Neuroergonomics and Cognitive Engineering, Vol. 42, 2022, 65–72
https://doi.org/10.54941/ahfe1001820

Classifying Mental Workload Using EEG Data: A Machine Learning Approach

Şeniz Harputlu Aksu, and Erman Çakıt

Department of Industrial Engineering, Gazi University, Ankara 06570, Turkey

ABSTRACT

The objectives of this study were two-fold: (1) to investigate the relationship among electroencephalography (EEG) features, task difficulty levels and subjective self-assessment (NASA-Task Load Index (TLX)) scores and (2) to develop machine learning algorithms for classifying mental workload using EEG features. Seventy EEG features (5 frequency band power for 14 channels) were selected as independent variables. One output variable reflecting the difficulty level of n-back memory task was classified. Prefrontal and frontal theta, prefrontal beta-high, occipital, parietal and temporal gamma and occipital alpha activities were found to be the most effective parameters. The results obtained for the four classes of classification problem reached the accuracy of ~68% with Random Forest (RF) algorithm. In addition, maximum accuracy of ~87% was reached with 2-class-based (low and high mental workload) estimation model along with Gradient Boosting Machines (GBM) algorithm. The results from the analysis indicate that EEG signals play an important role in the classification of mental workload. Another remarkable result was high classification performance of GBM, LightGBM and extreme gradient boosting (XGBoost) algorithms that have been developed in the recent past and therefore not frequently used in studies on this subject in the literature.

Keywords: Mental workload, EEG, Machine learning, Classification

INTRODUCTION

Mental workload analysis helps to recognize the mental fatigue, evaluate the human performance of different level tasks and adjust cognitive sources for safe and efficient human-machine interactions. Excessive levels of mental workload can lead to errors or delays in information processing. Physiological techniques have frequently used to investigate mental workload during human-machine interaction because they are objective, do not depend on what the participant perceives, and can be used in real time thanks to continuous signals. Especially, monitoring brain activity has been verified to be sensitive and consistent reflector of mental workload changes. In recent years, with the need to analyze continuous and large-scale data obtained by physiological methods, the use of machine learning algorithms has become widespread in estimating and classifying mental workload.

Within the scope of this study, it was aimed to investigate whether the change in mental workload during n-back memory tasks is related to brain waves, to compare it with the results obtained in the literature before, and to develop a model estimating mental workload based on machine learning algorithms by using EEG data. In the literature, the studies where this topic is addressed as a classification problem show that the problem becomes more difficult as the number of class increases and therefore the performance of the model is reduced (Grimes et al., 2008; Borys et al., 2017). Therefore, it was tried to obtain a model that predicts the mental workload with the highest possible accuracy according to four difficulty levels. Unlike other studies; GBM, LightGBM and XGBoost algorithms which are not used in the literature especially in the estimation of mental workload by EEG method were included in the study as well as traditional machine learning techniques. Model performances were examined and compared with other algorithms. The models were corrected by determining the best hyper-parameter values and the results were compared for all algorithms.

This study focuses on the classification rate of mental workload level based on EEG features and the use of relatively new machine learning algorithms for this target. In addition, it is aimed to shed light on further studies with the important findings of the study considering as a pilot application.

MENTAL WORKLOAD AND EEG

Among various neurophysiological markers, EEG has received much attention because of its ease for implementation with a portable measure and high temporal resolution (Yin & Zhang, 2017). EEG permits an objective workload assessment and can provide real-time evaluation, thus allowing the system designer to quickly and accurately identify usability problems as they occur. In general, the power spectrum analyses have been mostly applied in existing experimental studies with EEG measurement (Choi et al., 2018). Power spectrums are created and analyzed by looking at the amplitudes of the different frequency ranges. In particular, alpha and theta activity has been confirmed to be effective in discriminating mental workload levels (Zammouri et al., 2018). Experimental results show that the averaged accuracy of distinguishing changes in the theta [4–7 Hz] (θ) band is 79%. For the alpha band [8–11 Hz] (α) the averaged accuracy reached 78%. In a study, it was observed that the theta power in frontal brain area increases while the alpha power in the parietal and occipital sites decreases under high mental workload condition (Holm et al., 2009). The same authors reported that the task difficulty level was positively related to the frontal theta/parietal alpha ratio. There are also studies showing that central, parietal and occipital beta power is associated with changes in mental workload (Yin & Zhang, 2017; Plechawska-Wojcik et al., 2019), and also increased power of the high-frequency beta and gamma waves in posterior cortex in response to changing task performance (Chuang et al., 2012).

Traditional machine learning techniques applied to the problem of cognitive workload classification are k nearest neighbor (kNN), support vector machines (SVM), artificial neural networks (ANN) and random forest (RF). Borghetti et al. (2017) presented a model that maps neurological observations (EEG) to the operator's mental workload. For the study, the RF algorithm was chosen. Wilson et al. (2003) performed two-class classification based on

ANN, achieving 86% of accuracy. ANN were also used in another study (Zarjam et al., 2015), where the authors classified seven levels of cognitive workload with features extracted on the basis of wavelet entropy, achieving 83% classification accuracy. Plechawska-Wojcik et al. (2019) performed a three-class classification of cognitive workload based on EEG spectral data, achieving ~91% of accuracy with kNN algorithm. In another three-class classification, Borys et al. (2017) also reached 73% accuracy with kNN algorithm based on eye tracking features. In the same study, based on only EEG features, maximum accuracy of ~51% could be reached. Grimes et al. (2008) reported 99% classification accuracy for two classes and 88% for four classes (both results achieved for eight subjects). The study of Grimes et al. (2008) shed light on this study in terms of showing that high classification performance can be achieved with a small number of participants.

ANALYSIS OF MENTAL WORKLOAD

Methodology

In this study, Millisecond software which is a program that includes cognitive, social and neurophysiological online experiments was used. In the experiments, n-back tasks, which are commonly used in the literature, were applied (Ke et al., 2015; Liu et al., 2017; Tjolleng et al., 2017). N-back memory tests were performed at 4 different difficulty levels. As the number of "n" increases, the difficulty of the task increases. Participants are asked to keep the letters that appear on the screen for a certain period of time in their memory, and to press the "A" key when they see the letter M in the 0-back task, if not, not to press any keys. In the 1-back condition, the participant was asked to press the A key if the letter on the screen is the same as the previous one, otherwise not to press any key. In the 2-back condition, a letter was the match if it was shown two screens back. In the 3-back condition, a letter was the match if it was shown three screens back.

Four participants (2 female, 2 male) took part in the experiment. The mean age of the people examined was 36 years. They were not under pharmacological treatment. During the experiment, EEG waves were recorded with EMOTIV EPOC X device. The resolution of the device is 14 bits and the sampling rate is 128 Hz. The EPOC X device has 14 channels, and the sensor placements according to the international 10-20 system (Homan et al, 1987) are given in Figure 1. Recordings were started after checking contact quality and EEG quality level.

Each session included 12 n-back blocks, 3 from each condition. The block order was applied in such a way that the difficulty level was random. Each letter appears on the screen for 0.5 seconds and it takes 2.5 seconds until the next letter appears on the screen. After each block, participants were asked to subjectively evaluate the mental workload they experienced while performing the tests, using the NASA-TLX scale. Therefore, a participant filled the NASA-TLX scale 12 times in total, once at the end of each block. In addition, the participants were asked to prioritize the 6 sub-dimensions of NASA-TLX through the pairwise comparison tables given to them. Weighted total NASA-TLX scores were obtained by using the weights obtained

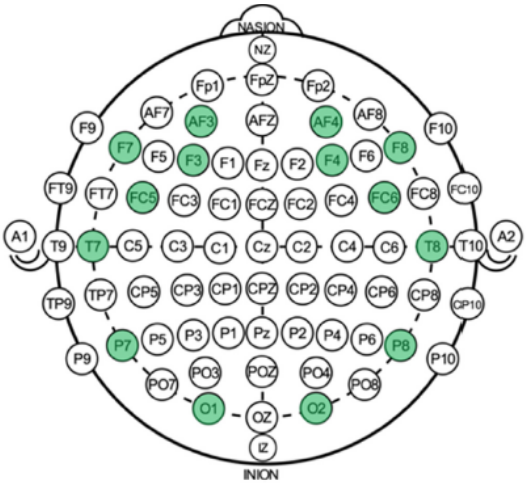

Figure 1: EMOTIV's electrode placement (Sareen et al., 2020). (Emotive electrodes are highlighted in green color over the 10-20 electrode system.)

as a result of 15 pairwise comparisons. In this work, we extracted for each EEG channel the band powers of 4–8, 8–12, 12–16, 16-25 and 25–45 Hz bandwidths by using EMOTIV Pro software. This was performed at a single stimulus level, forming a feature vector of 5 bands $\times$ 14 channels = 70 length for each of the 3 blocks $\times$ (15+16+17+18) stimuli = 198 sample epochs for each subject.

Results

Statistical analysis was performed using SPSS 21.0 software. Outliers in the data set consisting of 792 observations were removed from the data set, and the analysis was continued with 785 observations. The Kolmogorov-Smirnov test found all features with non-normal distribution, so the non-parametric statistical methods were applied. Spearman correlation tests were used to investigate whether there was a relationship between EEG variables and NASA-TLX and task difficulty levels. As a result of the correlation tests, it was seen that the weighted NASA-TLX score and the level of difficulty were strongly positively correlated (rho: 0.84, $p < 0.01$). It was determined that the EEG variables showing the highest correlation with the task difficulty level were the theta waves received from the AF3, F7, FC5, FC6, F8 and AF4 channels (rho > 0.3, $p < 0.01$). These results show that theta power in the prefrontal, frontal and frontal center regions increases as the task difficulty increases. It is also observed that high beta (betaH) power in the prefrontal and frontal regions is significantly negatively correlated with task difficulty. Correlation tests also demonstrated that the EEG variables showing the highest correlation with the weighted NASA-TLX total score were theta power in the F8 region, betaH power in the AF3 region, and gamma power in the O1, O2, P7, F8 regions (rho > 0.3, $p < 0.01$). As the perceived mental workload increases, frontal theta, frontal gamma, parietal gamma and occipital gamma increase. In contrast to theta power, betaH power in the prefrontal

Table 1. Accuracies of classification models.

Model	Accuracy for 4-class problem	Accuracy for 2-class problem
kNN	0.50	0.78
SVM	0.63	0.84
ANN	0.64	0.84
RF	0.68	0.84
XGBoost	0.66	0.84
GBM	0.64	0.87
LightGBM	0.63	0.85

area decreases as perceived mental workload increases. With Kruskal Wallis test, it was examined whether there was a significant difference between task difficulty levels in terms of EEG variables. For 31 EEG features, statistically significant difference was determined in at least one group distribution with the significance level of 1%. It was also revealed that 23 EEG features differed at 99.9% significance level according to difficulty levels. When the rank values of the variables were examined, AF3-theta, AF3-betaH, F7-theta, F3-betaH, FC5-theta, O2-gamma, P8-gamma, FC6-theta, F8-theta, AF4-theta variables were observed to increase or decrease smoothly according to the task difficulty changes. In the observations corresponding to the high task difficulty level, it was observed that the theta power in the frontal and frontal central line was higher, the high beta power in the same regions was lower, and the gamma power in the parietal and occipital areas was higher.

In the stage of constructing the model estimating mental workload level by using EEG data, Python 3.8 software was used. Machine learning algorithms were applied to obtain the classification model that determines task difficulty level most accurately according to the EEG features. For modeling, kNN, SVM, ANN, RF, GBM, LightGBM and XGBoost algorithms were used. Accuracy was calculated for each model in which 70 EEG features were considered as independent variables and "Task Difficulty Level" consisting of 4 classes as dependent variable. The algorithm that gave the best result was the RF with an accuracy rate of 68% (see Table 1). In other words, the class of 68% of all observations in the test data was correctly predicted. In the tuning step of the RF model, different hyper-parameters were tried to achieve the best result. When the number of maximum features is 4, minimum samples for split is 2 and the number of trees is 500, 68% success has been achieved. When the confusion matrix was examined, it was seen that the misclassification estimation occurred mostly between 0-1 and 2-3 classes. A total of 12 errors were made in the 0-1 class predictions made by the model, and a total of 22 errors were made in the 2-3 class predictions. In order to investigate how reducing complexity of the problem with fewer classes would affect the results, the observations in the 0-1 class and the 2-3 class were combined and the models were retested. In the 2-classes prediction model, in which the mental workload can be evaluated as low and high, the accuracy of 87% was reached with GBM algorithm (see Table 1). The model correctly predicted the classes of 171 of 197 randomly selected observations as the test data set.

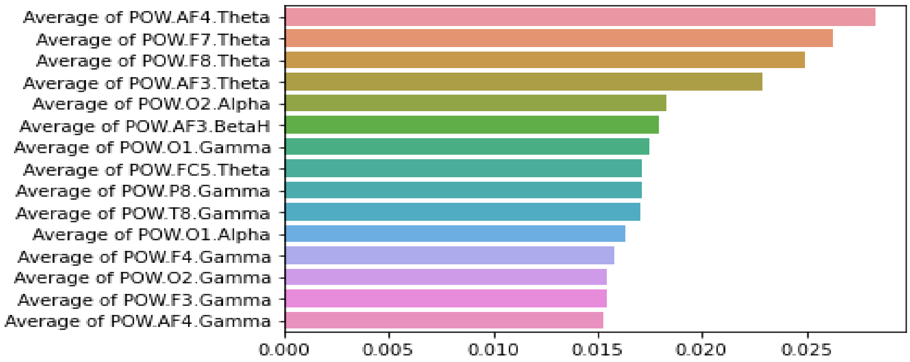

Figure 2: Importance levels of EEG features (x: importance scores, y: features).

When the feature selection method embedded in the RF algorithm was applied, the variable priorities were listed as in Figure 2.

In this figure, the most important 15 EEG features were given according to RF algorithm. Prefrontal, frontal and center frontal theta, prefrontal beta-high, frontal, temporal, parietal and occipital gamma and occipital alpha activities were found to be the most important parameters contributing to modeling the estimated mental workload levels.

CONCLUSION

In this paper, we used n-back task inducing four different levels of workload to investigate workload discrimination using EEG signals. The outcome of the study provides the fact that increased task difficulty determines the rise of theta activity in prefrontal and frontal regions. Especially frontal theta power and parietal and occipital gamma power were found positively related to perceived workload scores obtained via NASA-TLX. Prefrontal beta-high activity is, on the other hand, negatively related to self-assessment workload ratings. The Kruskal-Wallis analysis of ranks indicated the statistically significant difference in at least one group distribution for 31 EEG features with the significance level of 1%. The Random Forest algorithm achieved the highest accuracy (~68%) for 4-class classification problem. We observed that, prefrontal and frontal theta, prefrontal beta-high, occipital, parietal and temporal gamma and occipital alpha activities are the most important parameters contributing to model performance. Further, maximum accuracy of ~87% was reached with 2-class-based (low and high mental workload) estimation model along with GBM algorithm. From our study, it may be noted that, EEG signals alone play an important role in estimation of mental workload. Another remarkable result is high classification performance of GBM, LightGBM and XGBoost algorithms, which have been developed in the recent past and therefore not frequently used in studies on this subject in the literature. This result supports the use of these algorithms in studies to be conducted in this field. However, it is considered that further improvement can be achieved both with the increase in the number of observations and with detailed analysis to be made by testing different feature and hyper-parameter subsets.

ACKNOWLEDGMENT

This work was funded by Gazi University Scientific Research Projects Unit under Grant No. 6483.

REFERENCES

Borghetti, B.J., Giametta, J.J., and Rusnock, C.F. (2017), "Assessing Continuous Operator Workload with a Hybrid Scaffolded Neuroergonomic Modeling Approach." *Human Factors*, 59(1), 134–146.

Borys M., Plechawska-Wojcik M., Wawrzyk M., and Wesołowska K. (2017), "Classifying Cognitive Workload Using Eye Activity and EEG Features in Arithmetic Tasks." doi: 10.1007/978-3-319-67642-5_8.

Choi, M. K., Lee, S. M., Ha, J. S., and Seong, P. H. (2018), "Development of an EEG-based workload measurement method in nuclear power plants." *Annals of Nuclear Energy*, 111, 595–607.

Chuang S.W., Ko L.W., Lin Y.P., Huang R.S., Jung T.P., and Lin C.T. (2012), "Co-modulatory spectral changes in independent brain processes are correlated with task performance." *Neuroimage*, 62(3), 1469–77.

Grimes, D., Tan, D., Hudson, S., Shenoy, P., and Rao, R. (2008). "Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph." *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, 835–844.

Holm, A., Lukander, K., Korpela, J., Sallinen, M., and Müller, K.M.I. (2009), "Estimating brain load from the EEG." *The Scientific World Journal*, 9, 639–651, doi: 10.1100/tsw.2009.83.

Homan, R.W., Herman, J., and Purdy, P. (1987), "Cerebral location of international 10–20 system electrode placement." *Electroencephalography and clinical neurophysiology*, 66, 376–382.

Ke Y., Qi H., Zhang L., Chen S., Jiao X., Zhou P., Zhao X., Wan B., and Ming D. (2015), "Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression." *Int. J. Psychophysiol.*, 98 (2), 157–166.

Liu Y., Ayaz H. , and Shewokis P. A. (2017), "Multisubject "Learning" for Mental Workload Classification Using Concurrent EEG, fNIRS, and Physiological Measures." *Frontiers in Human Neuroscience*, 11(389).

Plechawska-Wojcik, M., Tokovarov, M., Kaczorowska, M., and Zapała, D. (2019), "A Three-Class Classification of Cognitive Workload Based on EEG Spectral Data." *Applied Sciences*, 9, doi: 5340. 10.3390/app9245340.

Sareen, E., Singh, L., Varkey, B., Achary, K., and Gupta, A. (2020), "EEG dataset of individuals with intellectual and developmental disorder and healthy controls under rest and music stimuli." *Data in Brief*, 30.

Tjolleng A., Jung K., Hong W., Lee W., Lee B., You H., Son H., and Park S. (2017), "Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals." *Applied Ergonomics*, 59, 326–332.

Wilson, G.F., Russell, C.A. (2003), "Operator functional state classification using multiple psychophysiological features in an air traffic control task." *Hum. Factors J. Hum. Factors Erg.*, 45, 381–389.

Yin, Z., and Zhang, J. (2017), "Cross-session classification of mental workload levels using EEG and an adaptive deep learning model." *Biomedical Signal Processing and Control*, 33, 30–47.

Zammouri, A., Abdelaziz, A.M., and Mebrouk, Y. (2018), "Brain-computer interface for workload estimation: Assessment of mental efforts in learning processes." *Expert Systems with Applications*, 112.

Zarjam, P., Epps, J., and Lovell, N. (2015), "Beyond Subjective Self-Rating: EEG Signal Classification of Cognitive Workload." *IEEE Transactions on Autonomous Mental Development*, 7, doi: 10.1109/TAMD.2015.2441960.