Driver Profile and Driving Pattern Recognition for Road Safety Assessment: Main Challenges and Future Directions

DIMITRIOS I. TSELENTIS® **AND ELEONORA PAPADIMITRIOU**

Faculty of Technology, Policy, and Management, Safety and Security Science Section, Delft University of Technology, 2628 BX Delft, The Netherlands

CORRESPONDING AUTHOR: D. TSELENTIS (e-mail: d.tselentis@tudelft.nl)

ABSTRACT This study reviews the Artificial Intelligence and Machine Learning approaches developed thus far for driver profile and driving pattern recognition, representing a set of macroscopic and microscopic behaviors respectively, to enhance the understanding of human factors in road safety, and therefore reduce the number of crashes. It provides a definition of the two scientific fields in terms of safety, and identifies the most efficient approaches used regarding methodology, data collection and driving metrics. Results show that K-means and Neural Networks are the most commonly used methodologies for driver profile identification, and Dynamic Time Warping for driving pattern detection. Most studies discovered driver profiles related to aggressiveness, considering mainly speed and acceleration as driving metrics. Based on the gaps and challenges identified, this paper provides a new framework for combining microscopic and macroscopic driving behavior analysis, instead of examining them separately as is the state-of-the-art. Such combined results can potentially improve the development of traffic risk models, which could be exploited in applications that monitor drivers in real-time and provide feedback. These models will represent human behavior more accurately, which can eventually lead to the recognition of "optimal" human driving patterns that Automated Vehicles (AV) could 'mimic' to become safer.

INDEX TERMS Driver profiles, driving behavior, driving patterns, artificial intelligence, machine learning, naturalistic driving data.

I. INTRODUCTION

ACCORDING to WHO, road traffic injuries cause the premature death of over 80,000 people every year and therefore constitute a major public health problem in the WHO European Region [1]. Approximately 2.7 million people are seriously injured each year in road crashes. These cause a substantial economic loss to society: up to 3% of the gross domestic product of any given country. The main cause of road crashes is persistently attributed to human factors, with a percentage of 65%–95% [2]. It is therefore crucial to deeply understand those factors in order to suggest new effective approaches to shape safe driving behaviors.

The review of this article was arranged by Associate Editor Chongfeng Wei.

Driver behavior analytics contribute to this direction through the monitoring of driver behavior in real time and fine resolution. They have important applications in several business fields including insurance, autonomous vehicles and road network management. In the current era of naturalistic driving, Big Data availability and advances in modelling techniques, there are considerable opportunities for statistical, econometric, Machine Learning (ML) and Artificial Intelligence (AI) applications as a basis for driving behavior analysis [3], [4]. Considerable opportunities are also present in terms of the usage of new data [5] such as driver physiological indicators, variables of driving time and conditions, congestion, road surface and environment conditions, detailed weather and spatial information [6].

The recognition of existing driver profiles and driving patterns could be an approach that takes into account all

those factors and contributes to the understanding of driving behavior for the improvement of road safety. Driver profiles and driving patterns are obviously related to the way the driver interacts with the environment. However, it is not always possible or available to collect related data through experiments.

A driver profile is defined as a group of drivers having similar driving behaviors and characteristics, whereas driving pattern is a specific driving behavior that is repetitively occurring by one or more drivers. Since these are two different fields that reflect a macroscopic and a microscopic behavior respectively, they require different approaches in terms of methodology, data collection and usage and therefore, they need to be studied separately.

Previous literature [7], [8] showed the feasibility of addressing the problems of driver profile and driving patterns identification based on driving metrics collected from inertial sensors (speed, acceleration, braking, steering etc.) across time and space [5], [6], [9], [10]. It is indicated though in recent research [9] that the temporal evolution of driving behavior can be rapid. Therefore, in order to capture these shifts and understand their safety implications, drivers should be continuously monitored at high-resolution and their behavior should be often re-evaluated.

One example of a relatively new concept that captures the temporal evolution of driving behavior and is incorporated in driving behavior analysis studies is the driving "pulse", which is defined as the time period that a vehicle is in motion, bounded by two adjacent stops. Some studies have found that this is a much more promising microscopic level of analysis [11], [12] that seems to yield significant results in driving pattern recognition. However, the characteristics, insights and added value of different methods and analysis scales for driver profile and driving pattern recognition have not been systematically explored.

This study will review the different methodological approaches that exist for driver profile and driving pattern recognition for traffic safety analysis purposes. It will also propose a new framework integrating both scales of analysis, exploiting methodological attributes from the two fields, and gaining potential for deeper understanding of driving behavior. Studies on driver profile and driving pattern recognition will be reviewed separately in terms of definitions, methodologies and data used. The primary focus is to reveal the best practices, identify future directions for driver profile and driving pattern identification for safety assessment, and determine what each field could potentially "learn" from the other and how both fields can be optimally integrated.

It is clarified that this research focuses only on studies related to driver profiles and driving patterns and not generally on the broader concept of individual driving behavior. For more details on studies related to personalized driving behavior analysis, e.g., using reinforcement learning, readers could refer to [13], [14], [15]. Some of these studies have developed more advanced simulation and co-simulation platforms to analyze personalized behavior and support personalized driving research.

Moreover, it focuses on safety-related studies, and excludes studies on the relationship between driving patterns and, e.g., efficiency, sustainability. Although these are important aspects of driver profiles and driving patterns detection [16], [17], [18], [19], they will not be examined herein.

The rest of the paper is structured as follows: Section II describes the methodological approach followed for this literature review; Section III presents the results of the review; the synthesis of the results is performed in Section IV; the conclusions drawn from this analysis and the new framework proposed are revealed in Section V. This research is based on the Rhapsody H2020 research project [20].

II. METHOD

A. RESEARCH QUESTIONS

The research question addressed in this research are:

RQ1) What are the existing methodologies for the identification of driver profiles in terms of safety?

RQ2) What are the existing methodologies for driving pattern recognition in terms of safety and what is their focus?

RQ3) What are the Artificial Intelligence (AI) techniques applied in these field?

RQ4) Which are the data sources and driving metrics used in the analyses?

RQ5) What are the main issues and challenges research has encountered, which are the gaps?

RQ6) What are the future opportunities and synergies arising in these two fields?

B. KEYWORDS AND SOURCE SELECTION

The search terms used were "driver profile" and "driving pattern". The keyword "driving style" was also tested, and it was observed that this term is usually mentioned in studies focusing on driver recognition and not to the identification of repetitive driving patterns, which is the focus of this research. The keyword "driving behavior" returned an excessively high number of studies on all human factors related to driving. A combination of the words "safe" AND ("driver" OR "driving" OR "profile" OR "pattern") was finally selected.

The online databases from which scientific literature was selected included the Scopus, Science Direct, Google scholar and IEEE Xplore search engines.

C. SEARCH STRATEGY AND STUDY SELECTION CRITERIA

The key papers presented in this systematic literature review (SLR) are selected based on some defined search criteria and filters applied in sequence. These criteria included the paper language, which should be English, the relevance of the keywords to the research topic of this review, the publication date range considered for the reviewed articles and access to full papers.
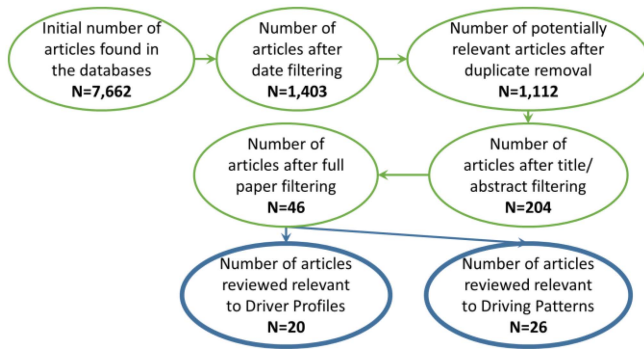
FIGURE 1. Paper search strategy and exclusion steps.

The selection steps are illustrated in Fig. 1. Initially, the combined keyword search lead to the identification of a total number of 7,662 studies.

Regarding the date range, the key relevant studies published within the past two decades were only taken into consideration – since, one of the pioneer studies that followed the "driving pulse" approach for the identification of driving patterns was [11], which was published in 2002. This review ended in late June 2022 and included research published before June 2022. The date criterion filtered out approximately 6,259 papers. A number of approximately 291 duplicate records were found and removed reaching thus the number of 1,112 studies.

At this point, it was noticed that there is ambiguity in the way the terms 'driver profiles" and "driving patterns" are used in literature. For instance, in some studies driver profiling is defined as the driving behavior observed in each recorded time step [10] whereas in others, driver profile is a group of drivers that have common behavioral characteristics [7]. ''Driving style" is used for driver recognition and classification purposes rather than for driving pattern recognition [21]. Other studies refer to "personalized driving behavior", but mostly focus on the identification of driver-specific patterns for the purpose of providing personalized feedback and/or adjustment of ADAS functionalities [22]. It is therefore highlighted that in this study:

• Driver profile is defined as a group of drivers having similar driving behavior and characteristics.

• A driving pattern can be defined as a driving behavior characteristic, such as a driving manoeuver like a harsh braking event, that is occurring repetitively either by the same driver or by different drivers in a population. Hence, in our study driving pattern is a more microscopic aspect than personalized driving behavior.

More specifically, driver profile detection investigates the existence of i) groups of people that behave similarly while driving, ii) common macroscopic characteristics among drivers, iii) methodologies to identify the "strongest" behavioral characteristics that separate these drivers into different groups. Driving pattern recognition investigates the existence of i) repetitive patterns in the data, ii) anomalous patterns in the data, iii) patterns that are uniquely

representative of the data, iv) methodologies to separate the data naturally into different regimes.

All the above were taken into consideration as criteria to filter out those studies that were not relevant to the primary focus of this research. This filtering was applied to the remaining 1,112 articles based on their title and/or abstract and determined whether an article was relevant to the research questions of this review. This narrowed down the pool of studies to 204. Subsequently, the exact filtering procedure was followed based on the full paper, which leads to the final number of 20 papers for driver profile identification and 26 papers for driving pattern recognition for which, the data collection and analysis methodologies will be reviewed.

III. RESULTS

This section presents the main features of the selected studies on driver profile and driving pattern recognition, in terms of the methodological approaches, the most preferred data sources and driving metrics used in the analyses. The ML and AI techniques applied so far in these fields and their main findings are presented in Table 1 and Table 4 and will also be discussed below. The data collection methodologies and driving metrics used for each driver profiling study can be found in Table 2 and Table 3 respectively, whereas Table 5 and Table 6 illustrate the data collection methodologies and driving metrics used in driving pattern recognition studies.

A. DRIVER PROFILE IDENTIFICATION STUDIES

Reference [7] proposed a methodological framework for the evaluation of driving safety efficiency based on Data Envelopment Analysis (DEA). This ML approach was tested on a sample of 56 drivers and resulted in the identification of three groups of drivers namely the non-efficient, weakly efficient and most efficient drivers. Results indicated that inefficient drivers present considerable differences in driving characteristics compared to the groups of weakly efficient and most efficient drivers with the difference of the two latter being less significant. It was found that the number of harsh braking events is an attribute is considered more significant for the characterization of a driver as aggressive or not. The percentage of speeding and the mobile phone usage were also identified as key factors for the estimation of the safety efficiency index of a driver. The temporal evolution of driving safety efficiency was studied separately in urban and rural road types by [9] with the aim to acquire insights useful for both driving behavior profiling. Driving safety efficiency was estimated for 200 drivers in consecutive rolling windows. Time-series characteristics were analyzed and several metrics were estimated and included in the profiling of drivers namely, the average driving safety efficiency, the volatility in driver's behavior and efficiency, and the stationarity and trend of the time-series. The k-means ML algorithm was employed to perform the clustering analysis and the optimal number of clusters was chosen using the elbow method. This

TABLE 1. Driver profile identification studies – AI and ML methodologies.

Paper	Methodology	Profiles discovered
[9]	K-means	Moderate drivers, unstable drivers and cautious drivers
[23]	RNN - Time Series Encoding	Classification of driving behavior as normal or aggressive on a route level
[7]	Data Envelopment Analysis	Three driver groups (non-efficient, weakly efficient and most efficient) based on driving safety efficiency
[10]	K-means	Six driving states - Driver profile is not stable
[25]	K-means	Five driver groups including aggressive, conservative, conservative making aggressive turns, low speed
[26]	Gaussian model	Calm, aggressive and drowsy behaviors
[27]	PCA and K-means is used	2-6 clusters depending on the combination of feature and signal. Profiles were not interpreted.
[28]	LSTM (RNN)	Three driving states - Normal/ Aggressive/ Drowsy
[5]	Random waypoint principle	Six driving styles - two driving styles depend on velocity - the rest depend on acceleration and deceleration
[29]	PCA - Hierarchical clustering	Five levels of aggressiveness, a tendency to drive with high speed in the city, and a more moderate speed driving
[30]	ANOVA	Driver performance profiles in response to the braking pace car
[6]	DBP scoring system	Mainly methodological contributions. Revealed profiles not discussed.
[31]	Neural Fuzzy system	Mainly methodological contributions. Revealed profiles not discussed.
[32]	PCA - Hierarchical clustering - Random Forest	3 personality types, i.e. the average, the positive and the aggressive mood
[33]	Spherical KMeans Clustering	Mainly methodological contributions. Revealed profiles not discussed.
[37]	Reinforcement learning	3 drivers profiles (easily stressed, normal, and resilient) with different sensitivity to vehicle's velocity and angular velocity
[34]	PCA – Random Forest – k-means	6 risky driver profiles based on the 14 most significant crash contributors
[35]	RNN	4 profiles - high-risk, medium-risk, low-risk and no risk drivers
[36]	Rule-based approach	2 profiles, non-aggressive and aggressive driving
[38]	Statistical analysis	Revealed profiles not discussed.

study concluded to three main driving groups, namely moderate drivers, unstable drivers and cautious drivers, which present considerable differences in terms of driving characteristics and performance especially between the cautious drivers and the rest. Both [9] and [7] used a similar data collection framework based on Smartphone sensors to collect naturalistic driving data. The driving metrics used in the analysis to measure safety efficiency were the distance travelled, the acceleration and braking per second as well as the harsh acceleration and braking events that occurred, the percentage driving over the speed limits and the percentage of mobile phone usage during each trip. Those metrics were captured using smartphone sensors that were recording at 1Hz frequency.

A methodology to classify driving behavior as normal or aggressive on a route level was developed by [23]. To this end, authors considered a hybrid AI classification methodological framework and employed both Recurrent Neural Networks (RNN)-guided time-series encoding and rule-guided event detection. It was shown that despite the fact that both long-short-term memory (LSTMs) and Gated recurrent units (GRUs) achieved a similar accuracy in the driving behavior classification task, GRUs were more efficient at the training stage. Their conclusions regarding driver profiling were that although all drivers drive 'normally' at the time-slice level, the time-slices characterized as 'aggressive' appear more frequently when aggressive drivers are driving. Data used for the analyses were derived from the UAH (University of Alcala) naturalistic driving dataset, which is a public collection of data captured by DriveSafe, which is a driving monitoring smartphone application recording driving behavior of various testers in different environments [24]. The data sample included 6 drivers and used recordings of 500 minutes in total. Aiming to characterize acceleration, braking and turning as medium, high, low using thresholds, the authors used the acceleration metric in their analysis.

With the goal to identify unsafe driving behaviors and provide relevant trip characterizations, a two level k-means ML clustering was applied by [10] using smartphone data. The aggressive driving characteristics within trips were identified by the first level clustering, whereas the second level of clustering for the same trips detected the additional unsafe behavior while driving, which is the inappropriate speeding and distraction. Overall, the authors concluded that there is no stable driving profile but instead, there are 6 driving states, namely safe behavior, aggressive behavior, risky behavior, distracted behavior, aggressive/risky behavior and aggressive/distracted behavior, among which drivers transition. More specifically, it was revealed that every driver shows risky behavior but with a different frequency. The database exploited included data from 129 unique drivers that drove a total of 10,212 trips within 10 months. The driving metrics used in their analysis were the number of harsh acceleration events occurred per km, the number of harsh braking events occurred per km, the acceleration smoothness indicator, the standard deviation of acceleration, the percentage of mobile usage and the percentage of speeding and the frequency of data collection was 1Hz.

Based on tens of thousands of driver logs from smartphone data, [25] developed supervised and unsupervised models based on ML and statistics to provide insights into different driver behavior and groups of drivers. Data from a naturalistic driving experiment recorded through smartphone sensors were used to understand traffic speed across the city of San Francisco and driving manoeuvers in different areas of the city. The authors also established a driver norm for each street and road segment, and abnormal driving behaviors

were flagged. Driver groups identification was based on the K-means clustering algorithm, which resulted to five clusters. Among others, the driver groups identified were those of the aggressive, conservative, conservative making aggressive turns and low speed drivers. The dataset recorded trips from 300 unique drivers who drove over 5000 rides. The sensors used were the GPS and magnetometer that recorded data at 1 Hz frequency, the accelerometer and gyroscope that recorded at 10 Hz, and the smartphone's camera.

In order to perform driving behavior analytics and identify differences between driver groups or states, [26] exploited smartphone data collected during a naturalistic driving experiment from a sample of 20 older drivers who participated for 1 to 2 weeks. To this end, a Gaussian model was developed based on the penalties obtained for the seven events (acceleration, braking, steering, weaving, drifting, overspeeding and car following) that were collected and taken into account. This study revealed 3 different behaviors, i.e., calm, aggressive and drowsy behaviors. Several driving metrics were used, including distance, hour, duration, speed, acceleration, braking, steering, weaving, drifting, overspeeding and car following, seatbelt wearing, hands on the wheels, smoking, objects manipulation and distractions. Inference of driver behaviors, among the three different classes of calm, drowsy and aggressive driving as well as the inference of a global score for each trip was also used.

A new methodology for near-real-time analysis and driver behavior ML-based classification was proposed by [27], who aimed at driver profile identification based on a combination of features and signals, including brake pedal pressure, gas pedal position, revolutions per minute (RPM), speed, steering wheel angle, steering wheel momentum, frontal acceleration and lateral acceleration. Data were collected from an uncontrolled experiment with 54 people, from over 2000 trips that took place within a 55 days duration. The research included four different steps, feature extraction, feature normalization, dimensionality reduction and an unsupervised learning approach for clustering. Several statistical features were created to support the analysis such as moving mean and median and the standard deviation, and dimensionality reduction was achieved using Principal Component Analysis (PCA). Clustering was performed based on the K-means algorithm and using seven different features from the metrics recorded by the data collection system, with a distributional approach. The results of this study indicated that an optimal number of clusters can be identified for each different combination of signal-feature, which ranges from two to six depending on the combination. Nonetheless, in this study the driver profiles were not interpreted and discussion focused on the methodological contributions of this paper.

A novel approach to improve driving behavior classification based on stacked LSTM Recurrent Neural Networks was proposed by [28]. In this study, the driving behavior classification problem was formulated as a time-series classification task based on AI, by exploiting data from a naturalistic driving experiment. The time-series data used in the classification

TABLE 2. Data collection methodologies for driver profile identification.

Paper	Experimental setup	Experiment characteristics	Data collection frequency
[9]	Naturalistic experiment - smartphone sensors	200 drivers - 7 months	1Hz
[23]	Naturalistic experiment - smartphone sensors	6 drivers - 500 minutes	-
[7]	Naturalistic experiment - smartphone sensors	56 drivers - 7 months	1Hz
[10]	Naturalistic experiment - smartphone sensors	10,212 trips - 129 drivers	1Hz
[25]	Naturalistic experiment - smartphone sensors	over 5,000 trips - 300 unique drivers	1 Hz - 10 Hz
[26]	Naturalistic experiment - smartphone sensors	20 Drivers - 1 to 2 weeks	-
[27]	Naturalistic experiment - vehicles with data loggers	54 people - over 2000 trips - 55 days	-
[28]	Naturalistic experiment - smartphone sensors	6 drivers - 500 minutes - 2 road types	-
[5]	Naturalistic driving database	several hundred vehicles - 12 months	-
[29]	Naturalistic experiment - Gipix	25 drivers	1Hz
[30]	Simulator experiment	40 drivers	30Hz
[6]	Naturalistic experiment - smartphone sensors	106 drivers - 10 week	1Hz
[31]	Simulator experiment	16 drivers	10Hz
[32]	Simulator experiment	28 participants	-
[33]	Naturalistic experiment	6 drivers	-
[37]	Autonomous simulator experiment	3 driver profiles simulated	-
[34]	Naturalistic experiment	204 drivers	0.1Hz
[35]	Naturalistic driving database	1,500 drivers	-
[36]	Naturalistic experiment	16 drivers	-
[38]	Questionnaire	100 participants	-

training were coming from nine internal sensors and were captured using a smartphone device. This research showed that it is possible to accurately classify driving behavior into three distinctive driving behavior classes, i.e., as normal, aggressive or drowsy driving, based on a window sequence of fused feature vectors of sensor data at any time step of a driving trip. The approach proposed achieved significantly improved results on the UAH DriveSet compared to previous studies, which featured much higher true positive rate and lower false positive rate. Data collection was based mainly on the GPS sensor and included metrics such as timestamp

TABLE 3. Driving metrics used for driver profile identification.

Paper	Timestamp	Coordinates	Distance	Duration	Acceleration	Braking	RPM	Pedal position	Speed	Distraction	Steering	Yaw/ Pitch/ Roll	Drifting	Other
[9]			√		√	√			√	√				
[23]					√									
[7]			√		√	√			√	√				
[10]					√				√	√				
[25]	√	√			√				√			√		Camera data
[26]	√		√	√	√	√			√	√	√		√	Car following, hands on the wheels, objects manipulation, seatbelt wearing, weaving
[27]					√		√	√	√		√			
[28]	√	√			√				√			√		Altitude, course and course variation, horizontal/ vertical GPS accuracy
[5]					√	√			√					
[29]	√	√							√					
[30]								√	√					Distance from other vehicles
[6]	√	√							√					
[31]					√	√			√		√		√	
[32]		√						√	√		√	√	√	Distance from other vehicles
[33]	√	√			√				√			√	√	Distance/time from other vehicles, altitude, car angle relative to lane curvature (degrees), road width, course and variation
[37]									√					Angular speed
[34]									√					Glances, secondary task activity, and hands-on/off wheel
[35]														respiratory rate, heart rate, eye closure rate, EEG factor
[36]					√	√			√					
[38]			√											Questionnaire data
Total	6	6	4	1	12	6	1	3	17	4	4	4	4	

(precise date and time of each recording), speed, coordinates, altitude, vertical accuracy, horizontal accuracy, course and course variation. Data from other sensors were also used, e.g., acceleration in X, Y and Z axis filtered by KF, roll, pitch and yaw. The driving sessions recorder took place in two road types, motorway and secondary road, by 6 different drivers and vehicles.

Using a database recorder from a large German car insurer with several hundred vehicles within a period of 12 months, [5] developed a stochastic driver profile identification model that takes into account speed, acceleration and deceleration. The modelling approach followed is based on the random waypoint (RWP) principle which, according to the authors, describes the movement behavior of a driver-vehicle unit in a given system area. The total number of driving styles that this study distinguished through the actual simulation is six. Two of the driving style groups depend on the speed process, which can be further distinguished to those styles with lower and those with higher speed values. Reference [29] modelled personal driving styles based on several driving parameters collected from various vehicle drivers through real-time experiments with the purpose of classifying drivers according to their risk-proneness. At first, PCA is performed on those driving data and resulted to 5 levels of aggressiveness based on the first principal component. Finally, hierarchical clustering was used and showed that there exist six clusters of drivers of which, two are the extreme clusters. Driving behavior metrics of 25 drivers were collected from a Gipix, which is a real-time vehicle tracking system that records information such as GPS coordinates, time and speed values with a frequency of 1Hz. An average of 9 tracks were collected under similar conditions per driver, over a period of 2-5 working days.

The effects of hands-free cell phone conversations on simulated driving was examined by [30]. Data from a driving simulator experiment with 40 subjects (20 older and 20 younger adults) were used to construct driver distraction profiles based on statistical analysis techniques. In order to provide an overall measure of driver performance as a function of experimental conditions, the authors used a multivariate analysis of variance (MANOVA). A univariate analyses was also performed on each of the dependent measures using a two by two split-plot analysis of variance (ANOVA), taking into account the age (younger versus older) and the task (single versus dual). Furthermore, driver performance profiles in response to the braking pace car, which was ahead of the participants, were examined to better understand how driving performance changes with age and cell phone use. Those profiles were created through the extraction of epochs of 10 seconds duration that were time locked to the onset of the brake lights of the pace car. The driving measures that were collected with 30Hz

frequency and stored for analysis were those of real-time driving performance, including driving speed, distance from other vehicles, and brake inputs.

One of the first studies that introduced the concept of driver behavior profiles (DBPs) was [6], which developed such an approach to enable driver behavior evaluation as a function of casualty crash risk and presented the results of an investigation into the factors associated with this risk. This approach was based on the estimation of common risk scores, which can potentially be used for driving performance comparison across time and space. This was done by applying multilevel models and temporal and spatial identifiers (TSIs) to control for the spatiotemporal environment. To this end, data from global positioning system (GPS) devices were recorded at a 1Hz rate from 106 drivers during a 10 week pay-as-you-drive (PAYD) study and supplemented with spatiotemporal information. The information collected included speed, coordinates, time and date driving. The authors concluded that DBPs can account for the complexity of the driving task. The revealed driver profiles are not discussed and the discussion focused mainly on the methodological contributions of this study.

A methodology to estimate car driver profiles in nonstationary contexts was developed by [31]. The model was based on an adaptive resource allocation neural fuzzy system and showed how the dynamic part of a driver's profile can be modelled as a multivariate time series prediction problem. Information regarding the driving behavioral variables such as the deviation of the car trajectory from the middle of the road, steering wheel angle, thrust and braking acceleration, speed, were collected from a driving simulator experiment that was conducted with 16 subjects under a variety of driving situations. The above variables were sampled with a sampling frequency of 10Hz, creating thus five time-series with samples kept every 10ms. This was achieved by developing AI techniques for emotion-related states induction, which affects performance on several tasks, including boredom and drowsiness, excitement, frustration and irritation, calm versus pressurized states, and those states that are affected by mental loads. It was shown that driving behavior varies significantly among drivers and that it is possible to identify these variations. Therefore, there are common driving characteristics among drivers and those can be represented by selecting the most representative profile from a pool of typical profiles. Revealed profiles were not discussed and the discussion of the paper focused on the methodological contributions.

Simulator data were used by [32] to look into driving style, mood states, and personality traits combined. 28 participants between 20-40 years old drove one urban and one highway scenario and went through a mood check using a questionnaire. PCA was applied for feature selection of the mood data and Hierarchical Clustering was used to cluster drivers with similar personalities. The analysis was based on vehicle coordinates, yaw, pitch, roll, speed, throttle and brake position, steering angle, distance from lane center and the distance from surrounding vehicles. Three personality types were discovered of which, type 1 personalities had the most average personality traits and mood states, and demonstrated more sedate driving. The second personality type had the most positive mood state whereas type 3 had the lowest positive mood states and drove more aggressively. A prediction model was trained based on random forest and validated, showing that (1) driving style can be predicted using mood states and personality traits and (2) personality types can be predicted using driving style and mood states.

Naturalistic driving data were also used by [33], [34], [35], [36] with almost all of them using speed as a driving metric to be analyze. Acceleration was also used by some of these studies. Physiological indicator data such as heart rate and eye movement were also collected by [34] and [35]. Studies [37] and [38] was differentiated by collecting data from an autonomous simulator experiment and an administered questionnaire, respectively. The number of participants ranges from less than 10 to 1,500 depending on the size of the experiment. Most studies discovered 2 to 6 profiles depending on the type of profiles created such as in terms of risk, speed, mood or aggressiveness.

B. DRIVING PATTERN RECOGNITION METHODOLOGIES

A methodological approach for the classification of driving behavior using Convolutional Neural Networks (CNNs) was presented by [39] and [21]. This AI-based approach was based on driving data obtained from an internal CAN-bus system from four different classes of driving types, i.e., a private car, a waste collection vehicle, a truck and a sweeper vehicle. Data for 27 different attributes were collected through on-board diagnostics (OBD) experiments and included information such as coordinates, speed, acceleration and its derivatives. The sample used comprised of more than 10.000 measurement vectors. They trained GRU, LSTM, 1D CNN and a 2D CNN models for driving class identification in order to classify the different vehicle types. The results of [39] revealed that the 2D CNN model outperformed the rest in terms of prediction accuracy, whereas in [21] the GRU model obtained the highest overall classification accuracy. The need to build a unique driver behavior fingerprint among different types of driving style was proposed as future research through the combination of driving style information in different situations and driving phases. Using a variation of the datasets used in the two previous studies, [40] attempted also to classify vehicle type among different driving types. The authors followed three popular classification approaches, i.e., k-nearest neighbors (k-NN), Support Vector Machines (SVM) and decision trees to train their models and test the results using the OBD experiment data. They concluded that based on the proposed features, the decision tree approach achieved the highest classification accuracy and it outperformed RNN-based approaches.

With the goal to measure the similarities among individual driving patterns, [41] proposed a classification model based on a combination of ML and AI techniques to recognize the

TABLE 4. Driving pattern recognition studies – AI and ML methodologies.

Paper	Methodology	Patterns discovered
[39]	GRU, LSTM, 1D CNN, 2D CNN	Different driving patterns of a private car, a waste collection vehicle, a truck and a sweeper vehicle
[21]	GRU, LSTM, 1D ConvNet	Different driving patterns of a private car, a waste collection vehicle, a truck and a sweeper vehicle
[41]	Dynamic time warping (DTW) and Hierarchical clustering analysis, Euclidean distance measures, Symbolic Aggregate approXimation (SAX)	Stable driving pattern varies with drivers and driving events
[42]	Dynamic Time Warping and Euclidean distance between time series	Distance measurement between distracted and normal driving
[43]	Supervised LSTM and combination of unsupervised LSTM autoencoder and supervised LSTM classifier	Driving manoeuver identification (Aggressive Acceleration/ Brake/ Right Lane Change/ Left Lane Change/ Left Turn/ Right Turn, Non Aggressive)
[44]	DTW, Hierarchical clustering, Gaussian mixture model, LeNet, ResNet, LSTM	Steering angle and Speed
[45]	A combination of Euclidean distance, DTW, Partitional Around Medoid clustering, Dynamic Barycenter Averaging method, the K − Shape method, the SAX method, hierarchical clustering with Ward's criterion	Speed time-series segmentation
[40]	k-NN, SVM, decision trees	Different driving patterns of a private car, a waste collection vehicle, a truck and a sweeper vehicle
[46]	Semi-supervised learning method, a danger-level estimation model	Driving danger level recognition
[47]	SAS, matrix profile	Driving patterns in manoeuvers such as turning, slowing, accelerating and parking regarding driver foot pedal behavior
[48]	HMM, Monte-Carlo methods	Driving manoeuver identification (Straight drive, Left turn, Right turn)
[48]	Support Vector Clustering	Driving style (Aggressive, Normal, Defensive)
[50]	LSTM, 1D CNN	Driving style (Aggressive, Mild, Gentle)
[12]	Dynamic Bayesian Network	Probabilistic behavioral model for normal and harsh acceleration, braking and cornering
[51]	Neural networks	Driving behavior normalization
[51]	Fully Convolutional Networks	Mainly methodological contributions. Revealed patterns not discussed.
[53]	Discrete Wavelet Transform, SVM, Dynamic Time Warping	Mainly methodological contributions. Revealed patterns not discussed.
[54]	Multivariate Dynamic Time Warping, k-nearest neighbour, kernel density estimation classifier	Mainly methodological contributions. Revealed patterns not discussed.
[55]	Dynamic Time Warping	Mainly methodological contributions. Revealed patterns not discussed.
[57]	Singular Value Decomposition, Q-measure, Hierarchical clustering	Mainly methodological contributions. Revealed patterns not discussed.
[11]	Fuzzy-logic-based approach	Mainly methodological contributions. Revealed patterns not discussed.
[58]	Deep NN, RNN and CNN are trained	Safe or Aggressive pattern recognition
[59]	Hierarchical, Kmeans, BIRCH, Gaussian mixture	Gear prediction
[60]	Neural networks	Detection of the driver, the driving style and the path
[61]	SimpleRNN, LSTM, and GRU	Detection of drivers' actions
[62]	LSTM-FCN	Drowsy pattern detection

individuals' stable driving patterns. In order to confirm that drivers have their own driving pattern, the authors applied a hierarchical clustering analysis that was performed using two approaches. The first one used the Dynamic time warping (DTW) method to measure the similarity between two original time series data, whereas the second one used the measure of the Euclidean distance based on a bag-of-pattern (BOP) representation. Consequently, a bi-directional LSTM layer was used and six driver classes were specified. In this research, the time-series data was expressed symbolically using a method called Symbolic Aggregate approximation (SAX), which is used for mapping a symbol or an alphabet to equal-sized segments of time-series data. It was shown that stable driving patterns vary with drivers and driving events, meaning that, e.g., driver A's driving pattern may be similar to that of driver B in a sudden-stop event but similar to driver C in the curve section. This methodology was applied on data collected from a driving simulator experiment with 6 participants who drove a test road of 6.7 km

that can be divided into five sections according to the events. Driving metrics were collected with a frequency of 20 Hz and included speed, acceleration, the depth of brake and acceleration pedal, the angle of the steering wheel and the distance from the road lane and the lead vehicle.

In order to detect important and potentially dangerous deviations from the norm in real-time, [42] trained ML models on the basis of driving simulator data that were able to understand normal driving behavior. The assumption that some "normal" driving behavior traces could potentially serve as a baseline to compare it with the actual observed behavior, enabled the definition of a distance measure between distracted and normal driving. The authors initially applied time series analysis techniques to assess the impact of cognitive distraction on drivers. Then, they defined a coarse- and a fine-grained distance measure between the time-series segments within a driving session, which are based on a combination of DTW and Euclidean distance between time-series. The driving metrics were collected

from an experiment with 16 participants and included speed, brake, accelerator, gear change, clutch, steering, lateral and longitudinal acceleration, and RPM. The simulator also captured other data such as the road and lane position as well as the path along the route. Finally, the authors also used basic physiology data by fitting drivers with a commercial bio-harness with heart rate monitoring capability, that is usually used in fitness applications.

A combination of a supervised learning model and a semi-supervised transfer learning model based on Artificial Intelligence was introduced by [43] for the classification of driving manoeuvers into aggressive acceleration, aggressive brake, aggressive right lane change, aggressive left lane change, aggressive left turn, aggressive right turn, and non-aggressive maneuver. This methodology was validated using smartphone's accelerometer and gyroscope data from a naturalistic driving experiment with 4 trips from 2 drivers, using domain specific knowledge data of the driving environment, such as data changing rules of various driving manoeuvers and temporal features. The front view of the subject vehicle was also recorded by a camera during the event. Class functions were used for the seven driving manoeuver types considered, which were converted into binary feature vectors. The models used were based on a supervised LSTM model and a combination of an unsupervised LSTM autoencoder and a supervised LSTM classifier. Results indicated that the supervised model performed better than the semi-supervised model. Nonetheless, it would be more beneficial to use a semi-supervised model in applications where the process of capturing labelled driving manoeuvers data is hard or insufficient.

With the aim to discover repetitive patterns related to the steering angle and speed during similar traffic situations, [44] used various machine leaning techniques to analyze driving data coming from a naturalistic driving experiment. The driving manoeuvers data were obtained from a data collection system in the car, which recorded speed, the accelerator and brake pedal position, steering angle in rad, driver power demand in kW as well as environmental factors such as speed limit and GPS data, as a set of time series with a collection frequency of 100 Hz. More specifically, an unsupervised learning and data mining techniques were first used to discover driving patterns and develop a labelling scheme. The discovery of driving patterns took place using Piecewise Aggregate Approximation, the SAX method and a classifier trained to recognize the current driving situation using the discovered patterns and labels from the previous steps. As for the data analysis techniques used, DTW is used to compare the similarity between time series and after computing the distance matrix, hierarchical clustering using Ward's minimum variance method and Gaussian mixture model clustering is implemented to achieve the grouping of the time-series. Finally, in order to achieve classification of time-series, four different AI network architectures, namely t-leNet, residual neural network (ResNet), LSTM and stacked LSTMs, and their performances on the specific problem are compared.

TABLE 5. Data collection methodologies for driving pattern recognition.

Paper	Data	Experiment	Frequency
[39]	Naturalistic experiments - OBD	14,288 samples - 4 driving types	
[21]	Naturalistic experiments - OBD	More than 10K measurement vectors from 4 driver classes	
[41]	Driving simulator	6 drivers - 6.7 km	20 Hz
[42]	Driving simulator	16 participants	
[43]	Naturalistic driving dataset	4 trips - 2 drivers	
[44]	Naturalistic experiment		100 Hz
[45]	Autonomous driving numerical simulations		
[40]	Naturalistic experiments - OBD	More than 10K measurement vectors from 4 driver classes	
[46]	Driving simulator	14 participants - 40 sessions - 20 min each	
[47]	Naturalistic experiment	4 weeks - 34 participants	10Hz
[48]	Driving simulator and naturalistic experiment	4 drivers	10Hz
[48]	Naturalistic experiment	3 drivers - 12 trips - total 1,106 km	Up to 77GHz
[50]	Naturalistic experiment using CAN Bus system		
[12]	Naturalistic experiment		
[51]	Naturalistic driving experiment	18 drivers	
[51]	Naturalistic driving data	105 hours of bus loggings - 9 cars - 16 days - 4 countries	
[53]	Naturalistic driving data from enhanced CAN bus		50Hz
[54]	Naturalistic datasets		
[55]	Naturalistic experiment		
[57]	Naturalistic experiment		
[11]	Naturalistic experiment	7 months - 6,000 trips	
[58]	Naturalistic experiment	30 drivers	
[59]	Naturalistic experiment		
[60]	Naturalistic experiment	16 drivers min – more than 1000 min driving	
[61]	Naturalistic dataset	89 drivers - 10,108 events	
[62]	Naturalistic experiment	18 drivers	

Reference [45] developed a three-step ML-based methodological framework that was applied on time-series data produced through autonomous driving numerical simulations. Those steps included the automatic segmentation of each time-series, the construction of the regime dictionary and the clustering of the produced categorical sequences.

TABLE 6. Driving metrics used for driving pattern detection.

Paper	Timestamp	Coordinates	Acceleration	Braking	RPM	Pedal position	Speed	Steering	Yaw/Pitch/Roll	Drifting	Surrounding vehicles distance	Other
[39]		√	√	√			√					
[21]		√	√	√			√					
[41]			√			√	√	√		√	√	
[42]			√	√	√	√	√	√		√		Driving path, physiological data, gear change
[43]			√						√			Camera data
[44]	√	√				√	√	√	√			
[45]			√								√	Driver behavior road & weather conditions, vehicle state
[40]			√	√			√					
[46]			√	√		√	√	√		√	√	
[47]	√	√	√	√	√	√	√		√	√		Duration
[48]			√	√		√	√	√	√			
[48]					√	√	√					Instantaneous fuel consumption
[50]			√	√	√	√	√	√				Lane change, wheel base, actual retarder % torque, % load at current speed, actual engine % torque, inlet air mass flow rate, fuel rate
[12]	√	√					√		√			Altitude, GPS signal strength data
[51]					√	√	√					Gear change
[51]							√		√			Perception from a front camera
[53]			√	√	√		√		√			
[54]												Comfort rates during lane changing
[55]	√	√							√			
[57]						√	√					
[11]							√					Distance
[58]			√	√			√	√			√	
[59]					√		√					
[60]			√	√					√			
[61]	√	√	√	√			√					
[62]												facial landmark sequences
Total	5	7	14	13	7	10	20	7	8	5	4	

Several methods were used including a polynomial regression mixture, clustering using Levenshtein distance in categorical sequence space as well as a combination of DTW with the Dynamic Barycenter Averaging (DBA) method, the K - Shape method and the SAX method. The hierarchical clustering procedure with Ward's criterion produces the final clusters based on this representation and associated distance. Five data categories were produced namely, environmental parameters including road characteristics, weather conditions and driver behavior, car physics such as weight distribution, engine capacities, sensor data, control law that triggers reacting to specific conditions such as when in close distance to nearby vehicles.

Reference [46] dealt with the interesting topic of driving danger level recognition using data collected from a driving simulator. They attempted to tackle this problem by developing a danger-level estimation model based on a semi-supervised ML method, which mined the safe and dangerous driving patterns considering time-series data with limited information on labelling, such as dangerous (time zones with incidents) and safe zones. Results were compared with other classification-based approaches including the hidden Markov model (HMM) and the conditional random field (CRF) algorithms that were trained for the same purpose of danger-level estimations. The driving metrics used in the analysis were dynamic parameters recorded from the vehicle including speed, acceleration, braking, steering, lateral lane position, throttle and braking pedal position as well as the minimum range between the driver and all vehicles in the driver's direction. A total of 14 participants participated in the experiment and each of them performed two to three sessions (40 conducted overall), which were 20 minutes long.

The purpose of [47] was to combine SAX and matrix profile methods to identify geo-spatial driving patterns, in terms of driver foot pedal behavior, focusing on those that typically involve accelerations such as turning, slowing, accelerating and parking. Data were collected from a 4-weeks naturalistic driving experiment with 34 drivers using OBD and GPS devices, which captured vehicle signals related to position, speed and acceleration, with a 10 Hz frequency as well as video recordings of the drivers' face, feet and front view. According to the authors, the matrix profile method runs as fast as, or faster than SAX in finding motifs, without sacrificing data resolution. Other advantages of matrix profile method compared to SAX and random projections is that it

is significantly shorter and simpler, and enables comparison between different datasets.

An imitative learning approach to intellectual cognition based on HMM and Monte-Carlo methods was proposed by [48] for driving manoeuver identification namely straight drive, left turn, and right turn. Using driving time-series data from an actual vehicle, and driver and environmental state from a driving simulator, they segmented the time-series into driving patterns that are assigned a proto-symbol. These data were classified into HMMs, the parameters of which were optimized using the classified driving patterns. The authors concluded that by adopting this approach, vehicles can grow their intelligence while observing expert's driving by storing driving patterns as symbols. Data collected from the actual vehicle were recorded from 4 participants with a sampling rate of 10Hz and included driving data of 9 kinds of metrics, roll, pitch, and yaw rates, vehicle accelerations in all 3 axis, vehicle speed, driver's accelerator stroke and steering angle. This framework using data from the driving simulator, which contained 4 kinds of time series measurements, i.e., speed, yaw rate, driver's accelerator stroke and the steering angle.

Reference [49] attempted to identify the differences between the variations of driving pattern and developed an objective driver classification using support vector clustering (SVC) and PCA. It was also confirmed by the results that a positive relationship exists between driving aggressiveness and fuel consumption and that driving style variations can be caused by weather condition, time of the day and driver's eagerness. The ML methodology was developed and validated using naturalistic driving data from 3 drivers that overall travelled 12 separate trips and a distance of 1106km. Information regarding the vehicle state was collected including speed, engine speed, pedal position and fuel consumption.

Reference [50] collected driving data from a naturalistic driving experiment using CAN Bus system to develop an AI-based classifier that identifies differences among aggressive, mild and gentle driving. This research employed LSTM and 1D CNN for this classification task with the aim to estimate the behavioral profile of the driver using 12 different driving parameters. It was found that despite the fact that both the LSTM and CNN based models performed with moderate success rates, the 1D CNN model performed more successfully. Driving metrics recorded and used were initial and final speed, engine speed, turn speed and wheel base in rad but CAN Bus messages also included wheel based vehicle and engine speed, brake and accelerator pedal position, actual retarder and engine percent torque, lateral and longitudinal acceleration, percent load at current speed, steering wheel angle, inlet air mass flow rate and fuel rate.

The power of Dynamic Bayesian Networks (DBN) was exploited by [12] to develop a driver behavior profiling model that recognizes acceleration, braking and cornering patterns, taking into account naturalistic driving data from GPS sensors such as timestamp, speed, altitude, direction and GPS signal strength data. This model had a probabilistic nature that was able to provide the probability of a behavior to be classified as normal or harsh event in terms of acceleration, braking and cornering. To achieve this, each trip was broken down into 230 time slices based on which the model was trained. It was shown that the nature of the driver's operational environment is determined by the model, which focuses on the road terrain. The suitability of the probabilistic methodologies to determine driving styles and operational environments for vehicle driver profiling was proved using the 2-Time-slice Bayesian Network (2TBN). No particular driving patterns were identified, but only the characteristics that influence the probability of an event to be harsh or normal.

A methodological framework, based on Neural Networks (NN), e.g., the radial basis function, for quantitative evaluation of driving styles was proposed by [51]. This approach established individual driver models and its results were validated using data from a naturalistic driving experiment with 18 participants. Driving metrics used included throttle position, brake pressure, vehicle speed, gear, engine speed, taking into account the various combinations of driving styles, road situations, and vehicle types. All participants were originally classified by technicians as highly, moderately or mildly aggressive. An aggressiveness index, which was based on energy spectral density analysis and the normalized driving behavior found in the previous step, was developed to achieve the goals set by this study. According to the authors, this index is very useful in applications where driving style may play an important role such as vehicle calibrations and intelligent transportation.

Reference [52] suggested that rule-based scenario detection of driving patterns should be complemented with a data-driven approach. To this end, they used rule-based detections as labels to train Fully Convolutional Networks (FCN) in a weakly supervised setup. The NN used were trained to detect the patterns of lane changes and cut-ins using 105 hours of naturalistic driving data in the form of bus loggings from 9 cars of 3 different models. The problem was formulated as a time series segmentation problem and the disagreement between the rule based method and the learned detection method was analyzed to find wrong or missing detections. The discussion of this study focused on its methodological contributions rather than on the revealed driving patterns. The conclusion was that the FCNs employed provided did not necessarily need large amounts of ground truth information to assess the quality of this rule based scenario detection, which showed their scalability. Driving metrics used were retrieved from the inertial measurement unit and included the ego velocity and yaw rate and data from a front camera to capture lane markings and other traffic participants.

A time series clustering approach was followed by [53] who exploited naturalistic driving data from enhanced CAN bus with a 50Hz collection frequency, for RPM driving pattern detection. The sensors used included the camera, radar and gyroscope. More than 33 variables were selected for the

analysis such as speed in x and y axis, rotational accelerations and real-time RPM. They used and compared several different algorithmic combinations based on ML, such as Discrete Wavelet Transform (DWT) with SVM, DTW with k-NN, the so-called KDK model, and LSTM. DTW was used to compare time series with multiple dimensions, i.e., multivariate time series. The DBA method was then exploited to find the representative average driving behavior and to estimate the similarities between driving behaviors. Finally, K-means was used to cluster these time-series into groups of similar behaviors. The results of this study indicated that the combination of DWT and SVM provided fast and precise results, while DTW with k-NN gave comprehended results and comparison of different driving behaviors as well as shorter calculation time but lower precision. The performance of the LSTM model was found to be between DWT and SVM and KDK. In this paper, the revealed driving patterns were not discussed.

Reference [54] aimed to provide an objective assessment of Advanced Driver Assistance Systems (ADAS) based on Multivariate DTW (MDTW), k-nearest neighbor classifier and the kernel density estimation. It provided a comparison between the subjective perception of different ADAS calibrations and the objectively measurable variables using datasets of comfort rates for different frequency and amplitude of an oscillation, comfort assessment during lane changing and comfort perception of three ACC scenarios. Again, the discussion highlighted the methodological contributions of this paper and did not focus on the revealed driving patterns. This study exploited the datasets of the Lane Keeping Assistant, Lane Change Assistant and the Adaptive Cruise Control, which include data with comfort rates for different frequency and amplitude of an oscillation, comfort assessment during lane changing as well as comfort perception of three ACC scenarios.

Reference [55] attempted to tackle the problem of real time identification of driving patterns, and more specifically of speeding and illegal overtaking patterns, using DTW. These two dangerous driving patterns were monitored and detected using vehicle motion data from gyroscope and GPS sensor, during a naturalistic driving experiment. DTW was applied on a moving window and its ability to match time-series sequences that are misaligned and stretched in the time domain [56] was exploited, which is very advantageous in this context of overtaking detection. The system developed was able to measure the similarity of the time-series coming from the live sensor data stream and a pre-recorded data sequence of an overtaking. DTW was proved to detect overtaking patterns regardless of how fast or slow vehicles were moving. Although this study contributed to the better understanding of the factors that influence the detection of this specific pattern, information such as the frequency of occurrence of this driving pattern was not revealed or discussed.

A methodological approach based on time-series segmentation and clustering was proposed by [57] to deal with the driving manoeuver classification problem. In order to

segment the multivariate time-series, they used Singular Value Decomposition of the segment matrices as a cost function in order to detect changes in the correlation structure among several variables. After segmenting time-series, the recognized segments and used the Q-measure to assess their homogeneity. Finally, recurring patterns of segment sequences are retrieved, capturing thus the evolution of multiple parameters over a time period. This study's methodology was validated using sensor data from a real-life experiment. The characteristics of each driving pattern discovered are not discussed in detail. The driving metrics of speed and accelerator angle were collected during a naturalistic driving experiment and data collection took place using a real-life sensor recording information from different vehicles during car drives.

Algorithms based on a fuzzy-logic-based technique as well as on driving cycles classification was developed by [11] for trip data recognition. Compositional summaries of vehicle usage were successfully generated and due to this, a systematic detailed analysis on vehicle performance was possible. The authors concluded that the time or distance normalization of driving pattern recognition profiles offers side-by-side comparison among trips, including the comparison between non-standard, randomly-generated-in-the-field driving cycles and standard ones. This could applied in this case study where previously analyzed results from standard driving cycles could be compared with field data for validation purposes. The characteristics of the revealed driving cycle patterns were not explicitly discussed. Speed and distance metrics used in this study were collected from electric vehicles under real-world driving conditions, in the form of time-series data per trip. Those data were recorded from more than 6,000 trips within a 7-months period.

Data from naturalistic driving experiments are also used from [58], [59], [60], [61], [62] to discover the existing driving patterns. The number of participants ranges from 16 to 89 in these studies and the mostly used driving metrics were speed, acceleration and braking, as in the rest of the studies, followed by RPM and yaw/pitch/roll. A differentiation is observed in [62] that uses facial landmark sequences for drowsy pattern detection. All these 5 studies made use of Neural network approaches, except [59] that used a clustering approach. Patterns discovered were related to driver's actions, aggressiveness, gear prediction and drowsiness.

IV. DISCUSSION
A. MAIN FINDINGS
Among the Artificial Intelligence methodologies that are used for driver profile identification, it seems that K-means is most commonly used, followed by NN-based models. The extended use of NN approaches also appears in recent studies [63]. Statistical and optimization methodologies are also utilized, whereas PCA is used in several studies to reduce dimensionality of the datasets used.

On the other hand, since driving pattern recognition involves mainly analyses of time-series data, the majority

of the studies reviewed herein make use of NN-based models. Several different NN approaches are used, mainly from the family of RNN, such as LSTM, GRU, standard RNN, CNN and fuzzy NN. Those are employed both for supervised (classification) and unsupervised tasks (clustering) depending on the type of the dataset. Moreover, DTW is a methodology that is very much preferred for pattern detection that usually complements other clustering methodologies such as Hierarchical and K-Medoid clustering. It was found that classification methodologies such as SVM, k-NN, and decision trees are exploited as a standalone approach for pattern identification more frequently than clustering methodologies, e.g., support vector clustering. Recent studies also confirm the efficiency of the combination of clustering and classification approaches [43]. Semi-supervised learning methodologies are also adopted by two studies. Other methodologies that were found useful for pattern detection were the HMM, DBN and the SAX method, which performs time-series segmentation by creating symbolic time-series. The common data analytics methodologies that was utilized across studies of both fields studied were the NN-based approaches. The main common characteristic of these studies is that they worked with time-series data, where the performance of RNN is proved to be significantly high.

Regarding the actual driver profiles discovered, it appears that the most commonly identified driver profile is that of aggressive drivers. This is reasonable since most studies are based on speed and acceleration driving metrics. One study is focusing solely on the identification of five different levels of aggressiveness. Several studies have also identified the group of "normal" drivers. Other researchers also discovered groups of drowsy drivers as well as calm, cautious and conservative drivers. Finally, there are also other characterizations of driver groups in literature, for instance in terms of their efficiency, or the consistency and stability of their temporal behavioral characteristics. There were also several studies that did not discuss the driver profiles discovered and focused mainly on the methodology used.

Regarding the results of the driving pattern recognition studies, again a common driving pattern discovered is the aggressive driving. This is usually discretized among other driving patterns such as normal, non-aggressive, defensive, stable, mild and gentle driving, or other manoeuvers such as normal acceleration and braking (events), turning, lane changing and parking. Other studies identify different patterns among several vehicle types [21], [39], [40]. Many of studies focused on the methodological contributions of their work and did not discuss the specific patterns discovered.

As aforementioned, speed and positive acceleration are the two driving metrics that were mostly collected and used for driver profiling, followed by negative acceleration (braking), timestamp, driver distraction that is usually measured through mobile phone usage or eye-tracking, and GPS coordinates. In driving pattern studies, pedal position and pressure is a metric that is strongly preferred followed by

braking, RPM, angular velocities and steering. The common use of speed and acceleration metric demonstrates the high importance that these two metrics play in the safety assessment of individual driving risk. On the other hand, there are important metrics such as videos, hands-on-wheel and distance from surrounding vehicles that are less frequently used until now mainly due to the fact that their recording requires technological advancements that have emerged just recently. They were found to play an important role in several studies and therefore, they are expected to become a strong preference in future research.

The vast majority of driver profiling studies exploited data from naturalistic driving experiments that were collected either from mobile phone sensors or from instrumented vehicles. The number of participants ranges from 6 to 300, whereas the duration ranges from a few minutes to 1 year. The data collection frequency is not provided in all papers but most papers used a 1Hz collection frequency. This reveals that 1Hz is an acceptable frequency that balances between the collection of noisy data and insufficient information and therefore, it provides an adequate amount of information for macroscopic analysis. Only two studies used data from driving simulator experiments and data were collected with higher frequencies (10Hz and 30Hz) compared to the naturalistic driving experiments.

Likewise, the majority of driving pattern recognition research exploited data from naturalistic driving experiments collected from OBD devices or smartphone sensors, whereas only 4 of them used data from driving simulators. This indicates that driving simulators are appropriate for personalized driving behavior analysis, but not so much for driver profiling and driving pattern recognition, which requires a very large data sample of drivers and trips. On the other hand, personalized driving behavior analysis typically examines fewer drivers for short time periods. The number of participants ranged from 4 to 34 and the data collection frequency from 10Hz to 100Hz. The collection frequency appears to be higher in driving pattern recognition study revealing thus that a higher granularity of information is required so that microscopic behavior can be captured and analyzed. A significantly lower frequency level is probably considered inadequate and may lead to the acquisition of insufficient information for microscopic analysis. One of the studies reviewed, analyzed data derived from numerical simulations conducted. Finally, it is highlighted that apart from driver profiling and driving pattern recognition studies, there are studies that make use of co-simulation platforms to support personalized driving research [15].

Regarding the data analysis and management approach, a conclusion drawn based on this review is that driver profiling studies should include the steps of data management, dimension reduction, feature importance and selection, profile clustering and assessment of results [34]. This suggestion can serve as a basis for future studies to build data strategies and methodologies on it.

B. MAIN CHALLENGES

The main research challenges identified during the conduction of this review are outlined below and discussed afterwards:

- The terms "driver profile" and "driving pattern" are used in an ambiguous way
- Absence of a robust methodology for the identification of driver profiles and the recognition of driving patterns
- Absence of a methodological approach combining both driver profiles and driving patterns
- The quality of the data collected through the naturalistic experiments

It was noticed in this review that the terms driver profile and especially driving pattern are used in an ambiguous way in the literature (see Section II-C), for different scopes and analysis purposes.

Another important research gap discovered is that there is no robust methodology to identify macroscopic driver safety profiles and microscopic driving safety patterns and understand their relationship with road risk. The methodologies developed so far have focused on grouping behaviors without identifying clear connections with safety. This would be extremely important as it would enable the provision of feedback to decrease driving risk even in real-time. It would also assist in predicting the future states of driving behavior that is entering into a new driver profile, having acquired knowledge on how the behavior of other drivers of the same profile was evolved in the past.

It was also highlighted that there is no study thus far that combines the two scientific objectives of driver profiles and driving patterns recognition. It is very important for an integrated methodology to be developed since in reality, microscopic and macroscopic aspects of driving behavior significantly influence each other. In other words, the identification of a driver profile should consider how the microscopic driving characteristics are evolving during each trip, and vice-versa, driving pattern recognition should consider the group of drivers that is investigated. For instance, a repetitive driving pattern such as an increased number of harsh acceleration events occurring only in specific parts of a trip or parts of a day (e.g., morning driving to work) could be a common characteristic among a group of drivers, forming thus a specific driver profile. On the other hand, when observing a specific driver profile, e.g., frequently distracted drivers, it is important to understand the microscopic characteristics of distraction, such as its duration, as well as whether other driving characteristics that increase risk co-exist such as over-speeding.

Another important challenge that should be faced is the quality of the data collected through the naturalistic experiments [64], [65]. This is because compared to a driving simulator, there are much more uncontrolled factors in naturalistic driving data such as the quality of the sensors recording behavior, the strength of the signal, the engagement in visually distracting secondary tasks and road traffic.

Finally, we highlighted that the identification of the more acceptable and safe behaviors is a major concern that is yet to be addressed in literature. The definition of optimal driving behavior, that could provide answers to this, is a methodological challenge by itself, which will be pursued in our further research. Initial insights on how it can be defined can be found in [66]. As also stated in [67], [68], it is important to keep improving the safety aspects of intelligent transportation systems as we move forward into the next generation of intelligent vehicles.

C. SUGGESTED FUTURE DIRECTIONS

The future directions suggested to tackle the challenges discussed earlier are outlined below and discussed afterwards:

- Standardization of the terminology for driver profiles and driving patterns
- Examination of the concept of "driving pulse"
- Application of a combined macroscopic and microscopic approach for driver profiling and pattern detection
- Focus on the collection of best-quality data that represent the efficient metrics identified by this review

The target of this review was not to analyze individualized driving behavior but rather to focus on micro and macroscopic methods and characteristics that could be used to identify common behaviors among drivers. It is therefore very important to explain the terminology and make clear what is the research objective in this case. From a road safety perspective, this is a research gap that was answered in this study and is summarized in the following paragraph.

Based on our research, driver profile can be defined as a group of drivers having similar driving behaviors and characteristics (e.g., aggressive/ non-aggressive, cautious/ distracted/ normal). A driving pattern on the other hand is a specific driving behavior that is *repetitively* occurring by one or different drivers, and this should be identified at a more disaggregate level, i.e., over very short time frames (within seconds) of driving. On a separate note, the term "personalized driving behavior" refers to the investigation and analysis of the behavior of an individual driver, e.g., for driver detection or for providing personalized feedback [22].

A significant contribution towards this direction is the concept of the driving pulse that was relatively recently introduced [11]. This is based on the concept that each trip should be segmented into shorter time-series in order to investigate the relationships among these segments and how they evolve over time either during the same trip or among trips of the same driver.

Fig. 2 illustrates the connection and interaction between the concepts of driver profiling and driving pattern recognition in a conceptual graph. This may serve as a methodological framework for future research, combining and integrating the macroscopic and microscopic approaches. A detailed description of this framework is provided below.
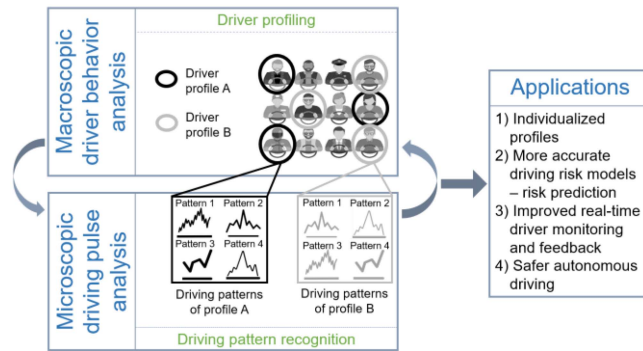
FIGURE 2. Conceptual graph illustrating the interaction between driver profiling and driving pattern recognition.

The first part of the methodological framework proposed includes the detection of the driver profiles through the analysis of the macroscopic driving behavior. This analysis will lead to groups of drivers with similar safety behavior and common driving characteristics. The macroscopic characteristics to be considered in this analysis should be those that are most commonly used for the evaluation of the safety performance that are found to affect crash risk such as speed limit violation, smartphone usage and harsh event performance. The outputs of this analysis will be driver profiles such as cautious or distracted drivers, and their connection to crash risk.

The second part of the methodological framework proposed includes the analysis of the microscopic characteristics of each driver profile identified. Those should be further investigated through a microscopic analysis of the driving pulses in each trip. This will lead to a further breakdown of the main profiles discovered using the macroscopic approach or to the recognition of sub-groups or sub-profiles that have very specific microscopic patterns in each driver profile. Example of patterns that may be detected in this analysis may be the investigation of the speed patterns before harsh maneuvers such as harsh braking, acceleration and cornering.

These are important aspects, since they will allow a deeper understanding of the driving characteristics of each group of drivers that influence risk and therefore safety. In this respect, the connections among several driving behavior safety indicators as well as their connection with driving risk should be further investigated. Driving risk here can be defined as the probability of causing a collision due to a high-risk driving action. The connection between driving behavior indicators and driving risk can be recognized through several means such as the accident history of the drivers or some indirect metrics that express risk of collision or very aggressive behavior. To summarize, it is recommended for future research to use the proposed integrated methodological framework, that combines the macroscopic and the microscopic driving behavior approach that should constantly be updated [34]. This interaction between the two scientific objectives of driver profiles and driving patterns will provide a more complete understanding of driving behavior.

The potential applications of the framework described above will enable a more detailed driver profiling that considers the complete picture of a driver's behavior on a micro and macro level as well as its evolution in time. This will help to avoid either looking at the "big picture" of macroscopic behavior and missing important information hidden in the microscopic driving features, or focusing only on the microscopic features without being able to draw conclusions on the overall driving safety performance of a driver.

Based on this, the driving risk identification and prediction will be certainly improved, as it will be possible to develop individualized driving risk models with higher accuracy. These models will allow the interpretation of driving safety behavior into probability of collision by focusing on the individual behavioral characteristics of each driver on a micro and macro level [69], [70]. These individualized driving risk models will be a more accurate representation of human behavior, which can be exploited in order to form an "optimal" driving model, by obtaining the safest driving characteristics under different driving and environmental conditions. The latter may potentially be used for the improvement of autonomous vehicles' (AVs) safety by mimicking human behavior to handle cases under which humans behave in an optimal way. Moreover, it could be used in the development of improved applications that monitor driving in real-time and provide personalized feedback to drivers to become less risky.

Finally, regarding the quality, efficiency and frequency of data metrics collected, it is suggested to focus on the indicators of speed, acceleration, distraction and pedal pressure are the most important driving metrics, which should be collected from naturalistic driving experiments with a frequency of at least 1Hz.

V. CONCLUSION

This research thoroughly reviewed the AI and ML approaches used thus far in driver profile and driving pattern recognition studies for traffic safety analysis purposes. The objective was to identify the best approaches in terms of methodology and data collection, and propose future directions to enhance the understanding of the macroscopic and microscopic aspects of driving behavior and therefore, road safety.

One of the main findings of this study, was the ambiguity in the definition of the two scientific fields. It also discovered the most efficient driving metrics that should be used in similar research and that data collection frequency should depend on the level of analysis. Moreover, it indicated that the levels of analysis that is used to identify groups of common behaviors, could be categorized as macroscopic, mesoscopic (e.g., [7], [9]) and microscopic depending on the level of information they use. The absence of a clear methodological framework for the identification of macroscopic driver profiles and microscopic driving patterns is suggested to be tackled through a methodology that combines macroscopic and microscopic driving metrics in order

to capture how these two aspects interact with each other as well as using the different approach of the "driving pulse" for microscopic research.

Regarding the contributions and innovations of this study, it provided the definition of these two scientific fields in terms of safety and assisted in the understanding of the most efficient methodologies, metrics and data collection methodologies used in these two fields. Finally, it provided suggestions and ideas on how microscopic driving patterns should be investigated and provided a methodological framework that combines both the macroscopic and microscopic driver behavior analysis.

The most important limitation of this review is that several studies are found that deal with the driver profiling and driving pattern recognition problem only from a methodological perspective without describing and discussing in details the findings of their studies. Another limitation is that this review did not explicitly look at the interaction between the driver with the environment and that only a small number of studies were found to use related data such as the distance from the surrounding vehicles.

Regarding future research, since microscopic and macroscopic aspects of driving behavior are interconnected, future research should focus on the analysis of the two scientific fields of driver profiles and driving patterns combined to obtain more promising results. To this end, interesting emerging concepts, such as the concept of the driving pulse that was recently introduced, should also be widely adopted in microscopic analysis of driving pattern identification. Additionally, there were papers that focused on driver profiling from a different perspective other than safety, such as eco-driving. It might be useful though to study the methodological approaches of those studies in the future to test how these could be applied in the safety analysis of driver profile and driving pattern recognition. Finally, it is suggested to also include other good predictors of a crash injury severity outcome such as weather status, road surface conditions, on-site damage type, lighting conditions, young age, weekday, off-peak, and vehicle type are also good predictors of a crash injury severity outcome [71], [72].

REFERENCES

[1] *European Regional Status Report on Road Safety*, WHO Europe, Geneva, Switzerland, 2019.

[2] E. Papadimitriou, C. Schneider, J. A. Tello, W. Damen, M. L. Vrouenraets, and A. ten Broeke, "Transport safety and human factors in the era of automation: What can transport modes learn from each other?" *Accid. Anal. Prevent.*, vol. 114, Sep. 2020, Art. no. 105656.

[3] D. Bhatt et al., "CNN variants for computer vision: History, architecture, application, challenges and future scope," *Electronics*, vol. 10, no. 20, p. 2470, 2021.

[4] C. Patel et al., "DBGC: Dimension-based generic convolution block for object recognition," *Sensors*, vol. 22, no. 5, p. 1780, 2022.

[5] W. Weidner, F. W. G. Transchel, and R. Weidner, "Telematic driving profile classification in car insurance pricing," *Ann. Actuarial Sci.*, vol. 11, no. 2, pp. 213–236, 2017.

[6] A. B. Ellison, S. P. Greaves, and M. C. Bliemer, "Driver behaviour profiles for road safety analysis," *Accid. Anal. Prevent.*, vol. 76, pp. 118–132, Mar. 2015.

[7] D. I. Tselentis, E. I. Vlahogianni, and G. Yannis, "Driving safety efficiency benchmarking using smartphone data," *Transp. Res. C, Emerg. Technol.*, vol. 109, pp. 343–357, Dec. 2019.

[8] E. Papadimitriou, A. Argyropoulou, D. I. Tselentis, and G. Yannis, "Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving," *Safety Sci.*, vol. 119, pp. 91–97, Nov. 2019.

[9] D. I. Tselentis, E. I. Vlahogianni, and G. Yannis, "Temporal analysis of driving efficiency using smartphone data," *Accid. Anal. Prevent.*, vol. 154, May 2021, Art. no. 106081.

[10] E. G. Mantouka, E. N. Barmpounakis, and E. I. Vlahogianni, "Identifying driving safety profiles from smartphone data using unsupervised learning," *Safety Sci.*, vol. 119, pp. 84–90, Nov. 2019.

[11] B. Y. Liaw, K. P. Bethune, and C. S. Kim, "Time-series field trip data analysis using adaptive recognition approach. Analysis on driving patterns and vehicle usage for electric vehicles," in *Proc. 19th Electr. Veh. Symp. (EVS-19)*, 2002, pp. 19–23.

[12] J. I. Obuhuma, H. O. Okoyo, and S. O. McOyowo, "Driver behaviour profiling using dynamic Bayesian network," *Int. J. Modern Educ. Comput. Sci.*, vol. 10, no. 7, pp. 50–59, 2018.

[13] X. Qi, Y. Luo, G. Wu, K. Boriboonsomsin, and M. Barth, "Deep reinforcement learning enabled self-learning control for energy efficient driving," *Transp. Res. C, Emerg. Technol.*, vol. 99, pp. 67–81, Feb. 2019.

[14] Z. Wang et al., "Driver behavior modeling using game engine and real vehicle: A learning-based approach," *IEEE Trans. Intell. Veh.*, vol. 5, no. 4, pp. 738–749, Dec. 2020.

[15] X. Liao et al., "Game theory-based ramp merging for mixed traffic with unity-sumo co-simulation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 9, pp. 5746–5757, Sep. 2022.

[16] M. C. González, C. A. Hidalgo, and A. L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[17] M. Lin and W. J. Hsu, "Mining GPS data for mobility patterns: A survey," *Pervasive Mobile Comput.*, vol. 12, pp. 1–16, Jun. 2014.

[18] E. Azadfar, V. Sreeram, and D. Harries, "The investigation of the major factors influencing plug-in electric vehicle driving patterns and charging behaviour," *Renew. Sustain. Energy Rev.*, vol. 42, pp. 1065–1076, Feb. 2015.

[19] P. Ahmadi, S. H. Torabi, H. Afsaneh, Y. Sadegheih, H. Ganjehsarabi, and M. Ashjaee, "The effects of driving patterns and PEM fuel cell degradation on the lifecycle assessment of hydrogen fuel cell vehicles," *Int. J. Hydrogen Energy*, vol. 45, no. 5, pp. 3595–3608, 2020.

[20] *RHAPSODY—Recognition of HumAn PatternS of Optimal Driving for safetY of Conventional and Autonomous Vehicles of the European Union's Horizon 2020 Research and Innovation Programme Under the Marie Skłodowska-Curie Scheme (2021–2023)*, TUDelft, Delft, The Netherlands, 2021.

[21] D. Mantzekis, M. Savelonas, S. Karkanis, and E. Spyrou, "RNNs for classification of driving behaviour," in *Proc. 10th Int. Conf. Inf. Intell. Syst. Appl. (IISA)*, Jul. 2019, pp. 1–2.

[22] E. G. Mantouka and E. I. Vlahogianni, "Deep reinforcement learning for personalized driving recommendations to mitigate aggressiveness and riskiness: Modeling and impact assessment," *Transp. Res. C, Emerg. Technol.*, vol. 142, Sep. 2022, Art. no. 103770.

[23] M. Savelonas, D. Mantzekis, N. Labiris, A. Tsakiri, S. Karkanis, and E. Spyrou, "Hybrid time-series representation for the classification of driving behaviour," in *Proc. 15th Int. Workshop Semantic Soc. Media Adapt. Personalization (SMA)*, Oct. 2020, pp. 1–6.

[24] E. Romera, L. M. Bergasa, and R. Arroyo, "Need data for driver behaviour analysis? Presenting the public UAH-DriveSet," in *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, Nov. 2016, pp. 387–392.

[25] J. Warren, J. Lipkowitz, and V. Sokolov, "Clusters of driving behavior from observational smartphone data," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 3, pp. 171–180, Jun. 2019.

[26] L. M. Bergasa et al., "Naturalistic driving study for older drivers based on the DriveSafe app," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1574–1579.

[27] U. Fugiglando et al., "Driving behavior analysis through CAN bus data in an uncontrolled environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 737–748, Feb. 2019.

[28] K. Saleh, M. Hossny, and S. Nahavandi, "Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2019, pp. 1–6.

[29] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, "Driving style analysis using data mining techniques," *Int. J. Comput. Commun. Control*, vol. 5, no. 5, pp. 654–663, 2010.

[30] D. L. Strayer and F. A. Drew, "Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers," *Human Factors*, vol. 46, no. 4, pp. 640–649, 2004.

[31] T. Balomenos, M. Pertselakis, E. McMahon, and N. Tsapatsoulis, "On the estimation of car driver profiles," in *Proc. Int. Conf. Sci. Comput. Comput. Eng. (SCCE)*, 2004.

[32] X. Liao, S. Mehrotra, S. Ho, X. Gorospe, X. Wu, and T. Misu, "Driver profile modeling based on driving style, personality traits, and mood states," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2022, pp. 709–716.

[33] A. Ouardini, I. E. O. Ech-Chahedy, A. Bouhoute, I. Berrada, and M. E. Kamili, "Analyzing driving behavior: Towards dynamic driver profiling," in *Proc. Int. Conf. Ad Hoc Netw.*, Nov. 2020, pp. 179–190.

[34] R. P. Payyanadan and L. S. Angell, "A framework for building comprehensive driver profiles," *Information*, vol. 13, no. 2, p. 61, 2022.

[35] R. Nouh, M. Singh, and D. Singh, "SafeDrive: Hybrid recommendation system architecture for early safety predication using Internet of Vehicles," *Sensors*, vol. 21, no. 11, p. 3893, 2021.

[36] S. N. Abdulwahid, M. A. Mahmoud, N. Ibrahim, B. B. Zaidan, and H. A. Ameen, "Modeling motorcyclists' aggressive driving behavior using computational and statistical analysis of real-time driving data to improve road safety and reduce accidents," *Int. J. Environ. Res. Public Health*, vol. 19, no. 13, p. 7704, 2022.

[37] C. Chronis, C. Sardianos, I. Varlamis, D. Michail, and K. Tserpes, "A driving profile recommender system for autonomous driving using sensor data and reinforcement learning," in *Proc. 25th Pan-Hellenic Conf. Inform.*, Nov. 2021, pp. 33–38.

[38] N. Jamaluddin, S. A. S. M. Rahim, A. Roslan, N. Z. Haron, and A. Jamaluddin, "Driver profile and travel distance among e-hailing drivers: An exploratory study," *Int. J. Road Safety*, vol. 2, no. 1, pp. 24–29, 2021.

[39] E. Spyrou, I. Vernikos, M. Savelonas, and S. Karkanis, "An image-based approach for classification of driving behaviour using CNNs," in *Proc. Conf. Sustain. Urban Mobility*, Jun. 2020, pp. 263–271.

[40] M. Savelonas, S. Karkanis, and E. Spyrou, "Classification of driving behaviour using short-term and long-term summaries of sensor data," in *Proc. 5th South-East Europe Des. Autom. Comput. Eng. Comput. Netw. Soc. Media Conf. (SEEDA-CECNSM)*, Sep. 2020, pp. 1–4.

[41] M. S. Lee, J. Park, Y. J. Jang, W. Kim, and D. Yoon, "Individual stable driving pattern analysis for evaluating driver readiness at autonomous driving levels 2 and 3," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 315–319.

[42] M. Garcia-Constantino, P. Missier, P. Blythe, and A. W. Guo, "Measuring the impact of cognitive distractions on driving performance using time series analysis," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 2047–2053.

[43] S. Sarker, M. M. Haque, and M. A. A. Dewan, "Driving maneuver classification using domain specific knowledge and transfer learning," *IEEE Access*, vol. 9, pp. 86590–86606, 2021.

[44] K. B. Mendoza and W. S. Vermelin, "Discovering patterns in driving data," M.S. thesis, Dept. Math. Sci., Chalmers Univ. Technol., Gothenburg, Sweden, 2019.

[45] E. Goffinet, M. Lebbah, H. Azzag, and L. Giraldi, "Autonomous driving validation with model-based dictionary clustering," in *Proc. ECML/PKDD*, 2020, pp. 323–338.

[46] J. Wang, S. Zhu, and Y. Gong, "Driving safety monitoring using semisupervised learning on time series data," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 728–737, Sep. 2010.

[47] C. Schwarz, "Time series categorization of driving maneuvers using acceleration signals," in *Proc. 9th Int. Driving Symp. Human Factors Driver Assess. Training Veh. Des.*, Jun. 2017, pp. 270–276. [Online]. Available: https://doi.org/10.17077/drivingassessment.1646

[48] W. Takano, A. Matsushita, K. Iwao, and Y. Nakamura, "Recognition of human driving behaviors based on stochastic symbolization of time series signal," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 167–172.

[49] Y. Feng, S. Pickering, E. Chappell, P. Iravani, and C. Brace, "Driving style analysis by classifying real-world data with support vector clustering," in *Proc. 3rd IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Sep. 2018, pp. 264–268.

[50] A. Cura, H. Küçük, E. Ergen, and I. B. Öksüzoglu, "Driver profiling using long short term memory (LSTM) and convolutional neural network (CNN) methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6572–6582, Oct. 2021.

[51] B. Shi et al., "Evaluating driving styles by normalizing driving behavior based on personalized driver modeling," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 12, pp. 1502–1508, Dec. 2015.

[52] P. Elspas, Y. Klose, S. T. Isele, J. Bach, and E. Sax, "Time series segmentation for driving scenario detection with fully convolutional networks," in *Proc. VEHITS*, 2021, pp. 56–64.

[53] X. Zhu, L. Zhang, B. Zhou, S. Zhao, and Y. Zhai, "Research on driving behaviors based on machine learning algorithms," in *Proc. IEEE 4th Inf. Technol. Netw. Electron. Autom. Control Conf. (ITNEC)*, vol. 1, Jun. 2020, pp. 20–25.

[54] U. D. Moser, "Multivariate time series clustering and classification for objective assessment of automated driving functions," Ph.D. dissertation, Dept. Maschinenbau und Verfahrenstechnik, Universität Duisburg-Essen, Duisburg, Germany, 2020.

[55] A. Bussooa and A. Mungur, "Driving behaviour analysis using IoT," in *Information Systems Design and Intelligent Applications*. Singapore: Springer, 2019, pp. 233–243.

[56] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. KDD Workshop*, vol. 10, Jul. 1994, pp. 359–370.

[57] S. Spiegel, J. Gaebler, A. Lommatzsch, E. De Luca, and S. Albayrak, "Pattern recognition and classification for multivariate time series," in *Proc. 5th Int. Workshop Knowl. Discov. Sens. Data*, Aug. 2011, pp. 34–42.

[58] W. A. Al-Hussein, L. Y. Por, M. L. M. Kiah, and B. B. Zaidan, "Driver behavior profiling and recognition using deep-learning methods: In accordance with traffic regulations and experts guidelines," *Int. J. Environ. Res. Public Health*, vol. 19, p. 1470, Jan. 2022. [Online]. Available: https://doi.org/10.3390/ijerph19031470

[59] P. R. A. Chandra and V. A. Narayanan, "Gear prediction using unsupervised machine learning to aid driver profile detection and correction," in *Advances in Automation, Signal Processing, Instrumentation, and Control*. Singapore: Springer, 2021, pp. 1431–1440.

[60] F. Martinelli, F. Marulli, F. Mercaldo, and A. Santone, "Neural networks for driver behavior analysis," *Electronics*, vol. 10, no. 3, p. 342, 2021.

[61] A. A. Lima, S. C. Das, and M. Shahiduzzaman, "Driver behavior analysis based on numerical data using deep neural networks," in *Proc. Int. Conf. Data Sci. Appl.*, 2022, pp. 211–219.

[62] S. Zu, Y. Jin, D. Yang, and H. Xu, "DrowsyNet: Multivariate time series classification for embedded driver drowsiness detection," in *Proc. 8th Int. Conf. Control Autom. Robot. (ICCAR)*, Apr. 2022, pp. 442–451.

[63] S. Mukherjee, A. M. Wallace, and S. Wang, "Predicting vehicle behavior using automotive radar and recurrent neural networks," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 254–268, 2021.

[64] D. Shrestha, D. J. Lovell, and Y. Tripodis, "Hardware and software for collecting microscopic trajectory data on naturalistic driving behavior," *J. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 202–213, 2017.

[65] S. G. Klauer, M. Perez, and J. McClafferty, "Naturalistic driving studies and data coding and analysis techniques," in *Handbook of Traffic Psychology*. London, U.K.: Academic, 2011, pp. 73–85.

[66] E. Papadimitriou, A. P. Afghari, D. Tselentis, and P. van Gelder, "Road-safety-II: Opportunities and barriers for an enhanced road safety vision," *Accid. Anal. Prevent.*, vol. 174, Sep. 2022, Art. no. 106723.

[67] R. De Iaco, S. L. Smith, and K. Czarnecki, "Universally safe swerve maneuvers for autonomous driving," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 482–494, 2021.

[68] Y. Liu and J. H. L. Hansen, "A review of UTDrive studies: Learning driver behavior from naturalistic driving data," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 338–346, 2021.

[69] K. Yang, C. Al Haddad, G. Yannis, and C. Antoniou, "Classification and evaluation of driving behavior safety levels: A driving simulation study," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 111–125, 2022.

[70] D. I. Tselentis, G. Yannis, and E. I. Vlahogianni, "Innovative motor insurance schemes: A review of current practices and emerging challenges," *Accid. Anal. Prevent.*, vol. 98, pp. 139–148, Jan. 2017.

[71] A. Jamal et al., "Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study," *Int. J. Injury Control Safety Promotion*, vol. 28, no. 4, pp. 408–427, 2021.

[72] M. Ijaz, L. Lan, M. Zahid, and A. Jamal, "A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw," *Accid. Anal. Prevent.*, vol. 154, May 2021, Art. no. 106094.

ELEONORA PAPADIMITRIOU received the Civil/Transportation Engineering Diploma degree from the National Technical University of Athens (NTUA) in 2001, the M.Sc. degree in transport from the Ecole des Ponts ParisTech in 2003, and the Ph.D. degree in transport safety from NTUA in 2010. She is an Assistant Professor with the Safety and Security Science Section, Faculty of Technology, Policy, and Management, Delft University of Technology. She has more than 20 years of research experience mostly within international research projects of the European Commission and other international organizations (e.g., OECD/ITF, WHO, UNECE, and CEDR). She has published more than 200 highly cited papers. Her research interests lie in transport safety, risk assessment and modelling, user behaviour with emphasis on vulnerable road users, transport safety management, and ethical issues.

DIMITRIOS I. TSELENTIS received the Diploma degree in civil engineering and the Ph.D. degree from the National Technical University of Athens, Greece, 2012 and 2018, respectively. He is currently a Marie-Curie Postdoctoral Fellow with the Safety and Security Science Section, Faculty of Technology, Policy, and Management, Delft University of Technology. He has more than ten years of experience, working with research and industry projects related to data and analytics and has published several scientific papers in peer-reviewed journals and international conference proceedings. His primary research fields are transport safety, driving behavior, transportation data analysis, machine learning and artificial intelligence applications, statistics and econometrics, traffic simulation and forecasting, and energy efficient vehicles.