Cariou, M., Mélan, C.: What is the relationship between mental workload factors and cognitive load types? Int. J. Psychophysiol. **83**(3), 269–275 (2012)
4. Salvucci, D.D., Lee, F.J.: Simple cognitive modeling in a complex cognitive architecture. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 265–272 (2003)
5. Abdul, A., von der Weth, C., Kankanhalli, M., Lim, B.Y.: COGAM: measuring and moderating cognitive load in machine learning model explanations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020)
6. Domingos, P.: A few useful things to know about machine learning. Commun. ACM **55**(10), 78–87 (2012)
7. Brunken, R., Plass, J.L., Leutner, D.: Direct measurement of cognitive load in multimedia learning. Educ. Psychol. **38**(1), 53–61 (2003)
8. Gevins, A., et al.: Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum. Factors **40**(1), 79–91 (1998)
9. Chen, S.: Cognitive load measurement from eye activity: acquisition, efficacy, and real-time system design. University of New South Wales (2014)
10. Klami, A., Saunders, C., de Campos, T.E., Kaski, S.: Can relevance of images be inferred from eye movements? In: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, pp. 134–140 (2008)
11. Eivazi, S., Bednarik, R.: Predicting problem-solving behavior and performance levels from visual attention data. In: Proceedings of Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI, pp. 9–16 (2011)
12. Li, X., Younes, R., Bairaktarova, D., Guo, Q.: Predicting spatial visualization problems' difficulty level from eye-tracking data. Sensors **20**(7), 1949 (2020)
13. Behroozi, M., Parnin, C.: Can we predict stressful technical interview settings through eye-tracking? In: Proceedings of the Workshop on Eye Movements in Programming, pp. 1–5 (2018)
14. Appel, T., et al.: Predicting cognitive load in an emergency simulation based on behavioral and physiological measures. In: 2019 International Conference on Multimodal Interaction, pp. 154–163 (2019)
15. Joseph, A.W., DV, J.S., Saluja, K.P.S., Mukhopadhyay, A., Murugesh, R., Biswas, P.: Eye tracking to understand impact of aging on mobile phone applications (2021)
16. Mao, Y., He, Y., Liu, L., Chen, X.: Disease classification based on eye movement features with decision tree and random forest. Front. Neurosci. **14**, 798 (2020)
17. Salojärvi, J., Puolamäki, K., Simola, J., Kovanen, L., Kojo, I., Kaski, S.: Inferring relevance from eye movements: feature extraction. In: Workshop at NIPS 2005, in Whistler, BC, Canada, 10 December 2005, p. 45 (2005)
18. Breiman, L.: Random forests Leobreiman and Adele Cutler. Random Forests-Classification Description (2015)

19. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
20. Zhang, Z.: Naïve Bayes classification in R. Ann. Transl. Med. **4**(12) (2016)
21. Wang, S.C.: Artificial neural network. In: Wang, S.C. (ed.) Interdisciplinary Computing in java Programming, pp. 81–100. Springer, Boston (2003). https://doi.org/10.1007/978-1-4615-0377-4_5
22. Abadeh, M.S., Habibi, J., Soroush, E.: Induction of fuzzy classification systems via evolutionary ACO-based algorithms. Computer **35**, 37 (2008)
23. Guan, F., Shi, J., Ma, X., Cui, W., Wu, J.: A method of false alarm recognition based on k-nearest neighbor. In: 2017 International Conference on Dependable Systems and Their Applications (DSA), pp. 8–12. IEEE (2017)
24. Behroozi, M., Parnin, C.: Can we predict stressful technical interview settings through eye-tracking? In: Proceedings of the Workshop on Eye Movements in Programming, pp. 1–5 (2018)
25. Krol, M., Krol, M.: A novel approach to studying strategic decisions with eye-tracking and machine learning. Judgm. Decis. Mak. **12**(6), 596 (2017)
26. Richstone, L., Schwartz, M.J., Seideman, C., Cadeddu, J., Marshall, S., Kavoussi, L.R.: Eye metrics as an objective assessment of surgical skill. Ann. Surg. **252**(1), 177–182 (2010)
27. Cai, Y., Huang, H., Cai, H., Qi, Y.: A k-nearest neighbor locally search regression algorithm for short-term traffic flow forecasting. In: 2017 9th International Conference on Modelling, Identification and Control (ICMIC), pp. 624–629. IEEE (2017)
28. Pouw, W.T., Eielts, C., Van Gog, T., Zwaan, R.A., Paas, F.: Does (non-) meaningful sensori-motor engagement promote learning with animated physical systems? Mind Brain Educ. **10**(2), 91 (2016)
29. Dubé, A.K., McEwen, R.N.: Do gestures matter? The implications of using touchscreen devices in mathematics instruction. Learn. Instr. **40**, 89–98 (2015)
30. Joseph, A.W., Murugesh, R.: Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. J. Sci. Res. **64**(1) (2020)