



Driver Classification Using Self-reported, Psychophysiological, and Performance Metrics Within a Simulated Environment

Vishal Chandra Kummetha¹ · Umair Durrani² · Justin Mason³ · Sisinnio Concas¹ · Alexandra Kondyli⁴

Received: 22 September 2022 / Revised: 22 February 2023 / Accepted: 4 April 2023 / Published online: 12 April 2023
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2023

Abstract

Driver classification provides an efficient approach to isolating unique traits associated with specific driver types under various driving conditions. Several past studies use classification to identify behavior and driving styles; however, very few studies employ both measurable physiological changes and environmental factors. This study looked to address the shortcomings in driver classification research using a data-driven approach to assess driving tasks performed under varying mental workloads. Psychophysiological and driving performance changes experienced by drivers when engaged in simulated tasks of varying difficulty were coupled with machine-learning techniques to provide a more accurate estimate of the ground truth for behavioral classification. A driving simulator study consisting of six tasks was carefully designed to incrementally vary complexity between individual tasks. Ninety drivers were recruited to participate in both the subjective and driving components of this research. Positive and Negative Affect Schedule (PANAS), Cognitive Reflection Task (CRT), Interpersonal Reactivity Index (IRI), Empathy Assessment Index (EAI), Psychological Entitlement Scale (PES), 18-point Need for Cognition (NFC), and a basic demographic survey were administered. Time-series clustering using the Dynamic Time Warping (DTW) score within hierarchical clustering was applied to determine difference in driving styles. The use of pre-driving psychometric questionnaires to determine the most suitable metrics for predicting driving style outside of the automobile suggest promising results. The results of the binary logistic regression indicate that an individual's annual driving mileage, CRT score, age, IRI fantasy score, and EAI affective response score, contribute significantly towards predicting their driving style.

Keywords Driving simulator · Mental workload · Time-series clustering · Engagement level · Psychometric questionnaires

Introduction

Driver classification is a common strategy used to easily group individuals based on similar behavioral traits, driving style, or demographic characteristics. Classification provides an efficient approach to isolating unique characteristics associated with specific driver types under various driving

conditions. Not only can this process be used to develop better driver behavioral models for car-following or lane changing, but also to identify risk-averse driving behaviors; thus, providing context for targeted safe-driving education.

Existing research indicates that driver classification has primarily been performed using aggregate or continuous driving performance metrics and driving-centric questionnaires. Furthermore, very few studies incorporate the use of physiological (defined as the information pertaining to a driver's bodily functions such as cognition, heart rate, vision-related changes, skin conductance, perspiration, and others (Lin et al. 2006)) changes and environmental factors when assigning driving styles. To achieve a more holistic classification, this study seeks to address the above shortcomings using a data-driven approach to assess driving tasks performed under varying traffic conditions resulting in measurable changes to mental workloads (defined as the proportion of mental capacity required by an individual to perform

✉ Vishal Chandra Kummetha
kummetha@usf.edu

¹ Center for Urban Transportation Research, University of South Florida, Tampa, FL, USA

² Department of Civil and Environmental Engineering, University of Windsor, Windsor, ON, Canada

³ National Advanced Driving Simulator, University of Iowa, Coralville, IA, USA

⁴ Department of Civil, Environmental, and Architectural Engineering, University of Kansas, Lawrence, KS, USA

a task (Brookhuis et al. 1991)). We also go a step further to assess the overall performance under complex tasks such as distracted driving. The main goals of this research are to classify drivers based on their simulated driving and psychophysiological (defined as physiological changes resulting from variations of psychological processes or behavior (Kummetha 2020)) performance using a machine-learning approach and predict their classification outcome (i.e., driving profile) from the scores of various psychometric questionnaires administered prior to their driving assessment, as independent variables within logistic regression. The novelty of this research stems from the extensive driving performance and psychophysiological time-series data collected across various task complexities and for a relatively large pool of ninety participants.

The manuscript begins by providing a comprehensive overview on the existing research on subjective and quantitative driver classification. The section detailing the data and methods is then presented, diving into the participant selection and simulation, psychophysiological measures, data preparation, and methodological approach to driver classification and statistical modeling. The results, discussion, and conclusions are then presented.

Background

Several studies have employed classification to identify behavior and driving styles. Some rely on continuous measures of driving performance or physiological parameters while others take on a subjective approach relying on questionnaires. Aoude et al. developed algorithms to classify drivers by considering their behavior at intersections (Aoude et al. 2012). The authors collected roadside data which included vehicle speeds, range, and lateral positions. A GPS antenna also recorded video from all intersection approaches including the traffic signal phases. They implemented two machine-learning algorithms called Support Vector Machine (SVM) and Hidden Markov Model (HMM) to classify whether the driver was compliant or violating the signal indication. The developed model was then validated on another naturalistic dataset.

However, roadside data alone may not accurately classify drivers. Driving performance data from inside the vehicle can provide richer information for driver classification. Kondyli and Eleftheriadou observed behavioral trends within a naturalistic setting suggesting three driver behavioral types: aggressive, average, and conservative. Aggressive drivers were observed to drive at high speeds (> 15 mph speed limit), perform six discretionary lane changes, and aggressive lane changing (Kondyli and Eleftheriadou 2011). Average drivers did not exceed 10 mph over posted speed limits while conservative drivers maintained speeds within

5 mph of posted limits. Fernandez et al. used several data elements such as gas/brake pedal usage, speed indicators, and acceleration collected from a 3-D virtual driving simulator with a lifelike cockpit to group drivers (Fernández and Ito 2016). They developed a Fuzzy rule-based system that can profile the drivers into five groups (very passive, passive, normal, aggressive, and dangerous). Ly et al. built both supervised (SVM) and unsupervised (K-means clustering) machine-learning models using the inertia data collected from an instrumented vehicle's Controller Area Network (CAN) bus (Ly et al. 2013). The study showed that features associated with the acceleration did not contribute much to classify drivers, rather the braking and turning events were prominent features to differentiate between drivers. Zhang et al. also developed a similar SVM model to classify drivers based on individual characteristics sensed by only low-level sensors (Zhang et al. 2016). Low-level sensor data include data gathered from sensors that are available on smartphones along with the sensor information available through the vehicle's on-board diagnostics (OBD-II) port. Acceleration and jerk were used as main parameters instead of speed, as these variables were more closely related to driving behavior (Zhang et al. 2016). One of the shortcomings of the study was that the classifier model worked best with fewer drivers, which may not be a realistic case for real world-driving conditions. A general framework to classify drivers through mathematical strategies was proposed by Jensen et al. (2011). In-vehicle data and survey data were both collected to build the framework to classify drivers into six classes ranging from 'timid' to 'aggressive'. The in-vehicle data were collected through the OBD-II recorder and included vehicle speed, acceleration, jerk, brake pressure, steering wheel angle, and yaw rate. The survey was mostly limited to questionnaires regarding demographic characteristics of the drivers.

Further, researchers have also explored the use of physiological measures in driver classification. Lin et al. classified ten drivers by analyzing physiological measures in response to an unexpected obstacle in a virtual simulator. Drivers were classified using driving trajectory, speeds, and event-related potentials (ERP) to form two categories: aggressive and gentle (Lin et al. 2006). Zeng et al. utilized Electroencephalography (EEG) signals to classify driver's mental state (Zeng et al. 2018). Physiological data (EEG signal, blink rate, and heart rate) were collected through a driving simulator experiment. They proposed two deep convolutional neural network-based models called EEG-Conv and EEG-Conv-R, which can predict the mental state of the drivers (whether the driver is drowsy or not). The experiment and validation concluded that the developed models were more reliable and accurate for binary classification of the driver's mental state than traditional machine-learning models like SVM.

Subjective questionnaires are frequently used in driving research to capture information not readily available or captured in performance datasets. The Driver Behavior Questionnaire (DBQ) was initially created to assess whether a questionnaire could distinguish between four main factors, i.e., unfocused errors, emotional errors, reckless violations, and confused errors. The questionnaire could explain up to 33% of the variance in aggressive driving taking into account various errors and violations (Hong et al. 2014). Berdoulat et al. investigated the potential effects of aggressiveness and transgressive behavior of the drivers while driving (Berdoulat et al. 2013). They utilized five different types of scales and questionnaires to assess driving anger, aggressiveness, and impulsiveness. DBQ was used to classify three types of behaviors: “violations”, “slips and lapses”, and “Errors”.

Tao et al. investigated the role of personality traits and driving experience in risky driving behaviors among drivers in China (Tao et al. 2017). Personality traits were measured by Eysenck Personality Questionnaire (EPQ) which has 12-item subscales assessing four personality constructs such as: extraversion (e.g., “Are you a talkative person?”), neuroticism (e.g., “Does your mood often go up and down?”), psychoticism (e.g., “Do you enjoy co-operating with others?”) and lie (e.g., “Do you say some bad words casually?”). The risky driving behaviors were analyzed by the Chinese version of DBQ which modified the original four factors of DBQ into: ordinary errors and violations (15 items), lapses (6 items), aggressive violations (3 items), and speeding (3 items). Another questionnaire that has been moderately used in driving behavior studies is “The Driving Vengeance Questionnaire (DVQ)”. It was created and used to gauge how often drivers avenge themselves in typical driving circumstances (Wiesenthal et al. 2000). Wiesenthal et al. further validated the DVQ where they showed that DVQ scores were predicted by driver violence and vengeful drivers had violence into their past acts. In situations of heavy traffic, vengeful drivers were more likely to display moderate aggression.

Other psychometric evaluation techniques used in behavioral studies to assess mood and personality, cognitive engagement, and empathy and social decision-making, could also be applied to link general behavioral characteristics to driving styles. The most popular questionnaires/metrics as found in the literature are summarized in Table 1.

From the literature, it is evident that past studies have primarily performed classification using one of the three categories of data, i.e., driving performance, physiological measures, or driving-centric questionnaires. However, most metrics are directly associated with driving, without regard for other behavioral traits outside the realm of the automobile but still influence the tasks associated with driving. In addition, most of the previous driver-classification studies, focused on uncovering extreme behaviors (e.g.,

errors or risky behaviors) and their correlation with safety and crash occurrence. However, they do not investigate how driver variability and different driver styles are affected by the roadway/traffic environment and complexity, and consequently, how this relationship may impact the operation of the traffic stream. Therefore, the main objective of this study was to investigate driver variability and potential driver classes while completing driving tasks under different roadway environments and normal/crash-free operations, by combining a variety of complex psychophysiological and driving performance measures. A secondary objective was to assess whether static driver information obtained from subjective questionnaires and demographic data can equivalently be used to predict the derived driver classes. This is achieved by:

- Data fusion of psychophysiological and driving performance changes experienced by drivers when engaged in simulated tasks of varying level of difficulty, coupled with machine-learning techniques to provide a more accurate estimate of the ground truth for behavioral classification, and
- Employing pre-driving psychometric questionnaires to determine the most suitable subjective metrics for predicting driving style. If psychometric questionnaires have strong predictive qualities, this may save resources and time for defensive driving educators, researchers, driving rehabilitation specialists, and drivers.

Data and Methods

In this study, we utilize time-series datasets of complex psychophysiological and driving performance measures as the ground truth to complement a set of psychometric and demographic questionnaires to assign and predict driver classes. This section details the participant selection and driving simulation, psychophysiological measures employed, and data preparation and methodological framework.

Participants and Driving Simulation

The established methodology involves collecting driving performance and psychophysiological data from six tasks in a driving simulator setting as shown in Fig. 1. A fixed-base simulator, in a half cab, utilizing three forward screens to provide a 170° horizontal field of view and a rear screen was used in this research. Ninety participants ranging from 18 to 64 years of age were recruited for the study (mean age = 31.4 years, standard deviation (SD) = 14.2 years, Male driver to Female driver ratio of 1). PANAS, CRT, IRI, EAI, PES, 18-point NFC, and a basic demographic survey (i.e., age; gender; annual mileage traveled; crash history;

Table 1 Summary of common psychometric evaluators

Name	Description
Positive and Negative Affect Schedule (PANAS)	This is a self-report measure to assess negative affect (i.e., distressed or nervous) and positive affect (i.e., excited or proud), with ten consistent and uncorrelated items for each subscale (Watson et al. 1988). Items are rated on a five-point Likert scale: 1 = “Very slightly or not at all” to 5 = “Extremely.” The total PANAS score is computed by adding the positive and negative subscale results
Cognitive Reflection Task (CRT)	Consists of a set of three questions designed to measure an individual’s ability to suppress an intuitive and spontaneous wrong answer in favor of a reflective and deliberative right answer (Frederick 2005; Kummetha et al. 2020). CRT score is used as a measure of inhibition, an important component of executive function, which is crucial for safe driving. The questions include: Qn 1: “A bat and ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?” Qn 2: “If it takes 5 machines 5 min to make 5 widgets, how long do 100 machines take to make 100 widgets?” Qn 3: “In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?”
Interpersonal Reactivity Index (IRI)	Questionnaire used to measure individual differences in empathy and consists of four subscales (i.e., perspective taking-PT, fantasy-FS, empathic concern-EC, and personal distress-PD) with seven items in each (Davis 1983). Including empathy as a variable could be key in understanding driver behavior especially when sharing the road with other users or in instances that require cooperation such as merging or lane changes
Empathy Assessment Index (EAI)	Measure based on cognitive neuroscience conceptualization of empathy consisting of a 50-item questionnaire with five sub-dimensions (i.e., affective response-AR, self-other awareness-SA, emotion regulation-ER, perspective taking-PT, and empathic attitudes-EA) (Lietz et al. 2011). The total EAI score can be computed by adding the sum of the five sub-dimensions
Psychological Entitlement Scale (PES)	This scale measures the stable and pervasive sense that one deserves more than others (Campbell et al. 2004; Kummetha 2020). This sense of entitlement will also be reflected in desired or actual behaviors. The concept of psychological entitlement is intrapsychically pervasive or global and does not necessarily refer to entitlement that is situation specific (e.g., “I deserve an ‘A’ because I performed well in class”) (Campbell et al. 2004). Rather, psychological entitlement is a sense of entitlement that is experienced across all situations. The scale consists of nine self-reported measures. This questionnaire has not been used in driving research, but entitlement could play a vital role in predicting aggressive driving behaviors
Need for Cognition (NFC)	This test is designed to assess the tendency to engage in and enjoy effortful cognitive endeavors (Cacioppo et al. 1984). Driving is an instrumental activity of daily living that requires cognitive resources of the driver and a tendency to engage in cognitive endeavors may predict assertive (i.e., not passive) driver behaviors. The test entails of both a short and a long form that consist of 18 and 34 self-assessment questions, respectively

Fig. 1 KU driving simulator in action

insurance type; traffic violation history; cell phone use while driving; and education level) were administered to drivers prior to their participation in the simulated drive. Drivers

were compensated for their participation in the study with a \$50 gift card. Table 2 presents a summary of the demographic properties of the recruited drivers.

Table 2 Summary of participant demographics

Variable	<i>n</i>	Mean	SD
Age			
18 to 24	45	20.3	1.4
25 to 49	29	35.0	8.0
50 to 65	16	56.3	3.7
Gender			
Male	45	na	na
Female	45	na	na
Annual mileage			
1000 to 1999	6	1150.0	281.1
2000 to 4999	4	2500.0	1000.0
5000 to 9999	23	6760.9	1166.5
10,000 to 19,999	42	12,035.7	2445.5
> 19,999	15	27,066.7	9027.5
Crash history			
0 to 1	82	na	na
2 to 3	8	na	na
> 3	0	na	na
Insurance type			
Liability	36	na	na
Comprehensive	11	na	na
Collision	41	na	na
Did not answer	2	na	na
Traffic violation history			
0 to 1	83	na	na
2 to 3	4	na	na
> 3	3	na	na
Cell phone use			
Scale of 1 to 10 (1-never to 10-always)	90	3.8	2.4
Education level			
High school	5	na	na
College student	51	na	na
Bachelor's degree	18	na	na
Graduate student	4	na	na
Graduate degree or higher	12	na	na

Participants' eligibility for the study was determined based on the following criteria:

- Aged between 18 and 65 years and in possession of a valid U.S. driver's license;
- At least 1 year of driving experience with an annual mileage of not less than 1,000 miles; and
- In good health (no eye conditions apart from corrective lenses, free from any heart conditions and seizures, no history of inner ear/balance problems, and little to no known history of severe motion sickness).

For the simulated drive, a tutorial and six tasks consisting of select external conditions as detailed in Table 3 were created. The tutorial comprised of familiarizing participants with the simulator environment and equipment, particularly gas and brake pedal responsiveness, steering wheel sensitivity, speed and gap adjustment, and lane changing activity. The tutorial also served as a tool to screen for simulator sickness and any participants that showed signs of severe simulator sickness (i.e., vertigo, nausea, sweating, dizziness, fatigue, stomach awareness, vomiting, and general discomfort) were advised to forfeit with a partial monetary compensation.

Each task was 8 km long with a posted speed limit of 70 mph (112.7 km/h) and was designed to capture incremental changes to behavioral and driving performance measures. Tasks varied in traffic density, but drivers were free to maintain any gap or speed. Emphasis was placed in scenario design to carefully control in-vehicle body movements, visual and auditory discrepancies, and task uniformity, as these significantly affect data quality of psychophysiological measures. The task sequence for each participant was completely randomized to eliminate order bias. Further, psychophysiological measures were continuously collected during the drive by capturing micro-level changes to vision and attention. At the end of each task, participants were required to complete the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) (Hart 1986) and Situation Awareness Rating Technique (SART) (Selcon and Taylor 1990) questionnaires to provide a well-rounded subjective baseline.

Tasks 4, 5, and 6 shared the same roadway configuration as shown in Fig. 2. Task 6 additionally required participants to attempt a visual secondary task on a 10-in. Windows tablet, up to four times at random intervals no longer than 20 s each.

The tablet was placed approximately 30° to the lower right from drivers' line of sight and was used to simulate the act of driving while using the media center or entering text into an on-board navigation system as shown in Fig. 3 (Kummetha 2020; Kummetha et al. 2020). The data collection zone was limited to the middle 6.4 km to minimize data irregularities arising from the beginning and ending of the drive, shown in Fig. 2.

Psychophysiological measures

For the continuous measurement of vision-related metrics such as pupil diameter, gaze point vectors, blinks, and fixations, a Fovio FX3 eye tracker with a data collection frequency of 60 Hz was utilized (EyeTracking-Inc. 2019). Mental workload was established from the eye tracker in one-second increments using a patented pupillometric technique called Index of Cognitive Activity (ICA) (Vogels et al.

Table 3 Summary of common psychometric evaluators

	Composition	Work zone type	Traffic density	Distraction
Task 1 (baseline)	4-lane divided highway 0% heavy vehicles	None	14–16 pc/km/ln (LOS B/C)	None
Task 2	4-lane divided highway 0% heavy vehicles	None	22–24 pc/km/ln (LOS D/E)	None
Task 3	4-lane divided highway 10% heavy vehicles	Inactive: left shoulder closed	22–24 pc/km/ln (LOS D/E)	None
Task 4	10-lane divided freeway 20% heavy vehicles	Active on both sides: 3 lanes closed	14–16 pc/km/ln (LOS B/C)	None
Task 5	10-lane divided freeway 20% heavy vehicles	Active on both sides: 3 lanes closed	22–24 pc/km/ln (LOS D/E)	None
Task 6	10-lane divided freeway 20% heavy vehicles	Active on both sides: 3 lanes closed	22–24 pc/km/ln (LOS D/E)	Secondary task

*Level of service (LOS) indicates freeway quality of service on a scale of A–F (Highway Capacity Manual 2010)

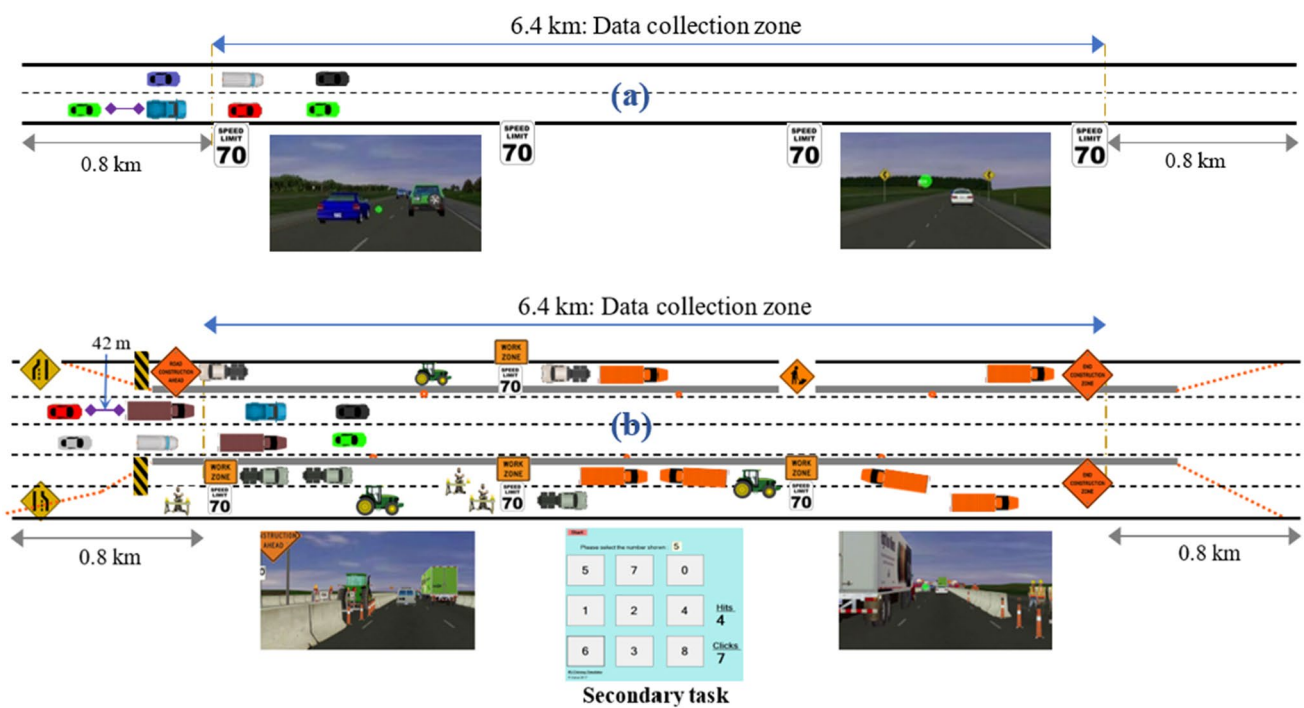
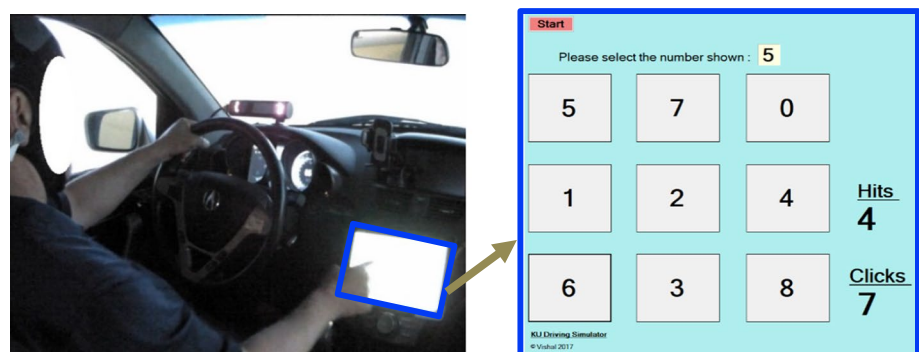


Fig. 2 Layout of tasks **a** Tasks 1 and 2, **b** Tasks 4, 5, and 6 (Kummetha et al. 2021)

Fig. 3 Secondary task used for distracted driving (Kummetha et al. 2021)



2018). ICA was chosen over the raw pupil diameter metric because of its ability to disentangle pupil response arising from changing lighting conditions especially prominent in virtual environments (Vogels et al. 2018). ICA scores vary continuously between 0 and 1 indicating the lower and higher spectrum of mental workloads, respectively.

Engagement level (EL), defined as the level of engagement/arousal experienced in response to a particular task, was an important component in this research (Pope et al. 1995). Enobio 8 portable Bluetooth-based system was used to capture participants baseline and time-correlated EEG at 500 Hz (Neuroelectrics 2019). Eight EEG electrode positions (i.e., P3, PZ, P4, CZ, T7, T8, O1, and FZ) aligned with the 10–20 international mapping layout and findings of Pope et al. were used as shown in Fig. 4 (Pope et al. 1995). An ear clip was also used to add two reference electrodes, i.e., common mode sense (CMS) and driven right leg (DRL). Driver EL was calculated using the power spectral density (PSD) relationship $\beta/(\alpha + \theta)$ established by Pope et al. and Prinzel III et al. to assess alertness and engagement, mental effort, and attention investment; with theta band (4–8 Hz), alpha band (8–13 Hz): linked to mental workload, and cognitive fatigue, and beta band (13–22 Hz): linked to visual attention and short-term memory (Pope et al. 1995; Prinzel 2001).

Data were first system time-synchronized to identify critical analysis zones using MATLAB. The EEG data were isolated using unique identifiers and split into 10-s events. This resulted in approximately twenty 10-s events for each task with the goal of 20-point estimates of EL per task. Post-processing was performed using the EEGLAB v14.1.1 toolbox in MATLAB (Delorme and Makeig 2004). A custom MATLAB script was developed to convert data from the native Neuroelectrics-Enobio 8 (.easy) format to EEGLAB (.set). The EEG data were then resampled from their native 500 to 240 Hz to ensure noise and processing time reduction. The dataset was filtered by applying a high-pass filter at 1.6 Hz and a notch filter at 60 Hz. An independent component analysis was performed by identifying and pruning artifacts from blinks and muscle movement to improve

signal quality. This was followed by the application of a low-pass filter at 40 Hz. The filtered signal was then split into 2-s epochs with 5% overlap. The Darbelai plugin in EEGLAB was applied to perform a fast Fourier transformation (FFT) using a 2-s window to generate PSD values for each electrode position. Using $\beta/(\alpha + \theta)$ at PZ, CZ, P3, and P4 electrode locations, a single combined PSD value representing driver EL was obtained every 10-s of the drive. PSD combined values closer to 0 indicate lower EL.

Further, a Polar H10 chest strap was used to monitor drivers heart rate (HR) at 1 Hz as a surrogate measure of mental workload. HR data were manually synchronized with the frames of the driving simulator output. Baseline HR data were collected prior to the start of the drive and HR variability within participants was normalized with respect to the baseline.

Data Preparation and Methods Summary

Due to the vast data collection techniques used in the study design, data fusion is employed to merge all output elements. The data fusion framework used to uniformly synchronize all data variables to 10 Hz is shown in Fig. 5.

Figure 6 summarizes the methodological approach with respect to five main stages, i.e., psychometric questionnaires, simulator study, data fusion, driver classification, and the prediction model.

Results

Task-Specific Analysis

As each task was designed to capture performance and behavioral variability, changes in mental workload and situation awareness between individual tasks were assessed using a series of repeated measure analysis of variances (ANOVAs) with task 1 as the baseline. The ANOVAs were used to test the null hypothesis that the driver-reported

Fig. 4 EEG electrode positions used (yellow) (Neuroelectrics 2019)

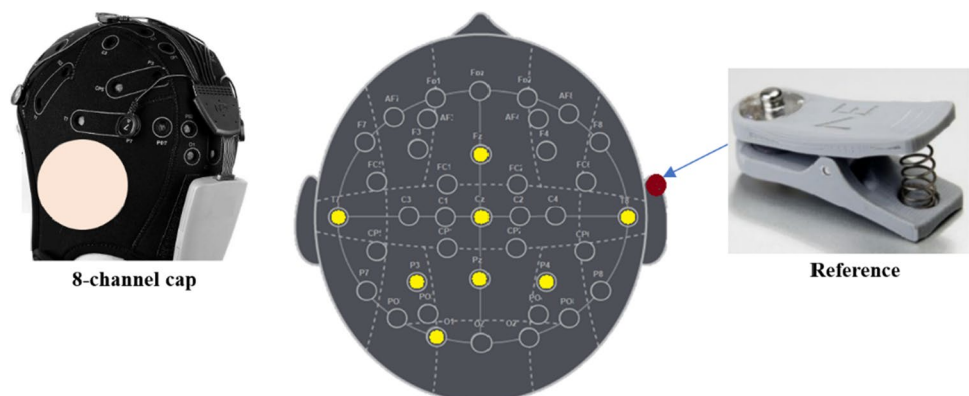


Fig. 5 Data fusion framework

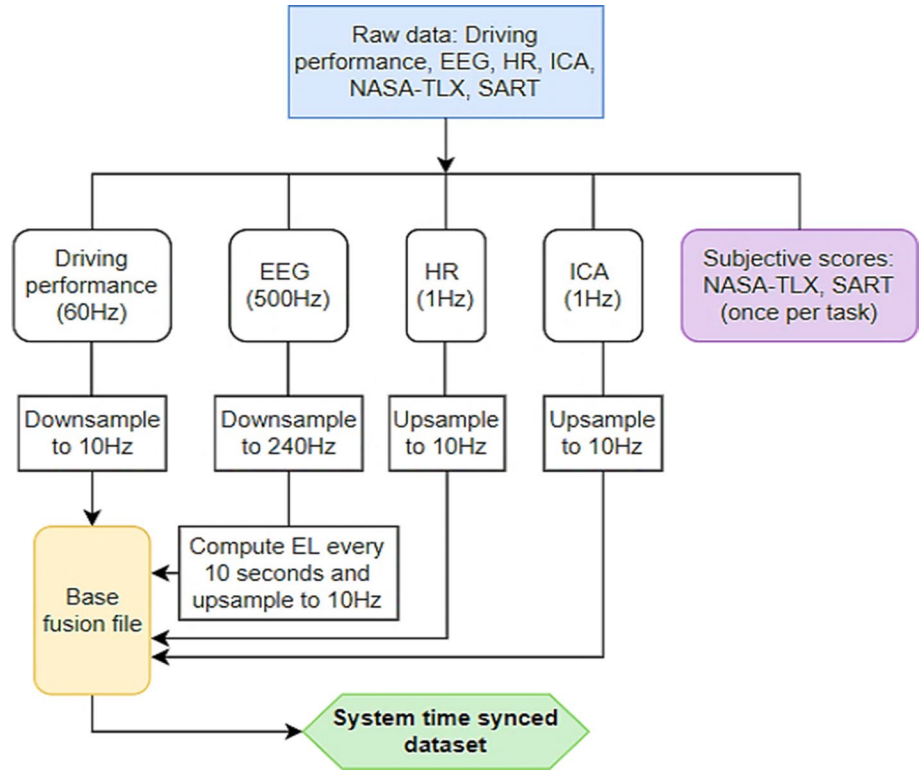
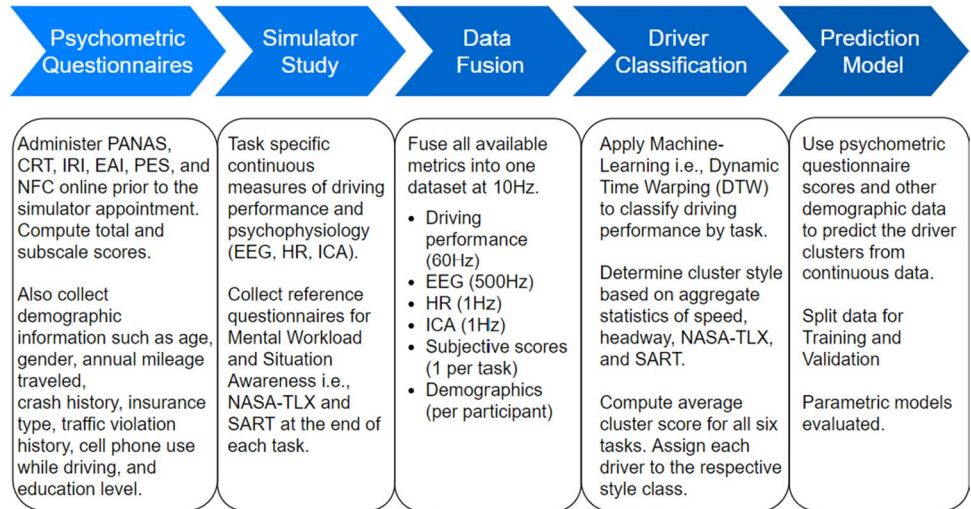


Fig. 6 Summary of methodological workflow



means of the average NASA-TLX and SART scores across the six tasks were similar and to identify the presence of any statistically significant differences, providing quantitative validity for the experimental design in terms of the varying complexity of the attempted driving tasks. The average NASA-TLX scores showed no significant differences ($\alpha = 0.05$) between the mean scores of task 2 and task 1 (baseline). However, significant differences in scores were observed between the baseline and tasks 3 ($F(1, 83) = 4.087, p = 0.046, \eta_p^2 = 0.047$), 4 ($F(1, 83) = 16.298, p < 0.001, \eta_p^2 = 0.164$), 5 ($F(1, 83) = 35.230, p < 0.001,$

$\eta_p^2 = 0.298$), and 6 ($F(1, 83) = 201.257, p < 0.001, \eta_p^2 = 0.708$). The significant increases in NASA-TLX scores suggest that the developed tasks successfully captured various levels of mental workload, shown in Fig. 7.

Similar trends across tasks were observed with the SART scores. No significant differences were observed between the baseline and the mean scores of the task 2 and task 3; however, task 4 ($F(1, 83) = 7.448, p = 0.008, \eta_p^2 = 0.082$), task 5 ($F(1, 83) = 7.840, p = 0.006, \eta_p^2 = 0.086$), and task 6 ($F(1, 83) = 26.794, p < 0.001, \eta_p^2 = 0.244$)

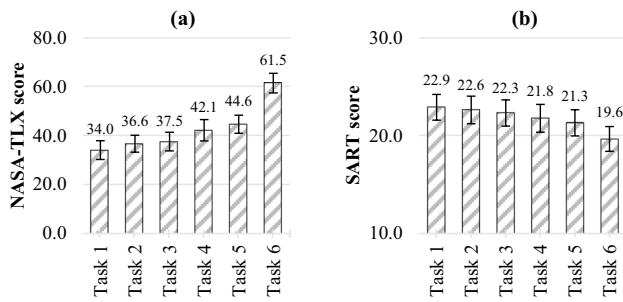


Fig. 7 Average a NASA-TLX and b SART scores

showed significant differences. Overall, with greater task complexity an increase in subjective mental workload was observed along with a decrease in situation awareness (Kummetha and Kondyli 2022).

Time-Series Clustering

As tasks were constant between participants, time-series clustering using the Dynamic Time Warping (DTW) algorithm was selected for the analysis. The DTW algorithm provides a robust method of measuring and matching similar time-series datasets, such as trajectory and psychophysiological data in our case. The DTW algorithm also allows us to compare time-series datasets of different lengths, or even with time lag between two pairs of interest, especially useful in driving studies where no two events can be completely identical (Giorgino 2009; Mueen and Keogh 2016). The DTW score (dissimilarity) for the two time-series ($D1, D2$) can be calculated using Eq. 1. The higher the DTW score the larger the dissimilarity and the closer the score to

zero the more identical the observed pattern in the selected variable across the pair of drivers (Durrani and Lee 2019).

$$DTW(D1, D2) = \frac{1}{K} \sqrt{\sum_{k=1}^K d_k}, \tag{1}$$

where d_k is the distance in cell k in the shortest path. K is the total number of cells in the shortest path.

The same approach is employed to compare more than two time-series trends. In this study, acceleration (m/s^2), steering wheel angle (degrees), headway (m), mental workload as ICA from left eye, and EL, were used for time-series clustering using the hierarchical clustering algorithm with DTW score as the distance metric (Yan et al. 2019). HR was not included as ICA was deemed to be more accurate for measuring the mental workload. Hierarchical clustering was selected due to its ability to provide intuitive visualization of the clustering process and unsupervised identification of cluster numbers (Johnson 1967). Hierarchical clustering was performed with various iterations of the number of clusters (i.e., $n = 2$ to 5), resulting in an optimal solution consisting of two clusters. The Silhouette coefficient was also calculated for various n values to verify the optimal solution and the best/largest coefficient value (0.33) was obtained using $n = 2$ (Rousseeuw 1987). Hierarchical clustering does have some limitations associated with sensitivity to the distance metric and inability to handle missing data points, however, with careful selection of the distance metric (i.e., DTW score) the effects were minimized.

The summary of the aggregate metrics for the two clusters established for each task is shown in Table 4. The two established clusters did not exhibit qualities of aggressive driving as evident from the average speeds and headways. Average speeds were within 6 km/h of the posted speed

Table 4 Cluster properties

	Cluster ID	N	Avg License Age	Avg Headway (m)	Avg Speed (km/h)	Avg NASA-TLX	Avg SART
Task 1	1	72	12.9	88.2	118.2	34.6	22.7
	2	13	18.2	209.9	114.7	32.3	24.8
Task 2	1	60	13.4	91.6	115.2	36.9	22.6
	2	25	14.9	109.9	114.7	35.8	22.6
Task 3	1	40	12.7	96.0	114.4	37.9	22.0
	2	45	15.3	89.5	113.0	37.0	22.8
Task 4	1	28	15.1	78.6	114.4	41.6	22.4
	2	56	13.7	103.6	112.0	42.0	21.7
Task 5	1	43	13.2	68.4	114.0	43.5	20.8
	2	41	15.8	104.1	111.4	45.5	21.9
Task 6	1	47	13.4	85.5	110.7	57.5	20.6
	2	38	14.9	143.1	107.7	65.7	18.6

*Cluster ID—1: Moderate; 2: Conservative

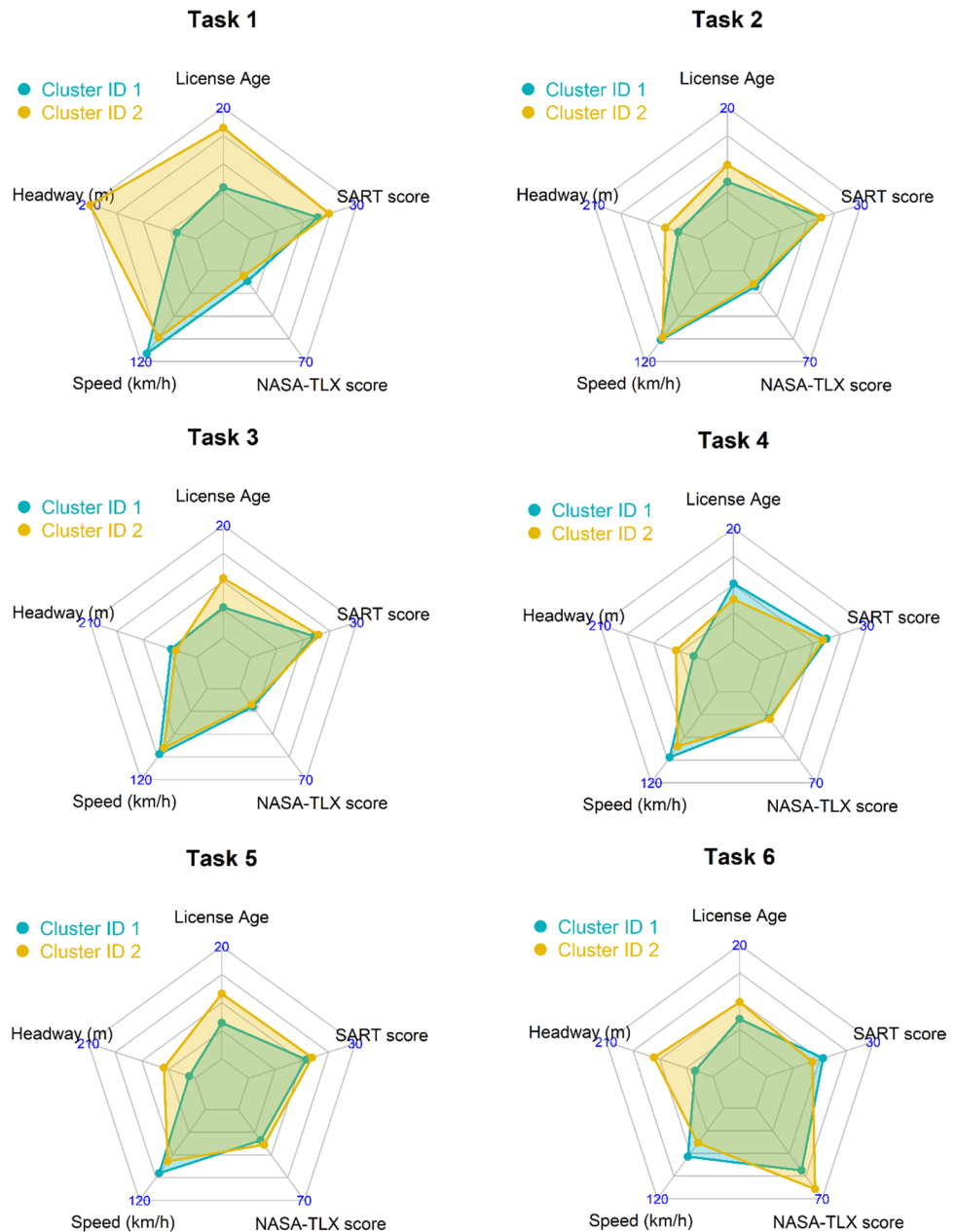
limit while the smallest headway was 68.4 m (2.2 s). Drivers showed distinct properties within these clusters, including differences in self-reported measures of mental workload (NASA-TLX score) and situation awareness (SART score).

Cluster outputs from the six driving tasks were aggregated to individually classify drivers as having moderate or conservative driving styles i.e., if a driver was classified to represent moderate driving style in three or more of the tasks, they were classified as moderate, else conservative. This was necessary because although hierarchical clustering classifies drivers, it is not apparent as to what each cluster represents, and the aggregate cluster properties help differentiate the behavioral composition within them. Figure 8

shows the mean variable spread of the conservative and moderate clusters across the six tasks. Cluster ID 2 in almost all instances has a larger surface spread, with the exception of the speed axis. This surface spread explains the differences between the two clusters and suggests that moderate drivers (i.e., Cluster 1) were younger with less driving experience and on average maintained shorter headway, at higher but safe speeds, and exhibited lower mental workload and higher situation awareness compared to conservative drivers.

The recorded increase in self-reported mental workload and decrease in situation awareness across both driver clusters could be a result of driving under extreme/non-typical conditions (such as Task 6 with the presence of a

Fig. 8 Radar charts showing the mean spread across the two cluster ids



distraction). Out of 90 drivers, 84 drivers completed all the components of the questionnaires and driving tasks. 20 drivers were classified as overall conservative across all tasks while 64 drivers were classified as moderate.

Logistic Regression

A 70/30 training–testing stratified split was applied to each of the obtained cluster IDs before performing further analysis. Binary logistic regression, shown in Eq. 2, is a form of statistical modeling appropriate for independent binary outcome variables, such as in classification problems to distinguish one class from another. Further, binary logistic regression is capable of handling both continuous and categorical independent variables and provides informative outcomes expressed as odds ratios. In this application, binary logistic regression was performed to distinguish between conservative (indicated by the value 0) and moderate (indicated by the value 1) drivers and predict their driving styles, using the psychometric metric and demographic questionnaires data.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n, \quad (2)$$

where $\log()$ is the natural logarithm. p is the probability of the dependent variable. β_0 is the intercept of the constant term. β_1, \dots, β_n are the coefficients assigned to the independent variables X_1, \dots, X_n

Training

A total of 10 iterations were performed using the following independent variables from the survey and questionnaires: PANAS (subscales: positive affect, negative affect), CRT, IRI (subscales: perspective taking, fantasy, empathic concern, and personal distress), EAI (subscales: AR, SA, ER, PT, EA), PES, 18-point NFC, and other demographics (i.e., age; gender; annual mileage traveled; crash history; insurance type; traffic violation history; cell phone use while driving; and education level). A balance between over-fitting the

model and classification accuracy was achieved by excluding correlated parameters within the questionnaires. The backward stepwise selection was implemented to sequentially remove statistically insignificant variables in the model. Manual checks were performed to ensure that the included independent variables were contributing substantially to the beta coefficient (B) i.e., $\text{Exp}(B)$ and 95% confidence intervals deviating considerably from 1. The results of the logistic regression are shown in Table 5.

The results from the binary logistic regression indicate that an individual’s annual mileage, CRT score, Age, IRI fantasy (FS), and EAI affective response (AR) can significantly help in predicting their driving style (moderate or conservative) with a training prediction accuracy of 87.7%. Figure 9 shows the scatter plots of the training dataset with respect to the significant variables in the logistic regression model. Results indicate a weak to moderate correlation between Annual Mileage and Age ($r=0.39$) and CRT score and EAI AR ($r=-0.32$), suggesting no substantial effect of multicollinearity.

Testing

Results from the confusion-matrix of the test sample, shown in Table 6, indicate a higher-than-average classification accuracy of 65.4% and a relatively good recall, precision score, and F_1 score, suggesting a good predictive model (especially for moderate drivers). However, for conservative drivers the percent of correct predictions is relatively low at 33%. This may be attributed to factors such as the relatively low sample size of the conservative drivers, selected training–testing data split, and the existence of other non-linear relationships between the variables not accounted for within the binary logistic regression.

Discussion

An important objective of this research was to investigate driver variability under different roadway/traffic environments and task complexity. This was achieved through the

Table 5 Logistic regression summary

Variable	<i>B</i>	S.E	Wald	<i>df</i>	Sig	Exp(<i>B</i>)	95% C.I. Exp(<i>B</i>)	
							Lower	Upper
Annual mileage in 1000 miles	0.436	0.135	10.372	1	0.001**	1.547	1.186	2.017
CRT score	−0.925	0.460	4.036	1	0.045*	0.397	0.161	0.978
Age	−0.110	0.040	7.413	1	0.006**	0.896	0.828	0.970
IRI FS	−0.151	0.071	4.556	1	0.033*	0.860	0.748	0.988
EAI AR	−0.337	0.125	7.230	1	0.007**	0.714	0.559	0.913
Constant	9.904	3.424	8.369	1	0.04*	20,019.582		

* <0.05, ** <0.01

Fig. 9 Scatter plots and respective Pearson correlation coefficients

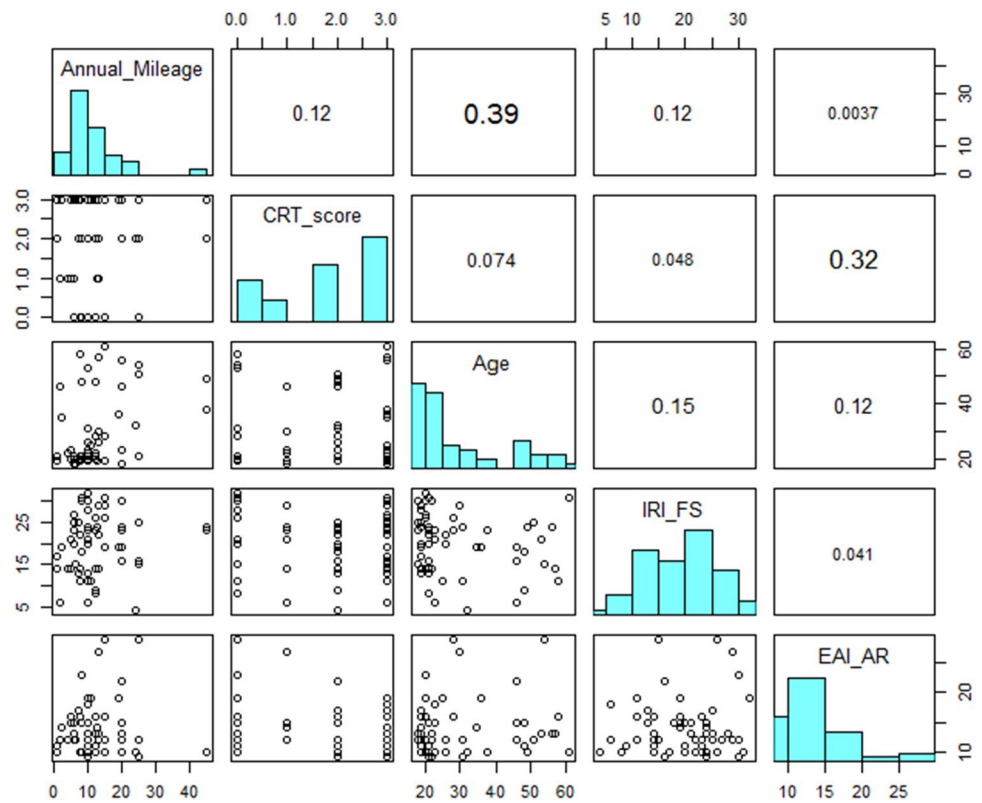


Table 6 Confusion matrix

Confusion-Matrix	Moderate (1)	Conservative (0)	% Correct (%)
Moderate (1)	TP=15	FP=5	75.0
Conservative (0)	FN=4	TN=2	33.3
Classification accuracy 65.4%; Precision=0.750; Recall=0.789; $F_1=0.769$			

TP true positive, *TN* true negative, *FP* false positive, *FN* false negative

testing of six driving simulator scenarios (Table 3) with multiple levels of task difficulty by varying the traffic, geometry, and control conditions, and presence of distraction. It was found that, as task difficulty increased, mental workload increased and situation awareness decreased, supporting the linear relationship between these variables and ensuring that our data collection methods were appropriate to study these relationships.

Time-series clustering using the hierarchical approach with the DTW score as the distance metric was also performed to classify drivers by utilizing a holistic approach where both performance metrics (acceleration, headway, steering wheel angle) and psychophysiological characteristics (mental workload, situation awareness) were adopted. The two derived clusters (i.e., moderate and conservative) indicate distinct differences in driver age, driving

experience, average headway, average speed, and average subjective scores for mental workload and situation awareness, similar to the observations made by Lin et al. (2006) and Manjunatha et al. (2019).

The second modeling approach of binary logistic regression was used to investigate whether we can use offline, static driver information to accurately cluster drivers. The results of the logistic regression indicate that annual driving mileage, CRT score, Age, IRI fantasy score, and EAI affective response score, can be used to predict driving styles with a decent accuracy and precision. The odds ratios ($Exp(B)$) for the significant variables shown in Table 5 were assessed for more insights into the findings. The odds ratio for annual mileage suggests that an increase in reported mileage increases the chance of being a moderate driver. To be specific, every 1000 miles increase in the reported annual mileage is associated with a 54.7% increase in the odds of being a moderate driver. The CRT score can be used as a measure of intuitive thinking and objective assessment, an important component of executive function, which is crucial for safe driving. The results suggest that an increase in the CRT score (indicating higher level of objective assessment) by one point decreases the odds of being a moderate driver by 60.7%. Similarly, for every 1-year increase in age, the odds of being classified as a moderate driver decreases by 10.4%.

The IRI FS score, which represents an individual's tendency to transpose into imaginative feeling and characters, was also found to be significant. A one-point increase in IRI FS score decreases the odds of being classified as a moderate driver by up to 14%. Although, not directly intuitive to driving, inference can be made from the fact that a high IRI FS score indicates the general tendency to use more cognitive resources engaging in thoughts while driving suggesting a slower behavioral adaptation to changes in conditions, typically associated with decreased driving performance (Davis 1983).

EAI AR, which is a subcomponent of empathy derived from affective response, was also found to be significant. An increase in the EAI AR score by one point suggests up to 28.6% decrease in the likelihood of being classified as moderate driver. This result seems plausible and similar to the past research (Wiesenthal et al. 2000) as an increase in empathy could indicate less vengeance and lead to more conservative driving behavior such as letting motorists cut in front or allowing merging at on/off ramps.

Overall, the results show similarities with past research utilizing other forms of questionnaires with similar subcomponents of personality traits. However, the differences lie within the variety of psychometric and psychophysiological data collected, multi-complexity experimental design, clustering approach, and prediction methods. The combined accuracy of predicting driving style from the logistic regression performed on the training and testing datasets was 76.6%, which is higher than similar studies using non-driving performance metrics for classification (Liao et al. 2022; Razmara et al. 2018).

Conclusions

This study aimed at addressing the shortcomings in driver classification research using a data-driven approach and assess driving tasks performed under varying mental workloads. A simulator study was carefully designed to achieve variability in complexity between individual tasks. Ninety drivers were recruited to participate in both the subjective and driving components of this research. The main goals of this research were achieved by successfully performing data fusion of psychophysiological and driving performance. Time-series clustering using the DTW score as the distance metric within the hierarchical clustering was performed to provide a more accurate and holistic estimate of the ground truth for behavioral classification.

Evaluating the use of pre-driving psychometric questionnaires to determine the most suitable metrics for predicting driving style outside of the automobile suggest promising results. Binary logistic regression summary

indicates that an individual's annual driving mileage, CRT score, Age, IRI fantasy score, and EAI affective response score, can be directly attributed to identifying their driving style. The results provide a promising outlook towards the practical strength of using psychometric evaluators in driver profiling with respect to personality, attitudes, and habits, risk assessment to develop targeted interventions for road safety improvement, deeper understanding of driver behavior and decision making processes that affect the conscious and subconscious components of driving, development of training and education programs targeted towards pre-screened individuals, and design/development of intelligent and autonomous transport systems to tailor behavioral needs and preferences of individual drivers.

Although the research goals were achieved, the limitations of simulator-based driving data should be acknowledged. However, the use of complex physiological equipment such as EEG and presence of hazardous conditions (in-vehicle distractions) might not be feasible for real-world testing. A larger sample size would provide more concrete evidence of the observed findings. Lastly, large sample sizes are also needed to investigate the attributes of the aggressive driving style cohort, since the existing dataset only captured moderate and conservative driving styles.

Based on the conclusions derived from this study, future research can utilize a similar methodology to study how psychometric evaluators can be used to predict driver behavior during partial (i.e., adaptive cruise control and lane centering) and full automation. Another aspect to explore would be utilizing psychometric evaluators to subjectively predict individual driving styles during travel-altering emergencies such as pandemics (i.e., coronavirus disease 2019 (COVID-19) travel restrictions/lockdowns) and disaster evacuations (i.e., hurricanes, wildfires, floods, and other types of severe weather).

Acknowledgements The authors would like to thank the Mid-America Transportation Center for funding and supporting this research initiative. They would also like to thank Dr. Evangelia Chryssikou from Drexel University for aiding with the questionnaires and methodological developments.

Author Contributions Conceptualization: VCK, AK; methodology: VCK, UD, JM; formal analysis and investigation: VCK, UD, JM, SC; writing—original draft preparation: VCK, UD; writing—review and editing: VCK, UD, JM, SC, AK; funding acquisition: AK. All authors read and approved the final manuscript.

Funding This work was supported by the Mid-America Transportation Center (MATC) under the Grant number 69A3551747107.

Availability of Data and Materials The datasets generated during and/or analyzed during the study are not publicly available due to them containing sensitive information that could compromise research participant privacy/consent but are available from the corresponding author [VK] on reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Aoude GS, Desaraju VR, Stephens LH, How JP (2012) Driver behavior classification at intersections and validation on large naturalistic data set. *IEEE Trans Intell Transp Syst* 13(2):724–736. <https://doi.org/10.1109/TITS.2011.2179537>
- Berdoulat E, Vavassori D, Sastre MT (2013) Driving anger, emotional and instrumental aggressiveness, and impulsiveness in the prediction of aggressive and transgressive driving. *Accid Anal Prev* 50:758–767. <https://doi.org/10.1016/j.aap.2012.06.029>
- Brookhuis KA, de Vries G, de Waard D (1991) The effects of mobile telephoning on driving performance. *Accid Anal Prev* 23(4):309–316. [https://doi.org/10.1016/0001-4575\(91\)90008-s](https://doi.org/10.1016/0001-4575(91)90008-s)
- Cacioppo JT, Petty RE, Feng Kao C (1984) The efficient assessment of need for cognition. *J Pers Assess* 48(3):306–307. https://doi.org/10.1207/s15327752jpa4803_13
- Campbell WK, Bonacci AM, Shelton J, Exline JJ, Bushman BJ (2004) Psychological entitlement: interpersonal consequences and validation of a self-report measure. *J Pers Assess* 83(1):29–45. https://doi.org/10.1207/s15327752jpa8301_04
- Davis MH (1983) Measuring individual differences in empathy: evidence for a multidimensional approach. *J Pers Soc Psychol* 44(1):113–126. <https://doi.org/10.1037/0022-3514.44.1.113>
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134(1):9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Durrani U, Lee C (2019) Investigating heterogeneity in car-following behavior due to driving style and lead vehicle type: implications for the wiedemann model. In: 98th transportation research board annual meeting, Washington, DC
- EyeTracking-Inc. (2019) EyeWorks 3 User Manual
- Fernández S, Ito T (2016) Driver classification for intelligent transportation systems using fuzzy logic. In: 2016 IEEE 19th international conference on intelligent transportation systems (ITSC), pp 1212–1216
- Frederick S (2005) Cognitive reflection and decision making. *J Econ Perspect* 19(4):25–42. <https://doi.org/10.1257/089533005775196732>
- Giorgino T (2009) Computing and visualizing dynamic time warping alignments in R: the dtw package. *J Stat Softw* 31(7):1–24. <https://doi.org/10.18637/jss.v031.i07>
- Hart SG (1986) NASA task load index (TLX)
- Highway Capacity Manual (2010) Fifth edition. Washington, D.C.: Transportation Research Board, c2010-. <https://search.library.wisc.edu/catalog/9910110589002121>
- Hong J-H, Margines B, Dey AK (2014) A smartphone-based sensing platform to model aggressive driving behaviors. In: Proceedings of the SIGCHI conference on human factors in computing systems, Toronto, Ontario, Canada. <https://doi.org/10.1145/2556288.2557321>
- Jensen M, Wagner J, Alexander K (2011) Analysis of in-vehicle driver behaviour data for improved safety. *Int J Veh Saf* 5(3):197–212. <https://doi.org/10.1504/ijvs.2011.042850>
- Johnson SC (1967) Hierarchical clustering schemes. *Psychometrika* 32(3):241–254
- Kondyli A, Elefteriadou L (2011) Modeling driver behavior at freeway-ramp merges. *Transp Res Rec* 2249(1):29–37. <https://doi.org/10.3141/2249-05>
- Kummetha VC, Kondyli A (2022) Simulator-based framework to incorporate driving heterogeneity via a biobehavioral extension to the intelligent driver model. *Transport Res Rec* 1(1):03611981221134630. <https://doi.org/10.1177/03611981221134630>
- Kummetha VC, Kondyli A, Chrysikou EG, Schrock SD (2020) Safety analysis of work zone complexity with respect to driver characteristics — A simulator study employing performance and gaze measures. *Accident Anal Prev* 142:105566. <https://doi.org/10.1016/j.aap.2020.105566>
- Kummetha VC, Kondyli A, Devos H (2021) Evaluating driver comprehension of the roadway environment to retain accountability of safety during driving automation. *Transport Res Part F Traffic Psychol Behav* 81:457–471. <https://doi.org/10.1016/j.trf.2021.07.007>
- Kummetha VC (2020) Incorporating biobehavioral architecture into car-following models: a driving simulator study, University of Kansas]. <https://www.proquest.com/openview/59eab8d718164956a0befca0e85df097/1?cb1=18750&diss=y&pq-origsite=gscholar>
- Liao X, Mehrotra S, Ho S, Gorospe Y, Wu X, Mistu T (2022) Driver profile modeling based on driving style, personality traits, and mood states. In: 2022 IEEE 25th international conference on intelligent transportation systems (ITSC)
- Lietz CA, Gerdes KE, Sun F, Geiger JM, Wagaman MA, Segal EA (2011) The Empathy Assessment Index (EAI): a confirmatory factor analysis of a multidimensional model of empathy. *J Soc Soc Work Res* 2(2):104–124. <https://doi.org/10.5243/jsswr.2011.6>
- Lin C, Liang S, Chao W, Ko L, Chao C, Chen Y, Huang T (2006) Driving style classification by analyzing EEG responses to unexpected obstacle dodging tasks. In: 2006 IEEE international conference on systems, man and cybernetics
- Ly MV, Martin S, Trivedi MM (2013) Driver classification and driving style recognition using inertial sensors. *IEEE Intell Veh Symp (IV)* 2013:1040–1045
- Manjunatha P, Kummetha VC, Kondyli A, Elefteriadou L (2019) Validating the task-capability extension to the Intelligent Driver Model (IDM) using driving simulator data. In: 98th annual meeting of the transportation research board
- Mueen A, Keogh E (2016) Extracting optimal performance from dynamic time warping. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, California, USA. <https://doi.org/10.1145/2939672.2945383>
- Neuroelectrics (2019) Neuroelectrics User Manual-Enobio 8
- Pope AT, Bogart EH, Bartolome DS (1995) Biocybernetic system evaluates indices of operator engagement in automated task. *Biol Psychol* 40(1):187–195. [https://doi.org/10.1016/0301-0511\(95\)05116-3](https://doi.org/10.1016/0301-0511(95)05116-3)
- Prinzel LJ (2001) Empirical Analysis of EEG and ERPs for Psychophysiological Adaptive Task Allocation. <https://books.google.com/books?id=oYg9AQAAMAAJ>
- Razmara A, Aghamolaei T, Madani A, Hosseini Z, Zare S (2018) Prediction of taxi drivers' safe-driving behaviors based on the theory of planned behavior: the role of habit. *J Educ Health Promotion*. https://doi.org/10.4103/jehp.jehp_61_18
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
- Selcon, S. J., & Taylor, R. (1990). Evaluation of the Situational Awareness Rating Technique (SART) as a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations* 8 p(SEE N 90-28972 23-53).
- Tao D, Zhang R, Qu X (2017) The role of personality traits and driving experience in self-reported risky driving behaviors and accident risk among Chinese drivers. *Accid Anal Prev* 99(Pt A):228–235. <https://doi.org/10.1016/j.aap.2016.12.009>

- Vogels J, Demberg V, Kray J (2018) The index of cognitive activity as a measure of cognitive processing load in dual task settings. *Front Psychol* 9:2276. <https://doi.org/10.3389/fpsyg.2018.02276>
- Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 54(6):1063
- Wiesenthal DL, Hennessy D, Gibson PM (2000) The Driving Vengeance Questionnaire (DVQ): the development of a scale to measure deviant drivers' attitudes. *Violence Vict* 15(2):115–136
- Yan F, Liu M, Ding C, Wang Y, Yan L (2019) Driving style recognition based on electroencephalography data from a simulated driving experiment [original research]. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2019.01254>
- Zeng H, Yang C, Dai G, Qin F, Zhang J, Kong W (2018) EEG classification of driver mental states by deep learning. *Cogn Neurodyn* 12(6):597–606. <https://doi.org/10.1007/s11571-018-9496-y>
- Zhang C, Patel M, Buthpitiya S, Lyons K, Harrison B, Abowd GD (2016) Driver classification based on driving behaviors. In: Proceedings of the 21st international conference on intelligent user interfaces, Sonoma, California, USA. <https://doi.org/10.1145/2856767.2856806>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.