Applied Ergonomics 119 (2024) 104305

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Cognitive workload classification of law enforcement officers using physiological responses

David Wozniak, Maryam Zahabi *

Wm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Keywords:
Machine learning algorithm
Law enforcement officers
Adaptive technology

ABSTRACT

Motor vehicle crashes (MVCs) are a leading cause of death for law enforcement officers (LEOs) in the U.S. LEOs and more specifically novice LEOs (nLEOs) are susceptible to high cognitive workload while driving which can lead to fatal MVCs. The objective of this study was to develop a machine learning algorithm (MLA) that can estimate cognitive workload of LEOs while performing secondary tasks in a patrol vehicle. A ride-along study was conducted with 24 nLEOs. Participants performed their normal patrol operations while their physiological responses such as heartrate, eye movement, and galvanic skin response were recorded using unobtrusive devices. Findings suggested that the random forest algorithm could predict cognitive workload with relatively high accuracy (>70%) given that it was entirely reliant on physiological signals. The developed MLA can be used to develop adaptive in-vehicle technology based on real-time estimation of cognitive workload, which can reduce the risk of MVCs in police operations.

1. Introduction

Motor vehicle crashes (MVCs) are one of the most prevalent causes of death in the U.S. About 46,000 people lost their lives in car crashes and roughly 5.2 million people were seriously injured due to crashes in 2022 alone (NSC, 2022). MVCs are also the leading cause of line-of-duty deaths for public safety workers and more specifically law enforcement officers (LEOs) (BLS, 2020). Compared to firefighters and emergency medical services workers, LEOs (people responsible for, among other duties, enforcing state and local law via patrolling and responding to emergency situations in their vehicles) are involved in a significantly higher number of fatal MVCs (BLS, 2019). These crashes account for around 30–40% of LEOs' fatal work injuries (NLEMF, 2020, 2023). For example, in 2023, 37 LEOs have died due to traffic-related crashes in the U.S. (NLEMF, 2023). Additionally, compared to all other occupations, LEO MVCs are 2.5 times more than the national average (Maguire et al., 2002). Primary reasons for these crashes include the frequent use of in-vehicle technology while driving (Yager et al., 2015), fatigue (Vila and Kenney, 2002), and lack of sufficient training in handling high-demand situations (e.g., pursuit situations, multi-tasking) (Hembroff et al., 2018). LEOs and more specifically novice LEOs (nLEOs) with less than 5 years of patrol experience were selected as the focus of this study because they are at the highest risk among all emergency responders to be involved in crashes (Maguire et al., 2002). Novice LEOs tend to be at higher risk due to experiencing higher workload caused by having to review more chunks of data to come to a decision and having more frequent saccades, fixations, and time to detect road anomalies (Park et al., 2024)().

Police in-vehicle technology include the technology that civilian drivers interact with frequently such as cell phones and global positioning systems (GPS) as well as LEO-specific technology such as mobile computer terminals (MCTs) (a laptop that provides real-time navigation and case information to LEOs) and dispatch radios. In prior investigations (Park et al., 2020; Shupsky et al., 2021; Zahabi and Kaber, 2018a, 2018b; Zahabi et al., 2020), the MCT and radio were found to be the most important and frequently used in-vehicle technologies for LEOs while driving, primarily used for tasks such as researching case information, communicating with dispatch officers, and navigation. Use of these technologies has increased LEOs' distraction and cognitive workload while driving (Shahini et al., 2020). However, research on the development of technology to aid LEOs that incorporates their mental workload or performance has been insufficient, with most studies focused on analyzing workload through simulator studies or on analyzing the devices used while LEOs complete the patrol task (Zahabi et al., 2021; Shupsky et al., 2021; Zahabi et al., 2021). Some studies introduced adaptive technology features to the MCT (i.e., changing the

* Corresponding author. Wm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, College Station, TX, USA
E-mail address: mzahabi@tamu.edu (M. Zahabi).

https://doi.org/10.1016/j.apergo.2024.104305
Received 1 December 2023; Received in revised form 18 April 2024; Accepted 1 May 2024
Available online 10 May 2024

information presentation based on the context) (Kurkinen et al., 2010; Streefkerk et al., 2006). However, these adaptations were limited to specific MCT tasks, did not consider driver cognitive state, and have not been implemented in current MCTs used by police departments. Other studies developed adaptive technology solutions for civilian drivers, under normal driving situations, and with relatively simple secondary tasks (e.g., Park and Kim, 2015). However, there are several differences between LEOs and civilian drivers such as the temporal demands placed on the officers due to the need for real-time information access, complexity of in-vehicle technology (e.g., MCT), and driving situations (e.g., driving in pursuit conditions) (Zahabi et al., 2021). Furthermore, LEOs' in-vehicle technologies are designed assuming that the user performing a task is an expert and do not make mistakes or decisions that are not optimal (Zahabi and Kaber, 2018b). This assumption has prevented the technologies from being effectively designed around the cognitive processes of a novice or focusing on how those processes differ from expert cognition. Novices deal with higher cognitive workload (the mental effort an individual exerts to complete a task) compared to experienced drivers and they are more vulnerable to the risk of MVCs while driving (Moray, 2013). To better aid in the design of technology to reduce cognitive workload, novice behavior must be more effectively accounted for.

Police in-vehicle technology should adjust its function or presentation in response to the current state of the officer. This could involve displaying less information on a user interface when the officer has a high cognitive workload or giving the officer warnings to indicate that they are operating in a high-workload condition and making recommendations accordingly. Knowing the cognitive workload (CW) of LEOs based on their physiological responses such as heartrate or eye movement is one way to allow adaptive technology to respond to the state of the driver. This study used the data from a ride-along study with LEOs as a basis for developing an algorithm to provide an adaptive technology that can be used in police vehicles. While other naturalistic driving studies have been conducted with civilian drivers (Tivesten and Dozza, 2015; Williamson et al., 2015), the findings may not be generalizable to LEOs due to the differences in in-vehicle technology and driving conditions between the LEOs and civilian drivers.

1.1. Cognitive workload classification

To understand how the CW for novice LEOs can be classified, the differences between novices and experts have to be understood in a cognitive context. These differences can be summarized using Wickens' human information processing model (Wickens, 2008). With regards to attentional resources, novices are more likely to be impaired by distractions due to higher attentional resource demands, while experts are less likely to be impaired and can rely on non-visual signals more easily (Regan et al., 1998; Mourant and Rockwell, 1970). With regards to memory, the chunking process for novices is less effective compared to experts, and novices tend to attempt to make decisions before they finish processing all the information (Bruer, 1993). This concept in particular is relevant to LEOs that are required to both search for and recall key pieces of information about cases they respond to while completing the driving task through visual and auditory modalities, sometimes simultaneously. Novices exhibit higher CW than experts when faced with critical decisions similar to those needed to prevent a MVC. (Ouddiz et al., 2020). In contrast, experts have better recall than novices, allowing them to more effectively rely on their long-term memory and experiences to make decisions and better manage their overall CW (Horswill and McKenna, 2004). One key example of this would be in officer response to incoming calls while in the middle of a stop. Novices spend more time deciding how to respond to multiple requests for action compared to experts that can quickly refer to their experiences and take action more effectively.

Technologies and models that target novices specifically have rarely been able to capture all the fundamental differences between novices and experts to effectively model or predict their CW in various driving situations (Islam et al., 2020; Son et al., 2013). While there have been plenty of attempts to model driver CW using machine learning algorithms (MLAs) (such as random forest and support vector machines), which estimate workload based on data fed to them in real time, in the past, these approaches relied on physiological variables that would be cumbersome to implement for naturalistic driving tasks or rely entirely on driving simulator data to develop their MLAs (Islam et al., 2020; Lee et al., 2024; Son et al., 2013). The approach taken in this study is novel in that it relies on physiological variables captured while nLEOs are performing their normal patrol duties. These physiological variables were collected using unobtrusive devices and included heartrate variability (HRV), percentage change in pupil size (PCPS), blink rate (BR) and galvanic skin response (GSR). As HRV decreases, cognitive workload increases, whereas increases in all other measures mentioned are generally indicative of higher cognitive workload. These physiological variables have all been validated as effective indicators of workload, and when combined under a single algorithm can be used to effectively determine an individual's cognitive workload at a given moment (McDonald et al., 2019; Singh et al., 2013; Zahabi et al., 2022). These measures were selected for their relative consistency in evaluating workload while being unobtrusive to the wearer, allowing for participation in normal work activities (Fuhl et al., 2016; Schuurmans et al., 2020).

1.2. Problem statement and research objectives

LEOs and more specifically novice LEOs are at a significantly higher risk of MVCs compared to other occupations. Taking advantage of technologies that accounts for the CW of LEOs might help reduce these crash rates. Therefore, there is a need to detect and predict CW for nLEOs in real-time and provide that information to in-vehicle technology. The objective of this study was to develop an MLA that could predict the CW of nLEOs using features that could be measured in real-time with unobtrusive devices while the patrol task is being performed.

2. Method

2.1. Participants

Twenty-four (24) LEOs were recruited (age: M = 30.76 yrs, SD = 5.07 yrs; gender: 6 females, 18 males). To qualify for this study, participants needed to have normal or corrected-to-normal vision without glasses, have less than 5 years of primary patrol experience (Filtness et al., 2013; Hillerbrand, 1989), and have more than 1.5 years of regular driving experience to control for workload increases caused by inexperience with the driving task. The intent of this study was to observe how secondary tasks could affect the CW of nLEOs while they are performing their duties in the vehicle, not to observe inexperience with the task of driving in general. From this pool of participants, four participants were excluded from the final count due to data collection issues or ride-along had to be stopped due to emergencies. All participants read and signed the provided informed consent form before participating in the study. As the study took place during the participants' normal working hours, they were not compensated for their time. The study protocol was approved by Texas A&M Institutional Review Board (IRB 2021-0757D).

2.2. Equipment

An Empatica E4 (Empatica) watch was used to measure the HRV and GSR data from the participant. To measure the pupillometry data, the Pupil Labs eye tracking glasses (Pupil Labs) were used. These devices are validated for use in measuring these physiological measures and were synchronized before data were collected by plugging the E4 watch into the laptop used to run the eye tracking software (Fuhl et al., 2016; Schuurmans et al., 2020). The ride-alongs were also recorded using a dash camera attached behind the front seats of the police vehicle (Fig. 1).

2.3. Study procedure

Upon arriving at the police station, the researcher presented the participant with an informed consent form. Once this form was signed, the Empatica E4 was attached to the wrist of the participant (Fig. 2) and activated to give the device time to calibrate while other set-up procedures were completed.

While the E4 was calibrating, the participant filled out a demographic questionnaire. The researcher used this time to set up the eye tracking glasses and the dash camera in the participant's police vehicle as shown in Fig. 1. Then, the participant put on the eye tracking device and a calibration procedure was executed where the participant was asked to look at each of the four apriltags placed on the windshield without blinking as directed by the researcher. Following this step, a baseline pupil diameter was collected by running the eye tracking software for 2 min (Zahabi et al., 2021) while the participant remained seated in the vehicle. Another baseline pupil data was also collected at the end of the ride-along and before the officer left the vehicle.

Once the calibration was completed, the study was initiated. A unique synchronization technique explained in the following section was performed to ensure that data from the E4, dash camera, and the eye tracking glasses could be synchronized after the data collection. The participant was then instructed to perform their normal patrol duties (such as monitoring civilian traffic for infractions and responding to emergency calls requiring a police presence) while wearing the eye tracking glasses and Empatica E4. The researcher did not initiate interactions with the participant to ensure that the patrol was as naturalistic as possible. Fig. 3 illustrates the set-up for the experiment, with a participant in the driver's seat on the left and a researcher on the right (passenger seat). Note the myriad technologies that the officer has to interact with during their patrols and the apriltags that can be seen on the MCT for tracking eye movements.

Data collection continued until at least 3 h of data were collected or the participant chose to stop the experiment for any reason. A 3-h time frame was chosen to ensure enough data could be collected to train an effective MLA and has been used in previous ride-along studies with LEOs (Zahabi et al., 2022). Data were collected during daylight hours due to eye tracking glasses not functioning during nighttime conditions. When participants were required to stop and exit their vehicle to do their police duties, data collection was paused, and the participant removed their eye tracking glasses (but not the Empatica E4). Once the participant was ready to drive again, the synchronization technique was repeated, and the naturalistic observation resumed.

Once the study was concluded, the participant returned to their police station and the equipment used for the observation was removed. A Driver Activity Load Index (DALI) questionnaire was given to participants to evaluate their CW during the driving part of their patrol

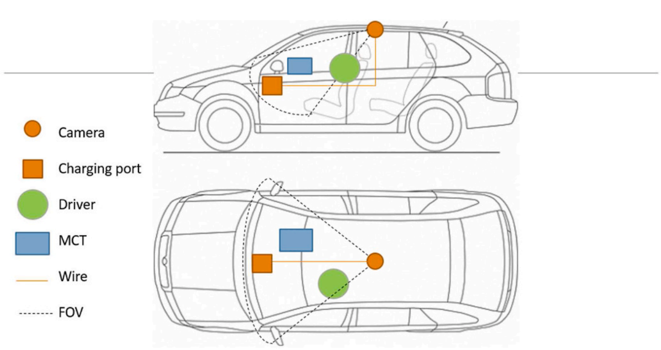

Fig. 1. Ride-along Study Set-up (Note: MCT: Mobile computer terminal, FOV: Field of view).

(Pauzié, 2008b). DALI is a revised version of the NASA-Task Load Index (NASA-TLX) questionnaire specifically adapted to accommodate the driving task. The most important difference among the methods is the unique set of DALI demand components to promote applicability to the driving domain. For example, the 'physical demand' component of the NASA-TLX is not very relevant to the driving activity where maneuvers are not physically demanding especially in modern cars. In addition, the 'cognitive demand' component of the NASA-TLX refers to both perceptual and cognitive aspects of workload but DALI identifies these various modalities in the context of driving.

Then, the participant was given a copy of the informed consent form for their records and thanked for their participation. Fig. 4 outlines how data were collected and moved from the devices used for data collection and transformed into a useable format.

2.4. Synchronization technique

The first step in the synchronization process was ensuring that the dash camera and eye tracking software were turned on to record and a timestamp was taken with the E4 by pressing the button as shown in Fig. 2. Doing this caused a red LED to flash on the E4 for 3 s. This procedure was done in view of both the dash camera and the world camera of the eye tracking software. A file within the E4's data storage was used to hold the timestamp that occurred each time the button was pressed. In post-processing, the data collected by all devices before this timestamp could be discarded to ensure that all data were synchronized. When the participant had to stop the observation to conduct police activities, a similar procedure was executed. The sampling rate for Empatica E4 for collecting GSR data was 4 Hz and for collecting the BVP data was 60 Hz. The eye-tracking device sampling frequency was varied with a cap of 130 Hz. More information about the synchronization approach can be found from Wozniak et al. (2022).

2.5. Data analysis

Data points that fell within periods where the police vehicle was stopped or occurred before or after the start and end timestamps respectively were removed so only active patrol times were included. Rows of data were found in 5-min intervals starting from where the observation began. This interval was chosen because it is the standard interval used for collecting root mean squared standard deviation (RMSSD) data (Electrophysiology, 1996).

RMSSD and the low frequency/high frequency (LF/HF) ratio for each participant were calculated using HRV data. From the GSR data statistics, the skin conductance level (SCL) and skin conductance response (SCR) in the form of the SCLm (SCL magnitude), SCLc (SCL change), SCRh (index of SCR habituation), SCRa (SCR amplitude), and SCRr (SCR response rise) were extracted (McDonald et al., 2019; Singh et al., 2013; Zahabi et al., 2022). Blink rate was calculated within each 5-min interval using the number of blinks recorded during that period by the Pupil Labs eye tracking software. The PCPS was also calculated using Pupil Labs data and the baseline pupil diameters recorded before and after the ride-along.

For instances where one of the four raw data streams could not be collected, the 5-min intervals associated with that set of data were discarded. In the case of missing only one or two raw data streams for only some intervals within a participant, the missing values were approximated using a method based on accepted decision tree imputation methods (Rahman and Islam, 2011) . This decision tree looked at all of the values for the missing data value from the other 5-min intervals and made an estimate of the missing value using the other data collected from that participant. This process was only used for data values within participants due to the differences that exist between participants with regards to average physiological values.

The final step in data analysis before MLA development could begin was to establish ground truth workload values for each of the 5-min

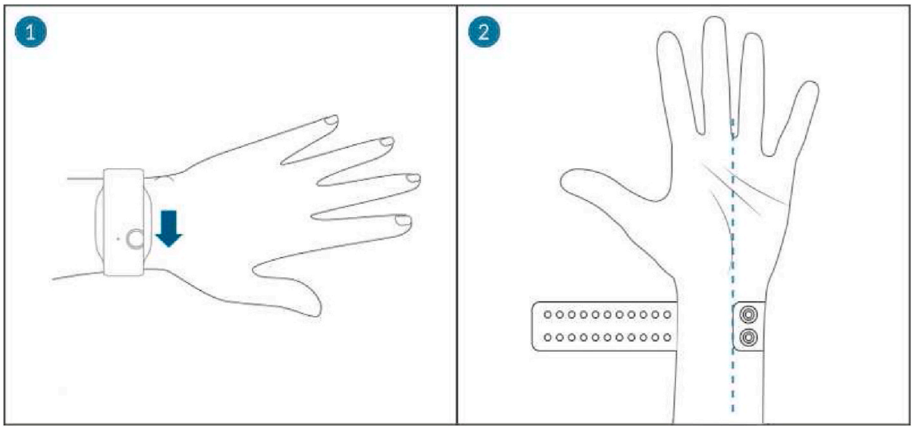

Fig. 2. Empatica E4 attachment procedure (Empatica, 2020).

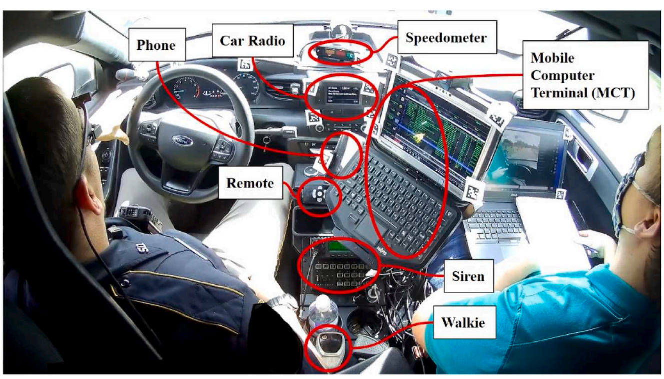

Fig. 3. Police in-vehicle technologies.

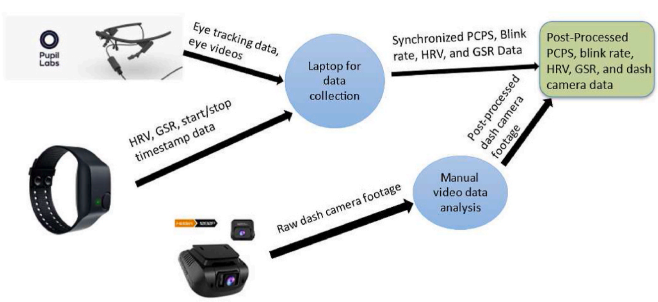

Fig. 4. Data Movement Chart (Note: HRV: Heart rate variability, GSR: Galvanic skin response, PCPS: Percentage change in pupil size).

intervals. Because it would be unreasonable to assign DALI values taken for the overall ride-along as the CW for each interval, these ratings were weighted against the physiological features themselves to establish a ground truth and get around the limitations of naturalistic observation studies. The classification of CW was divided into two groups, “high” and “low” CW. This number of groups was chosen based on a fuzzy logic analysis in MATLAB to separate the collected physiological data into two CW groups, three CW groups, and five CW groups to see which separation resulted in the most even split of data (based on pre-established thresholds for each physiological feature). An even split of data would imply that there was significant variation between the groups and that an MLA would be warranted. It was found that having two CW groups resulted in the most even split of the data, with 40.1% of data rows being classified as low CW. This was also considered to be reasonable due to the naturally high CW that was expected to be experienced by nLEOs during their patrol task. Similar approaches have been used in previous studies to find the optimal number of CW classes (Park et al., 2023).

Each feature, including DALI ratings, was assigned to be either a high impact, medium impact, or low impact feature for establishing ground truth CW. High impact variables included DALI, RMSSD, SCLm and SCLc due to their resistance to environmental factors, high number of validating studies, and ability to detect minute changes in workload (Cinaz et al., 2013; Fallahi et al., 2016; Mehler et al., 2010, 2011; Pauzié, 2008a, 2008b; Reimer and Mehler, 2011; Shimomura et al., 2008; Zakerian et al., 2018). These features were given a weight of 0.125 (or 12% in Fig. 5) for determining the ground truth workload, with the thresholds for these physiological variables were established by previous studies (Abhishekh et al., 2013; Abusharha, 2017; Arthur, 1990; de Waard, 1996; Pfleging et al., 2016; Zahabi et al., 2021). Medium impact variables included the LF/HF ratio, SCRh, SCRa, and SCRr, due to lower resilience to environmental factors and high correlation to other physiological measures (Cinaz et al., 2013; Fallahi et al., 2016; Hsu et al., 2015; Novak et al., 2011; Rodriguez Paras, 2015; Verwey and Veltman, 1996). These features were given a weight of 0.075 each (or 8% in Fig. 5) in determining ground truth workload. Finally, PCPS, BR, and average GSR were assigned as low impact features due to the nature of data collection impeding the quality of eye tracking data and the noise factor associated with raw average GSR values (Cardona and Quevedo, 2014; Faure et al., 2016; Iqbal et al., 2005; Johns et al., 2014; Kahng and Mantik, 2002; Kosch et al., 2019; Pfleging et al., 2016; Stern et al., 1994). These features were assigned an importance weight of 0.0667 each (or 7% in Fig. 5). The specific weights chosen for each group were selected to keep the weight gap between feature groups relatively low while still maintaining a significant difference between the high impact and low impact features. While there is no established guideline

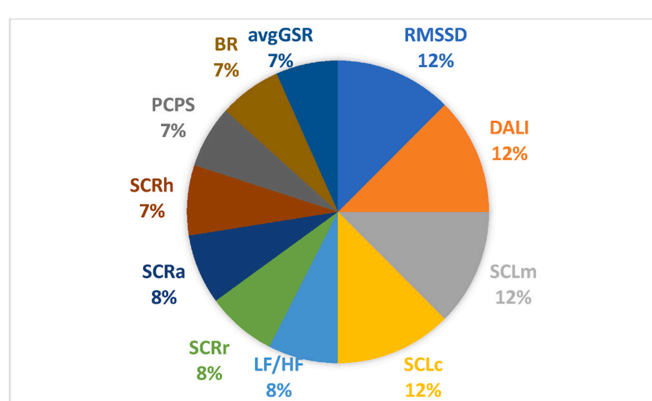

Fig. 5. Weights assigned to features for establishing the ground truth.

exploring how these features should be prioritized over each other, the weights selected here were inferred based on the above literature and thorough examination of the collected data. Fig. 5 illustrates the breakdown in feature weight assignment between all of the features. Once these weights were applied to all of the 5-min intervals, ground truths could be established, and an MLA could be developed.

2.6. MLA development

Development of an MLA to predict CW of nLEOs was completed first by developing MLAs that were found to be prevalent in the prediction of CW in the driving domain. The quality of each of these MLAs was assessed across multiple seeds on the metrics of receiver operating characteristic (ROC) area under the curve (AUC), precision, and accuracy. The classification dataset used for validation and testing consisted of a randomly selected set of 20% of the overall collected data, with the remaining 80% of the data being used to train the model.

3. Results

3.1. Data screening

After grouping the raw data into 5-min intervals, a total of 769 rows of data were initially created. Of those rows, only 328 had all metrics filled in with no missing values. Due to the nature of collecting data from a naturalistic setting, this amount of missing data was roughly expected. Sampling rate was based on the fastest recorded metric, which was usually eye tracking and had a variable Hz rate around 30–60 Hz. Before missing data could be filled in, outliers in the data had to be removed. This was done by finding for each column all rows that had a value more than two standard deviations larger or smaller than the mean for that column with the value in question removed. The next step in the filtering process was to remove rows for which there were no values to extrapolate average values within the participant for that column. As 229 rows of data needed extrapolation to fill in missing values, a decision tree algorithm was created to predict the missing values for each missing data entry for these rows. To prevent individual differences from confounding the predictions for these missing values, the algorithm only considered values within participants when filling in data. After removing outliers and filling in missing values, a total of 557 rows of data were captured. Once ground truths had been assigned to each data row, there were 228 rows with the high workload classification and 329 rows with the low workload classification, meaning that approximately 59% of the rows were classified as being low workload.

3.2. MLA performance

Based on previous studies on effective MLAs in the driving domain that relied on physiological variables, the following MLAs were selected to be trained by the collected data: decision trees (DT), random forests (RF), naïve bayes algorithms, and support vector machines (McDonald et al., 2019). Table 1 summarizes the best performance of these algorithms on the basis of their accuracy in classifying a randomly selected set of 20% of the total dataset with the rest being used to train the algorithm. For the MLAs besides naïve bayes, hyperparameters were selected and tuned using 10-fold cross validation sets from the training data repeated three times, with the best set of hyperparameters in terms of test data accuracy being selected. The support vector machine MLAs are split into radial and polynomial kernels indicated by SVMr and SVMp respectively. These two kernels were selected based on their use in a previous study evaluating MLAs in the driving domain using physiological variables only (McDonald et al., 2019). Additionally, the no information rate (NIR) refers to the rate of success at guessing the classification of a row of data with no other information available. Tuned hyperparameters for each seed included minimum n value, cost complexity, and tree depth.

A trained MLA was considered successful if it had a higher accuracy than the NIR with at least 95% confidence. The results displayed the average of 5 seeds that performed the best from a total of 50 seed tests for each MLA to showcase their most effective performance. If an MLA had fewer than 5 seeds perform better than the NIR within a 95% confidence interval (CI), then only seeds that met this condition were considered when averaging results. Overall, success rates were 42% for RF, 34% for SVMr, 8% for naïve bayes, 6% for SVMp, and 4% for DT. It was found that the RF model performed the best both in terms of high accuracy (i.e., 73%) and consistent performance when compared to the NIR across multiple seeds.

Additional metrics that were evaluated to determine the best MLA include the AUC and precision. AUC is a measure of model performance at any given threshold that evaluates the predictive ability of learning algorithms (Huang and Ling, 2005) while precision refers to the degree of difference between various samples. For both metrics, high values indicate a more effective model. It was found that the RF model performed the best on average for AUC while NB performed the best on precision when looking at the best performing seeds overall. The results of the AUC and precision comparisons are shown in Table 2. Note that no AUC was calculated for the naïve bayes MLA because no hyperparameters were manipulated.

The MLAs were also compared in terms of the training and test times (i.e., the amount of time on average it took to train and run test data through the MLAs respectively). Once again, the RF MLA outperformed the other MLAs in test time with an average test time roughly 0.06 s faster than the second fastest MLA. Table 3 below displays the average training time and testing time for each MLA. Note that training time is in minutes and testing time is in seconds. These testing times in particular are important because they carry implications for how well each MLA might be able to perform when actually implemented into adaptive technology. In addition, classification of officers' cognitive workload is a time sensitive task and should be performed in real-time.

Table 1
Accuracy results for most successful seeds of each MLA trained on physiological data.

Algorithm	Metric		
	Accuracy (%)	NIR (%)	95% CI (%)
RF	73.21	59.24	(64, 81.1)
SVMr	67.7	56.25	(58.22, 76.20)
SVMp	68.62	57.4	(59.2, 77.02)
DT	62.5	52.23	(52.86, 71.45)
NB	71.4	53.57	(57.81, 82.69)

Note: RF = Random Forest, SVMr = Support Vector Machine radial, SVMp = Support Vector Machine Polynomial, DT = Decision Tree, NB = Naïve Bayes, NIR=No information rate.

Table 2
Precision and AUC results for most successful seeds of each MLA trained on physiological data.

Algorithm	Metric	
	AUC	Precision (%)
RF	0.79	73.16
SVMr	0.70	68.12
SVMp	0.72	71.99
DT	0.65	66.33
NB	N/A	76.03

Note: RF = Random Forest, SVMr = Support Vector Machine radial, SVMp = Support Vector Machine Polynomial, DT = Decision Tree, NB = Naïve Bayes, AUC = Area Under the Curve.

Table 3
Average training time and testing time for each MLA.

Machine Learning Algorithm	Training Time (minutes)	Test Time (seconds)
RF	6.97	0.02
SVMr	1.48	0.28
SVMp	2.34	0.31
DT	2.20	0.15
NB	0.62	0.08

Note: RF = Random Forest, SVMr = Support Vector Machine radial, SVMp = Support Vector Machine Polynomial, DT = Decision Tree, NB = Naïve Bayes.

3.3. Feature importance

Table 4 summarizes the importance of features for each of the tested MLAs. These feature importance ratings, generally speaking, can be thought of as the amount of relative weight that each MLA puts on the value of a feature when deciding how to classify workload, with more positive values indicating higher weight. Note that for SVMp and SVMr some feature importance ratings were negative, indicating that those features were not useful in predicting CW. For every MLA except for NB, SCRr was found to be the most important feature, with features such as SCRa and SCLa being considered less important. The most important features for the best model (i.e., RF) were SCRr, SCLc, LF/HF, PCPS, and SCLm respectively.

4. Discussion

4.1. MLA selection

The objective of this study was to develop an MLA that could predict the CW of nLEOs using features that could be measured in real-time with unobtrusive devices while the patrol task is being performed. Out of all of the MLAs tested, the RF algorithm consistently performed better than the NIR rate in terms of accuracy while meeting the precision and ROC AUC guidelines found for creating effective MLAs, which include precision ratings on average of at least 0.7 as well as ROC AUC values of around 0.85 (Lee et al., 2010; Pencina et al., 2008). Specific values for good accuracy are not standardized, so 0.7 was used as a general benchmark in line with the precision recommendation. Under these guidelines, it can be assumed with confidence that the RF algorithm can perform well when fed new or real-time test data. These values are general guidelines, as the effectiveness of a MLA is primarily determined by its ability to learn as it obtains more data, meaning that future data collection should be able to improve this MLA to validate its effectiveness in predicting cognitive workload (El Naqa and Murphy, 2015). Note that these MLAs were not compared to other studies due to this data being pulled from a naturalistic study rather than simulator or experiment data. Other recent studies have been conducted with similar objectives and were able to obtain higher classification accuracy (>90%) primarily due to being able to collect data in controlled simulator settings (Lee et al., 2024). Of the tested MLAs, the DT algorithm performed the worst, failing to perform significantly better than the NIR rate roughly 96% of the time. Naïve bayes and SVMp algorithms performed poorly as well. As naïve bayes assumes full independence of features and it was reasonable to assume that at least some of the features in this dataset were related to each other, this result was expected for naïve bayes (Lewis, 1998). However, it was anticipated that DT and SVMp would perform much better than they actually did. One possible reason for this discrepancy would be the use of data collected in a naturalistic setting rather than in a laboratory or in a simulator study. This could also be due to the stringent requirements placed on tested MLAs to be considered successful in terms of accuracy with regard to the NIR.

The important features found for this MLA are also of note. In particular, the GSR features were found to be the best for RF regarding predicting cognitive workload. This carries important implications for real-world implementation due to the E4 being the least obtrusive device LEOs were required to wear. While eye-tracking glasses are limited by the time of day and the places being driven, GSR data can be collected at any time without potentially intruding on officer performance. While it would be ideal to maintain all features used for subsequent data collection, knowing that the potential to truncate feature collection in a pinch opens up opportunities for data collection in other naturalistic settings that were not considered for this study, such as night patrols.

To validate the selection of the RF MLA, several factors were considered, most notably the accuracy, precision, and ROC AUC on average of all the seeds tested with the RF MLA. As the RF MLA performed better than its other competitors, it was the final MLA selected. Note that while specific seeds might have had other MLAs perform better than the RF, the RF MLA provided the most consistent results. RF algorithms are also popular in the driving domain and have been used in several studies in the past for purposes such as driver behavior profiling, making incorporation to other technologies and comparisons easier to make (Das and Khilar, 2019; Ferreira et al., 2017; Rahman et al., 2019). These studies have looked at several different avenues of application such as using phones or in-vehicle GPS for the implementation of adaptive technology, and the advantages of such an algorithm would be invaluable to the development of these technologies.

Table 4
Feature importance for best MLA results.

Feature	Algorithm				
	RF	DT	SVMp	SVMr	NB
SCRr	0.21	0.41	3.90	3.90	0.26
LF/HF	0.11	0.16	0.54	0.54	0.0052
SCLm	0.10	0.15	0.11	0.11	8.86E-06
Blink Rate	0.092	0.080	−0.49	−0.49	0.41
SCRa	0.073	0.044	−3.63	−3.63	0.0045
SCRh	0.060	0.042	−0.90	−0.90	0.18
RMSSD	0.067	0.039	0.63	0.63	0.0041
SCLc	0.12	0.030	0.053	0.053	0.00097
avgGSR	0.067	0.025	0.59	0.59	0.10
PCPS	0.11	0.024	0.20	0.20	0.022

Note: RF = Random Forest, SVMr = Support Vector Machine radial, SVMp = Support Vector Machine Polynomial, DT = Decision Tree, NB = Naïve Bayes, SCLm = SCL magnitude, SCLc = SCL change, SCRh = index of SCR habituation, SCRa=SCR amplitude, and SCRr = SCR response rise, LF/HF = Low frequency/ High frequency Ratio, RMSSD = Root Mean Squared Standard Deviation, avgGSR = Average Galvanic Skin Response, PCPS = Percentage Change in Pupil Size.

4.2. Real-time workload classification

The primary reason that physiological variables alone were considered in this ride-along study design was because one of the end goals for the developed MLA was to be able to classify workload in real-time. The applications of this classification would be to implement them into technology that can use the current workload of the user to adjust the in-vehicle technology to accommodate them. Given that novice drivers are more prone to high workload and this higher workload can lead to more potentially fatal mistakes, understanding when these risks might happen using MLAs is critical. To offset the potential issues in accuracy of the MLA, new samples need to consistently be taken to have the workload update as frequently as possible. Individual differences also need to be accounted for by having the MLA be trained specifically with data collected for an individual participant and supplemented by the already collected data. To test this, a real-time algorithm was developed in python and tested in a lab setting to see if data could be recorded and run through the developed MLA in real-time. This was proven to be the case and the developed MLA was able to predict CW in real-time without having to stop data collection. This finding is crucial when considered in tandem with the test times for the developed MLAs. In addition to being the most accurate on average, the RF MLA was by far the fastest in calculating the output when fed the same amount of test data as the other MLAs. Though the difference of a few hundredths of a second might seem insignificant, the longer it takes an MLA to output results in real-time, the higher the risk that the output it provides will be too late to be useful. To prevent this, an MLA that can calculate output as quickly as possible is essential, giving RF more credence as an optimal choice among the tested MLAs.

4.3. Technology applications

In this study, adaptive technology refers to in-vehicle technology that detects the driver's workload and responds accordingly, by adjusting the amount/format of information that is presented to the driver. While performance on tasks has been shown to be most effective when an optimal level of arousal is maintained under the Yerkes-Dodson law (Yerkes and Dodson, 1908), the goal of this technology is to reduce the impact on the primary driving task by minimizing attention needed for the secondary task of using in-vehicle technology. Adaptive technology, therefore, has the dual goal of maximizing driver safety and making information as accessible as possible for the driver when their level of CW allows for it.

Adaptive technology has the potential to take full advantage of the real-time workload classification MLA developed in this study. An example of this can be found in the information that police officers receive about a suspect on their MCTs. LEOs are frequently confronted with overwhelming amounts of text-based information that obscure important information and make it difficult to determine the nature of the situation they are driving to while focusing on the road. At the same time, this information might be important to the case and the LEO needs to be aware of it before arriving at the scene. To address this issue, adaptive technology based on LEOs' workload level can be implemented. Figs. 6 and 7 illustrate a prototype of a heads-up display that would automatically adjust its appearance based on the CW of a LEO. If the LEO is experiencing high workload, then a low clutter display (i.e., a summary page with icons showing the most important pieces of information and their status) will be displayed (Fig. 6). Note that the red pictures indicate the most relevant points of information for the LEO (i.e., the violations). More detail on how to implement the MLA in an adaptive head-up display is provided in Nadri et al. Although prior research (Kurkinen et al., 2010; Streefkerk et al., 2006) introduced some adaptive technology features to the MCT (i.e., changing the information presentation based on the context), these adaptations were limited to specific MCT tasks and did not consider driver cognitive state.

Conversely, when the LEO's CW is low, more information can be made available to them (i.e., a high clutter display similar to the current MCT interface in police vehicles which is text-based) without increasing the risk of a MVC (Fig. 7). This is an example of how adaptive technology could be integrated with real-time CW classification from the developed MLA. Future studies should validate the proposed MLA with additional user-testing and then implement it in real-world scenarios to evaluate the effectiveness of this technology in reducing CW. To that end, the feature importance findings of this study can be useful. Given the difficulty of implementing a multitude of physiological measures into in-vehicle technology, prioritizing GSR in accordance with the findings on feature importance is also suggested.

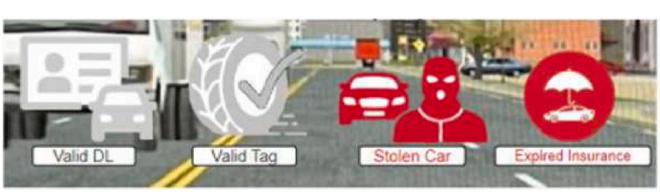

Fig. 6. Low clutter display.

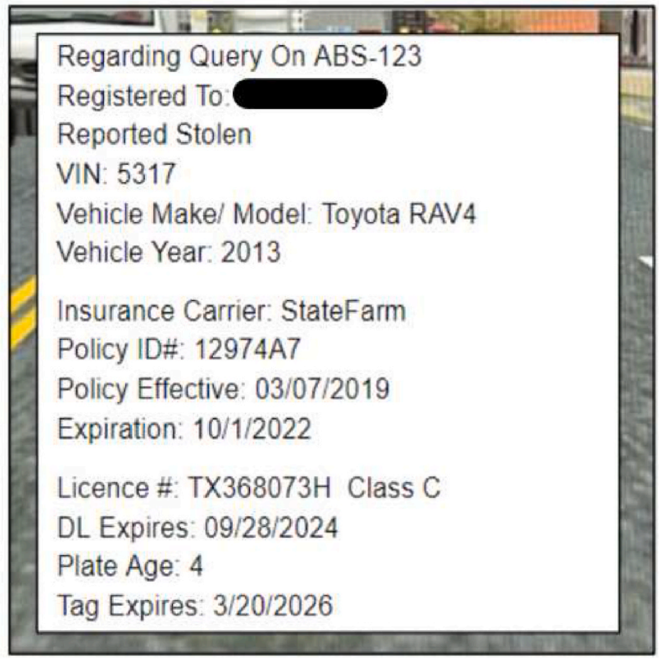

Fig. 7. High clutter display.

4.4. Limitations and future work

The nature of naturalistic observation caused significant amounts of data to be lost over the course of the experiment and necessitated retaining all features in spite of the risk of overfitting MLAs. Mitigation techniques that were found to be effective to maximize the amount of data collected included performing data collection only in the daytime, targeting cloudy days to avoid the amount of eye data lost to sunlight glare, and ensuring that all devices are properly secured to avoid data loss that goes undetected until the experiment is over.

Future studies need to continue testing the developed MLA with different scenarios to see how effective it is at predicting CW in scenarios other than the ones identified in this study. This could be done with driving simulator studies or further naturalistic observation studies. As this study was focused on novice LEOs, the recommendations and algorithm are mainly for novice LEOs. Future studies should assess the accuracy of the model in predicting cognitive workload for expert officers. Expert officers might also need adaptive technology, but based on our prior ride-along observations (Shahini et al., 2020; Zahabi et al., 2022), expert officers rely less on MCTs and are more experienced in handling high demand situations than novice LEOs. However, novel/difficult situations can cause experts to have similar behavior as novices meaning that adaptive technology for novices is likely to be broadly applicable to the situations where experts would need to take advantage of it. Although our proposed algorithm and adaptive system idea was not intended for training purposes, they can be used to train the officers to be more aware of the cognitive demands posed by different in-vehicle technologies.

One of our future goals is to incorporate other measures such as driving and secondary task performance to more accurately predict LEOs' CW. The results of these future experiments should provide a more robust MLA that can take advantage of the driving performance of LEOs to adjust how adaptive in-vehicle technology interacts with the driver. Ideally, wearable devices that are less intrusive without sacrificing effectiveness should be employed. While the Pupil Labs eye tracking device did not impede the patrol task of LEOs significantly, implementing a wireless version of the glasses or one that functions as sunglasses that most officers wear while on duty would improve the quality of implemented MLAs while reducing any induced cognitive load by the system on LEOs. Development and implementation of this technology would greatly improve the quality of data collection and real-time MLA implementation overall.

5. Conclusion

A machine learning algorithm for classifying cognitive workload was developed based on a ride-along study with novice LEOs. This MLA uses physiological responses in the form of HRV, GSR, BR, PCPS, and extracted features from these metrics to classify workload. The results can be used to develop technology that can predict the workload of nLEOs in real-time and adapt the functions of a vehicle accordingly, either to emphasize or deemphasize secondary tasks. Incorporating the developed MLA with adaptive technology can help nLEOs to better manage their tasks in the vehicle and can improve their safety in police operations.

CRediT authorship contribution statement

David Wozniak: Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Maryam Zahabi:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding for this research was provided by the National Science Foundation (No. IIS-2041889). The views and opinions expressed are those of authors.

References

Abhishekh, H.A., Nisarga, P., Kisan, R., Meghana, A., Chandran, S., Raju, T., Sathyaprabha, T.N., 2013. Influence of age and gender on autonomic regulation of heart. J. Clin. Monit. Comput. 27 (3), 259–264.
Abusharha, A.A., 2017. Changes in blink rate and ocular symptoms during different reading tasks. Clin. Optom. 9, 133.
Arthur, E., 1990. Physiological Metrics of Mental Workload: A Review of Recent Progress.
BLS, 2019. Table A-6. Fatal Occupational Injuries Resulting from Transportation Incidents and Homicides by Occupation, vol. 2015. all United States, p. 2016.
BLS, 2020. Fatal Occupational Injuries to Emergency Responders.
Bruer, J.T., 1993. The mind's journey from novice to expert. Am. Educat. 17 (2), 6–15.
Cardona, G., Quevedo, N., 2014. Blinking and driving: the influence of saccades and cognitive workload. Curr. Eye Res. 39 (3), 239–244.
Cinaz, B., Arnrich, B., La Marca, R., Tröster, G., 2013. Monitoring of mental workload levels during an everyday life office-work scenario. Personal Ubiquitous Comput. 17 (2), 229–239.
Das, R., Khilar, P.M., 2019. Driver behaviour profiling in VANETs: comparison of ensemble machine learning techniques. In: Paper Presented at the 2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP).
El Naqa, I., Murphy, M.J., 2015. What Is Machine Learning? *Machine Learning in Radiation Oncology*. Springer, pp. 3–11.
Electrophysiology, T.F.O.T.E.S.O.C.T.N.A.S.O.P., 1996. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 93 (5), 1043–1065.
Empatica, 2020. Wear your E4 wristband. Retrieved from. https://support.empatica.com/hc/en-us/articles/206374015-Wear-your-E4-wristband.
Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A.R., Miyake, S., 2016. Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. Appl. Ergon. 52, 95–103.
Faure, V., Lobjois, R., Benguigui, N., 2016. The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. Transport. Res. F Traffic Psychol. Behav. 40, 78–90.
Ferreira, J., Carvalho, E., Ferreira, B.V., de Souza, C., Suhara, Y., Pentland, A., Pessin, G., 2017. Driver behavior profiling: an investigation with different smartphone sensors and machine learning. PLoS One 12 (4), e0174959.
Filtness, A., Mitsopoulos-Rubens, P.E., Lenne, M., 2013. The impact of interface modality on police officers' visual behaviour when using an in-vehicle system. In: Paper Presented at the Proceedings of the 3rd International Conference on Driver Distraction and Inattention (DDI 2013).
Fuhl, W., Tonsen, M., Bulling, A., Kasneci, E., 2016. Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art. Mach. Vis. Appl. 27 (8), 1275–1288.
Hembroff, C.C., Arbuthnott, K.D., Krätzig, G.P., 2018. Emergency response driver training: dual-task decrements of dispatch communication. Transport. Res. F Traffic Psychol. Behav. 59, 222–235.
Hillerbrand, E., 1989. Cognitive differences between experts and novices: implications for group supervision. J. Counsel. Dev. 67 (5), 293–296.
Horswill, M.S., McKenna, F.P., 2004. Drivers' hazard perception ability: situation awareness on the road. A cognitive approach to situation awareness: Theory and application 155–175.
Hsu, B.-W., Wang, M.-J.J., Chen, C.-Y., Chen, F., 2015. Effective indices for monitoring mental workload while performing multiple tasks. Percept. Mot. Skills 121 (1), 94–117.
Huang, J., Ling, C.X., 2005. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans. Knowl. Data Eng. 17 (3), 299–310.
Iqbal, S.T., Adamczyk, P.D., Zheng, X.S., Bailey, B.P., 2005. Towards an index of opportunity: understanding changes in mental workload during task execution. In: Paper Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
Islam, M.R., Barua, S., Ahmed, M.U., Begum, S., Aricò, P., Borghini, G., Di Flumeri, G., 2020. A novel mutual information based feature set for drivers' mental workload evaluation using machine learning. Brain Sci. 10 (8), 551.
Johns, M., Sibi, S., Ju, W., 2014. Effect of cognitive load in autonomous vehicles on driver performance during transfer of control. In: Paper Presented at the Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.
Kahng, A.B., Mantik, S., 2002. Measurement of inherent noise in EDA tools. In: Paper Presented at the Proceedings International Symposium on Quality Electronic Design.
Kosch, T., Karolus, J., Ha, H., Schmidt, A., 2019. Your skin resists: exploring electrodermal activity as workload indicator during manual assembly. In: Paper Presented at the Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems.
Kurkinen, E., Sullivan, H.T., Häkkinen, M.T., Lauttamus, M., 2010. Optimizing mobile social media interfaces for rapid internal communication by emergency services. In: Paper Presented at the ISCRAM.
Lee, B.K., Lessler, J., Stuart, E.A., 2010. Improving propensity score weighting using machine learning. Stat. Med. 29 (3), 337–346.
Lee, C., Shin, M., Eniyandunmo, D., Anwar, A., Kim, E., Kim, K., Yoo, J.K., Lee, C., 2024. Predicting Driver's mental workload using physiological signals: a functional data analysis approach. Appl. Ergon. 118, 104274.
Lewis, D.D., 1998. Naive (Bayes) at forty: the independence assumption in information retrieval. In: Paper Presented at the European Conference on Machine Learning.
Maguire, B.J., Hunting, K.L., Smith, G.S., Levick, N.R., 2002. Occupational fatalities in emergency medical services: a hidden crisis. Ann. Emerg. Med. 40 (6), 625–632.
McDonald, A.D., Ferris, T.K., Wiener, T.A., 2019. Classification of driver distraction: a comprehensive analysis of feature generation, machine learning, and input measures. Hum. Factors 0 (0), 0018720819856454. https://doi.org/10.1177/0018720819856454.
Mehler, B., Reimer, B., Coughlin, J.F., 2010, September. Physiological reactivity to graded levels of cognitive workload across three age groups: An on-road evaluation. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Vol. 54, No. 24,. Sage, CA: Los Angeles, CA, pp. 2062–2066.
Mehler, B., Reimer, B., Wang, Y., 2011. A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In: Paper Presented at the Driving Assesment Conference.
Moray, N., 2013. Mental Workload: its Theory and Measurement, vol. 8. Springer Science & Business Media.
Mourant, R.R., Rockwell, T.H., 1970. Mapping eye-movement patterns to the visual scene in driving: an exploratory study. Hum. Factors 12 (1), 81–87.
de Waard, D., 1996. The measurement of drivers' mental workload [Thesis fully internal (DIV), University of Groningen]. s.n.
Nadri, C., Deng, J., Chauhan, A., Wozniak, D., & Zahabi, M. (Accepted). Police in-vehicle technology adaptation based on officer cognitive state. Paper presented at the IISE Annual Conference and Expo, Montreal, Canada..
NLEMF, 2020. Causes of law enforcement deaths. Retrieved from. https://nleomf.org/facts-figures/causes-of-law-enforcement-deaths.
NLEMF, 2023. 2023 end-of-year preliminary law enforcement officers fatalities report. Retrieved from. https://nleomf.org/memorial/facts-figures/latest-fatality-reports/.
Novak, D., Mihelj, M., Munih, M., 2011. Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction. Robotica 29 (3), 367–374.
NSC, 2022. Preliminary Semiannual Estimates.
Ouddiz, S., Paubel, P.-V., Lemercier, C., 2020. How do novice and expert drivers prepare for takeover when they are drivengers of a level 3 autonomous vehicle? Investigation of their visual behaviour. Trav. Hum. Le. 83 (4), 361–378.
Park, J., Berman, J., Dodson, A., Liu, Y., Armstrong, M., Huang, H., Kaber, D., Ruiz, J., Zahabi, M., 2023. Assessing workload in using electromyography (EMG)-based prostheses. Ergonomics 1–17.
Park, H.S., Kim, K.H., 2015. Adaptive Multimodal In-Vehicle Information System for Safe Driving. ETRI J. 37 (3), 626–636.
Park, J., McKenzie, J., Shahini, F., Zahabi, M., 2020. Application of cognitive performance modeling for usability evaluation of emergency medical services in-

vehicle technology. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Park, J., Wozniak, D., Zahabi, M., 2024. Modeling novice law enforcement officers' interaction with in-vehicle technology. Appl. Ergonom. 114, 104154.

Pauzié, A., 2008a. Evaluating driver mental workload using the driving activity load index (DALI). In: Paper Presented at the Proc. Of European Conference on Human Interface Design for Intelligent Transport Systems.

Pauzié, A., 2008b. A method to assess the driver mental workload: the driving activity load index (DALI). IET Intell. Transp. Syst. 2 (4), 315–322.

Pencina, M.J., D'Agostino Sr, R.B., D'Agostino Jr, R.B., Vasan, R.S., 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat. Med. 27 (2), 157–172.

Pfleging, B., Fekety, D.K., Schmidt, A., Kun, A.L., 2016. A model relating pupil diameter to mental workload and lighting conditions. In: Paper Presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.

Rahman, M.G., Islam, M.Z., 2011. A decision tree-based missing value imputation technique for data pre-processing. In: *The 9th Australasian Data Mining Conference: AusDM 2011*. Australian Computer Society Inc, pp. 41–50.

Rahman, A.A., Saleem, W., Iyer, V.V., 2019. Driving behavior profiling and prediction in KSA using smart phone sensors and MLAs. In: Paper Presented at the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT).

Regan, M.A., Deery, H.A., Triggs, T.J., 1998. October. Training for attentional control in novice car drivers: A simulator study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* Vol. 42 (20), 1452–1456. Sage, CA: Los Angeles, CA.

Reimer, B., Mehler, B., 2011. The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. Ergonomics 54 (10), 932–942.

Rodriguez Paras, C., 2015. Exploring Physiological Measures for Prediction and Identification of the Redline of Cognitive Workload.

Schuurmans, A.A., de Looff, P., Nijhof, K.S., Rosada, C., Scholte, R.H., Popma, A., Otten, R., 2020. Validity of the Empatica E4 wristband to measure heart rate variability (HRV) parameters: a comparison to electrocardiography (ECG). J. Med. Syst. 44 (11), 1–11.

Shahini, F., Zahabi, M., Patranella, B., Mohammed Abdul Razak, A., 2020, December. Police officer interactions with in-vehicle technologies: an on-road investigation. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Vol. 64, No. 1,. Sage, CA: Los Angeles, CA, pp. 1976–1980.

Shimomura, Y., Yoda, T., Sugiura, K., Horiguchi, A., Iwanaga, K., Katsuura, T., 2008. Use of frequency domain analysis of skin conductance for evaluation of mental workload. J. Physiol. Anthropol. 27 (4), 173–177.

Shupsky, T., Lyman, A., He, J., Zahabi, M., 2021. Effects of mobile computer terminal configuration and level of driving control on police officers' performance and workload. Hum. Factors 63 (6), 1106–1120.

Singh, R.R., Conjeti, S., Banerjee, R., 2013. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. Biomed. Signal Process Control 8 (6), 740–754.

Son, J., Oh, H., Park, M., 2013. Identification of driver cognitive workload using support vector machines with driving performance, physiology and eye movement in a driving simulator. Int. J. Precis. Eng. Manuf. 14 (8), 1321–1327.

Stern, J.A., Boyer, D., Schroeder, D., 1994. Blink rate: a possible measure of fatigue. Hum. Factors 36 (2), 285–297.

Streefkerk, J.W., van Esch-Bussemakers, M.P., Neerincx, M.A., 2006. Designing personal attentive user interfaces in the mobile public safety domain. Comput. Hum. Behav. 22 (4), 749–770.

Tivesten, E., Dozza, M., 2015. Driving context influences drivers' decision to engage in visual–manual phone tasks: evidence from a naturalistic driving study. J. Saf. Res. 53, 87–96.

Verwey, W.B., Veltman, H.A., 1996. Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. J. Exp. Psychol. Appl. 2 (3), 270.

Vila, B., Kenney, D.J., 2002. The prevalence and potential consequences of police fatigue. NIJ Journal 248, 17–21.

Wickens, C.D., 2008. Multiple resources and mental workload. Hum. Factors 50 (3), 449–455.

Williamson, A., Grzebieta, R., Eusebio, J.E., Zheng, W.Y., Wall, J., Charlton, J., Lenne, M., Haley, J., Barnes, B., Rakotonirainy, A., 2015. The australian naturalistic driving study: from beginnings to launch. In: Paper Presented at the Proceedings of the 2015 Australasian Road Safety Conference (ARSC2015).

Wozniak, D., Park, J., Nunn, J., Maredia, A., Zahabi, M., 2022. Measuring cognitive workload of novice law enforcement officers in a naturalistic driving study. In: Paper Presented at the HFES 66th International Annual Meeting, Atlanta, GA.

Yager, C., Dinakar, S., Sanagaram, M., Ferris, T.K., 2015. Emergency vehicle operator on-board device distractions. Texas A&M Transportation Institute Technical Report 2015, 1–50.

Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. Punishment: Issues and experiments 27–41.

Zahabi, M., Kaber, D., 2018a. Effect of police mobile computer terminal interface design on officer driving distraction. Appl. Ergon. 67, 26–38.

Zahabi, M., Kaber, D., 2018b. Identification of task demands and usability issues in police use of mobile computing terminals. Appl. Ergon. 66, 161–171.

Zahabi, M., Nasr, V., Mohammed Abdul Razak, A., Patranella, B., McCanless, L., Maredia, A., 2021. Effect of secondary tasks on police officer cognitive workload and performance under normal and pursuit driving situations. Hum. Factors, 00187208211010956.

Zahabi, M., Pankok Jr, C., Park, J., 2020. Human factors in police mobile computer terminals: a systematic review and survey of recent literature, guideline formulation, and future research directions. Appl. Ergon. 84, 103041.

Zahabi, M., Shahini, F., Yin, W., Zhang, X., 2022. Physical and cognitive demands associated with police in-vehicle technology use: an on-road case study. Ergonomics 65 (1), 91–104.

Zakerian, S.A., Zia, G., Nasl Seraji, G., Azam, K., Mortezapour, A., 2018. Reliability and validity of the driver activity load index for assessing mental workload among drivers in production companies. Journal of Occupational Hygiene Engineering 5 (2), 65–71. *Volume*.