# Assessing the Effects of Various Physiological Signal Modalities on Predicting Different Human Cognitive States

Ayca Aygun[1,*], Thuan Nguyen[2], Matthias Scheutz[1]

**1** Department of Computer Science, Tufts University, Medford, MA, USA
**2** Department of Electrical & Computer Engineering, Worcester Polytechnic Institute, Worcester, MA, USA

* ayca.aygun@tufts.edu

## Abstract

Robust estimation of systemic human cognitive states is critical for a variety of applications, from simply detecting inefficiencies in task assignments, to the adaptation of artificial agents' behaviors to improve team performance in mixed-initiative human-machine teams. This study showed that human eye gaze, in particular, the percentage change in pupil size (PCPS), is the most reliable biomarker for assessing three human cognitive states including workload, sense of urgency, and mind wandering compared to electroencephalogram (EEG), functional near-infrared spectroscopy (fNIRS), respiration, and skin conductance. We used comprehensive multi-modal driving dataset to examine the accuracy of signals to assess these cognitive states. We performed comprehensive statistical tests to validate the performance of several physiological signals to determine human cognitive states and demonstrated that PCPS shows noticeably superior performance. We also characterized the link between workload and sense of urgency with eye gaze and observed that consecutive occurrences of higher sense of urgency were prone to increase overall workload. Finally, we trained five machine learning (ML) models and showed that four of them had similar accuracy in cognitive state classification (with one, random forest, showing inferior performance). The results provided evidence that the PCPS is a reliable physiological marker for cognitive state estimation.

## Introduction

A reliable determination of human systemtic cognitive states, including *cognitive workload*, *sense of urgency*, and *mind wandering*, would be helpful in many task-based settings, and even critical in some (e.g., detecting overloaded humans in safety-critical contexts who will be more prone to errors and oversight due to being overloaded, or detecting mind wandering in situations where human observers might miss important events due to attentional drifts). Being able to assess human systemic cognitive states can also be of great utility for artificial agents operating in human-agent teams in that it would allow artificial agents to adapt their behavior and the teams' task allocation to achieve a better load balance and increase the effectiveness of the whole team (rather than acting in ways that might, for example, increase human workload, see [11]). Similarly, it would be important in some contexts to track mind wandering, defined as the detachment of human attention from the ongoing operation, to prevent the decline in situational awareness of human performers [12]. Moreover, information about one's

sense of urgency could be used as a social signal causing the subject to pay attention to social cues in a human-machine interaction settings [13] and help artificial agents to better understand the needs of their human teammates.

In this paper, we investigated the extent to which the above three systemic human cognitive states—cognitive workload, sense of urgency, and mind wandering—could be inferred from physiological signals including eye gaze, EEG, fNIRS, skin conductance, and respiration rate. Specifically, we aimed to develop computational models which, when trained on a small set of human physiological data to predict systemic cognitive states, could be used to predict those states in general for different people. To develop our models and validate their performance, we utilized a dataset collected from an interactive multi-modal multi-task driving experiment presented in [14].

The two main contributions of this paper are thus as follows:

- We investigated the effects of various physiological signal modalities including pupillometry, EEG, fNIRS, skin conductance, and respiration for determining **(1)** workload, **(2)** sense of urgency, and **(3)** mind wandering through various statistical analyzes, determining the efficacy of using eye gaze signals, particularly, the PCPS, for assessing different levels of workload, sense of urgency, and mind wandering.

- We applied five state-of-the-art machine learning (ML) models: (1) $k$-Nearest Neighbor (k-NN), (2) Naive-Bayes (NB), (3) Random Forest (RF), (4) Support Vector Machines (SVM), and (5) Multiple Layer Perceptron (MLP) to evaluate the effectiveness of multiple physiological signal types for inferring different levels of the three cognitive state levels, showing again that the best model prediction results can be obtained with PCPS.

The remainder of this paper was structured as follows. We started with our motivation for considering the above three human systemic cognitive states as well as a brief description of the data we used for our analysis and modeling efforts, followed by a review of related work on using different sensory modalities to detect them. We then provided details on the employed data and how it was used, including information on data selection and preprocessing for each of the sensory modalities. The statistical analyses were then provided with the performance results of the different machine learning models. The discussion section put the implications of the results in the broader human-machine teaming context, also discussing limitations and future work. Finally, we concluded with the summary of our accomplishments and points to the utility of collecting and processing eye gaze for assessing human performance.

# Motivation

Human cognitive states such as workload, sense of urgency, but also mind wandering can modulate individual task performance and thus impact the performance of mixed initiative teams. To improve team performance, it is essential to successfully predict these individual cognitive states and ideally utilize the assessment to adapt interactions and team settings (e.g., task assignment, team processes, etc., see also [3]). For example, it might be important to reduce the thoughts of team members unrelated to the ongoing task (called *experimental noise* by [15]).

Moreover, it is also important to understand possible interactions among different systemic cognitive states as systemic cognitive states do not occur in isolation but can thus influence each other. Figure S1 Fig illustrates hypothesized relationships among workload, sense of urgency, and mind wandering.

**Fig 1. The relationship between Different Cognitive State Pairs.** An illustration of the hypothesized relationships among pairs of three systemic cognitive states: workload, sense of urgency, and mind wandering.

**Workload** has been described as a human-centered concept which arises from the interactions between task requirements, the environmental conditions under which the task is accomplished, the abilities, the attitudes, and the awareness of the person who operates the task [16]. Workload has also been associated with the reflection of cognitive demands enforced on human operators by the task requirements [17]. In particular, workload was considered as a multidimensional perception, experienced by an operator, which is moderated by the abilities and the achievements of the operator.

**Sense of urgency** can be defined as a perception that a given task must be accomplished quickly, usually within a certain period of time, which affects decision-making mechanisms in humans. For example, Yau *et al.* considered time-varying and evidence-free signals as a sense of urgency, and found that it is possible to determine EEG signals which are relevant to sense of urgency [18]. Another study examined the impacts of warning signals, including auditory, vibrotactile, and, multimodal, in assessing sense of urgency in different cognitive workload conditions [19].

**Mind wandering** is defined as a particular case of spontaneous thought, a cognitive state or a series of multiple cognitive states that proceed comparably freely due to a lack of intense restrictions on the subject matter of any mental state or on the transition between any of the two subsequent mental states ( [15]). Specifically, mind wandering was considered as a special case of spontaneous thought that is less-intentionally restricted than creative thought and objective-directed thinking but more-intentionally restricted than dreaming. Another study ( [20]) associated mind wandering with the independence of experiences from impressions and continuing activities, and highlighted the terms "task unrelated" [21] and "stimulus independent" [22] which represent the thoughts happen during mind wandering occurrences.

**Workload vs. mind wandering.** There have been research efforts focused on understanding the relationship between workload and mind wandering. Zhang *et al.* suggested that there is a negative correlation between mental workload and mind wandering during driving. Specifically, the participants who have lower cognitive load have additional mental capacity which may be assigned to other tasks irrelevant to the primary task and cause higher mind wandering [23]. Another study claimed that any individual has a fixed mental capacity, and thus, increased perceptual load would lead to diminished available mental capacity for task-irrelevant activities [24]. This is consistent with our hypothesis that higher mind wandering is prone to decrease with increased mental load caused by performing multiple activities at a time.

**Sense of Urgency vs. mind wandering.** Latinjak *et al.* proposed a model to examine the relationship between mind wandering and goal-assisted, spontaneous thinking [26]. In particular, the author(s) designed a competition task, settled in a laboratory and performed by athletes, to investigate the link between spontaneous thoughts of the participants and task-unrelated mental activity, and found out that there is a negative correlation between goal-assisted thinking and mind wandering. Considering that, by nature, mind wandering develops when individual's attention immediately switches from the current task to inner thoughts [5], it is acceptable to

hypothesize that higher mind wandering is negatively correlated with higher sense of urgency which is compatible with what was proposed in [26].

**Workload vs. sense of urgency.** In contrast to the relationship between workload and mind wandering, and sense of urgency and mind wandering, the relationship between workload and sense of urgency is not entirely clear. Even though there are studies (see Section ) which examined the correlation between workload and sense of urgency, those research efforts were mainly based on subjective sense of urgency ratings and do not provide moment-by-moment analyses of the impact of sense of urgency occurrences on overall workload. In this paper, we thus aimed to explore the effects of consecutive instantaneous task occurrences on overall workload. Intuitively, we may consider three cases to examine how workload is associated with sense of urgency occurrences. First, there may be occasional sense of urgency moments which might not affect overall task-dependent workload. Second, we may consider cases where tasks triggering sense of urgency, *e.g.*, such as braking events in a driving task, occur frequently and would thus potentially lead to increased overall workload as well. Third, even if we observe higher workload after frequent sense of urgency moments, we cannot conclusively claim that the increased overall workload is solely a result of the task-dependent frequent sense of urgency occurrences as other task-irrelevant activities might have contributed to the overall workload as well. Our future efforts will further investigate the association between workload and sense of urgency which is essential for the robust development of human-machine teaming setup.

**Experimental Data.** We utilized the data from a comprehensive interactive multi-modal multi-task driving experiment collected in [14] to validate the capability of eye gaze, particularly the percentage change in pupil size (PCPS), in assessing different human cognitive states including workload, sense of urgency, and mind wandering. We accessed the data in January 15 2024 for research purposes. Specifically, in [14], participants were instructed to perform accident-free driving in a driving simulator while completing different secondary tasks. Participants had to complete two sessions one after another with a short break time between them. For convenience, in this study, we denote the first part of the experiment in [14] as "Session 1" and the second part as "Session 2". For both sessions, participants were asked to drive on a four-lane highway (two-lane per direction) with a speed limit of 65 mph. During the first three minutes of each session, the only task was driving which helped the participants get ready for the rest of the experiment where other tasks were added. Specifically, participants were instructed to perform three additional tasks which we discussed two of them (used in this study) as follows:

1. Participants engaged in brief communication tasks at different times during the simulation. The communication task included 40 questions in total (20 questions per session) which included a series of "yes/no" and explanation dialogue events. Participants were instructed to respond to the questions approximately every 30 to 60 seconds during each session [14].

2. Secondly, the participants completed braking events that occurred ten times per session during both sessions. In the braking event, the front vehicle first came 200 m into the sight of the driver. Next, the driver reached that front vehicle until it was 75 m away and then followed it for 20 seconds at a constant distance of 75 m. Then, the front vehicle immediately decelerated for five seconds with activated brake lights. Finally, the front car accelerated again and moved away from the vehicle. Figure S2 Fig illustrates an example of a braking event.

**Fig 2. An illustration of a braking event.** An example of a braking event: (a) a participant drives on the right lane of a straight road, (b) a car appears in front of the participant's car, and (c) the front car immediately decelerates, then moves far away.

We considered multiple physiological signals including pupillometry, electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), skin conductance, respiration during the experiment [14, 45] to explore their effectiveness in assessing different types of human cognitive states such as workload, sense of urgency, and mind wandering.

# Related Work

In this section, we will review past research on using eye gaze, EEG, fNRIS, skin conductance, and respiration for assessing cognitive workload, sense of urgency and mind wandering. As will be clear from the summaries, there is no uniform approach to this in the literature. For one, there is no single experimental study that assesses all three systemic cognitive states. Nor is there a study that assesses any one single systemic cognitive state using all five physiological measures. Consequently, the resulting findings are difficult to compare across different experimental paradigms and often not consistent.In contrast, our subsequent analysis allows use to assess the potential for inferring all three systemic cognitive states using all five phyisological measures, thus providing the first of its kind comprehensive analysis and machine learning effort for isolating the most promising physiological measures for systemtic cognitive state inference in a consistent manner.

## Eye Gaze

### Eye gaze and workload

There have been several studies that consider different eye gaze features as an estimator of cognitive workload. One study recorded eye gaze and EEG from fifteen students in a virtual learning platform for the distinction of cognitive workload levels in two separate groups: the students who were able to complete the tasks successfully and the students who were not [27]. They suggested the k-Nearest Neighbor (kNN) as the best predictor over other classification methods. In contrast, we used state-of-the-art classification methods and also utilized neural network-based deep learning models to validate the efficiency of eye gaze in estimating workload. Another study investigated the efficiency of saccadic intrusion, which is considered as a specific type of fixational eye movement, as well as the pupil diameter in assessing cognitive workload [28]. Although their results demonstrated that pupil diameter has better accuracy in predicting workload compared to the saccadic intrusion, they did not consider the subject-based variations of pupil diameter, while our study aimed to eliminate subject-based discrepancy by leveraging the percentage change in pupil size (PCPS) rather than the pupil diameter itself. Bodala *et al.* explored the efficacy of saccadic gaze movements of sixteen subjects in determining cognitive workload while performing visual tasks [29]. Even though this study demonstrated a correlation between an increased workload and the mean peak saccadic velocity, they validated the performance of their technique with an inadequate number of participants. In our study, we used eye gaze features collected from a larger number of participants with 46 subjects to verify our classification models.

### Eye gaze and sense of urgency

There are only a few research papers that explore the efficiency of different physiological signals for the determination of the sense of urgency. Khalaf *et al.* proposed a method to estimate the participants' challenge and threat states when they proceed with three mental arithmetic tasks by leveraging different physiological signal types including electrocardiogram (ECG), continuous blood pressure, respiration, impedance cardiogram, and facial electromyography (EMG) [30]. Salkevicius *et al.* studied the assessment of the anxiety levels of the participants by leveraging blood volume pressure, galvanic skin response, and skin temperature [31]. To the best of our knowledge, there is no research effort that focuses on understanding the dynamics of human gaze parameters in estimating the sense of urgency levels.

### Eye gaze and mind wandering

There have been relatively few studies that investigate the efficiency of the human gaze in assessing mind wandering. Biller *et al.* [32] examined the combination of eye gaze and physiological features to predict mind wandering when the participants were reading instructional texts via the computer interface. Although this study explored mind wandering during a driving task, its simulation environment is too sophisticated for evaluating mind wandering since the drivers (participants) must perform too many tasks at the same time such as watching the traffic lights, arranging the speed, and paying attention to the other vehicles and pedestrians. He *et al.* [33] suggested that the participants showed few deficits in vehicle manipulation during mind wandering. Although this study examined the efficiency of eye gaze parameters including percentage gaze dwell time spent in the side mirrors and mean of the standard deviation of horizontal gaze position collected from eighteen subjects in estimating mind wandering, their dataset has an inadequate number of participants which might cause their model to overfit. Klesel *et al.* [34] proposed a study to investigate mind wandering in a virtual reality environment by using eye gaze features including fixation coin and fixation duration. Although this study introduced the procedures to examine the performance of eye gaze in assessing mind wandering, they did not provide any experimental finding which validates the efficiency of the human gaze in detecting mind wandering.

## EEG

### EEG and workload

EEG has been one of the most common physiological markers in predicting human cognitive workload [8,35–38]. In [37], the authors explored the effects of different frontal lobe frequency bands on assessing working memory workload and observed that decreased delta power was substantially linked to increased workload. Another study defined a workload index, representing the difficulty levels, by leveraging theta and alpha power spectra of EEG and examined the impact of different power bands on multiple areas. In particular, they found out that increased workload was associated with decreased alpha band power and increased theta band power in parietal areas and in prefrontal areas, respectively [38]. In [9], Berka *et al.* investigated the link between task engagement and human workload while completing learning and memory tasks by using EEG signals acquired from eighty healthy participants. Another study first utilized independent component analysis (ICA) to collect the independent features from EEG streams, then estimated the cognitive workload based on these features [39]. Despite we can determine EEG signals directly from the electrical exertion of cortical and sub-cortical neurons with higher temporal resolution, usually in milliseconds, the

EEG streams are non-stationary and sensitive to various types of noise including blinking, frequency intrusion, and sensor-based interference associated with motion [40].

### EEG and sense of urgency

One study investigated the proficiency of EEG and fNIRS in assessing sense of urgency levels by designing a dynamic decision making setup based on a stimulation of the participants by neutral to happy or sad faces. They found out that EEG is able to distinguish between sense of urgency levels which varies person to person [18]. Another research effort explored the effect of time pressure, in particular long-term (9 seconds) and short-term (5 seconds), on cognitive response by using mental arithmetic tasks while acquiring EEG signals [41]. They observed increased arousal levels under higher time pressure by using EEG signals.

### EEG and mind wandering

There have been a few studies which explored the efficacy of EEG in assessing mind wandering. In [42], Jin *et al.* developed ML classifiers to examine the state of mind wandering (in particular, either mind-wandering or on-task) by using EEG markers and they were able to separate two mind wandering levels with an average accuracy of 60%. Another work analyzed the association between mind wandering and online education by using the spectral power of EEG signals and event-related potential (ERP) waveforms [43]. Dong *et al.* proposed a method to differentiate attention states of the participants for both person-dependent and person-independent situations by leveraging ERPs acquired from the EEG signals and concluded that either non-linear or linear ML models can be used to determine mind wandering with the AUC level of 0.715 [44].

## fNIRS

### fNIRS and workload

There have been studies which focus on understanding the dynamics and capabilities of fNIRS signal in assessing of different workload levels. In [46], the authors proposed a method which combines the bi-variate functional brain connectivity (FBC) characteristics of three EEG frequency bands (delta, theta, and alpha) and two fNIRS features (oxyhemoglobin and deoxyhemoglobin) to assess multiple mental workload levels. They concluded that EEG and fNIRS have superior performance in different regions of the brain, e.g., fNIRS performed better in the right frontal region (AF8). Midha *et al.* explored the effectiveness of fNIRS signals, acquired from 20 healthy subjects (8 females and 12 males), in differentiating various human cognitive levels in the case of common office-like tasks including different levels of reading and writing tasks, and observed higher oxygenated haemoglobin (O2Hb) and lower deoxygenated haemoglobin (HHb) in the case of harder reading tasks [47]. In [48], Putze *et al.* examined the performance of fNIRS signals in automatic cognitive workload identification in a virtual reality (VR) setting and suggested that fNIRS signals were capable of detecting mental workload in immensely engaging VR tasks.

### fNIRS and sense of urgency

Holtzer *et al.* explored the impact of gender and perceived stress, alone or combined, in pre-frontal cortex (PFC) oxygenated hemoglobin ($HBO_2$) variation in the course of walking and observed that perceived stress impacted the task cost, which was specifically observed by $HBO_2$ levels of older men [49]. Mirelman *et al.* investigated the fluctuations in brain activity during the primary walking task while performing secondary cognitive

tasks by using fNIRS and showed that $HBO_2$ is capable of differentiating different stress levels including only walking or walking and counting together [50].

### fNIRS and mind wandering

There are relatively few research efforts focused on assessing mind wandering using fNIRS signals. One study investigated different ML and deep learning classifiers such as deep neural networks (DNNs), convolution neural networks (CNNs), and XGBoost to identify different mind wandering levels without interrupting the incoming task [52]. Another study analyzed the participants' mind wandering levels by exploring when the participants leave the primary task in the Sustained Attention to Response Task (SART) [53]. They acquired sixteen-channel fNIRS data from frontal cortices and observed increased activity in the area of the medial prefrontal cortex during the mind wandering episodes.

## Skin Conductance

### Skin conductance and workload

In [54], the authors analyzed the efficacy of the frequency domain features of skin conductance in evaluating mental workload. They modeled a mental task which included memorizing target letters and identifying them from distributed alphabet, and concluded that the frequency domain characteristics of skin conductance were capable of detecting small changes in cognitive workload. Lew *et al.* proposed a method to determine mental workload from the combination of skin conductance and pupil diameter by using wavelets and genetic programming (GP) and found out that GP had better performance in detecting task performance and difficulty compared to traditional linear discriminant analysis (LDA) [55]. In [56], *et al.* introduced a technique to examine the effectiveness of skin conductance, heart rate, and respiration, acquired from 121 young adults, in predicting mental workload in a simulated driving setting and found out that skin conductance was increased with higher task demand.

### Skin conductance and sense of urgency

Kurniawan *et al.* explored how to measure the change in galvanic skin response (GSR) during acute stress conditions by using multiple classification techniques and found out that in the case of three different workload settings (recovery,light, and heavy), SVM had the highest classification accuracy among four classification methods in differentiating different stress levels [57]. Bakker *et al.* examined GSR data, collected in real-world settings, in assessing stress and argued that the GSR data is ambiguous and further information is needed to make a confident determination of stress levels in addition to GSR signals [58].

### Skin conductance and mind wandering

We have found comparatively few studies which considered skin conductance as an indicator of different mind wandering levels. Blanchard et. al used the fusion of skin conductance and skin temperature to determine the degree of mind wandering by training various supervised classification methods [59]. Bixler et. al combined eye gaze, skin conductance, and skin temperature along with the contextual information including task difficulty and time on task to assess automatic multimodal detection of mind wandering [32].

## Respiration

### Respiration and workload

One study examined the association between multiple respiration features (such as inductance plethysmography and pressure of end-tidal carbon dioxide) and different workload levels in simulated flight setting [60]. Specifically, the authors used the NASA-TLX setup and collected respiration signals from seven pilots during one-session simulation test. They found out that ventilation has higher values in all tasks compared to the baseline. Another study analyzed the performance of multiple physiological signal modalities including heart rate, blood pressure, respiration, and human gaze in assessing mental workload. The authors concluded that the respiratory cycle duration was sensitive to the fluctuations in cognitive workload [61].

### Respiration and sense of urgency

McDuff *et al.* designed a model to capture cognitive stress via a digital camera by leveraging different physiological measures including breathing rate and heart rate variability, and found out that respiration rate increases with increased stress level [62]. Another research study investigated the effects of stress on multiple physiological signal indicators such as skin temperature, facial temperature, and respiration rate, and demonstrated that respiration rate rises with higher stress levels [63].

### Respiration and mind wandering

In [64], the authors introduced a method which utilized the temporal synchronization between the respiration signals and the fingertip force pulses, acquired from 12 participants, to measure an objective index to determine mind wandering and suggested that the proposed objective index is a reliable indicator in detecting mind wandering. In [65], Hasenkamp et. al defined four cognitive state intervals including mind wandering, awareness of mind wandering, shifting of attention, and sustained attention, and explored the performance of respiration rate in monitoring these states and found out that respiration rate had no influence in separating mind wandering levels.

## Workload vs. Sense of Urgency

There have been a few studies that attempted to characterize the relationship between cognitive workload and the sense of urgency. Wei *et al.* investigated the subjective mental workload of twenty-two residents while entering prescriptions in a computerized physician order entry system by the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) setup and suggested that workload tends to increase with increased sense of urgency level [66]. Keunecke *et al.* related the overall workload to emergency and non-emergency medical transport based on subjective sense of urgency ratings and proposed that workload increases with a higher sense of urgency rating [67]. However, there has been no research effort that investigated the sense of urgency and cognitive workload with the help of eye gaze. Hence, we specifically explored the differentiation of workload and sense of urgency in this study with regard to time duration by using the percentage change in pupil size (PCPS) as an indicator and demonstrated that consecutive occurrences of a higher sense of urgency lead to an increase in the overall workload.

# Methods

In this section, we first introduce the definitions of the three systematic cognitive together with the different levels we aim to distinguish using the physiological data. Then we review the comprehensive set of different physiological signals that we used to train our computational models and to predict the systematic states.

## Levels of the Systemic Cognitive States

We define different levels for all three systemic cognitive states: three levels for cognitive workload, two levels for sense of urgency and also two levels for mind wandering based on the experimental manipulations that were intended to increase or decrease those levels (see [10] for details).

### Workload Levels

We defined three cognitive workload levels: $\mathrm{l_{w_1}}, \mathrm{l_{w_2}}, \mathrm{l_{w_3}}$ shown in Figure S3 Fig. Specifically, $\mathrm{l_{w_1}}$ and $\mathrm{l_{w_2}}$ start with the $1^{\mathrm{st}}$ and the $5^{\mathrm{th}}$ braking events, respectively, and include four braking and six communication events each. In other words, we defined $\mathrm{l_{w_1}}$ and $\mathrm{l_{w_2}}$ by taking different segments during the driving simulation as each of them includes multiple consecutive braking events and dialogue tasks. Besides, $\mathrm{l_{w_3}}$ represents the initial period of the driving simulation before the start of the first braking event. Then, we produced a balanced dataset having a total of 276 samples (92 samples per cognitive workload level) taken from both of the sessions of 46 subjects.

**Fig 3. Evaluation of three cognitive workload levels.** By selecting different combinations of braking and dialogue events, three different cognitive workload levels were generated.

### Sense of Urgency Levels

We defined two sense of urgency levels: $\mathrm{l_{sou_1}}$ and $\mathrm{l_{sou_2}}$. Specifically, we used the first 2.5 seconds of the $1^{st}$ braking event to generate $\mathrm{l_{sou_2}}$. Then, we generated $\mathrm{l_{sou_1}}$ by taking a 2.5 seconds time frame, obtained from 30 seconds before the $1^{st}$ braking event, to assure that it does not include any secondary events except from the primary driving task. The intuition behind this is that there are circumstantial conditions or task demands which require the participants to react immediately and thus generate higher sense of urgency such as completing a dialogue task on time or sudden braking while performing a driving simulation [10].

### Mind Wandering Levels

We defined mind two wandering levels based on the first 1.5 minutes and the next second 1.5 minutes of the simulation, the idea being that during the first 1.5 minutes, the participants were attempting to acclimate to the driving environment which requires more attention, resulting in a low level of mind wandering which we called $\mathrm{l_{mw_1}}$. On the other hand, during the next 1.5 minutes, the participants were already familiar with the driving task, in addition, the external road and traffic circumstances were at a minimum level with no task requirement (*i.e.*, no communications, no braking demands) other than simply driving on the right side of a straight highway, thus, the less attention is required which potentially evokes higher level of mind wandering (we called high mind wandering level as $\mathrm{l_{mw_2}}$.).

**Table 1.** Notations

| Term | Description |
|---|---|
| PCPS | The percentage change in pupil size |
| $PCPS_i$ | The PCPS of the $i^{th}$ sample |
| APCPS | Average PCPS |
| FixNum | The number of fixations |
| FixDur | The fixation duration |
| BlinkNum | The number of blinks |
| BlinkDur | The blink duration |
| CM | The current measure of pupil diameter |
| BM | The baseline measure of pupil diameter |
| N | The total number of samples in time domain |
| BR | Breathing rate |
| $BR_{coeff}$ | The coefficient of variation of BR |
| IBI | The inter-breath interval |
| DC | Duty cycle |
| $DC_{coeff}$ | The coefficient of variation of DC |
| IT | The inspiration time |
| ET | The expiration period |
| IE | The ratio between sequential IT and ET |

### Workload vs Sense of Urgency

In this study, we explored how consecutive occurrences of the sense of urgency affect the overall workload. In particular, we considered the average PCPS (APCPS) signal during level $l_{w_1}$ of session 1 which included multiple subsequent braking tasks and a couple of dialogue interactions. Then, we annotated the onsets of dialogue and braking events and explored the impacts of secondary tasks on the APCPS signal. In particular, we explored the spikes of the signal at the onset points of braking and dialogue events during level $l_{w_1}$ of session 1 and hypothesized that we observe higher spikes at the onset points of the braking and dialogue tasks which causes increased cognitive workload during level $l_{w_1}$ of session 1.

## Physiological Signal Modalities

We evaluated five physiological signal types—pupillometry, EEG, fNIRS, skin conductance, and respiration—to assess the three different systemic cognitive states. Table 2 illustrates the characteristics of the signals used in this study including the sampling rate, the pre-processing steps, and the morphological features which we acquired for the evaluation of the human cognitive states. We summarized all the abbreviations used for these signals in Table 1.

### Pupillometry

The pupillometry signal was recorded with a sampling rate of 400 Hz. To calculate the pupil diameter, we used the left pupillometry signal considering that the left and the

**Table 2.** Summary of the employed physiological signal modalities including the sampling rate, the pre-processing steps, and the extracted morphological features.

| Signal Modality | Sampling Rate (Hz) | Pre-processing Steps | Morphological Features |
|---|---|---|---|
| Pupillometry | 400 | **1-** Amplitude thresholding between 0.8 mm – 10 mm<br>**2-** Linear interpolation<br>**3-** $5^{th}$ order Butterworth low-pass filter (LPF) with cutoff frequency of 10 Hz | **1-** Percentage change in pupil size (PCPS)<br>**2-** Average PCPS (APCPS)<br>**3-** Fixation number (per second)<br>**4-** Mean fixation duration<br>**5-** Blink number (per second)<br>**6-** Mean blink duration |
| EEG | 500 | **1-** $6^{th}$ order Butterworth band-pass filter (BPF) between 0.1 Hz - 10 Hz<br>**2-** Independent Component Analysis (ICA)<br>**3-** Kalman Smoother | **1-** Delta<br>**2-** Theta<br>**3-** Alpha<br>**4-** Beta<br>**5-** Gamma |
| fNIRS | 20 | **1-** p-chip interpolation with a sampling rate of 20 Hz<br>**2-** Linear-wise de-trending to find regions in which the variance of the signal is between $25^{th}$ percentile and $75^{th}$ percentile<br>**3-** High-pass filtering with a cutoff frequency of 1.7 Hz<br>**4-** Thresholding to eliminate the channels with the median variance above 1 uM | **1-** The mean of the cleaned channels<br>**2-** The variance of the cleaned channels |
| Skin Conductance | 20 | - | **1-** Mean skin conductance |
| Respiration | 20 | **1-** Upsampling the respiration signal from 20 Hz to 250 Hz<br>**2-** $5^{th}$ order Butterworth low-pass filter (LPF) with cutoff frequency of 1 Hz<br>**3-** Obtaining the local minimas and the local maximas for candidate inhale peaks and exhale troughs<br>**4-** Time thresholding of 0.6 seconds between consecutive inhale peaks and exhale troughs<br>**5-** Amplitude thresholding of 0.2 z-score between consecutive inhale peaks and exhale troughs | **1-** Breathing rate<br>**2-** Coefficient of variation of breathing rate<br>**3-** Inter-breath interval which is the period in seconds between two sequential peaks within the tidal volume waveform (TVW) signal<br>**4-** Duty cycle<br>**5-** Coefficient of variation of duty cycle<br>**6-** Inspiration time which is the period between a trough and peak within the TVW signal (in seconds)<br>**7-** Expiration period which is the time interval between a peak and trough within the TVW signal (in seconds)<br>**8-** I:E which is the ratio between the consecutive inspiration time and the expiration period |

right pupil diameters are synchronous. Next, we applied a three-step pre-processing technique to remove any out-of-band sensory noise and the blink artifact. Particularly, we first utilized amplitude thresholding to eliminate the signal segments below 0.8 mm and above 10 mm by assuming that any values lower than 0.8 mm are potential blink artifacts [68] and the measurable pupil dilation only expands up to at most 10 mm [69]. Secondly, we used linear interpolation to fix the extracted parts [68]. Thirdly, we applied a fifth-order Butterworth low-pass filter with a cutoff frequency of 10 Hz to remove baseline wander [70].

To eliminate the subject-based variations in pupil diameter, we calculated the percentage change in pupil size (PCPS) utilizing the following equation [71]:

$$\text{PCPS} = \frac{\text{CM} - \text{BM}}{\text{BM}} \times 100, \tag{1}$$

where CM and BM represent the current measure of pupil diameter and the baseline measure of diameter, respectively. Here, BM is determined by calculating the mean of a 10-seconds signal before the stimulus. Additionally, we calculated the average PCPS (APCPS) as follows [71]:

$$\text{APCPS} = \frac{1}{N} \sum_{i=1}^{N} \text{PCPS}_i, \tag{2}$$

where $\text{PCPS}_i$ denotes the percentage change in pupil size at $i^{th}$ sample and $N$ is the total number of samples in the time domain.

### Electroencephalography

We used three-step pre-processing to denoise the raw EEG signals which were recorded via eight channels with a sampling rate of 500 *Hz*. First, we applied sixth-order Butterworth band-pass filter between 0.1 *Hz* and 30 *Hz* to remove the out-of-band noise [72]. Then, we decomposed the combination of signal epochs to its statistically independent components by leveraging the common technique called independent component analysis (ICA). To manually remove the ICA components which are correlated with blink artifacts, we identified the blink artifacts with the instantaneous spikes in the amplitude. Finally, we applied a Kalman Smoother which is a well-known tool to assess the state of the dynamic linear structures in the presence of noise [73]. We used the Python library called "Tsmoothie" [74] to smooth the EEG signals.

To statistically analyze the performance of EEG on estimating different human cognitive states, we performed a power spectral density analysis (PSD) and calculated the power distributions of the EEG signals in the frequency domain. Specifically, we obtained the average power of EEG signal from five frequency bands: 1 *Hz* to 4 *Hz* (delta), 4 *Hz* to 8 *Hz* (theta), 8 *Hz* to 12 *Hz* (alpha), 12 *Hz* to 30 *Hz* (beta), 30 *Hz* to 120 *Hz* (gamma), respectively.

### Functional near-infrared spectroscopy

We utilized the cleaning process in [10] to clean the raw fNIRS signal. Next, the mean and variance of fNIRS are extracted which will be used as inputs for learning models in the next sessions.

### Skin Conductance

Using the skin conductance signal sampled at a rate of 20 Hz, we obtained one well-known and commonly used feature, the mean skin conductance (in microSiemens (muS)), from the raw skin conductance signal.

### Respiration

We applied five-step pre-processing to denoise the respiration signal which were recorded with a sampling rate of 20 Hz. First, we up-sampled the signal from 20 Hz to 250 Hz. Second, we applied fifth-order Butterworth low-pass filter with a cutoff frequency of 1 *Hz*. Third, we calculated the local minimas and the local maximas to obtain the candidate inhale peaks and exhale troughs. Fourth, we applied time thresholding of 0.6 seconds between the sequential inhale peaks and exhale troughs. Finally, we used amplitude thresholding of 0.2 z-score between the sequential inhale peaks and exhale troughs to acquire the final set of respiration beats. In this study, we collected eight morphological features from the respiration signal [51, 75] which we summarized as follows:

1. **The breathing rate (BR)** (in 1/sec) which represents the time between the inhale onsets.
2. **The coefficient of variation of the breathing rate ($\mathbf{BR_{coeff}}$)** (in 1/sec) which is proportion of the standard deviation of the difference between inhale onsets over the average difference between inhale onsets.
3. **The inter-breath interval (IBI)** (in sec) which represents the average time between the inhale onsets.
4. **The duty cycle (DC)** which represents the proportion of breath that is inhaled calculated as the proportion of the average inhale duration over the average inter-breath interval.
5. **The coefficient of variation of duty cycle ($\mathbf{DC_{coeff}}$)** which is the proportion of the standard deviation of inhale duration over the average inhale duration.
6. **The I:E (IE)** which represents the ratio between the sequential inspiration time and the expiration period.
7. **The inspiration time (IT)** (in sec) which is the time duration between a trough and a peak within the tidal volume waveform (TVW) signal.
8. **The expiration period (ET)** (in sec) which is the time duration between a peak and a trough within the TVW signal.

## Statistical Analyses of Systemic Cognitive State Levels

We performed comprehensive statistical analyses using the five physiological signal modalities to determine the best modalities for detecting different levels of cognitive workload, sense of urgency, and mind wandering. Specifically, we used two ANOVA statistical tests including the "Student–Newman–Keuls" (SNK) test [76] and Tukey's "honestly significant difference" (HSD) multiple pairwise comparison test [77] to investigate the variations of different physiological indicators for three workload levels ($\mathrm{l_{w_1}}, \mathrm{l_{w_2}}$, and $\mathrm{l_{w_3}}$), two sense of urgency levels ($\mathrm{l_{sou_1}}$ and $\mathrm{l_{sou_2}}$) , and two mind wandering levels ($\mathrm{l_{mw_1}}$ and $\mathrm{l_{mw_2}}$). In this part of the study, we used the combination of samples obtained from both session 1 and session 2.

### Workload

Figure S4 Fig shows the boxplot of the mean APCPS taken from both of the session to assess different human cognitive workload levels. The results demonstrated that the

**Table 3.** Results of SNK Test for different cognitive workload levels.

| Levels | Mean | SD | SE | Min | Max | Q25 | Q50 | Q75 |
|---|---|---|---|---|---|---|---|---|
| $l_{w_1}$ | 2.65 | 8.53 | 0.81 | -12.91 | 37.89 | -2.22 | 1.12 | 5.00 |
| $l_{w_2}$ | -0.83 | 6.32 | 0.81 | -22.71 | 14.32 | -3.93 | -1.13 | 2.12 |
| $l_{w_3}$ | -15.29 | 8.27 | 0.81 | -36.48 | 10.15 | -20.69 | -15.69 | -10.45 |

**Table 4.** Results of Tukey HSD multiple pairwise test and SNK multiple pairwise test for different pairs of cognitive workload levels.

| Tukey's HSD Test | p-value (in total) | p-value (Session 1) | p-value (Session 2) |
|---|---|---|---|
| $l_{w_1}$ - $l_{w_2}$ | $7.47 \times 10^{-3}$ | $8.74 \times 10^{-3}$ | $3.81 \times 10^{-1}$ |
| $l_{w_1}$ - $l_{w_3}$ | $2.23 \times 10^{-13}$ | 0.00 | $1.76 \times 10^{-14}$ |
| $l_{w_2}$ - $l_{w_3}$ | $2.23 \times 10^{-13}$ | $2.52 \times 10^{-14}$ | $2.75 \times 10^{-14}$ |

| SNK Test | p-value (in total) | p-value (Session 1) | p-value (Session 2) |
|---|---|---|---|
| $l_{w_1}$ - $l_{w_2}$ | $0.27 \times 10^{-2}$ | $3.09 \times 10^{-4}$ | 0.19 |
| $l_{w_1}$ - $l_{w_3}$ | 0.00 | 0.00 | 0.00 |
| $l_{w_2}$ - $l_{w_3}$ | 0.00 | 0.00 | 0.00 |

APCPS is a reliable indicator to separate multiple workload levels. We explored the effect of the APCPS in separating different cognitive workload levels by performing the SNK test seen in Table 3. The results show that the mean values for $l_{w_1}$, $l_{w_2}$, and $l_{w_3}$ are 2.65, -0.83, and -15.29, respectively. Then, we performed Tukey's HSD multiple pairwise comparison test and SNK multiple pairwise comparison test to examine the variation in the mean APCPS among the three cognitive workload levels. Table 4 shows the results of the p-values for all pairs of workload levels obtained from Tukey's HSD and SNK multiple pairwise comparison tests at a significance level of .95. The results demonstrate that there is a statistically significant difference between all pairs of workload levels (p-value <.05). Additionally, Fig  depicts the mean APCPS values related to separate sessions for different workload levels. The results demonstrate that we can differentiate session 1 and session 2 during $l_{w_1}$ by using the APCPS. On the other hand, the difference between the APCPS values of session 1 and session 2 during $l_{w_2}$ and $l_{w_3}$ are negligible. This is the evidence of the training effect of session 1 during $l_{w_1}$ which we required the participants to acclimate with different types of secondary events including braking tasks and dialogue communications.

**Fig 4. The change in the mean APCPS related to different cognitive workload levels.** Boxplot of the mean APCPS values corresponding to different cognitive workload levels including both sessions from all of the subjects.

We further investigated the efficacy of various physiological indicators in differentiation different cognitive workload levels in addition to the APCPS. Table 5 illustrates the p-values of Tukey HSD multiple pairwise test for 21 physiological markers obtained from five signal types including pupillometry, EEG, fNIRS, skin conductance, and respiration. The results demonstrated that the APCPS is the best feature in classifying all pairs of cognitive workload levels among other physiological features. We

**Table 5.** p-values of Tukey HSD multiple pairwise test for the morphological features obtained from different physiological signal types including pupillometry, EEG, fNIRS, skin conductance, and respiration for different pairs of cognitive workload levels.

| Signal Type | Feature | $l_{w_1}$ - $l_{w_2}$ | $l_{w_1}$ - $l_{w_3}$ | $l_{w_2}$ - $l_{w_3}$ |
|---|---|---|---|---|
| Pupillometry | APCPS | $7.47 \times 10^{-3}$ | $\mathbf{2.23 \times 10^{-13}}$ | $\mathbf{2.23 \times 10^{-13}}$ |
| | BlinkNum | 0.68 | 0.45 | 0.01 |
| | Mean BlinkDur | 0.86 | 0.89 | 0.99 |
| | FixNum | 0.99 | 0.99 | 0.99 |
| | Mean FixDur | 0.46 | 0.99 | 0.47 |
| EEG | Delta | 0.04 | 0.09 | $1.66 \times 10^{-5}$ |
| | Theta | 0.06 | 0.1 | $5.30 \times 10^{-5}$ |
| | Alpha | 0.03 | 0.54 | $6.99 \times 10^{-4}$ |
| | Beta | $8.03 \times 10^{-3}$ | 0.03 | $7.87 \times 10^{-8}$ |
| | Gamma | 0.02 | $2.70 \times 10^{-3}$ | $7.07 \times 10^{-9}$ |
| fNIRS | Mean | 0.99 | 0.91 | 0.93 |
| | Variance | 0.97 | 0.99 | 0.99 |
| Skin Conductance | Mean Skin Conductance | 0.99 | 0.99 | 0.98 |
| Respiration | $BR_{coeff}$ | 0.97 | 0.10 | 0.06 |
| | BR | 0.82 | $1.52 \times 10^{-3}$ | $1.23 \times 10^{-4}$ |
| | $DC_{coeff}$ | 0.92 | 0.06 | 0.02 |
| | DC | 0.20 | $8.47 \times 10^{-3}$ | 0.60 |
| | ET | 0.01 | $3.75 \times 10^{-5}$ | 0.26 |
| | IBI | 0.06 | $1.56 \times 10^{-3}$ | 0.57 |
| | IE | 0.01 | $1.53 \times 10^{-5}$ | 0.19 |
| | IT | $\mathbf{2.08 \times 10^{-3}}$ | $1.76 \times 10^{-4}$ | 0.79 |

observed that EEG is capable of separating $l_{w_2}$ and $l_{w_3}$ by using five different frequency bands. Specifically, the beta band was able to separate all pairs of workload levels. Despite the p-values were lower than .05, they were relatively higher than the p-values obtained from the APCPS. Moreover, some respiration metrics are able to classify $l_{w_1}$ and $l_{w_3}$ including BR, DC, ET, IBI, IE, and IT. In particular, BR was competent in classifying the pairs $l_{w_1}$-$l_{w_3}$ and $l_{w_2}$-$l_{w_3}$. However, there was no physiological marker, other than the PCPS, which was superior in differentiating all pairs of the cognitive workload levels.

## Sense of Urgency

Figure S5 Fig shows the boxplot of the mean APCPS values corresponding to lower and higher sense of urgency levels. The results demonstrated that the APCPS is capable of differentiating two sense of urgency levels. We observed that the APCPS is notable lower for less sense of urgency ($l_{sou_1}$) compared to higher sense of urgency levels ($l_{sou_2}$).

**Fig 5. The change in the mean APCPS related to different sense of urgency levels.** Boxplot of the mean APCPS corresponding to two sense of urgency levels including both sessions from all of the subjects. $l_{sou_1}$: Lower sense of urgency, $l_{sou_2}$: Higher sense of urgency.

Furthermore, we explored the effectiveness of other physiological indicators in differentiating two sense of urgency levels. Table 6 indicates the p-values taken from

Tukey's HSD pairwise test for 11 physiological markers obtained from four signal types. Here, we excluded eight respiration ($BR_{coeff}$, BR, $DC_{coeff}$, DC, ET, IBI, IE, and IT) and two eye gaze features (BlinkNum and mean BlinkDur) given that we explored 2.5-seconds signal segment which may not include enough number of respiration or blinking to calculate the related physiological markers. The results indicated that the APCPS is the best feature in separating two sense of urgency levels compared to other physiological signal indicators. On the other hand, there were three EEG features (including alpha, beta, and gamma) which were also capable of differentiating $l_{sou_1}$ and $l_{sou_2}$. In particular, the performance of alpha (EEG) was closely equivalent to the performance of the APCPS (pupillometry) which guided us to further investigate the performance of the fusion of pupillometry and EEG in classifying two sense of urgency levels.

**Table 6.** p-values of Tukey HSD multiple pairwise test for the morphological features obtained from different physiological signal types including pupillometry, EEG, fNIRS, and skin conductance for two sense of urgency levels.

| **Signal Type** | **Feature** | $\mathbf{l_{sou_1}}$ **-** $\mathbf{l_{sou_2}}$ |
|---|---|---|
| Pupillometry | APCPS | $\mathbf{5.25 \times 10^{-4}}$ |
| | FixNum | 0.30 |
| | Mean FixDur | 0.01 |
| EEG | Delta | $5.31 \times 10^{-2}$ |
| | Theta | 0.96 |
| | Alpha | $7.28 \times 10^{-4}$ |
| | Beta | $1.14 \times 10^{-2}$ |
| | Gamma | $1.03 \times 10^{-3}$ |
| fNIRS | Mean | 0.94 |
| | Variance | 0.96 |
| Skin Conductance | Mean Skin Conductance | 0.76 |

## Mind Wandering

Figure S6 Fig shows the boxplot of the mean APCPS values related to lower and higher mind wandering levels taken from both of the sessions. The results indicated that the mean APCPS is considerably lower in $l_{mw_1}$ (more mind wandering) than in $l_{mw_2}$ (less mind wandering where the participants were used to driving and were responsible for proceeding only with driving without additional secondary tasks). We also ran Tukey's HSD pairwise test and found out that the p-value is $3.60 \times 10^{-14}$ in differentiating $l_{mw_1}$ and $l_{mw_2}$ at a significance level of .95. The overall results demonstrated that the APCPS is a reliable indicator in predicting different mind wandering levels.

**Fig 6. The change in the mean APCPS related to different mind wandering levels.** Boxplot of the mean APCPS corresponding to two mind wandering levels including both sessions from all of the subjects. $l_{mw_1}$: Lower mind wandering, $l_{mw_2}$: Higher mind wandering.

We further analyzed the performance of other physiological metrics in classifying two mind wandering levels in addition to the APCPS. Table 7 shows the p-values obtained from Tukey's HSD pairwise test for 21 physiological markers acquired from five signal types. The results showed that the APCPS is the most effective indicator in

differentiating $l_{mw_1}$ and $l_{mw_2}$ compared to other physiological markers. Even though there were some other indicators including delta, theta, and alpha powers obtained from EEG and ET, IBI, IE, and IT acquired from respiration which have p-values lower than .05, they were significantly higher that the p-value obtained from the APCPS which makes the APCPS the most reliable physiological feature in assessing different mind wandering levels.

### Workload vs Sense of Urgency

Figure S7 Fig depicts the variation in PCPS taken from the session 1 of a random subject during $l_{w_1}$. In the figure, B and D represent the onset of the braking event and the onset of the dialogue event, respectively. We chose $l_{w_1}$ to analyze the relationship between the sense of urgency and the workload considering that during $l_{w_1}$ of session 1, the participants encountered the braking and the dialogue events for the first time which accounts for a potential increase in the sense of urgency.

**Fig 7. Illustration of the change in PCPS with different events.** Illustration of the change in PCPS with different events including braking and dialogue. The results include two sessions obtained from two different subjects. B and D represent the onset of the braking event and the onset of the dialogue event, respectively.

The results indicated that there were noticeable spikes after the onset points of the braking tasks and the dialogue interactions. We observed instantaneous spikes of PCPS in the middle of the braking events. The front car brakes in the middle of the braking event which explains the rapid increase in PCPS during the first braking event of session 1. We also observed instant peaks in PCPS during the second braking event thereupon the dialogue event occurs. The consecutive occurrences of multiple sense of urgency cases thus contribute to higher workload resulting in increased PCPS values during $l_{w_1}$ of session 1.

## Model Development and Evaluation

Given that the statistical analyses showed statistically significant differences between the different levels of the three systemic cognitive states, the question arises whether it is possible to train machine learning models on a subset of the data and use the models to predict those levels in new data that was not use for training. Put differently, the question is whether state-of-the-art machine-learning methodologies allow for the general prediction of different levels of the three systemic cognitive states based on the five types of physiological signals: pupillometry, EEG, fNIRS, skin conductance, respiration, and possibly combinations thereof (e.g., pupillometry and EEG). We considered five commonly used models: (1) *k*-Nearest Neighbor (k-NN), (2) Naive-Bayes (NB), (3) Random Forest (RF), (4) Support Vector Machines (SVM), and (5) Multiple Layer Perceptron (MLP) to determine the three human cognitive states (workload, sense of urgency, and mind-wandering) in four cases as follows (described in detail in Table 8):

1. **Three-class workload classification:** We performed the tasks $W_{S_1-1}$, $W_{S_2-1}$, and $W_{Comb-1}$.
2. **Two-class workload classification:** We performed the tasks $W_{S_1-2}$, $W_{S_2-2}$, and $W_{Comb-2}$.
3. **Two-class sense of urgency classification:** We performed the tasks $SoU_{S_1}$, $SoU_{S_2}$, and $SoU_{Comb}$.

**Table 7.** p-values of Tukey HSD multiple pairwise test for the morphological features obtained from different physiological signal types including pupillometry, EEG, fNIRS, skin conductance, and respiration for two mind wandering levels.

| Signal Type | Feature | $\mathbf{l_{mw_1}}$ **-** $\mathbf{l_{mw_2}}$ |
|---|---|---|
| Pupillometry | APCPS | $\mathbf{3.60 \times 10^{-14}}$ |
| | BlinkNum | 0.59 |
| | Mean BlinkDur | 0.53 |
| | FixNum | 0.77 |
| | Mean FixDur | 0.29 |
| EEG | Delta | 0.79 |
| | Theta | 0.03 |
| | Alpha | $5.27 \times 10^{-3}$ |
| | Beta | 0.12 |
| | Gamma | 0.20 |
| fNIRS | Mean | 0.58 |
| | Variance | 0.48 |
| Skin Conductance | Mean Skin Conductance | 0.24 |
| Respiration | $BR_{coeff}$ | 0.26 |
| | BR | 0.20 |
| | $DC_{coeff}$ | 0.77 |
| | DC | 0.16 |
| | ET | $2.52 \times 10^{-4}$ |
| | IBI | $2.43 \times 10^{-5}$ |
| | IE | $9.51 \times 10^{-5}$ |
| | IT | $8.49 \times 10^{-5}$ |

4. **Two-class mind wandering classification:** We performed the tasks $MW_{S_1}$, $MW_{S_2}$, and $MW_{Comb}$.

For the given classification task, the learning method, and the type of signal, we divided the data into a training set and a test set with a ratio of about 80% and 20%, *i.e.*, data from 37 subjects for the training set and data from 9 subjects for the test set, respectively. Next, the cross-validation method [78] was applied to the training set to select the best-learned model to apply to the test set. To mitigate the randomness in the training and test processes as well as to stabilize the testing performance, for a given classification task and learning method, we repeated the whole experiment 10 times and only report the average accuracy together with its standard deviation. First, we ran each of these tasks separately for 16 features acquired from five signal types which we described in detail in Section . Then, in Section , we combined pupillometry and EEG, and reported the performance of the combination of pupillometry and EEG on twelve classification tasks described in Section .

## Cognitive State Classification with One Signal Type

As previously described in Section , we generated three cognitive workload levels ($l_{w_1}$, $l_{w_2}$, and $l_{w_3}$), two sense of urgency levels ($l_{sou_1}$ and $l_{sou_2}$), and two mind wandering levels ($l_{mw_1}$ and $l_{mw_2}$) corresponding to different intervals of time and braking/dialogue events (see Figure S3 Fig). Then, we run six workload, three sense of urgency, and three

**Table 8.** Collection of workload, sense of urgency, and mind wandering classification tasks which includes the task ID, the classification type, the sessions that we used to do classification, and the levels were classified. $X_{y\text{-}z}$ represents the classification task of "X" (workload, sense of urgency, or mind wandering) using data from session(s) y (only session 1 ($S_1$), only session 2 ($S_2$), or the combination of them (Comb) which classifies different levels (1 and 2 represent the three-class classification and the two-class classification of workload, respectively; for mind wandering and sense of urgency, there are only two classes for all classification tasks). As an example, $W_{S_1\text{-}1}$ represents the classification of three workload levels obtained from session 1.

| Task ID: | Classification Type: | Data used from Session(s): | Levels to be classified: |
|---|---|---|---|
| $W_{S_1\text{-}1}$ | Workload | 1 | $l_{w_1}, l_{w_2}, l_{w_3}$ |
| $W_{S_2\text{-}1}$ | Workload | 2 | $l_{w_1}, l_{w_2}, l_{w_3}$ |
| $W_{Comb\text{-}1}$ | Workload | 1,2 | $l_{w_1}, l_{w_2}, l_{w_3}$ |
| $W_{S_1\text{-}2}$ | Workload | 1 | $l_{w_1}, l_{w_3}$ |
| $W_{S_2\text{-}2}$ | Workload | 2 | $l_{w_1}, l_{w_3}$ |
| $W_{Comb\text{-}2}$ | Workload | 1,2 | $l_{w_1}, l_{w_3}$ |
| $SoU_{S_1}$ | Sense of Urgency | 1 | $l_{sou_1}, l_{sou_2}$ |
| $SoU_{S_2}$ | Sense of Urgency | 2 | $l_{sou_1}, l_{sou_2}$ |
| $SoU_{Comb}$ | Sense of Urgency | 1,2 | $l_{sou_1}, l_{sou_2}$ |
| $MW_{S_1}$ | Mind Wandering | 1 | $l_{mw_1}, l_{mw_2}$ |
| $MW_{S_2}$ | Mind Wandering | 2 | $l_{mw_1}, l_{mw_2}$ |
| $MW_{Comb}$ | Mind Wandering | 1,2 | $l_{mw_1}, l_{mw_2}$ |

mind wandering classification tasks which we described in detail in Section , Section , Section , and Section .

### Three-level Workload Classification

First, we investigated the performance of the five ML algorithms in separating three cognitive workload levels by using various physiological markers. Table 9a, Table 9b, and Table 9c show the classification accuracies of the five algorithms related to the tasks $W_{S_1-1}$, $W_{S_2-1}$, and $W_{Comb-1}$, respectively. The results indicate that for the task $W_{S_1-1}$, which aims to distinguish three workload levels generated from session 1, APCPS achieves the highest accuracy of $72.57 \mp 8.75$ using the Naive Bayes algorithm which outperforms other tested signals by a large margin of nearly 25%. Similarly, for the task $W_{S_2-1}$, which aims to distinguish three workload levels created from session 2, APCPS also achieves the highest accuracy of $59.13 \mp 8.27$, exceeding the accuracy obtained by other signals by a margin of at least nearly 15%. Note that both $W_{S_1-1}$ and $W_{S_2-1}$ are three-class classification tasks, therefore, a random guess will likely provide an accuracy of 33% which is far below the performance achieved by all five tested algorithms on APCPS. Lastly, we run the five ML algorithms for the task $W_{Comb-1}$, which aims to distinguish three workload levels by using data from both session 1 and session 2, and found out that APCPS has the best classification accuracy among the 16 features in classifying three workload levels with the highest accuracy of $63.35 \mp 5.63$ using the Naive Bayes algorithm. The results demonstrate that the APCPS signal is the best physiological indicator in separating three cognitive workload levels.

### Two-level Workload Classification

Second, we examined the effectiveness of the five ML algorithms in differentiating two cognitive workload levels, $l_{w_1}$ and $l_{w_3}$, by using different physiological markers.

**Table 9.** Classification accuracies (in %) of the tasks **(a)** $W_{S_1\text{-}1}$, **(b)** $W_{S_2\text{-}1}$, and **(c)** $W_{\text{Comb-1}}$ described in Table 8.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | 71.03 ∓ 7.80 | **72.57** ∓ 8.75 | 63.30 ∓ 9.98 | 70.24 ∓ 10.65 | 72.02 ∓ 6.81 |
| | BlinkNum | 29.19 ∓ 5.58 | 34.13 ∓ 7.16 | 29.37 ∓ 8.21 | 35.32 ∓ 5.94 | 42.86 ∓ 4.62 |
| | Mean BlinkDur | 37.11 ∓ 4.60 | 28.05 ∓ 6.75 | 37.85 ∓ 2.79 | 41.40 ∓ 7.56 | 35.80 ∓ 2.83 |
| | FixNum | 35.40 ∓ 6.62 | 28.97 ∓ 4.27 | 26.59 ∓ 5.08 | 29.76 ∓ 6.73 | 26.63 ∓ 4.54 |
| | Mean FixDur | 26.19 ∓ 7.14 | 34.21 ∓ 8.04 | 32.00 ∓ 7.34 | 31.75 ∓ 8.32 | 32.71 ∓ 10.02 |
| **EEG** | EEG | 48.34 ∓ 8.15 | 44.15 ∓ 5.48 | 41.50 ∓ 6.07 | 43.90 ∓ 7.29 | 42.29 ∓ 9.55 |
| **fNIRS** | fNIRS | 33.41 ∓ 2.97 | 28.48 ∓ 3.37 | 33.81 ∓ 8.25 | 33.33 ∓ 7.34 | 31.24 ∓ 5.69 |
| **Skin** | Skin | 29.98 ∓ 7.74 | 27.43 ∓ 3.37 | 32.98 ∓ 7.92 | 31.75 ∓ 5.94 | 25.19 ∓ 5.48 |
| **Respiration** | $BR_{coeff}$ | 35.32 ∓ 9.59 | 36.11 ∓ 3.55 | 34.92 ∓ 4.99 | 41.27 ∓ 6.10 | 35.62 ∓ 6.24 |
| | BR | 34.92 ∓ 9.64 | 43.33 ∓ 7.34 | 40.87 ∓ 9.24 | 32.14 ∓ 6.52 | 26.51 ∓ 8.93 |
| | $DC_{coeff}$ | 36.90 ∓ 5.32 | 30.95 ∓ 5.32 | 42.86 ∓ 7.34 | 34.13 ∓ 10.12 | 39.74 ∓ 4.25 |
| | DC | 44.84 ∓ 7.74 | 40.87 ∓ 9.39 | 32.94 ∓ 6.69 | 45.24 ∓ 5.58 | 32.46 ∓ 8.56 |
| | ET | 43.65 ∓ 8.21 | 45.63 ∓ 2.81 | 39.29 ∓ 8.91 | 41.67 ∓ 6.73 | 33.33 ∓ 0.00 |
| | IBI | 41.67 ∓ 8.58 | 45.63 ∓ 6.69 | 34.52 ∓ 7.53 | 42.86 ∓ 5.83 | 31.36 ∓ 7.17 |
| | IE | 46.43 ∓ 8.07 | 41.67 ∓ 4.12 | 42.06 ∓ 8.55 | 46.83 ∓ 6.62 | 43.96 ∓ 5.83 |
| | IT | 37.70 ∓ 11.31 | 42.86 ∓ 5.58 | 33.33 ∓ 6.07 | 43.25 ∓ 6.40 | 36.83 ∓ 8.50 |

**(a)** Classification accuracies (in %) of $W_{S_1\text{-}1}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | **59.13** ∓ 8.27 | 55.61 ∓ 9.28 | 50.79 ∓ 6.02 | 55.16 ∓ 8.27 | 57.85 ∓ 7.30 |
| | BlinkNum | 36.11 ∓ 8.15 | 26.98 ∓ 5.86 | 28.57 ∓ 4.45 | 36.51 ∓ 10.88 | 40.65 ∓ 3.38 |
| | Mean BlinkDur | 26.05 ∓ 1.11 | 31.30 ∓ 10.33 | 31.62 ∓ 5.20 | 36.63 ∓ 2.15 | 26.27 ∓ 1.97 |
| | FixNum | 29.76 ∓ 8.07 | 28.57 ∓ 6.30 | 31.35 ∓ 6.02 | 34.52 ∓ 7.90 | 34.17 ∓ 5.12 |
| | Mean FixDur | 25.00 ∓ 5.05 | 33.33 ∓ 14.38 | 28.97 ∓ 9.29 | 37.30 ∓ 6.76 | 29.13 ∓ 6.37 |
| **EEG** | EEG | 45.50 ∓ 6.37 | 39.77 ∓ 5.54 | 42.65 ∓ 9.32 | 42.19 ∓ 8.23 | 40.76 ∓ 5.29 |
| **fNIRS** | fNIRS | 29.84 ∓ 7.16 | 31.62 ∓ 3.76 | 27.78 ∓ 4.05 | 33.73 ∓ 5.35 | 39.31 ∓ 3.11 |
| **Skin** | Skin | 31.43 ∓ 11.42 | 23.81 ∓ 4.45 | 27.41 ∓ 7.55 | 35.71 ∓ 4.12 | 28.1 ∓ 6.38 |
| **Respiration** | $BR_{coeff}$ | 33.73 ∓ 8.60 | 34.13 ∓ 4.79 | 26.98 ∓ 9.39 | 36.51 ∓ 7.10 | 37.71 ∓ 3.19 |
| | BR | 41.67 ∓ 9.07 | 42.06 ∓ 4.70 | 30.16 ∓ 5.86 | 34.13 ∓ 8.09 | 29.32 ∓ 3.30 |
| | $DC_{coeff}$ | 40.08 ∓ 7.67 | 36.90 ∓ 6.52 | 44.84 ∓ 3.81 | 37.70 ∓ 5.08 | 31.15 ∓ 5.32 |
| | DC | 36.51 ∓ 9.78 | 39.68 ∓ 8.32 | 39.29 ∓ 9.52 | 37.30 ∓ 11.05 | 25.15 ∓ 5.91 |
| | ET | 40.87 ∓ 7.55 | 39.90 ∓ 6.94 | 42.46 ∓ 7.03 | 38.89 ∓ 7.23 | 36.57 ∓ 2.31 |
| | IBI | 34.13 ∓ 5.35 | 37.30 ∓ 5.35 | 28.97 ∓ 6.83 | 34.92 ∓ 6.90 | 27.17 ∓ 11.23 |
| | IE | 38.89 ∓ 11.60 | 44.44 ∓ 8.09 | 32.54 ∓ 8.15 | 37.30 ∓ 6.96 | 38.57 ∓ 10.16 |
| | IT | 36.51 ∓ 9.19 | 39.68 ∓ 5.17 | 28.57 ∓ 7.14 | 42.46 ∓ 4.59 | 38.80 ∓ 7.13 |

**(b)** Classification accuracies (in %) of $W_{S_2\text{-}1}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | 62.78 ∓ 5.47 | **63.35** ∓ 5.63 | 55.64 ∓ 4.15 | **63.35** ∓ 4.43 | 63.25 ∓ 5.22 |
| | BlinkNum | 29.88 ∓ 5.05 | 32.89 ∓ 4.46 | 32.24 ∓ 5.66 | 33.83 ∓ 4.62 | 34.21 ∓ 4.86 |
| | Mean BlinkDur | 41.96 ∓ 6.52 | 26.97 ∓ 3.79 | 29.51 ∓ 7.06 | 32.05 ∓ 2.81 | 32.71 ∓ 2.96 |
| | FixNum | 29.14 ∓ 7.47 | 27.63 ∓ 5.51 | 28.85 ∓ 5.80 | 30.08 ∓ 4.35 | 35.18 ∓ 1.95 |
| | Mean FixDur | 30.92 ∓ 4.26 | 35.06 ∓ 4.95 | 30.36 ∓ 3.75 | 32.42 ∓ 4.93 | 32.33 ∓ 2.77 |
| **EEG** | EEG | 42.82 ∓ 4.04 | 49.94 ∓ 6.69 | 33.30 ∓ 4.90 | 40.51 ∓ 5.12 | 43.05 ∓ 5.11 |
| **fNIRS** | fNIRS | 30.83 ∓ 4.55 | 28.76 ∓ 3.70 | 29.98 ∓ 6.31 | 32.52 ∓ 1.79 | 33.18 ∓ 2.11 |
| **Skin** | Skin | 24.81 ∓ 4.48 | 28.29 ∓ 4.50 | 43.39 ∓ 7.87 | 30.92 ∓ 4.26 | 33.18 ∓ 1.46 |
| **Respiration** | $BR_{coeff}$ | 34.11 ∓ 4.77 | 37.50 ∓ 6.45 | 34.87 ∓ 4.68 | 37.12 ∓ 3.74 | 33.36 ∓ 2.78 |
| | BR | 38.82 ∓ 5.67 | 37.41 ∓ 6.33 | 32.14 ∓ 5.12 | 39.57 ∓ 5.35 | 33.65 ∓ 2.41 |
| | $DC_{coeff}$ | 44.46 ∓ 5.21 | 31.48 ∓ 4.06 | 39.76 ∓ 7.48 | 35.43 ∓ 3.54 | 35.06 ∓ 3.72 |
| | DC | 35.53 ∓ 4.91 | 41.34 ∓ 5.49 | 34.68 ∓ 5.88 | 45.11 ∓ 5.67 | 44.36 ∓ 5.03 |
| | ET | 40.98 ∓ 5.03 | 39.57 ∓ 6.56 | 38.35 ∓ 5.06 | 40.51 ∓ 6.78 | 34.21 ∓ 2.41 |
| | IBI | 39.10 ∓ 4.98 | 42.76 ∓ 4.32 | 33.55 ∓ 6.41 | 39.10 ∓ 5.34 | 40.70 ∓ 4.40 |
| | IE | 41.45 ∓ 5.87 | 40.51 ∓ 4.74 | 33.65 ∓ 5.32 | 41.17 ∓ 3.50 | 41.63 ∓ 4.85 |
| | IT | 35.62 ∓ 6.77 | 43.23 ∓ 3.04 | 37.41 ∓ 6.06 | 41.82 ∓ 4.68 | 36.28 ∓ 6.29 |

**(c)** Classification accuracies (in %) of $W_{\text{Comb-1}}$.

Table 10a, Table 10b, and Table 10c show the classification accuracies of five algorithms related to the tasks $W_{S_1-2}$, $W_{S_2-2}$, and $W_{Comb-2}$, respectively. For the classification tasks $W_{S_1-2}$ and $W_{S_2-2}$, that aim to discern between the lowest and the highest workload levels in session 1 and session 2, respectively, the APCPS outperforms other signals by a large margin and achieves impressive accuracies of $93.25 \mp 3.55$ ($W_{S_1-2}$) by using MLP and $88.89 \mp 5.23$ ($W_{S_2-2}$) by using NB. Similarly, the APCPS achieved the best classification accuracy ($90.18 \mp 3.42$) for the task $W_{Comb-2}$ which aims to separate $l_{w_1}$ and $l_{w_3}$ by using the data obtained from both session 1 and session 2. The observed numerical results in tasks $W_{S_1-2}$, $W_{S_2-2}$, and $W_{Comb-2}$ consistently demonstrate the superior performance of APCPS in workload classification tasks compared to other tested physiological signals.

### Two-level Sense of Urgency Classification

As previously mentioned in Section , we generated two different levels for sense of urgency for use with the five ML algorithms aimed at differentiating these two levels. Table 11a, Table 11b, and Table 11c provide the classification accuracies of five algorithms related to the tasks $SoU_{S_1}$, $SoU_{S_2}$, and $SoU_{Comb}$, respectively. Specifically, the highest classification accuracy for task $SoU_{S_1}$, task $SoU_{S_2}$, and task $SoU_{Comb}$ are 63.43%, 64.27%, and 60.88% achieved by FixNum, APCPS and Mean FixDur, respectively. Even though the classification performance may depend on the learning algorithm, it seems like APCSP and Mean FixDur are the best two signals for assessing the sense of urgency that not only provide higher accuracy but also more stability than other compared signals.

### Two-level Mind Wandering Classification

As previously described in Section , two different levels of mind wandering were defined based on the first 1.5 minutes ($l_{mw_1}$) and the second 1.5 minutes ($l_{mw_2}$) of the experiment in each session, respectively. We constructed three mind wandering classification tasks, $MW_{S_1}$, $MW_{S_2}$, and $MW_{Comb}$ which aimed to differentiate two mind wandering levels by utilizing the data acquired from session 1, from session 2, and from the combination of session 1 and session 2, respectively. We refer the readers to Table 8 for the details of all of the classification tasks. Table 12a, Table 12b, and Table 12c provide the mind wandering classification accuracies corresponding to three classification tasks $MW_{S_1}$, $MW_{S_2}$, and $MW_{Comb}$, respectively. The results indicate that for the task $MW_{S_1}$ which aims to classify between two mind-wandering levels in session 1, the APCPS achieves the highest accuracy of $76.22 \mp 4.17$ (with RF) which outperforms all other tested signals by at least 10%. Similarly, we observed the same situation where the APCPS provides significantly better performance compared to other signals on the classification task $MW_{S_2}$. Particularly, the highest classification accuracy between the two mind-wandering levels in session 2 is $86.08 \mp 5.55$ achieved by the APCPS using the RF algorithm. This outperforms other tested signals by a large gap of at least 20%. In addition, the APCPS has the best accuracy ($82.08 \mp 7.35$) in differentiating the two mind wandering levels by using the data obtained from both session 1 and session 2. The results of the mind wandering classification tasks demonstrate that the APCPS has the best performance in predicting two mind wandering levels regardless of which classification model is used.

**Table 10.** Classification accuracies (in %) of the tasks **(a)** $W_{S_1\text{-}2}$, **(b)** $W_{S_2\text{-}2}$, and **(c)** $W_{\text{Comb-}2}$ described in Table 8.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | 88.89 ∓ 6.77 | 90.06 ∓ 5.23 | 89.47 ∓ 6.56 | 92.98 ∓ 4.96 | **93.25** ∓ 3.55 |
| | BlinkNum | 53.86 ∓ 6.08 | 53.22 ∓ 5.79 | 49.77 ∓ 7.49 | 50.29 ∓ 8.63 | 50.40 ∓ 6.17 |
| | Mean BlinkDur | 55.72 ∓ 12.10 | 49.27 ∓ 4.70 | 59.12 ∓ 10.22 | 52.19 ∓ 12.23 | 56.17 ∓ 9.00 |
| | FixNum | 52.05 ∓ 8.02 | 47.37 ∓ 7.44 | 48.54 ∓ 7.35 | 42.11 ∓ 8.59 | 50.00 ∓ 0.00 |
| | Mean FixDur | 53.86 ∓ 7.85 | 43.86 ∓ 6.56 | 43.86 ∓ 14.25 | 42.69 ∓ 9.10 | 51.19 ∓ 6.52 |
| **EEG** | EEG | 61.27 ∓ 4.33 | 58.60 ∓ 5.51 | 57.46 ∓ 3.83 | 60.98 ∓ 7.35 | 60.18 ∓ 5.52 |
| **fNIRS** | fNIRS | 49.77 ∓ 7.07 | 59.49 ∓ 3.51 | 55.32 ∓ 4.60 | 50.29 ∓ 10.85 | 52.05 ∓ 8.02 |
| **Skin** | Skin | 46.26 ∓ 11.22 | 45.03 ∓ 5.03 | 45.03 ∓ 13.84 | 44.44 ∓ 8.27 | 49.21 ∓ 1.48 |
| **Respiration** | $BR_{coeff}$ | 56.14 ∓ 6.08 | 54.39 ∓ 7.85 | 52.63 ∓ 8.23 | 52.63 ∓ 11.90 | 50.00 ∓ 0.00 |
| | BR | 65.50 ∓ 8.98 | 66.08 ∓ 7.89 | 58.48 ∓ 8.75 | 56.14 ∓ 8.59 | 48.81 ∓ 3.37 |
| | $DC_{coeff}$ | 63.77 ∓ 7.89 | 63.22 ∓ 10.66 | 66.67 ∓ 6.56 | 56.73 ∓ 6.92 | 57.14 ∓ 8.25 |
| | DC | 61.99 ∓ 9.54 | 67.25 ∓ 11.03 | 52.05 ∓ 11.22 | 63.74 ∓ 9.09 | 68.25 ∓ 5.70 |
| | ET | 64.91 ∓ 7.85 | 64.33 ∓ 4.14 | 63.74 ∓ 10.66 | 67.84 ∓ 8.39 | 50.00 ∓ 0.00 |
| | IBI | 65.50 ∓ 7.07 | 66.67 ∓ 11.64 | 67.25 ∓ 5.96 | 68.42 ∓ 9.92 | 62.70 ∓ 6.10 |
| | IE | 56.73 ∓ 11.03 | 69.59 ∓ 9.21 | 59.06 ∓ 13.08 | 62.57 ∓ 6.77 | 63.89 ∓ 4.59 |
| | IT | 61.40 ∓ 10.53 | 65.49 ∓ 10.85 | 51.46 ∓ 8.87 | 68.42 ∓ 9.92 | 59.92 ∓ 8.55 |

**(a)** Classification accuracies (in %) of $W_{S_1\text{-}2}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | 86.55 ∓ 5.61 | **88.89** ∓ 5.23 | 84.80 ∓ 6.30 | 84.80 ∓ 4.60 | 84.52 ∓ 4.45 |
| | BlinkNum | 37.43 ∓ 7.21 | 46.20 ∓ 5.96 | 46.78 ∓ 6.30 | 46.20 ∓ 7.35 | 46.03 ∓ 6.17 |
| | Mean BlinkDur | 48.12 ∓ 7.37 | 52.86 ∓ 3.33 | 61.10 ∓ 5.55 | 46.60 ∓ 6.85 | 49.52 ∓ 2.38 |
| | FixNum | 48.54 ∓ 9.86 | 40.94 ∓ 8.15 | 46.78 ∓ 6.30 | 45.03 ∓ 8.27 | 50.00 ∓ 1.68 |
| | Mean FixDur | 48.01 ∓ 6.46 | 42.11 ∓ 8.95 | 47.95 ∓ 8.02 | 43.27 ∓ 8.87 | 48.41 ∓ 5.08 |
| **EEG** | EEG | 56.25 ∓ 9.19 | 52.30 ∓ 4.81 | 52.60 ∓ 5.08 | 58.15 ∓ 2.65 | 54.33 ∓ 3.17 |
| **fNIRS** | fNIRS | 42.16 ∓ 9.10 | 56.42 ∓ 1.65 | 58.83 ∓ 2.19 | 47.37 ∓ 3.51 | 53.00 ∓ 4.01 |
| **Skin** | Skin | 44.01 ∓ 7.35 | 43.27 ∓ 8.51 | 45.61 ∓ 15.49 | 42.11 ∓ 8.23 | 48.81 ∓ 3.37 |
| **Respiration** | $BR_{coeff}$ | 66.08 ∓ 9.64 | 53.80 ∓ 11.31 | 56.14 ∓ 8.23 | 51.46 ∓ 10.75 | 50.00 ∓ 0.00 |
| | BR | 69.00 ∓ 10.06 | 47.95 ∓ 12.76 | 59.06 ∓ 7.76 | 62.57 ∓ 10.66 | 51.19 ∓ 3.37 |
| | $DC_{coeff}$ | 54.97 ∓ 8.98 | 55.56 ∓ 5.61 | 49.71 ∓ 10.56 | 56.14 ∓ 10.53 | 59.52 ∓ 8.07 |
| | DC | 62.57 ∓ 6.77 | 60.82 ∓ 8.63 | 57.89 ∓ 6.07 | 60.82 ∓ 7.07 | 65.48 ∓ 9.07 |
| | ET | 55.56 ∓ 12.92 | 61.40 ∓ 10.53 | 54.39 ∓ 9.92 | 61.40 ∓ 11.90 | 50.00 ∓ 0.00 |
| | IBI | 56.73 ∓ 14.42 | 58.48 ∓ 10.06 | 42.69 ∓ 9.10 | 55.56 ∓ 9.96 | 58.73 ∓ 6.76 |
| | IE | 50.29 ∓ 10.56 | 57.31 ∓ 9.10 | 57.31 ∓ 9.75 | 61.99 ∓ 5.96 | 61.11 ∓ 9.44 |
| | IT | 60.23 ∓ 9.96 | 60.82 ∓ 8.27 | 46.20 ∓ 8.15 | 69.01 ∓ 7.62 | 55.16 ∓ 6.10 |

**(b)** Classification accuracies (in %) of $W_{S_2\text{-}2}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | **90.18** ∓ 3.42 | 87.91 ∓ 6.04 | 85.35 ∓ 5.97 | 88.76 ∓ 4.14 | 85.71 ∓ 4.41 |
| | BlinkNum | 49.22 ∓ 7.47 | 51.64 ∓ 7.59 | 49.79 ∓ 7.09 | 50.92 ∓ 7.71 | 51.03 ∓ 4.03 |
| | Mean BlinkDur | 61.09 ∓ 4.36 | 45.80 ∓ 5.15 | 44.10 ∓ 6.62 | 47.65 ∓ 4.82 | 48.68 ∓ 2.77 |
| | FixNum | 46.80 ∓ 5.62 | 43.39 ∓ 6.23 | 49.22 ∓ 7.37 | 49.36 ∓ 5.94 | 49.99 ∓ 1.00 |
| | Mean FixDur | 44.52 ∓ 7.04 | 43.53 ∓ 7.69 | 48.22 ∓ 7.40 | 48.93 ∓ 6.44 | 49.99 ∓ 5.50 |
| **EEG** | EEG | 60.95 ∓ 5.41 | 58.95 ∓ 5.83 | 54.49 ∓ 7.71 | 63.44 ∓ 5.64 | 63.44 ∓ 4.19 |
| **fNIRS** | fNIRS | 47.52 ∓ 5.56 | 45.09 ∓ 5.95 | 42.25 ∓ 7.08 | 49.34 ∓ 1.44 | 49.81 ∓ 2.45 |
| **Skin** | Skin | 40.26 ∓ 6.73 | 44.81 ∓ 4.32 | 28.67 ∓ 6.20 | 48.36 ∓ 6.84 | 49.25 ∓ 3.44 |
| **Respiration** | $BR_{coeff}$ | 59.74 ∓ 6.32 | 54.77 ∓ 10.44 | 59.60 ∓ 7.26 | 54.62 ∓ 5.16 | 50.0 ∓ 0.0 |
| | BR | 62.87 ∓ 8.26 | 59.46 ∓ 7.89 | 56.33 ∓ 7.19 | 65.43 ∓ 5.65 | 50.47 ∓ 3.22 |
| | $DC_{coeff}$ | 63.02 ∓ 6.33 | 52.35 ∓ 6.92 | 56.61 ∓ 6.88 | 58.03 ∓ 6.76 | 58.08 ∓ 5.06 |
| | DC | 64.01 ∓ 4.89 | 65.58 ∓ 6.37 | 56.19 ∓ 7.15 | 63.87 ∓ 6.15 | 67.11 ∓ 5.29 |
| | ET | 55.48 ∓ 7.36 | 63.30 ∓ 6.10 | 53.34 ∓ 6.67 | 62.73 ∓ 8.02 | 50.94 ∓ 3.24 |
| | IBI | 60.60 ∓ 5.28 | 63.44 ∓ 5.43 | 45.80 ∓ 8.11 | 64.30 ∓ 6.65 | 63.53 ∓ 4.86 |
| | IE | 59.88 ∓ 6.46 | 60.88 ∓ 6.28 | 54.06 ∓ 6.85 | 62.87 ∓ 5.81 | 61.28 ∓ 4.56 |
| | IT | 55.90 ∓ 8.04 | 60.46 ∓ 5.49 | 53.49 ∓ 5.45 | 61.88 ∓ 8.31 | 56.86 ∓ 7.31 |

**(c)** Classification accuracies (in %) of $W_{\text{Comb-}2}$.

**Table 11.** Classification accuracies (in %) of the tasks **(a)**$\text{SoU}_{\text{S}_1}$, **(b)** $\text{SoU}_{\text{S}_2}$, and **(c)** $\text{SoU}_{\text{Comb}}$ described in Table 8.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| | APCPS | $51.08 \mp 12.65$ | $55.96 \mp 7.48$ | $44.32 \mp 9.09$ | $49.86 \mp 9.56$ | $48.42 \mp 10.00$ |
| | FixNum | $50.80 \mp 10.56$ | $61.77 \mp 9.33$ | $52.91 \mp 12.37$ | $63.43 \mp 9.58$ | $52.57 \mp 9.39$ |
| **Pupillometry** | Mean FixDur | $61.88 \mp 15.13$ | $59.83 \mp 13.68$ | $52.63 \mp 10.10$ | $59.56 \mp 9.20$ | $59.06 \mp 10.42$ |
| **EEG** | EEG | $48.14 \mp 7.32$ | $51.52 \mp 11.90$ | $53.02 \mp 10.08$ | $49.80 \mp 7.30$ | $48.25 \mp 7.41$ |
| **fNIRS** | fNIRS | $44.75 \mp 11.43$ | $49.46 \mp 8.06$ | $48.63 \mp 8.19$ | $45.42 \mp 6.90$ | $43.14 \mp 5.50$ |
| **Skin** | Skin | $32.90 \mp 10.69$ | $44.60 \mp 7.32$ | $44.32 \mp 9.09$ | $50.42 \mp 7.88$ | $48.14 \mp 5.76$ |

**(a)** Classification accuracies (in %) of $\text{SoU}_{\text{S}_1}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| | APCPS | $63.82 \mp 7.05$ | $64.27 \mp 10.88$ | $55.40 \mp 10.01$ | $63.16 \mp 8.87$ | $58.39 \mp 7.71$ |
| | FixNum | $59.66 \mp 9.41$ | $49.03 \mp 9.96$ | $60.94 \mp 9.71$ | $44.88 \mp 8.08$ | $45.65 \mp 8.95$ |
| **Pupillometry** | Mean FixDur | $56.62 \mp 5.78$ | $56.23 \mp 11.20$ | $51.52 \mp 11.01$ | $54.29 \mp 7.65$ | $53.06 \mp 9.70$ |
| **EEG** | EEG | $56.73 \mp 9.55$ | $59.83 \mp 10.28$ | $57.89 \mp 13.00$ | $58.39 \mp 9.25$ | $57.11 \mp 11.50$ |
| **fNIRS** | fNIRS | $56.94 \mp 9.09$ | $57.31 \mp 10.47$ | $52.69 \mp 9.84$ | $54.56 \mp 9.96$ | $58.10 \mp 8.05$ |
| **Skin** | Skin | $42.49 \mp 8.34$ | $43.49 \mp 5.08$ | $46.81 \mp 8.69$ | $45.98 \mp 9.48$ | $46.20 \mp 4.60$ |

**(b)** Classification accuracies (in %) of $\text{SoU}_{\text{S}_2}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| | APCPS | $57.04 \mp 6.52$ | $60.46 \mp 9.29$ | $54.05 \mp 5.95$ | $57.04 \mp 8.41$ | $55.75 \mp 7.80$ |
| | FixNum | $55.90 \mp 7.62$ | $53.91 \mp 5.51$ | $57.33 \mp 7.15$ | $55.33 \mp 8.87$ | $50.49 \mp 4.14$ |
| **Pupillometry** | Mean FixDur | $56.61 \mp 8.77$ | $59.17 \mp 6.32$ | $52.63 \mp 7.09$ | $60.88 \mp 8.34$ | $58.17 \mp 8.17$ |
| **EEG** | EEG | $54.35 \mp 5.61$ | $59.60 \mp 9.74$ | $51.06 \mp 7.88$ | $51.34 \mp 5.03$ | $56.74 \mp 6.67$ |
| **fNIRS** | fNIRS | $50.92 \mp 6.52$ | $55.04 \mp 7.36$ | $45.66 \mp 8.45$ | $50.35 \mp 8.14$ | $53.61 \mp 9.96$ |
| **Skin** | Skin | $41.11 \mp 5.70$ | $46.23 \mp 7.28$ | $47.23 \mp 5.64$ | $47.51 \mp 6.70$ | $46.08 \mp 4.26$ |

**(c)** Classification accuracies (in %) of $\text{SoU}_{\text{Comb}}$.

## Cognitive State Classification with the Fusion of Pupillometry and EEG

We also investigated whether combining different types of physiological signal modalities could improve the performance of single modalities in workload, sense of urgency, and mind wandering classification tasks. Since we had 16 physiological signals, there were numerous combinations which may potentially result in a very high computational complexity. Therefore, we limited ourselves to only combining the APCPS and EEG given that they were the best two indicators for assessing different human cognitive state levels. I.e., even though respiration features including $\text{BR}_{\text{coeff}}$, BR, $\text{DC}_{\text{coeff}}$, and DC had slightly better accuracies compared to EEG, considering all of the twelve classification tasks (described in Section ), EEG is still the second best physiological indicator which guided us to combine the APCPS and EEG. Next, we combined the APCPS and EEG signals given in Section  to form a single input and performed all five ML algorithms (k-NN, NB, RF, SVM, MLP) for all six workload, three sense of urgency, and three mind wandering classification tasks (described in Table 8). The results were illustrated in Table 13. As seen, for all workload tasks ($\text{W}_{\text{S}_1\text{-1}}$, $\text{W}_{\text{S}_2\text{-1}}$, $\text{W}_{\text{Comb-1}}$, $\text{W}_{\text{S}_1\text{-2}}$, $\text{W}_{\text{S}_2\text{-2}}$, and $\text{W}_{\text{Comb-2}}$) and mind wandering tasks ($\text{MW}_{\text{S}_1}$, $\text{MW}_{\text{S}_2}$, and $\text{MW}_{\text{Comb}}$), combining the APCPS and EEG did not lead to better performance compared to APCPS alone. A similar phenomenon can be observed in a sense of urgency when combining APCPS and Mean FixDur which did not significantly improve the overall classification performance. For example, the highest accuracy for the APCPS in task $\text{W}_{\text{S}_1\text{-2}}$ is $93.25 \mp 3.55$ (see Table 10a). However, the highest accuracy in task $\text{W}_{\text{S}_1\text{-2}}$ by combining the APCPS and EEG is just $89.20 \mp 4.62$ (see Table 13) which is 4% lower.

**Table 12.** Classification accuracies (in %) of the tasks **(a)** $MW_{S_1}$, **(b)** $MW_{S_2}$, and **(c)** $MW_{Comb}$ described in Table 8.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | 73.10 ∓ 9.75 | 74.27 ∓ 6.35 | **76.22** ∓ 4.17 | 75.50 ∓ 8.33 | 75.38 ∓ 8.52 |
| | BlinkNum | 45.61 ∓ 9.92 | 51.00 ∓ 6.17 | 48.71 ∓ 6.52 | 39.25 ∓ 12.40 | 55.38 ∓ 1.95 |
| | Mean BlinkDur | 47.37 ∓ 10.53 | 52.85 ∓ 6.33 | 57.13 ∓ 9.44 | 39.52 ∓ 12.82 | 49.52 ∓ 4.75 |
| | FixNum | 51.46 ∓ 6.92 | 46.10 ∓ 3.90 | 49.66 ∓ 8.27 | 54.72 ∓ 4.98 | 59.50 ∓ 4.48 |
| | Mean FixDur | 51.46 ∓ 12.84 | 54.71 ∓ 13.04 | 50.00 ∓ 6.47 | 49.16 ∓ 3.81 | 55.49 ∓ 6.32 |
| **EEG** | EEG | 59.72 ∓ 2.09 | 58.27 ∓ 7.06 | 58.15 ∓ 4.27 | 57.87 ∓ 4.11 | 56.57 ∓ 3.44 |
| **fNIRS** | fNIRS | 48.32 ∓ 4.57 | 54.23 ∓ 9.46 | 53.49 ∓ 9.18 | 54.92 ∓ 4.06 | 50.00 ∓ 0.00 |
| **Skin** | Skin | 42.11 ∓ 7.44 | 50.10 ∓ 6.24 | 44.15 ∓ 4.44 | 52.26 ∓ 3.11 | 43.12 ∓ 2.33 |
| **Respiration** | $BR_{coeff}$ | 49.71 ∓ 12.18 | 44.70 ∓ 10.19 | 47.74 ∓ 4.10 | 48.36 ∓ 6.54 | 43.36 ∓ 9.96 |
| | BR | 46.78 ∓ 8.39 | 48.21 ∓ 4.33 | 43.04 ∓ 6.62 | 45.18 ∓ 7.75 | 46.01 ∓ 6.30 |
| | $DC_{coeff}$ | 46.78 ∓ 6.30 | 43.15 ∓ 3.30 | 46.88 ∓ 7.40 | 49.71 ∓ 3.54 | 48.72 ∓ 5.59 |
| | DC | 56.73 ∓ 11.58 | 54.32 ∓ 9.71 | 56.80 ∓ 8.11 | 54.44 ∓ 5.56 | 55.99 ∓ 6.30 |
| | ET | 54.97 ∓ 11.13 | 54.30 ∓ 1.17 | 58.70 ∓ 9.13 | 57.09 ∓ 10.05 | 53.07 ∓ 8.10 |
| | IBI | 67.25 ∓ 6.92 | 61.81 ∓ 7.37 | 65.56 ∓ 4.21 | 64.58 ∓ 1.25 | 64.75 ∓ 8.62 |
| | IE | 64.33 ∓ 9.21 | 65.05 ∓ 8.40 | 64.30 ∓ 3.66 | 61.44 ∓ 7.22 | 61.10 ∓ 14.76 |
| | IT | 61.99 ∓ 9.86 | 60.40 ∓ 5.45 | 60.53 ∓ 2.19 | 60.50 ∓ 5.29 | 59.80 ∓ 5.56 |

**(a)** Classification accuracies (in %) of $MW_{S_1}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | 85.38 ∓ 5.42 | 81.47 ∓ 5.20 | 86.08 ∓ 5.55 | 83.99 ∓ 4.46 | 84.13 ∓ 5.08 |
| | BlinkNum | 46.20 ∓ 11.03 | 44.60 ∓ 3.73 | 52.00 ∓ 5.30 | 48.10 ∓ 9.05 | 57.21 ∓ 13.70 |
| | Mean BlinkDur | 50.88 ∓ 11.37 | 51.15 ∓ 8.30 | 41.90 ∓ 12.17 | 51.60 ∓ 9.32 | 51.06 ∓ 3.34 |
| | FixNum | 42.11 ∓ 6.56 | 44.72 ∓ 3.15 | 45.21 ∓ 5.56 | 49.09 ∓ 10.26 | 46.69 ∓ 9.33 |
| | Mean FixDur | 45.03 ∓ 7.07 | 48.20 ∓ 6.60 | 42.12 ∓ 8.88 | 46.06 ∓ 5.03 | 49.35 ∓ 5.56 |
| **EEG** | EEG | 62.11 ∓ 5.35 | 66.41 ∓ 9.30 | 60.15 ∓ 3.34 | 63.90 ∓ 10.46 | 59.84 ∓ 4.37 |
| **fNIRS** | fNIRS | 56.35 ∓ 8.56 | 46.70 ∓ 3.04 | 56.00 ∓ 6.30 | 58.73 ∓ 6.66 | 56.37 ∓ 4.15 |
| **Skin** | Skin | 51.46 ∓ 14.42 | 52.77 ∓ 5.33 | 39.76 ∓ 12.14 | 51.01 ∓ 9.76 | 50.81 ∓ 4.45 |
| **Respiration** | $BR_{coeff}$ | 43.27 ∓ 6.92 | 51.33 ∓ 5.02 | 49.53 ∓ 8.22 | 45.28 ∓ 4.68 | 49.20 ∓ 7.99 |
| | BR | 45.61 ∓ 8.59 | 41.12 ∓ 11.10 | 49.17 ∓ 7.35 | 45.06 ∓ 6.67 | 48.15 ∓ 7.15 |
| | $DC_{coeff}$ | 50.88 ∓ 8.59 | 53.23 ∓ 9.11 | 51.08 ∓ 6.29 | 52.12 ∓ 4.52 | 52.08 ∓ 6.70 |
| | DC | 53.22 ∓ 10.66 | 55.02 ∓ 6.60 | 54.17 ∓ 7.63 | 51.99 ∓ 3.17 | 54.44 ∓ 7.23 |
| | ET | 60.82 ∓ 9.64 | 61.22 ∓ 12.04 | 58.70 ∓ 3.92 | 61.21 ∓ 7.76 | 60.02 ∓ 3.65 |
| | IBI | 57.31 ∓ 13.46 | 55.12 ∓ 6.72 | 54.35 ∓ 10.16 | 56.32 ∓ 10.11 | 56.60 ∓ 6.24 |
| | IE | 53.22 ∓ 11.22 | 56.12 ∓ 7.77 | 55.56 ∓ 8.21 | 58.25 ∓ 4.44 | 42.25 ∓ 9.39 |
| | IT | 57.31 ∓ 8.39 | 57.62 ∓ 4.30 | 58.70 ∓ 5.91 | 56.32 ∓ 5.40 | 56.35 ∓ 7.07 |

**(b)** Classification accuracies (in %) of $MW_{S_2}$.

| Types of signals | Features | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| **Pupillometry** | APCPS | **82.08** ∓ 7.35 | 80.51 ∓ 6.54 | 71.83 ∓ 7.61 | 77.24 ∓ 5.70 | 79.94 ∓ 6.98 |
| | BlinkNum | 42.96 ∓ 9.60 | 47.51 ∓ 5.84 | 45.66 ∓ 7.01 | 48.22 ∓ 8.33 | 47.51 ∓ 5.90 |
| | Mean BlinkDur | 53.34 ∓ 9.23 | 50.36 ∓ 5.35 | 54.77 ∓ 8.54 | 49.93 ∓ 4.42 | 48.93 ∓ 4.63 |
| | FixNum | 47.37 ∓ 8.34 | 47.51 ∓ 5.57 | 48.93 ∓ 5.03 | 45.95 ∓ 6.01 | 50.07 ∓ 2.68 |
| | Mean FixDur | 49.22 ∓ 7.62 | 48.93 ∓ 6.38 | 45.52 ∓ 8.65 | 45.23 ∓ 5.39 | 49.79 ∓ 4.32 |
| **EEG** | EEG | 59.36 ∓ 5.30 | 59.24 ∓ 6.63 | 61.62 ∓ 7.37 | 61.17 ∓ 6.92 | 63.73 ∓ 8.10 |
| **fNIRS** | fNIRS | 49.67 ∓ 5.28 | 48.52 ∓ 6.70 | 52.21 ∓ 7.23 | 52.81 ∓ 5.36 | 48.51 ∓ 5.30 |
| **Skin** | Skin | 43.95 ∓ 7.74 | 51.49 ∓ 6.23 | 43.24 ∓ 6.79 | 51.78 ∓ 7.76 | 50.78 ∓ 2.41 |
| **Respiration** | $BR_{coeff}$ | 50.07 ∓ 5.77 | 45.80 ∓ 6.04 | 51.78 ∓ 7.08 | 45.95 ∓ 8.59 | 50.36 ∓ 2.51 |
| | BR | 45.09 ∓ 7.13 | 53.06 ∓ 7.75 | 47.80 ∓ 8.81 | 45.23 ∓ 5.67 | 50.07 ∓ 2.03 |
| | $DC_{coeff}$ | 49.07 ∓ 5.63 | 48.08 ∓ 4.35 | 48.51 ∓ 6.04 | 46.37 ∓ 4.98 | 47.65 ∓ 3.84 |
| | DC | 54.34 ∓ 6.44 | 54.77 ∓ 5.87 | 54.62 ∓ 6.94 | 50.92 ∓ 4.58 | 53.63 ∓ 6.09 |
| | ET | 58.61 ∓ 6.96 | 61.45 ∓ 6.37 | 55.05 ∓ 5.35 | 62.59 ∓ 6.81 | 52.20 ∓ 7.18 |
| | IBI | 58.04 ∓ 6.98 | 60.88 ∓ 6.28 | 59.89 ∓ 4.98 | 59.03 ∓ 6.28 | 57.61 ∓ 7.94 |
| | IE | 56.47 ∓ 7.06 | 58.04 ∓ 6.22 | 55.48 ∓ 8.52 | 63.16 ∓ 7.55 | 62.30 ∓ 7.31 |
| | IT | 56.19 ∓ 7.26 | 61.31 ∓ 6.01 | 54.91 ∓ 6.91 | 59.74 ∓ 5.11 | 57.18 ∓ 7.71 |

**(c)** Classification accuracies (in %) of $MW_{Comb}$.

**Table 13.** Classification accuracies (in %) by combining APCPS and other physiological signals (EEG, Mean FixDur) for six workload, three sense of urgency, and three mind wandering classification tasks defined in Table 8 by different ML methods.

| Signal Modality | Tasks | k-NN | NB | RF | SVM | MLP |
|---|---|---|---|---|---|---|
| Workload (APCPS+EEG) | $W_{S_1\text{-}1}$ | $64.56 \mp 8.02$ | $66.50 \mp 4.54$ | $65.15 \mp 8.33$ | $67.12 \mp 3.22$ | $67.30 \mp 7.36$ |
| | $W_{S_2\text{-}1}$ | $57.92 \mp 5.80$ | $56.03 \mp 1.72$ | $52.44 \mp 4.67$ | $54.81 \mp 5.52$ | $57.80 \mp 6.27$ |
| | $W_{\text{Comb-1}}$ | $61.03 \mp 5.93$ | $60.97 \mp 4.63$ | $52.04 \mp 6.34$ | $63.26 \mp 5.63$ | $62.38 \mp 4.16$ |
| Workload (APCPS+EEG) | $W_{S_1\text{-}2}$ | $86.55 \mp 10.10$ | $88.63 \mp 9.21$ | $89.41 \mp 4.19$ | $85.97 \mp 6.35$ | $89.20 \mp 4.62$ |
| | $W_{S_2\text{-}2}$ | $83.94 \mp 11.26$ | $84.77 \mp 9.06$ | $85.05 \mp 6.15$ | $85.44 \mp 5.48$ | $84.64 \mp 4.10$ |
| | $W_{\text{Comb-2}}$ | $85.77 \mp 5.40$ | $86.34 \mp 3.62$ | $85.06 \mp 4.26$ | $87.90 \mp 6.43$ | $84.56 \mp 3.54$ |
| Sense of Urgency (APCPS+Mean FixDur) | $SoU_{S_1}$ | $51.08 \mp 10.23$ | $53.19 \mp 12.06$ | $51.24 \mp 9.17$ | $52.54 \mp 9.55$ | $49.52 \mp 12.26$ |
| | $SoU_{S_2}$ | $61.05 \mp 11.14$ | $62.43 \mp 8.27$ | $56.02 \mp 9.79$ | $61.77 \mp 7.98$ | $58.94 \mp 9.09$ |
| | $SoU_{\text{Comb}}$ | $58.19 \mp 7.66$ | $61.32 \mp 5.91$ | $52.91 \mp 5.37$ | $59.34 \mp 5.46$ | $58.31 \mp 5.95$ |
| Mind Wandering (APCPS+EEG) | $MW_{S_1}$ | $69.80 \mp 11.35$ | $73.16 \mp 8.92$ | $74.93 \mp 7.38$ | $73.02 \mp 5.55$ | $74.36 \mp 6.27$ |
| | $MW_{S_2}$ | $85.40 \mp 7.65$ | $83.29 \mp 9.43$ | $81.24 \mp 3.76$ | $83.91 \mp 2.15$ | $82.87 \mp 6.72$ |
| | $MW_{\text{Comb}}$ | $79.09 \mp 4.90$ | $77.23 \mp 4.30$ | $70.97 \mp 6.74$ | $78.09 \mp 4.49$ | $76.66 \mp 6.79$ |

However, it is worth noting that the performance of this combination largely outperforms the performance obtained by EEG only. For example, the highest accuracy for EEG in task $W_{S_2\text{-}2}$ (see Table 10b) is just $58.15 \mp 2.65$, however, the highest accuracy in task $W_{S_2\text{-}2}$ by combining the APCPS and EEG is $85.44 \mp 5.48$ (see Table 13). Again, this shows that APCPS, and not EEG, is really the important modality responsibility for the high classification performance.

Note that combining EEG and pupillometry to characterize the cognitive states has been previously considered in [1, 2, 27, 79], and a few of them stated that combining these two signals may not lead to any improvement in cognitive state classification accuracy (e.g., see [79]). Here, since combining the APCPS and EEG neither improves the classification accuracy reduces the performance in the rest, we do not see a clear gain of combining multiple physiological signals in our experiment.

## Statistical Performance Analyses

We further performed statistical analyses on the different performance of the the five ML algorithms to determine which of them had the overall highest performance. Table 14a, Table 14b, Table 14c, and Table 14d depict the results of the paired sampled t-tests performed on the accuracy arrays obtained from five ML classifiers corresponding to 3-class workload classification, 2-class workload classification, sense of urgency classification, and mind wandering classification, respectively. In particular, for a given type of signal (the APCPS) and a learning algorithm, we repeated the experiment 19 times (each had the 5-fold and provided one accuracy in the test) to collect 19 classification accuracies. Then, we run the paired sampled t-test and obtained the p-values (at a significance level of .95) to investigate the similarities between different pairs of ML algorithms. In addition, we presented the mean of the accuracy arrays acquired from five ML classifiers in Table 15 along with Table 14 to further determine the best and the worst algorithms. The results indicated that we can differentiate RF from all of the other four algorithms in the case of 3-class workload classification and mind wandering classification. On the other hand, there is no consistency in the p-values related to 2-class classification and sense of urgency classification. In overall, we observed that except for RF (in 3-class workload classification and mind wandering classification), there is no statistically significant

**Table 14.** The results of the paired sampled t-test on the accuracy arrays obtained from five ML classifiers to evaluate the similarities of their efficiencies on (a) $W_{\text{Comb-1}}$: 3-class workload classification, (b) $W_{\text{Comb-2}}$: 2-class workload classification, (c) $SoU_{\text{Comb}}$: sense of urgency classification, and (d) $MW_{\text{Comb}}$: mind wandering classification. Here, we marked the cells with red which include p-values less than .05 (statistically significant).

| | **k-NN** | **NB** | **RF** | **SVM** |
|---|---|---|---|---|
| **NB** | 0.77 | - | - | - |
| **RF** | $\mathbf{1.7 \times 10^{-4}}$ | $\mathbf{5.3 \times 10^{-4}}$ | - | - |
| **SVM** | 0.74 | 0.99 | $\mathbf{3.3 \times 10^{-5}}$ | - |
| **MLP** | 0.78 | 0.96 | $\mathbf{4.3 \times 10^{-4}}$ | 0.96 |

**(a)** $W_{\text{Comb-1}}$.

| | **k-NN** | **NB** | **RF** | **SVM** |
|---|---|---|---|---|
| **NB** | 0.21 | - | - | - |
| **RF** | **0.01** | 0.25 | - | - |
| **SVM** | 0.17 | 0.61 | $\mathbf{9.0 \times 10^{-2}}$ | - |
| **MLP** | $\mathbf{3.6 \times 10^{-3}}$ | 0.26 | 0.79 | $\mathbf{9.8 \times 10^{-2}}$ |

**(b)** $W_{\text{Comb-2}}$.

| | **k-NN** | **NB** | **RF** | **SVM** |
|---|---|---|---|---|
| **NB** | 0.21 | - | - | - |
| **RF** | 0.18 | **0.03** | - | - |
| **SVM** | 0.99 | 0.21 | 0.16 | - |
| **MLP** | 0.61 | 0.06 | 0.51 | 0.67 |

**(c)** $SoU_{\text{Comb}}$.

| | **k-NN** | **NB** | **RF** | **SVM** |
|---|---|---|---|---|
| **NB** | 0.46 | - | - | - |
| **RF** | $\mathbf{1.3 \times 10^{-3}}$ | $\mathbf{4.8 \times 10^{-3}}$ | - | - |
| **SVM** | $\mathbf{2.6 \times 10^{-2}}$ | 0.12 | **0.04** | - |
| **MLP** | 0.32 | 0.76 | $\mathbf{1.5 \times 10^{-3}}$ | 0.14 |

**(d)** $MW_{\text{Comb}}$.

difference between the accuracies of the ML algorithms used in this study for classifying different workload and mind wandering levels. Based on Table 15, we can claim that RF is the worst algorithm in achieving different cognitive state classification tasks.

## Summary of the Observations

Given the numerical results from the workload, sense of urgency, and mind-wandering classification tasks, we can summarize the main observations as follows:

1. There is no difference in performance among the five algorithms for a given type of signal and classification task except for RF which had lower accuracies for both 3-class workload classification and mind wandering classification tasks (see Tables 14a, 14d, and 15).

2. The APCPS outperforms all other physiological signals in differentiating three cognitive states by a large margin for all six workload, three sense of urgency, and three mind wandering levels regardless of learning algorithms.

**Table 15.** The mean of the accuracy arrays obtained from five ML classifiers to evaluate the similarities of their effects on (a) $W_{\text{Comb-1}}$: 3-class workload classification, (b) $W_{\text{Comb-2}}$: 2-class workload classification, (c) $SoU_{\text{Comb}}$: sense of urgency classification, and (d) $MW_{\text{Comb}}$: mind wandering classification.

| | **k-NN** | **NB** | **RF** | **SVM** | **MLP** |
|---|---|---|---|---|---|
| $W_{\text{Comb-1}}$ | 62.78 | **63.35** | 53.64 | **63.35** | 63.25 |
| $W_{\text{Comb-2}}$ | **90.18** | 87.90 | 85.35 | 88.76 | 85.71 |
| $SoU_{\text{Comb}}$ | 57.04 | 60.46 | 54.05 | 57.04 | 55.75 |
| $MW_{\text{Comb}}$ | **82.08** | 80.51 | 71.83 | 77.24 | 79.94 |

3. Combining the APCPS and EEG signals does not really boost the classification performances in workload, sense of urgency, and mind-wandering classification tasks.

It is worth noting that our numerical results from the machine learning models also agreed with our statistical analysis performed on the APCPS signal in Section . Indeed, as previously mentioned in Section , because of the training effect in session 1 where the participants first encountered the primary driving task and secondary tasks, leading to more attention and awareness, one may expect a higher workload in session 1 compared to session 2, specifically during $l_{w_1}$ where the participants were first instructed to complete the braking events and dialogue communications (see Figure S8 Fig). This has been observed in our modeling results, for example, by comparing the classification accuracy of the APCPS in task $W_{S_1-1}$ (Table 9a) and in task $W_{S_2-1}$ (Table 9b). As seen, for a given task of classifying between three workload levels, a classifier based on the APCPS signal extracted from session 1 can achieve the highest accuracy of $72.57 \mp 8.75$, while the highest accuracy of the APCPS signal extracted from session 2 is far below which is just $59.13 \mp 8.29$. A similar observation can be concluded by comparing the classification accuracy of APCPS in task $W_{S_1-2}$ (Table 10a) and in task $W_{S_2-2}$ (Table 10b). Particularly, for a given task of classifying between the highest and lowest workload levels, a classifier based on the APCPS signal extracted from session 1 can achieve the highest accuracy of $93.25 \mp 3.55$ in task $W_{S_1-2}$ which is, again, a bit higher than the highest accuracy of $88.89 \mp 5.23$ on the APCPS signal extracted from session 2 in task $W_{S_2-2}$.

**Fig 8. The change in the mean APCPS related to different cognitive workload levels for two sessions.** Boxplot of the mean APCPS values corresponding to different sessions of three cognitive workload levels.

# Discussion

We have explored in great detail the utility of multiple physiological signal modalities based on pupillometry, EEG, fNIRS, skin conductance and respiration for assessing three systemic human cognitive states—cognitive workload, sense of urgency, and mind wandering—by using multiple pairwise statistical tests and state-of-the-art machine learning models. Our statistical analyses results demonstrated that APCPS is the best physiological signal for distinguishing the three systemic cognitive states. We then examined the performance of various machine learning models (k-NN, NB, RF, SVM, and MLP) on the full set of physiological signal types for differentiating different levels of workload, sense or urgency, and mind wandering. Specifically, we performed four cognitive state classification tasks including 3-class workload, 2-class workload, 2-class sense of urgency, and 2-class mind wandering classification which showed that APCPS

is the most reliable physiological indicator for achieving high performance in all four classification tasks regardless of which machine learning method was used. In addition to investigating the performance of individual physiological signal modalities in assessing different human cognitive state levels, we also analyzed the efficiency of combining the best two signal types—eye gaze and EEG—for the prediction of different human cognitive states and observed a similar phenomenon: that fusing different physiological signal markers did not improve the classification performance compared to only using eye gaze.

Besides individual cognitive state estimation tasks, we also investigated the association between workload and sense of urgency. Particularly, we analyzed how sequential occurrences of the sense of urgency might impact the overall workload. We considered the APCPS as the most suitable physiological marker (based on its superior performance on cognitive state classification tasks) for exploring the relationship between workload and sense or urgency, and found that consecutive triggering of sense of urgency contributed to overall workload increase.

Finally, we performed a statistical comparison of different ML algorithms, repeating the experiment multiple times by using the APCPS (each time we performed 5-fold cross validation and obtained one accuracy in the test) for four classification tasks (defined in Section ) and ran paired t-test within five ML algorithms. We found that except for RF, the remainder of four ML algorithms show roughly equal performance in terms of 3-class workload and 2-class sense of urgency classification accuracy. Moreover, all of the five ML algorithms are comparable in accomplishing 2-class workload and 2-class mind wandering classification. The results indicate that APCPS is able to achieve high prediction performance irrespective of which ML method is used. This result is especially important for ambulatory humans in mixed-initiative teams where certain types of physilogical sensor instrumentations are not possible (e.g., EEG and fNIRS would cause large motion artifacts that might overshadow any changes in the signals that could reflect cognitive state changes). Being able to use eye gaze to track the human workload and sense of urgency as well as mind wandering will allow autonomous systems to better estimate human situational awareness, to adapt their interactions based on the detect levels of human systemic cognitive states, and to utilize opportunities to intervene to address possible degradation of human performance.

**Limitations and Future Work.** The outcomes of this study are important for future efforts to assess systemic human cognitive states in real-time considering that human gaze is easy to collect and fast to process compared to other physiological signals. To do so, future work will need to develop methods for real-time assessment and tracking of these human cognitive states, and possibly additional states such as interference and distraction which have been shown to impact human task performance as well. It is also worth pointing out that the dataset utilized in this study was collected from participants who were mostly undergraduate and graduate students with an average age of just 20. In future work, it would be important to verify that the results obtained based on this dataset also hold for a broader set of human participants for whom we expect to see even greater variations in the various cognitive state levels. It would be important to consider additional systemic cognitive states such as distraction (e.g., [80] or interference (e.g., [81]). Finally, while the characteristics of current task setting (ongoing manual actions based on visual perceptions, intermittent spoken dialogue interactions) was an instance of a fairly large class of tasks, it would be interesting to consider other tasks with different characteristics to see whether the results would transfer.

## Conclusions

The aim of this study was to investigate the utility of various physiological signal types such as pupillometry, EEG, fNIRS, skin conductance, and respiration for assessing different human systemic cognitive states like workload, sense of urgency, and mind wandering in a multimodal driving simulation environment. Our statistical analyses and classification results indicate that the "percent change in pupil size" is the single most reliable physiological marker for predicting different levels of workload, sense of urgency and mind wandering. We also explored the change in workload relative to changes in sense of urgency by utilizing the PCPS and determined that frequent and consecutive instances of higher sense of urgency are likely to increase the overall workload. Based on the significant differences in levels of systemic cognitive states, we examined the performance of five different state-of-the-art machine learning models to determine which would be best in predicting those levels and found that except for one ("random forest"), all other four models (k-neareast neighbor, naive Bayes, support vector machines, and muli-layer perceptron) are equally good at classifying the various levels of three systemic cognitive states, critically in data from novel subjects on which these models were not trained, thus showing their potential for state classification in general. While the results were obtained using data from only one task, we believe that they have the potential to generalize to other tasks and thus for improving human performance in safety-critical applications as well as in human-machine teams.

## Supporting information

**S1 Fig. The relationship between Different Cognitive State Pairs.** An illustration of the hypothesized relationships among pairs of three systemic cognitive states: workload, sense of urgency, and mind wandering.

**S2 Fig. An illustration of a braking event.** An example of a braking event: (a) a participant drives on the right lane of a straight road, (b) a car appears in front of the participant's car, and (c) the front car immediately decelerates, then moves far away.

**S3 Fig. Evaluation of three cognitive workload levels.** By selecting different combinations of braking and dialogue events, three different cognitive workload levels were generated.

**S4 Fig. The change in the mean APCPS related to different cognitive workload levels.** Boxplot of the mean APCPS values corresponding to different cognitive workload levels including both sessions from all of the subjects.

**S5 Fig. The change in the mean APCPS related to different sense of urgency levels.** Boxplot of the mean APCPS corresponding to two sense of urgency levels including both sessions from all of the subjects. $l_{sou_1}$: Lower sense of urgency, $l_{sou_2}$: Higher sense of urgency.

**S6 Fig. The change in the mean APCPS related to different mind wandering levels.** Boxplot of the mean APCPS corresponding to two mind wandering levels including both sessions from all of the subjects. $l_{mw_1}$: Lower mind wandering, $l_{mw_2}$: Higher mind wandering.

**S7 Fig. Illustration of the change in PCPS with different events.** Illustration of the change in PCPS with different events including braking and dialogue. The results include two sessions obtained from two different subjects. B and D represent the onset of the braking event and the onset of the dialogue event, respectively.

**S8 Fig. The change in the mean APCPS related to different cognitive workload levels for two sessions.** Boxplot of the mean APCPS values corresponding to different sessions of three cognitive workload levels.

## Acknowledgments

This project was in part supported by AFOSR grant #FA9550-18-1-0465.

## References

1. Christensen, J. C.,Estepp, J. R., Wilson, G. F., Russell, C. A. The effects of day-to-day variability of physiological data on operator functional state classification. NeuroImage. 2012; 59(1):57–63.

2. Rozado, D., Dunser, A. Combining EEG with pupillometry to improve cognitive workload detection. Computer. 2015; 48(10):18–25.

3. Scheutz M., DeLoach S., Adams J. A Framework for developing and using shared mental models in human-agent teams. Journal of Cognitive Engineering and Decision Making. 2017; 11(3): 203–224.

4. Howard Z. L., Innes R., Eidels A., Loft S. Using Past and Present Indicators of Human Workload to Explain Variance in Human Performance. Psychonomic Bulletin & Review. 2021; 28(6):1923–1932.

5. Girardeau, J. C., Ledru, R., Gaston-Bellegarde, A., Blondé, P., Sperduti, M., Piolino, P. The benefits of mind wandering on a naturalistic prospective memory task. Nature Publishing Group UK London. 2023; 13(1):e11432.

6. Grimes D., Tan D. S., Hudsom S. E., Shenoy P., Rao R. PN. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008;835–844.

7. Paas F., Tuovinen J. E., Tabbers H., Van Gerven P. WM. Cognitive load measurement as a means to advance cognitive load theory. Educational psychologist. 2003; 38(1):63–71.

8. Heard J., Harriott C. E., Adams J. A survey of workload assessment algorithms. IEEE Transactions on Human-Machine Systems. 2018; 48(5):434–451.

9. Berka C., Levendowski D. J., Lumicao M. N., Yau A., Davis G., Zivkovic V. T., Olmstead R. E., Tremoulet P. D., Craven P. L. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation, space, and environmental medicine. 2007; 78(5):8231–8244.

10. Scheutz M., Aeron S., Aygun A., de Ruiter JP., Fantini S., Fernandez C., Haga Z., Nguyen T., Lyu B. Estimating Systemic Cognitive States from a Mixture of Physiological and Brain Signals. Topics in Cognitive Science; 2023.

11. Mercado, J. E., Rupp, M. A., Chen, J. YC., Barnes, M. J., Barber, D., Procci, K. Intelligent agent transparency in human-agent teaming for Multi-UxV management. Human factors. 2016; 58(3):401–415.

12. Roy, R. N., Drougard N., Gateau, T., Dehais, F., Chanel, C. PC. How can physiological computing benefit human-robot interaction?. Robotics. 2020; 9(4):100.

13. Fiore, S. M., Wiltshire, T. J., Lobato, E. JC., Jentsch, F. G., Huang, W. H., Axelrod, B. Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. Frontiers in psychology. 2013; 4:859.

14. Aygun, A., Lyu, B., Nguyen, T., Haga, Z., Aeron, S., Scheutz, M. Cognitive Workload Assessment via Eye Gaze and EEG in an Interactive Multi-Modal Driving Task. Proceedings of the 2022 International Conference on Multimodal Interaction. 2022; 337–348.

15. Christoff, K., Irving, Z. C., Fox, K. CR., Spreng, R. N., Andrews-Hanna, J. R. Mind-wandering as spontaneous thought: a dynamic framework. Nature reviews neuroscience. 2016; 17(11):718–731.

16. Hart, S. G., Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology. 1988; 52:139–183.

17. Cain, B. A review of the mental workload literature. DTIC Document; 2007.

18. Yau, Y., Hinault, T., Taylor, M., Cisek, P., Fellows, L. K., Dagher, A. Evidence and urgency related EEG signals during dynamic decision-making in humans. Journal of Neuroscience. 2021; 41(26):5711–5722.

19. Biondi, F., Leo, M., Gastaldi, M., Rossi, R., Mulatti, C. How to drive drivers nuts: effect of auditory, vibrotactile, and multimodal warnings on perceived urgency, annoyance, and acceptability. Transportation research record. 2017; 2663(1):34–39.

20. Smallwood, J., Schooler, J. W. The science of mind wandering: Empirically navigating the stream of consciousness. Annual review of psychology. 2015; 66:487–518.

21. Giambra, L. M. Task-unrelated thought frequency as a function of age: a laboratory study. Psychology and aging. 1989; 4(2):136.

22. Antrobus, J. S., Singer, J. L., Greenberg, S., et al. Studies in the stream of consciousness: Experimental enhancement and suppression of spontaneous cognitive processes. Perceptual and Motor Skills. 1966; 23(2):399–417.

23. Zhang, Y., Kumada, T. Relationship between workload and mind-wandering in simulated driving. PloS one. 2017; 12(5):e0176962.

24. Lavie, N. Attention, distraction, and cognitive control under load. Current directions in psychological science. 2010; 19(3):143–148.

25. Latinjak, A. T. Goal-directed, spontaneous, and stimulus-independent thoughts and mindwandering in a competitive context. The Sport Psychologist. 2018; 32(1):51–59.

26. Girardeau, JC., Ledru, R., Gaston-Bellegarde, A., Blondé, P., Sperduti, M., Piolino, P. The benefits of mind wandering on a naturalistic prospective memory task. Scientific Reports. 2023; 13(1):11432.

27. Khedher, A. B., Jraidi, I., Frasson, C. Predicting learners' performance using EEG and eye tracking features. The Thirty-Second International Flairs Conference; 2019.

28. Tokuda, S., Obinata, G., Palmer, E., Chaparro, A. Estimation of mental workload using saccadic eye movements in a free-viewing task. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011; 4523–4529.

29. Bodala, I. P., Ke, Y., Mir, H., Thakor, N. V., Al-Nashash, H. Cognitive workload estimation due to vague visual stimuli using saccadic eye movements. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014; 2993–2996.

30. Khalaf, A., Nabian, M., Fan, M., Yin, Y., Wormwood, J., Siegel, E., Quigley, K. S., Barrett, L. F., Akcakaya, M., Chou, C., et al. Analysis of multimodal physiological signals within and between individuals to predict psychological challenge vs. threat. Expert Systems with Applications. 2020; 140:112890.

31. Šalkevicius, J., Damaševičius, R., Maskeliunas, R., Laukienė, I. Anxiety level recognition for virtual reality therapy system using physiological signals. Electronics. 2019; 8(9):1039.

32. Bixler, R., Blanchard, N., Garrison, L., D'Mello, S. Automatic detection of mind wandering during reading using gaze and physiology. Proceedings of the 2015 ACM on international Conference on Multimodal Interaction. 2015; 299–306.

33. He, J., Becic, E., Lee, Y., McCarley, J. S. Mind wandering behind the wheel: Performance and oculomotor correlates. Human factors. 2011; 53(1):13—21.

34. Klesel, M., Schlechtinger, M., Oschinsky, F. M., Conrad, C., Niehaves, B. Detecting mind wandering episodes in virtual realities using eye tracking. Information Systems and Neuroscience: NeuroIS Retreat. 2020; 163–171.

35. Zhou, Y., Huang, S., Xu, Z., Wang, P., Wu, X., Zhang, D. Cognitive workload recognition using EEG signals and machine learning: A review. IEEE Transactions on Cognitive and Developmental Systems; 2021.

36. So, W. KY., Wong, S. WH., Mak, J. N., Chan, R. HM. An evaluation of mental workload with frontal EEG. PloS one. 2017; 12(4):e0174949.

37. Zarjam, P., Epps, J., Chen, F. Characterizing working memory load using EEG delta activity. 2011 19th European signal processing conference. 2011; 1554–1558.

38. Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., Caltagirone, C., Kong, W., Wei, D., Zhou, Z., et al. Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. 2012 annual international conference of the IEEE engineering in medicine and biology society. 2012; 6442–6445.

39. Qu, H., Shan, Y., Liu, Y., Pang, L., Fan, Z., Zhang, J., Wanyan, X. Mental workload classification method based on EEG independent component features. Applied Sciences. 2020; 10(9):3036.

40. Reddy, A. G., Narava, S. Artifact removal from EEG signals. International Journal of Computer Applications. 2013; 77(13).

41. Cheng, S. Evaluation of effect on cognition response to time pressure by using EEG. Advances in Human Factors and Ergonomics in Healthcare and Medical Devices: Proceedings of the AHFE 2017 International Conferences on Human Factors and Ergonomics in Healthcare and Medical Devices. 2017; 45–52.

42. Jin, C. Y., Borst, J. P., Van V., Marieke K. Predicting task-general mind-wandering with EEG. Cognitive, Affective, & Behavioral Neuroscience. 2019; 1059–1073.

43. Conrad, C., Newman, A. Measuring mind wandering during online lectures assessed with EEG. Frontiers in Human Neuroscience. 2021; 15:697532.

44. Dong, H. W., Mills, C., Knight, R. T., Kam, J. WY. Detection of mind wandering using EEG: Within and across individuals. Plos one. 2021; 16(5):e0251490.

45. Aygun, A, Nguyen, T., Haga, Z., Aeron, S., Scheutz, M. Investigating Methods for Cognitive Workload Estimation for Assistive Robots. Sensors. 2022; 22(18):6834.

46. Cao, J., Garro, E. M., Zhao, Y. EEG/FNIRS based workload classification using functional brain connectivity and machine learning. Sensors. 2022; 22(19):7623.

47. Midha, S., Maior, H. A., Wilson, M. L., Sharples, S. Measuring mental workload variations in office work tasks using fNIRS. International Journal of Human-Computer Studies. 2021; 147:102580.

48. Putze, F., Herff, C., Tremmel, C., Schultz, T., Krusienski, D. J. Decoding mental workload in virtual environments: a fNIRS study using an immersive n-back task. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019; 3103–3106.

49. Holtzer, R., Schoen, C., Demetriou, E., Mahoney, J. R., Izzetoglu, M., Wang, C., Verghese, J. Stress and gender effects on prefrontal cortex oxygenation levels assessed during single and dual-task walking conditions. European Journal of Neuroscience. 2017; 45(5):660–670.

50. Mirelman, A., Maidan, I., Bernad-Elazari, H., Nieuwhof, F., Reelick, M., Giladi, N., Hausdorff, J. M. Increased frontal brain activation during walking while dual tasking: an fNIRS study in healthy young adults. Journal of neuroengineering and rehabilitation. 2014; 11:1—7.

51. Prinable, J., Jones, P., Boland, D., Thamrin, C., McEwan, A., et al. Derivation of breathing metrics from a photoplethysmogram at rest: machine learning methodology. JMIR mHealth and uHealth. 2020; 8(7):e13737.

52. Liu, R., Walker, E., Friedman, L., Arrington, C. M., Solovey, E. T. fNIRS-based classification of mind-wandering with personalized window selection for multimodal learning interfaces. Journal on multimodal user interfaces. 2021; 15:257–272.

53. Durantin, G., Dehais, F., Delorme, A. Characterization of mind wandering using fNIRS. Frontiers in systems neuroscience. 2015; 9(45).

54. Shimomura, Y., Yoda, T., Sugiura, K., Horiguchi, A., Iwanaga, K., Katsuura, T. Use of frequency domain analysis of skin conductance for evaluation of mental workload. Journal of physiological anthropology. 2008; 27(4):173—177.

55. Lew, R., Dyre, B. P., Soule, T., Ragsdale, S. A., Werner, S. Assessing mental workload from skin conductance and pupillometry using wavelets and genetic programming. Proceedings of the human factors and ergonomics society annual meeting. 2010; 54(3):254—258.

56. Mehler, B., Reimer, B., Coughlin, J. F., Dusek, J. A. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transportation research record. 2009; 2138(1):6–12.

57. Kurniawan, H., Maslov, A. V., Pechenizkiy, M. Stress detection from speech and galvanic skin response signals. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013; 209–214.

58. Bakker, J., Pechenizkiy, M., Sidorova, N. What's your current stress level? Detection of stress patterns from GSR sensor data. 2011 IEEE 11th international conference on data mining workshops. 2011; 573–580.

59. Blanchard, N., Bixler, R., Joyce, T., D'Mello, S. Automated physiological-based detection of mind wandering during learning. International conference on intelligent tutoring systems. 2014; 55–60.

60. Karavidas, M. K., Lehrer, P. M., Lu, S., Vaschillo, E., Vaschillo, B., Cheng, A. The effects of workload on respiratory variables in simulated flight: a preliminary study. Biological Psychology. 2010; 84(1):157—160.

61. Veltman, JA., Gaillard, AWK. Physiological workload reactions to increasing levels of task difficulty. Ergonomics. 1998; 41(5):656–669.

62. McDuff, D., Gontarek, S., Picard, R. Remote measurement of cognitive stress via heart rate variability. 2014 36th annual international conference of the IEEE engineering in medicine and biology society. 2004; 2957–2960.

63. Vinkers, C. H., Penning, R., Hellhammer, J., Verster, J. C., Klaessens, J. HGM., Olivier, B., Kalkman, C. J. The effect of stress on core and peripheral body temperature in humans. Stress. 2013; 16(5):520–530.

64. Zheng, Y., Wang, D., Zhang, Y., Xu, W. Detecting mind wandering: an objective method via simultaneous control of respiration and fingertip pressure. Frontiers in Psychology. 2019; 10:216.

65. Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., Barsalou, L. W. Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. Neuroimage. 2012; 59(1):750–760.

66. Wei, D., Gong, H., Wu, X. Residents' subjective mental workload during computerized prescription entry. Informatics for Health and Social Care. 2022; 47(3):283–294.

67. Keunecke, J. G., Gall, C., Birkholz, T., Moritz, A., Eiche, C., Prottengeier, J. Workload and influencing factors in non-emergency medical transfers: a multiple linear regression analysis of a cross-sectional questionnaire study. BMC health services research. 2019; 19(1):1–11.

68. Saeedpour-Parizi, M. R., Hassan, S. E., Shea, J. B. Pupil diameter as a biomarker of effort in goal-directed gait. Experimental Brain Research. 2020; 238(11):2615–2623.

69. Wildemeersch, D., Peeters, N., Saldien, V., Vercauteren, M., Hans, G. Pain assessment by pupil dilation reflex in response to noxious stimulation in anaesthetized adults. Acta Anaesthesiologica Scandinavica. 2018; 62(8):1050–1056.

70. Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., Carlson, J. M., Schooler, J. W. Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. Acta Anaesthesiologica Scandinavica. 2011; 6(3):e18298.

71. Zhao, M., Gao, H., Wang, W., Qu, J., Chen, L. Study on the identification of irritability emotion based on the percentage change in pupil size. Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing. 2020; 20–24.

72. Prieur-Coloma, Y., Reinoso-Leblanch, R. A., Mayeta-Revilla, L., Delisle-Rodríguez, D., Bastos, T., López-Delis, A., Balart-Fernández, L., Falk, T. H. Enhancing shoulder pre-movements recognition through EEG Riemannian covariance matrices for a BCI-based exoskeleton. 2020 IEEE International Conference on Human-Machine Systems (ICHMS). 2020; 1–3.

73. Kalman, R. E. A new approach to linear filtering and prediction problems. 1966; 35–45.

74. Cerliani, M. Tsmoothie. `https://github.com/cerlymarco/tsmoothie`. 2021.

75. Noto, T., Zhou, G., Schuele, S., Templer, J., Zelano, C. Automated analysis of breathing waveforms using BreathMetrics: a respiratory signal processing toolbox. Chemical senses. 2018; 43(8):583–597.

76. Abdi, H., Williams, L. J. Newman-Keuls test and Tukey test. Encyclopedia of research design. 2010; 2:897–902.

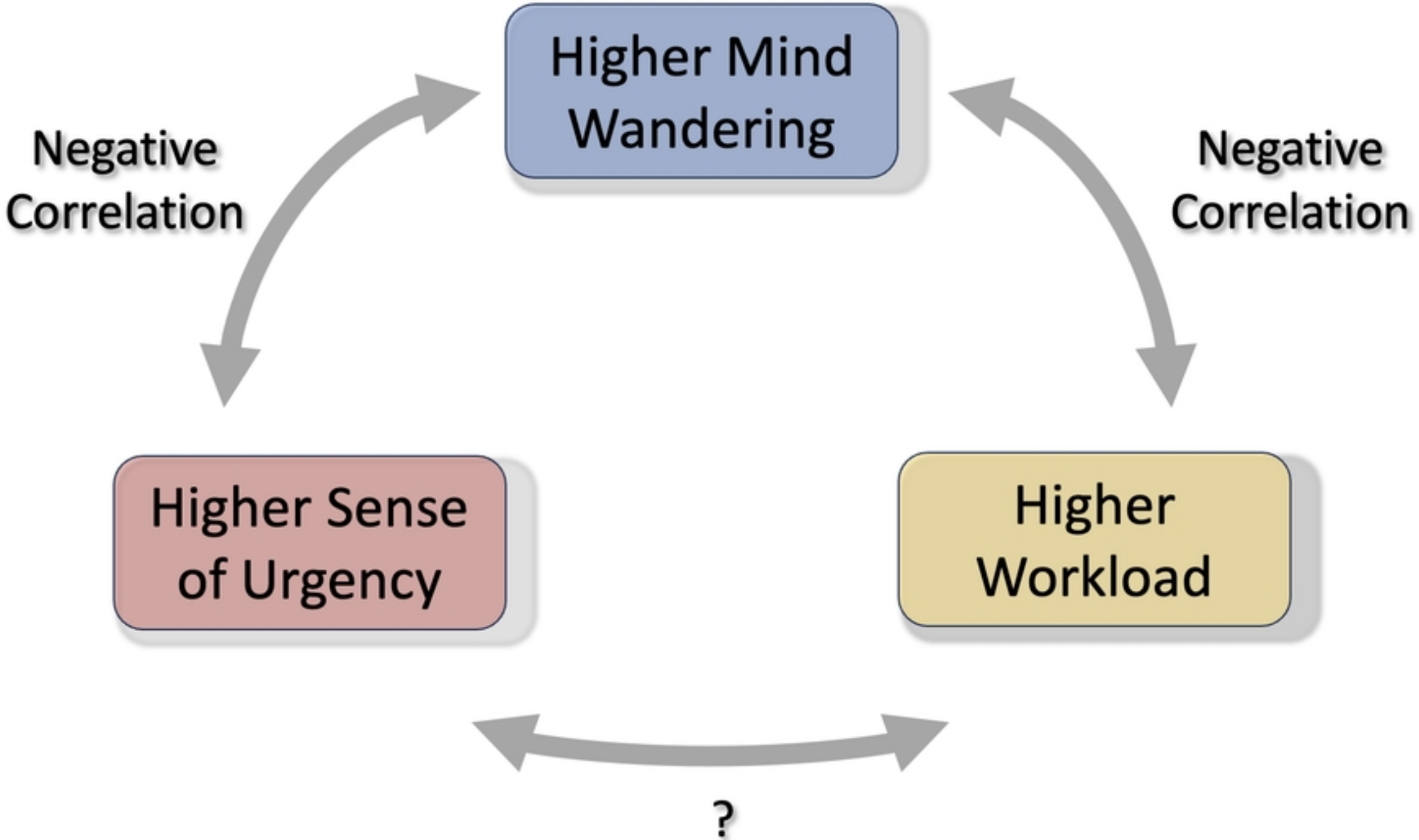
77. Jaccard, J., Becker, M. A., Wood, G. Pairwise multiple comparison procedures: A review. Psychological Bulletin. 1984; 96(3):589.

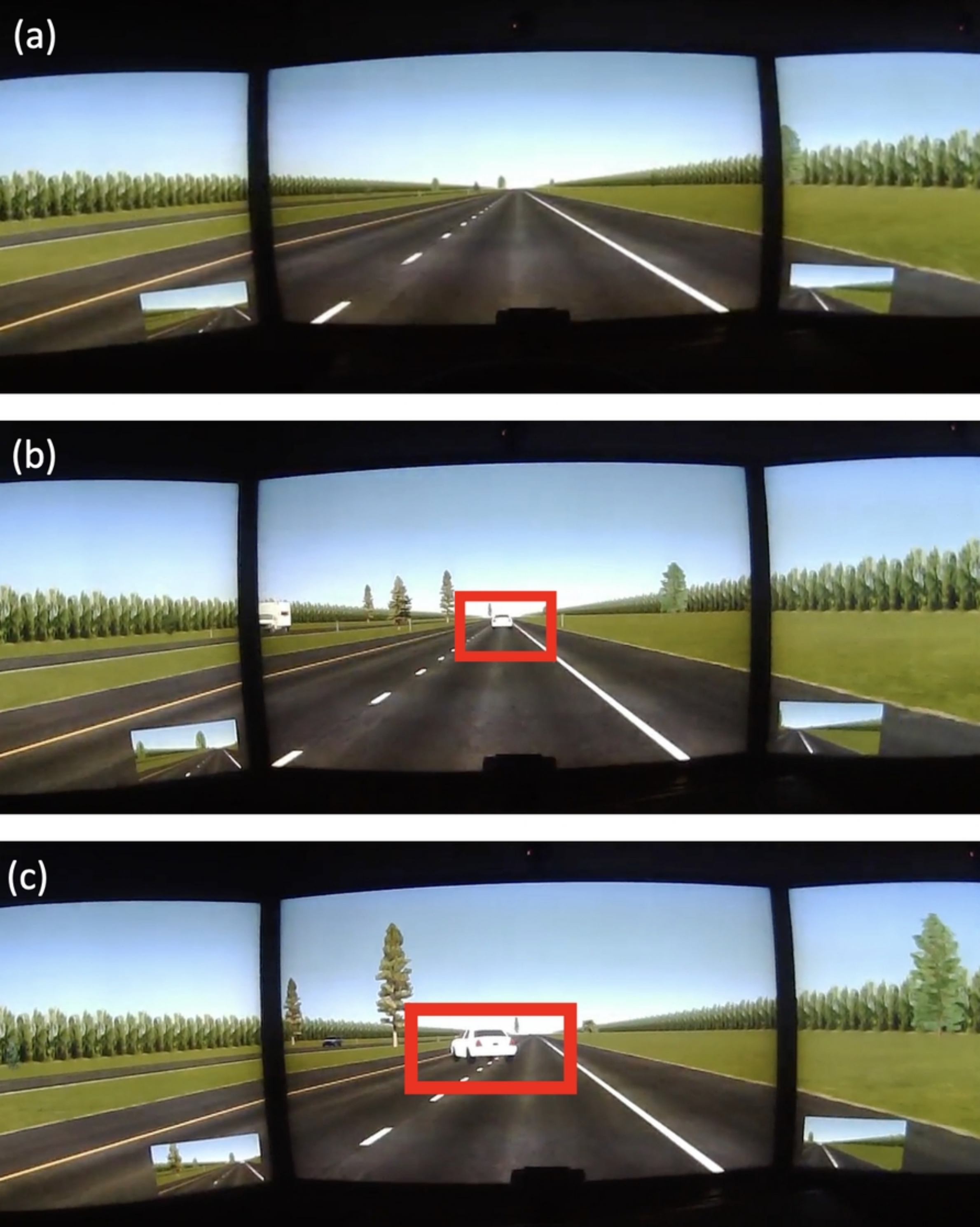
78. Stone, M. Cross-validation: A review. Statistics: A Journal of Theoretical and Applied Statistics. 1978; 9(1):127–139.

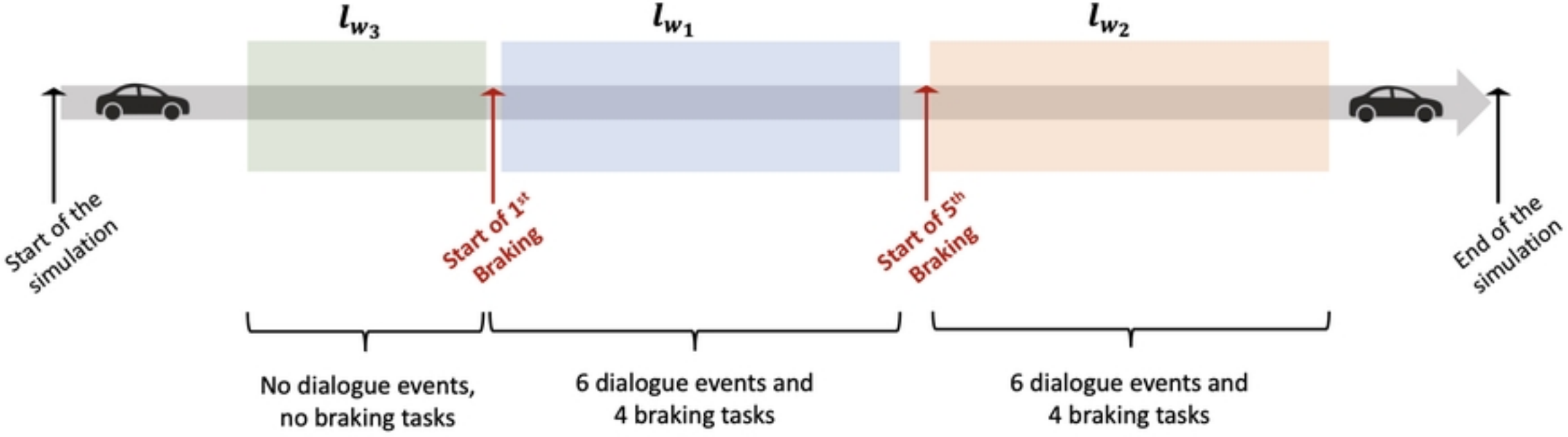
79. Borys, M., Plechawska-Wójcik, M., Wawrzyk, M., Wesołowska, K. Classifying cognitive workload using eye activity and EEG features in arithmetic tasks. Statistics: International conference on information and software technologies. 2017; 90–101.

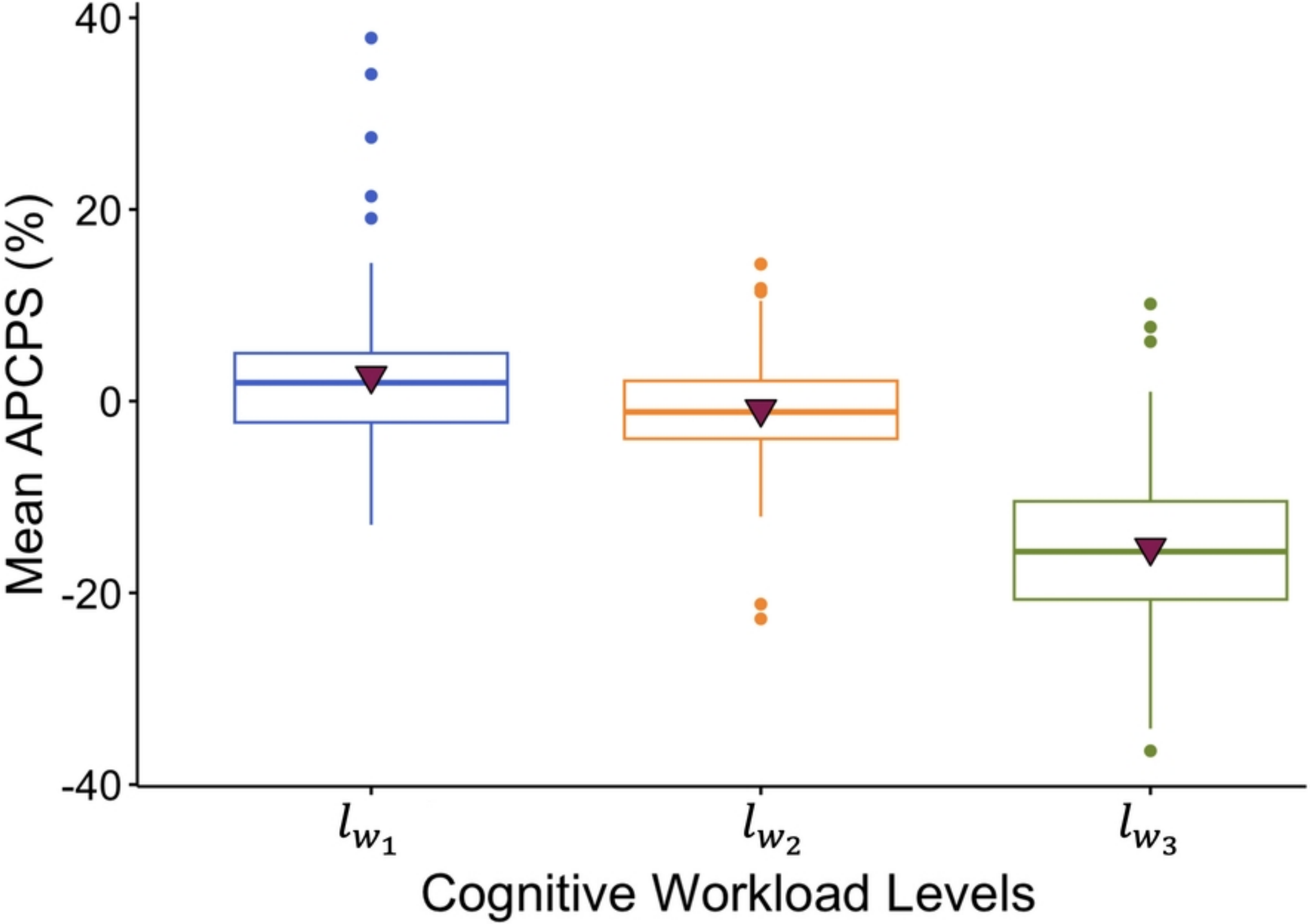
80. Liang, Y., Lee, J. D. Driver cognitive distraction detection using eye movements. Passive Eye Monitoring: Algorithms, Applications and Experiments; 2008.

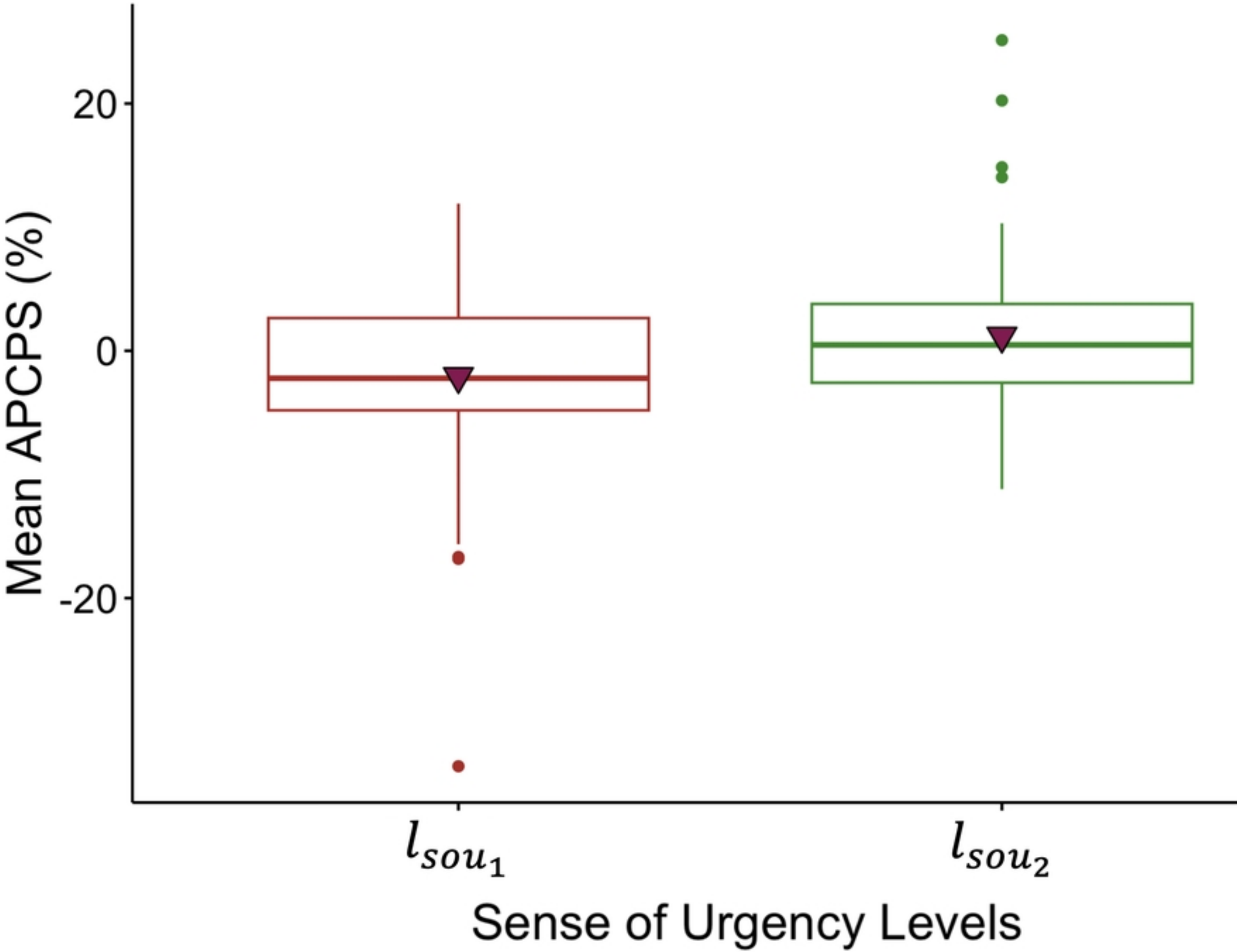
81. Chatham, C., Frank, M. J., Munakata, Y. Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. Proceedings of the National Academy of Sciences. 2009; 106(14):5529–5533.

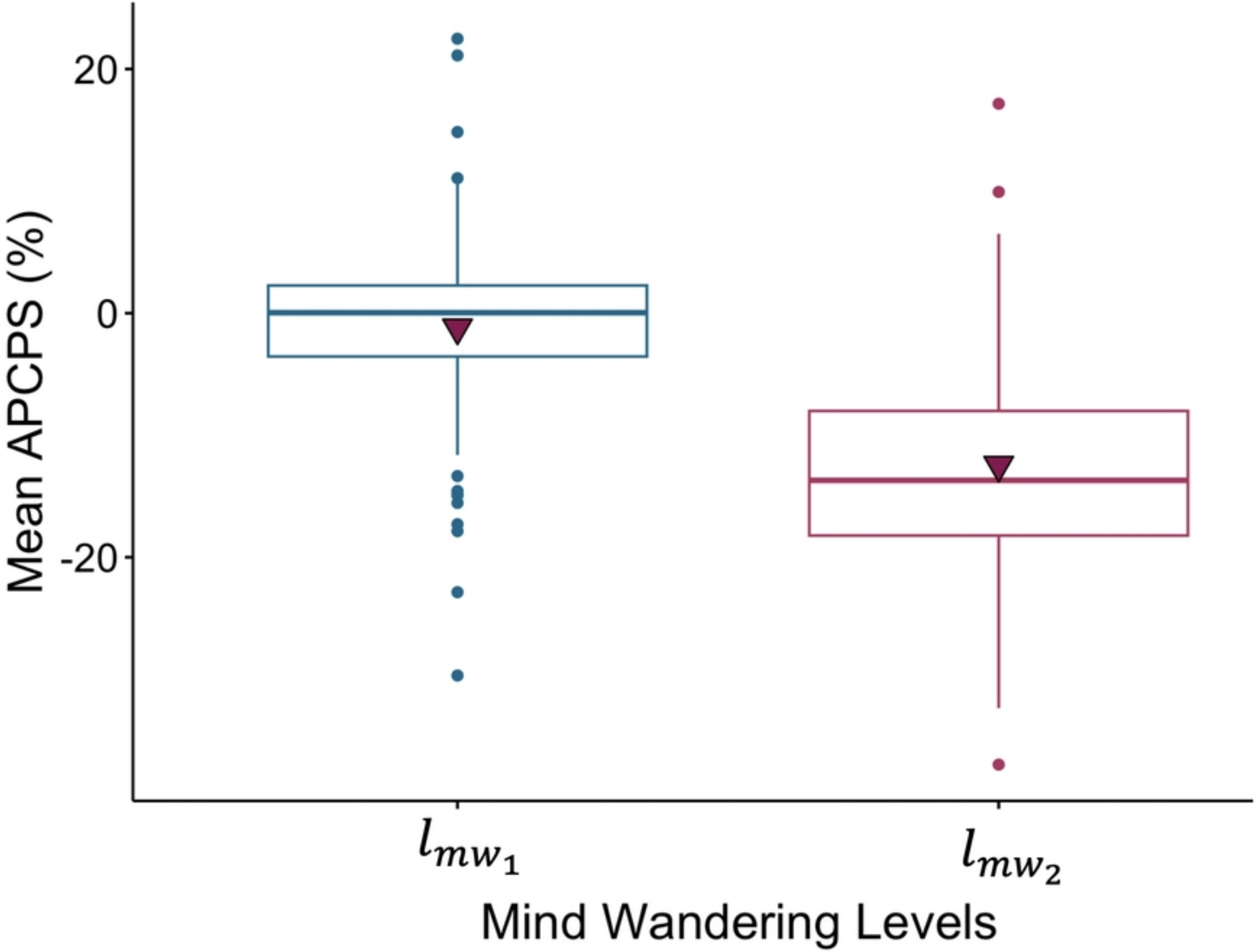

Higher Mind Wandering
Negative Correlation
Negative Correlation
Higher Sense of Urgency
Higher Workload
?

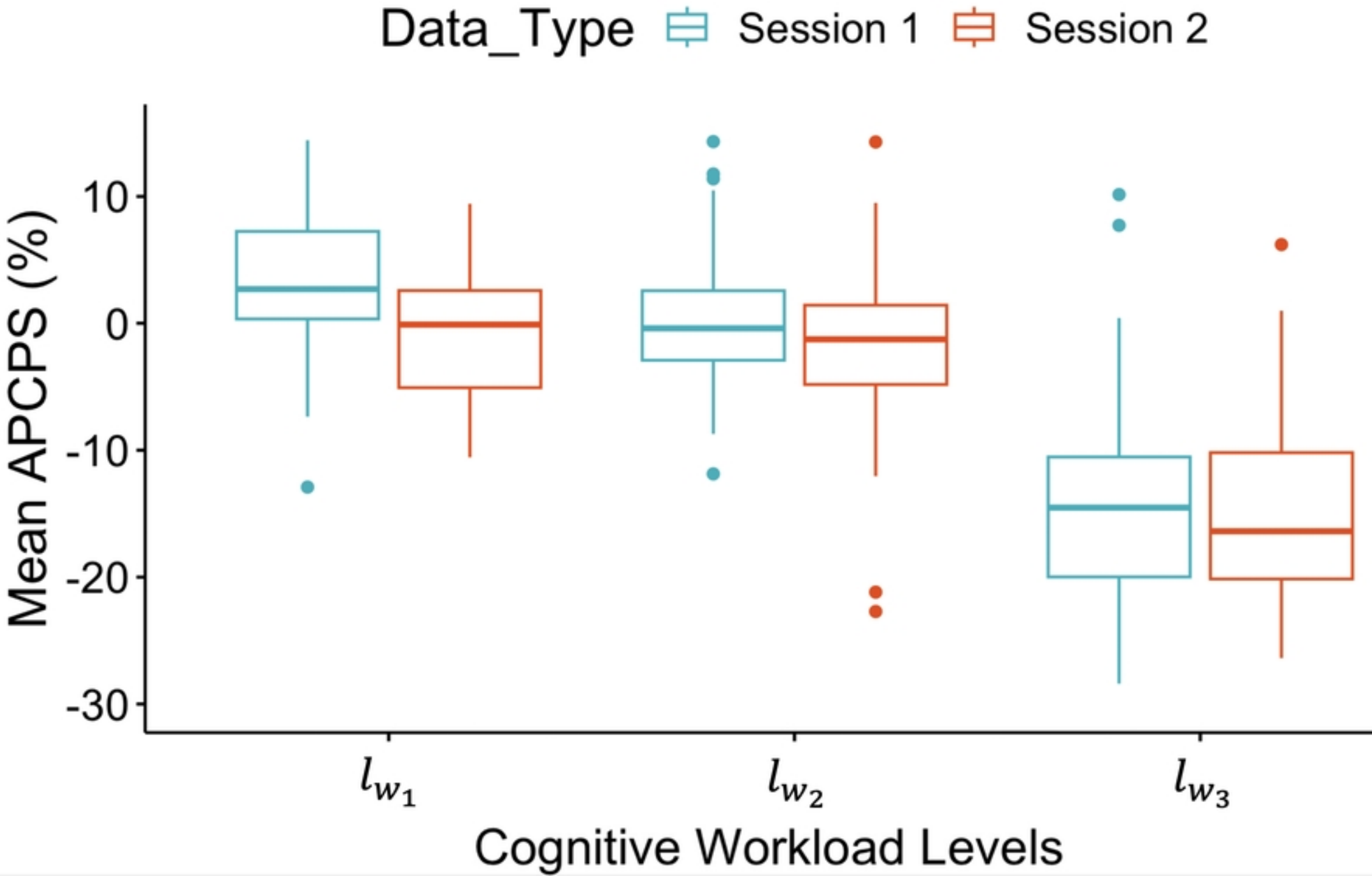

(a)
(b)
(c)

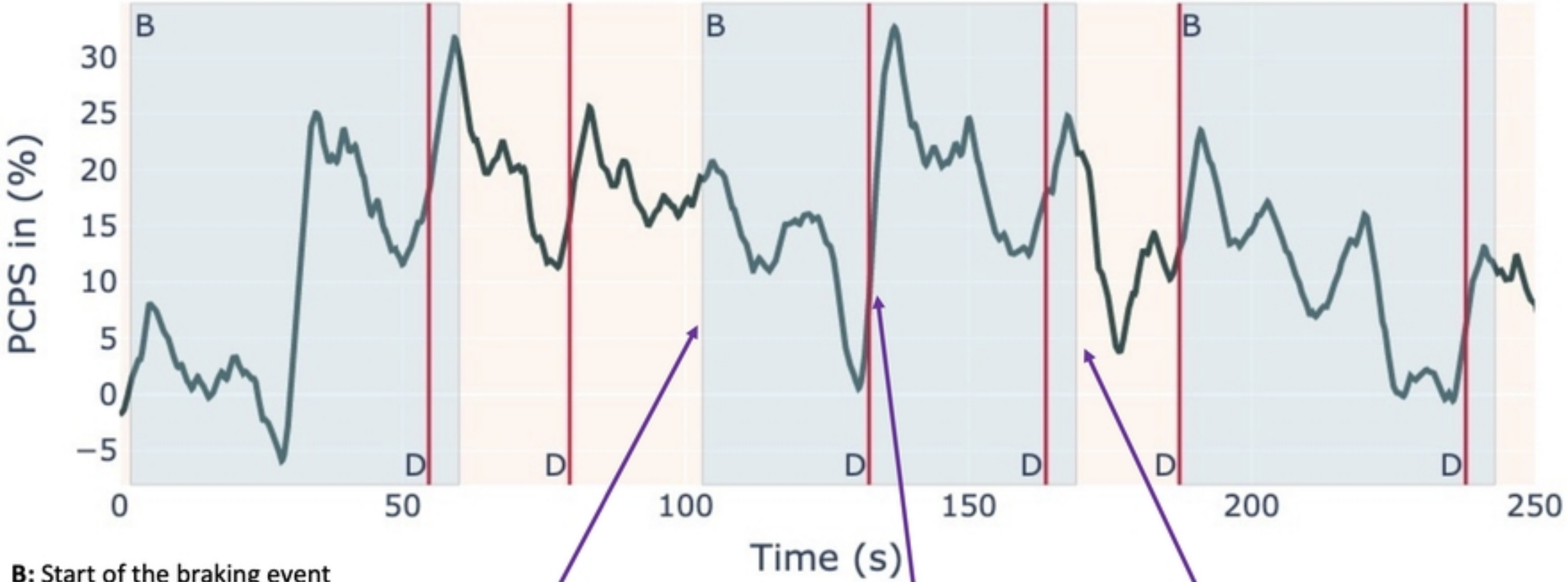

$l_{w_3}$
$l_{w_1}$
$l_{w_2}$
Start of the simulation
Start of 1st Braking
Start of 5th Braking
End of the simulation
No dialogue events, no braking tasks
6 dialogue events and 4 braking tasks
6 dialogue events and 4 braking tasks


40
20
0
-20
-40
Mean APCPS (%)
$l_{w_1}$
$l_{w_2}$
$l_{w_3}$
Cognitive Workload Levels


20
0
-20
Mean APCPS (%)
$l_{sou_1}$
$l_{sou_2}$
Sense of Urgency Levels


20
0
-20
Mean APCPS (%)
$l_{mw_1}$
$l_{mw_2}$
Mind Wandering Levels


Data_Type
Session 1
Session 2
Mean APCPS (%)
10
0
-10
-20
-30
$l_{w_1}$
$l_{w_2}$
$l_{w_3}$
Cognitive Workload Levels




**B:** Start of the braking event (the front car appears)

**D:** Start of the dialogue task (the question starts)

At the start of the braking, the front car appears

In the middle of the braking, the front car instantly brakes (we observe higher APCPS)

Through the end of braking, the front car gradually disappears