Classification of driver's cognitive workload for the transfer of control of Autonomous Vehicles

URS BUSCHAUER, Saarland University
PRIYANKA UPADHYAY, Saarland University
JUHI GUPTA, Saarland University

Additional Key Words and Phrases: driver cognitive load, machine learning classification, EEG dataset

1 INTRODUCTION AND MOTIVATION

The previous study on traffic safety imputes that the main reason for road crashes are due to human/driver error, ranging from 75% to 95% of the total cases. Among the reasons, the state of distraction and inattention play an important role in the drivers' errors that lead to road crashes. So to limit the negative consequences of distracting activities, one can use an electroencephalogram (EEG) to measure the driver's brain activity which defines driver's cognitive/mental workload. By knowing the driver being distracted (when driver's cognitive workload is elevated/high), an autonomous system could, e.g., offer to take over the control, when driver can not handle the situation, or estimate when the driver has the cognitive capacities (have low cognitive workload) to take the control back from the vehicle.

Passive Brain–computer interfaces (BCIs) provide information from user mental activity to a computerized application without the need for the user to control their brain activity. Hence, BCIs can evaluate users' states, e.g., their cognitive or affective states, from their brain signals, and use these estimations to adapt a human-computer interaction system accordingly [6]. Hence, we aim to put forward the best algorithm to detect mental workload from EEG signals. This report describes our research groups' efforts in tackling the mental workload prediction task from Electroencephalogram (EEG) signals.

2 DATASET

To achieve our aim, We used a publicly available passive BCI dataset [4] which provides EEG signals across multiple sessions for each user. As part of the dataset, for each subject EEG signals across three sessions were given. For 15 subjects the EEG signals were collected via 62 electrodes and then sampled at 500 Hz. For three independent experimental sessions each seven days apart from each other, the participants were invited to the lab. The signals were captured for two states which were the testing and the resting state. For one minute subjects rested with their eyes open and EEG was recorded during resting state. Epoch length was two seconds. Subjects performed a Multi-Attribute Task Battery-II (MATB-II) task [5]. The MATB-II is a computer-based task designed to evaluate operator performance and workload which here is used in the testing state where different workload levels were offered in a random manner. To put together the data, the following pre-processing steps were implemented to the data: epoching, high and low pass filtering by using finite impulse response (FIR) filter, referencing, and electrode rejection. Moreover, they also downsampled the signal to 250 Hz.

2.1 Levels of cognitive workload

The MATB-II developed by NASA is a task that elicits levels of mental and cognitive workload. It assesses task-switching and mental workload capacities. Several sub-tasks were performed simultaneously. Difficulty level differs depending on the condition and the number of sub-tasks.

For label 0, the easy condition, participants performed the TRACK and SYSTEM MONITORING task. The participant had to manage a target at the middle of a window and the task for monitoring required monitoring of 2 warning lights and 4 gauges.

For label 1, the medium condition, an additional task was added. A fuel management system was introduced to the participant for the RESOURCE MANAGEMENT to maintain a fuel level by activating and deactivating a set of pumps.

For the last label 2, the difficult condition, the task for COMMUNICATION was added to the three previous tasks: the participant responded to radio messages by modifying the frequencies of different radios. Moreover, in difficult condition, the tracking task was made tougher.

3 EXPERIMENT SETUP

In order to predict the cognitive load level, we use **Logistic Regression and Support Vector Machine classifier**. We consider two main models for Logistic Regression : (1) Classification using per participants, (2) Classification using all participants. And considered one model for SVM: Classification using all participants. So together we performed three experiments. To gather information on how important certain brain areas are for the prediction of cognitive load levels we did electrode selection.

3.1 Electrode selection

Electrode selection can be used to select a subset out of all electrodes of the pre processed data which is then used for training the model. We have an EEG dataset and understanding how these different electrode settings in the EEG data influence the classification is important. We choose multiple different feature sets (figure 5) from the given total electrode setting.

Based on the topography showcased in the Impact of Secondary Tasks Paper[2] and left part of figure 1 and 3 (where we can see all the electrodes placing to calculate brain EEG), we constructed multiple electrode groups. The topography shows the Wavelet-power-Delta (WPD), which is the difference in wavelet-power when comparing low and high cognitive load. For our groups electrodes with high changes in activity comparing low to high cognitive load were the most important. For the comparison, we had a total of 8 electrode groups.

Two Basic groups containing 13 frontal and all total electrodes, which showcase the standard approaches. Using frontal electrodes is based on the increased active thinking, happening in the prefrontal cortex, during high cognitive load, we used the electrode group given in the code example since it displays the standard approach. Using all electrodes is commonly used to simply utilize all available/collected Data in order to achieve accurate estimations.

Three Combination groups based on WPD. A 30 electrode group contained all Electrodes that had a Bandpower difference of more than 10% (Delta value above 0.1 or below -0.1). And the group of Electrodes that had a Bandpower increase of more than 10% (only Delta values above 0.1) was split up into an 11 Electrode group including all frontal Electrodes and a 12 electrode group including all non-frontal Electrodes.

Three Singular groups again based on the WPD. A 7 Electrode group only containing Electrodes that had a Bandpower decrease of more than 10%. A 14 Electrode group only containing electrodes Electrodes that had a Bandpower increase of more than 15%. And a 23 Electrode group only containing electrodes Electrodes that had a Bandpower increase of more than 10%.

3.2 Class/labels settings

We set up few experiments with different Classification problems using different class settings: (1) 3 Class-Classification using [easy, med, diff] class (2) 2 Class-Classifications using [easy, med] class and (3) 2 Class-Classifications using [easy, diff] class.

4 IMPLEMENTATIONS

Every brain has its own characteristics and differences therefore the common approach is to train the model on the data of its future user. We have performed experiments using two different approaches:
Approach 1-Using the pre-processed data of each electrode directly/Pre-Processed Data Only: Herefore we extracted the time series data of each epoch from the .set files and annotated them based on their cognitive load level.
Approach 2-Calculating the specific Bandpower for each frequency-range: In general we have four frequency-range: Beta[12-30Hz] , Alpha[8-12Hz], Theta[4-8Hz] and Delta[0-4Hz] for each electrode. Herefore we used the time series data of each epoch from the .set Files to calculate the Bandpower for all 4 Frequency ranges and save them together with the annotation based on their cognitive load level.
Experiment 1 and 2 (for logistic regression) have used both approaches to perform the experiments where as SVM only used approach 1:

4.1 Experiment 1: Logistic Regression Classifier using per Participant

Idea of Experiment 1: Training Machine Learning models in the field of EEG is usually done per participant. In our 5-fold cross validation for the electrode groups we already achieved high accuracy using this method. For each electrode setting the specific electrodes data was loaded and used to calculate a 5-fold cross validation accuracy over the set of each one session for all 15 participants. Model training and validation was done on each participants first session and testing was done on independent second session.

4.2 Experiment 2: Logistic Regression Classifier using all participants

Idea of Experiment 2: Find common features which can classify cognitive load automatically across all individuals. This way the model will learn multiple drivers' physiologically features to perform well in real-time scenarios where one driver's physiologically response to an elevated cognitive load could be the combination of multiple driver's response. We used 5-fold cross validation method here as well. Both sessions of participants 1 to 10 used for training and and validation and testing was done on both sessions of participants 11 to 15.

In both experiment implementation: for Approach 1 while perform model training and validation all the different class settings: [easy, med, diff], [easy, med] and [easy, diff] were used and we calculated validation accuracy thought it whereas only "3 class setting" [easy, med, diff] has been used to test the model and to get the test accuracy (because this class combination is most important for testing the model to analyze the performance). In Approach 2 we used "3 class setting" [easy, med, diff] to perform training and validation and same class setting to test the model.

4.3 Experiment 3: Support vector machine (SVM) Classifier using all participants

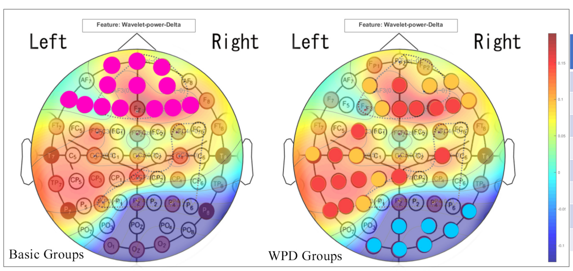
Idea of Experiment 3: This approach demonstrates SVM for multi-class classification(easy,med,diff) using the one-vs-one method for a general model type across all participants. The One-vs-One strategy splits a multi-class classification into one binary classification problem per each pair of classes.

5 RESULTS AND EVALUATION

We will discuss all experiments results into this section and discuss the evaluation.

5.1 Results and Evaluation from Experiment 1

For **Approach 1**: From figure 2 we can say that we achieved the **highest accuracy** with **Feature set5** and **lowest accuracy** with **Feature set7**. In the contrast of cross validation results in figure 1 (right part) which is not correlated with the number of electrodes, it seems like certain electrodes are influenced less by the changes of the new situation/session therefore containing more information about the actual cognitive load level and less contextual noise.



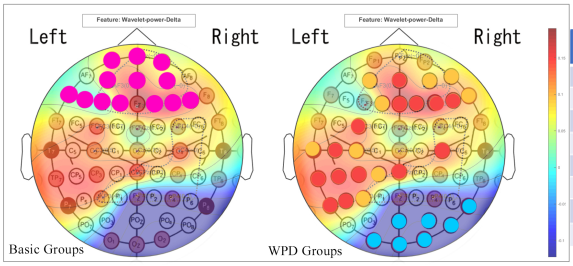
Electrodes Features Setting	Approach 1 3 Class [easy, med, diff]	Approach 1 2 Class [easy, med]	Approach 1 2 Class [easy, diff]	Approach 2 3 Class [easy, med, diff]
Feature set0 (13)	72.34%	83.04%	85.96%	48.16%
Feature set1 (62)	92.56%	95.53%	97.43%	60.24%
Feature set2 (30)	90.99%	94.36%	96.98%	57.66%
Feature set3 (11) (Frontal)	74.64%	85.52%	87.36%	49.74%
Feature set4 (12) (Non-Frontal)	79.73%	88.80%	90.90%	49.86%
Feature set5 (7)	72.38%	80.85%	88.63%	50.20%
Feature set6 (14)	80.84%	89.49%	92.54%	51.05%
Feature set7 (23)	87.71%	93.33%	94.89%	54.09%

Fig. 1. (a) Left: Electrodes placing to calculate brain EEG, (b) Right: Validation accuracy of Approach 1 with the different electrode groups (classification per participants)

Electrodes Features Setting		Approach 1 3 Class [easy, med, diff]	Approach 2 3 Class [easy, med, diff]
Feature set0 (13)		36.84%	43.01%
Feature set1 (62)		42.36%	45.04%
Feature set2 (30)		37.72%	43.67%
Feature set3 (11)	(Frontal)	36.99%	41.89%
Feature set4 (12)	(Non-Frontal)	33.41%	39.07%
Feature set5 (7)		42.76%	43.01%
Feature set6 (14)		38.37%	41.37%
Feature set7 (23)		36.18%	42.00%
Feature set minimal (2) [Fz,POz]		41.98%	47.38%

Fig. 2. Testing accuracy of Approach 1 and Approach 2 with the different electrode groups (classification per participants)

For **Approach 2**: From figure 2 we can say that we achieved the **highest accuracy** with **Feature set minimal** and **lowest accuracy** with **Feature set4**. In comparison to Approach 1, the accuracy are higher in general. Since only the Bandpower feature is extracted most of the contextual noise is not influencing the training of the model. The High performance of the **Feature set minimal** could be due to the two electrodes being in the center of frontal and back electrodes and might therefore be additionally less susceptible for contextual noise.



Electrodes Features Setting	Approach 1 3 Class [easy, med, diff]	Approach 1 2 Class [easy, med]	Approach 1 2 Class [easy, diff]	Approach 2 3 Class [easy, med, diff]
Feature set0 (13)	83.15%	92.44%	97.48%	40.94%
Feature set1 (62)	99.56%	99.35%	99.18%	42.84%
Feature set2 (30)	98.32%	95.80%	99.16%	43.51%
Feature set3 (11) (Frontal)	81.01%	92.44%	96.64%	41.54%
Feature set4 (12) (Non-Frontal)	85.96%	84.10%	98.32%	36.62%
Feature set5 (7)	71.51%	74.79%	87.50%	43.15%
Feature set6 (14)	91.10%	89.92%	98.23%	41.14%
Feature set7 (23)	96.11%	94.12%	99.14%	41.14

Fig. 3. (a) Left: Electrodes placing to calculate brain EEG, (b) Right: Validation accuracy of Approach 1 with the different electrode groups (classification using all participants)

5.2 Results and Evaluation from Experiment 2

For **Approach 1**: If we look at figure 4, we can say that we achieved the **highest accuracy** with **Feature set minimal** and **lowest accuracy** with **Feature set4**. In contrast to the cross validation result in figure 3(right part) testing accuracy does not seem to be correlated with the number of electrodes, it seems like certain electrodes are influenced less by the changes of the new situation/session, therefore containing more information about the actual cognitive load level and less contextual noise.

For **Approach 2**: from figure 4 we can say that we achieved the **highest accuracy** with **Feature set0** and **lowest accuracy** with **Feature set4**. In comparison to Approach 1 the accuracies are higher in general. Since only the Bandpower feature is extracted most of the contextual noise is not influencing the model training. The High Performance of the **Feature set0** is mainly since the prefrontal cortex has a higher activity during high cognitive load, which is exactly what the Bandpower feature is tracking.

Electrodes Features Setting	Approach 1 3 Class [easy, med, diff]	Approach 2 3 Class [easy, med, diff]
Feature set0 (13)	37.76%	44.56%
Feature set1 (62)	38.75%	42.63%
Feature set2 (30)	40.76%	40.89%
Feature set3 (11) (Frontal)	38.79%	42.37%
Feature set4 (12) (Non-Frontal)	36.47%	35.97%
Feature set5 (7)	37.63%	38.75%
Feature set6 (14)	40.09%	41.75%
Feature set7 (23)	39.19%	41.52%
Feature set minimal (2) [Fz,POz]	43.87%	42.73%

Fig. 4. Testing accuracy of Approach 1 and Approach 2 with the different electrode groups (classification using all participants)

As overall we can take some more observation from both experiments: Focusing on the 2 setups suggested in the task description [easy, med, diff] and [easy, diff] they give us important information on the importance of certain electrodes. As the results shown in figure 1 **set3 to set5** all outperformed the standard approach set0 while using less Electrodes. Whereas in figure 3 **set3 and set5** both outperformed the standard approach set0 while using less Electrodes. This indicates that when **"choosing a subset of electrodes the wavelet-power-delta is a good indicator on their**

importance". Since no electrode group under performed in comparison to the set0 standard approach all of them have been used in both experiments.

The low testing results in **figure 2** and **figure 4** show that the data contained a lot of contextual noise. Since the "Experiment 1: classification per participant model" was trained and validated only on data of the first session it has not been able to maintain its performance, when applied on the completely distinct second session. The "Experiment 2: classification using all participants model" or "generalized model" was trained and validated on both sessions of Participants 1 to 10 and was tested on both Sessions of Participants 11 to 15. It has also not been able to maintain its cross-validation performance.

When comparing both testing results the generalization approach (experiment 2) achieved similar results which indicates that it could be an actual competitor of the commonly used per participant approach. In order to reach higher accuracies a combination of more training data and more generalizing features like the Bandpower feature would be our approach for further development.

5.3 Results and Evaluation from Experiment 3

For experiment 3 with Approach 1, we achieved maximum of **50%** of accuracy for Support vector machine (SVM) classifier with Radial Basis Function (RBF) kernel from different configurations for the model covering all the participants. While with Linear kernel setting, the accuracy obtained is **34%**.

6 OVERVIEW OF THE ENTIRE SYSTEM

The driver distraction mitigation system which can be developed using our proposed techniques/strategies can diminish overload caused by drivers engaging in distracting tasks and to guide driver behavior to maintain safety. The driver monitors the driving environment and adopts appropriate actions to keep driving safe. The system collects the information about vehicle kinematics and driving environment. On the system identifying driver distraction, provides mitigation if the driver is distracted. The driver monitors the performance of the system and makes a decision of using it.

7 CONCLUSION

In this project, we have seen that machine learning techniques can be applied with driver EEG data to provide the elevated cognitive workload. We also reported results and evaluation of all experiments to show specific approach we followed. We can say that the classification results depend on multiple factors: Learning Algorithm, Amount of training data, Bias/unbias training data. Feature selection is always the key for any machine learning algorithm: Electrodes settings, choosing a subset of electrodes the wavelet-power-delta is a good indicator. Poorly designed in vehicle interfaces can lead to risky and fatal driving. Our project and experiments were motivated by the need of method to classify drivers' elevated cognitive workload to provide safe driving. Thus we can conclude that better accuracy can help us to achieve: better recognition of elevated cognitive load periods, can improve adaptive human interface performance and can help in the transfer of control of autonomous vehicles.

REFERENCES

[1] Tango, Fabio and Botta, Marco. (2013). Real-Time Detection System of Driver Distraction Using Machine Learning. Intelligent Transportation Systems, IEEE Transactions on. 14. 894-905. 10.1109/TITS.2013.2247760.

- **Feature set0 (13 electrodes)**

  ['F7',’F5’,'F3','F1','F2','F4','F6','AF3','AFz','AF4','FP1','FP2','FPz’]

- **Feature set1 (all 62 electrodes)**

- **Feature set2 (30 electrodes)**
  ['AF3','AFz','AF4','AF8','F1','Fz','F2','F4','F6','F8','FC3','T7','C5','C3','C4','TP7','CP5','CP3','CP1','CPz','P7','P5','P3','P8','POz','PO4','PO8','O1','OZ','O2']

- **Feature set3 (11 electrodes):**
  ['AF3','AFz','AF4','AF8','F1','Fz','F2','F4','F6','F8','FC3']

- **Feature set4 (12 electrodes):**
  ['T7','C5','C3','C4','TP7','CP5','CP3','CP1','CPz','P7','P5','P3']

- **Feature set5 (7 electrodes):** ['P8','POz','PO4','PO8','O1','OZ','O2’]

- **Feature set6 (14 electrodes)**
  ['AFz','Fz','F2','F4','F6','FC3','C5','C4','TP7','CP5','CP3','CPz','P7','P5’]

- **Feature set7 (23 electrodes):**
  ['AF3','AFz','AF4','AF8','F1','Fz','F2','F4','F6','F8','FC3','T7','C5','C3','C4','TP7','CP5','CP3','CP1','CPz','P7','P5','P3']

Fig. 5. Different feature sets used in Model training

[2] The Impact of Secondary Tasks on Drivers During Naturalistic Driving: Analysis of EEG Dynamics . Vahid Alizadeh and Omid Dehzangi, University of Michigan, Dearborn, USA, November 2016

[3] Compute the average bandpower of an EEG signal. Raphael Vallat. (2022) UC Berkeley https://raphaelvallat.com/bandpower.html (Accessed: February 2022)

[4] Dataset: https://zenodo.org/record/5055046.YbNzGr3MJPb

[5] MATB-II: Revised Multi-Attribute Task Battery: https://matb.larc.nasa.gov/

[6] Zander, T. O., Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. Journal of neural engineering, 8(2), 025005