IEEE Open Journal of
Intelligent Transportation
Systems

Received 24 July 2021; revised 31 December 2021; accepted 2 February 2022. Date of publication 7 February 2022; date of current version 17 February 2022.

Digital Object Identifier 10.1109/OJITS.2022.3149474

Classification and Evaluation of Driving Behavior Safety Levels: A Driving Simulation Study

KUI YANG[1], CHRISTELLE AL HADDAD[1], GEORGE YANNIS[2], AND CONSTANTINOS ANTONIOU[1]

[1]Chair of Transportation Systems Engineering, TUM School of Engineering and Design, Technical University of Munich, 80333 Munich, Germany

[2]Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, 15773 Athens, Greece

CORRESPONDING AUTHOR: K. YANG (e-mail: kui.yang@tum.de)

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme i–DREAMS under Grant 814761. A part of this work has been published in 7th International IEEE Conference on Models and Technologies for Intelligent Transportation Systems, online, Jun. 16–17, 2021 [DOI: 10.1109/MT-ITS49943.2021.9529309].

ABSTRACT The road traffic safety situation is severe worldwide and exploring driving behavior is a research hotspot since it is the main factor causing road accidents. However, there are few studies investigating how to evaluate real-time traffic safety of driving behavior and the number of driving behavior safety levels has not yet been thoroughly explored. This paper aims to propose a framework of real-time driving behavior safety level classification and evaluation, which was validated by a case study of driving simulation experiments. The proposed methodology focuses on determining the optimal aggregation time interval, finding the optimal number of safety levels for driving behavior, classifying the safety levels, and evaluating the driving safety levels in real time. An improved cross-validation mean square error model based on driver behavior vectors was proposed to determine the optimal aggregation time interval, which was found to be 1s. Three clustering techniques were applied, i.e., k-means clustering, hierarchical clustering and model-based clustering. The optimal number of clusters was found to be three. Support vector machines, decision trees and naïve Bayes classifiers were then developed as classification models. The accuracy of the combination of k-means clustering and decision trees proved to be the best with three clusters.

INDEX TERMS Driving behavior safety levels, driving simulation, clustering, support vector machine, decision tree.

I. INTRODUCTION

EVEN though the road accident frequency and the number of persons killed in road traffic accidents have fallen considerably over the last 20 years in the European Union (EU), the traffic safety situation is still severe. In 2019, around 22800 persons were killed in road accidents in the EU. 44.2% of the fatalities were passenger car drivers or passengers and 20.2% of them were pedestrians. Therefore, road safety improvement is one of the most important goals introduced by European Union in its "Zero Vision" [2]. Specifically, the goal is to cut European road fatalities and serious injuries down to zero. The percentage of crashes involving driver error or impairment before the crash was higher as higher as 94% (±2.2%) [3], which is the main factor reducing road safety. Therefore, more and more researchers and traffic managers have started to explore and better understand drivers' behavior with the help of driving simulation experiments and naturalistic driving experiments. Driving simulator can develop and simulate different driving scenarios to measure the behavior of a car driver in a simulated environment for human factor research. This allows the researchers to collect easily the driving behavior data in a safe room and to analysis driving behavior characteristics.

Given the advanced technology improvement, collecting driving behavior data and analyzing the driving behavior safety levels in real time have been possible in a real moving vehicle. They have been a research hotspot so as

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

to provide warnings or visual interventions for the drivers and Advanced Driver Assistance Systems (ADAS). Besides, driving behavior is postulated to belong to one or more safety levels or zones, ranging from "normal" to "dangerous" driving. Therefore, it is essential to evaluate the traffic safety level of driving behavior in real time. However, there is few studies to investigate it and how many driving behavior safety levels has not been deeply explored.

The goal of this paper is to propose a real-time classification and evaluation framework of driving behavior safety levels, which was validated by a case study based on a driving simulation experiment. Three clustering algorithms are proposed including k-means clustering [4], hierarchical clustering [5] and a model-based clustering [6], and the optimal number of clusters for each is also identified. The obtained clusters can be well visualized using advanced machine learning algorithms such as T-distributed Stochastic Neighbor Embedding (T-SNE) [7]. Support vector machines (SVM), decision trees (DT) and naïve Bayes (NB) classifiers are then used to develop models for real-time safety level classifications for new observations.

The contents of the paper will be structured as follows. First, the relaed work will be presented. Then the overall methodology will be introduced, including the overall framework, and the formulation of the used clustering and classification (modeling) algorithms. Thirdly, the simulation design environment, and the data collection and variables of interest are presented. Afterwards, the results are described and analyzed. Finally, a conclusion is given, focusing on the main contributions but also limitations and future work needed.

II. LITERATURE REVIEW

A. DRIVING SIMULATION STUDIES

The development and study of driving simulators in the automotive domain started in the 1960 [8]. For instance, in 1963, a moving belt driving simulator was employed to explore the effect of alcohols on driving performance [9]. Until today, driving simulators have become more popular with the development of cheap computer technology since simulators have advantages over other methods. Driving conditions are easily controllable, and reproducible and various, such as in heavy traffic, at night, in various types of weather, or in dangerous circumstances (i.e., collision avoidance, obstacles on the road). Besides, it can further reduce the costs of experiments and data collection, and there is no risk for subjects, which allows exploring the effects of special factors, e.g., alcohols, drugs, sleep deprivation or distraction. Additionally, driving simulators can provide more variables and measures related to driving behavior. Therefore, a number of literature on driving simulations has been produced.

Driving simulators have been applied in many different applications [10].

(i) Investigation of the driving behavior, e.g., the factors that affect perceived and observed (as measured, based on driving simulation experiments) aggressive driving behavior [11], the safety of raised pavement markers (RPMs) in a freeway tunnel [12];
(ii) Analyzing how secondary tasks (or distraction) affect driving performance, e.g., when dialing numbers or texting on a mobile phone [13];
(iii) Evaluating the impact of drugs, fatigue, or sickness on the driver, e.g., investigating the temporal patterns of variations in driving fatigue and driving performance [14];
(iv) Optimizing the interior design of a vehicle, e.g., surrogate in-vehicle information systems and driver behavior [15];
(v) Testing the driving ability, such as reaction time and perception [16];
(vi) Developing and testing effect and acceptance of new ADAS, e.g., a dangerous-driving warning system [17].

In summary, even though there is still some limitation of using the driving simulator, it is quite useful for research.

B. EVALUATION OF DRIVING BEHAVIOR SAFETY LEVELS

For the design of a vehicle control algorithm that monitors and corrects longitudinal driving behavior [18], it is essential to predict driving risks, and the major challenge of the research is how to discover the safe/dangerous driving patterns or driving risk levels [17]. To the best of our knowledge, related studies are not many. They have not deeply explore how many driving behavior safety levels is optimal and cannot also meet the requirement of the real-time evaluations during the whole driving with different environments.

Wang *et al.* [17] proposed a semisupervised learning method to utilize both the labeled and the unlabeled data, as well as their interdependence to build a proper danger-level function. The results show that the proposed method requires less training time and achieved higher prediction accuracy. However, this paper simply uses two states, namely safe/dangerous-driving state, which is not quite convictive.

Wang *et al.* [18] used the K-means clustering algorithm to classify longitudinal driving behavior according to driving and driver characteristics. The results of the study show that four main determinants of longitudinal driving behavior can be distinguished by using measurable parameters. The limitation of this paper is that it only focuses on the longitudinal direction without synthetically considering the lateral direction.

In order to develop algorithms for estimating driver behavior at road intersections, Aoude *et al.* [19] introduced two classes of algorithms that can classify drivers as compliant or violating. The classification algorithms are based on 1) support vector machines and 2) hidden Markov models. The results show significant performance improvements with the new algorithms compared with three traditional methods. The limitation is that it only focuses on a special road infrastructure, i.e., the road intersections.

In order to identify the risk level of each driver according to three levels based on both subjective and objective parameters, Eboli *et al.* [20] proposed a percentage of external points at borderline of the safety domain based on the kinematic parameter (i.e., acceleration) to define three levels (low, medium, and high risk). The thresholds are the average of the first quartiles (5%, 8%) for aggressive and cautious drivers. However, it cannot evaluate the driving safety level in real time since it is used to evaluate the risk level of each driver during the whole driving.

Zheng *et al.* [21] build a near-crash database and applied k-means cluster analysis to classify the near-crash cases into different driving risk levels using braking process features, namely maximum deceleration, average deceleration and percentage reduction in the vehicle kinetic energy. The results of clustered driving risk levels are low-risk group, moderate-risk group, and high-risk group. The limitation is that it only used three variables to cluster the driving risk levels and it did not explore why three clustered groups are determined.

According to the risk constraints under free-flow, car-following and lane-changing conditions, the average traffic flow risk index representing six risk levels and the safety threshold of the corresponding risk indicators were determined by Yan *et al.* [22] to predict the driving risk indicators and determine different risk levels under continuous tunnel environment. The result show that driving behaviors significantly vary in different tunnel risk feature points. However, this paper only focuses on the fixed point in the tunnel.

In existing related literature, k-means clustering [4], [18], [21] is utilized to cluster the driving risk level group since it is an unsupervised issue. It is a highly popular unsupervised learning algorithm that solves the well-known clustering problem. After obtaining a sampling data with labels, the classification models such as support vector machine and hidden Markov models [19] are widely applied to classify the new observations. However, there is not a clear framework to conduct driving behavior safety level classifications and estimations based on the historical and real-time collection of surveillance driving behavior data. Therefore, this paper tries to make contributions in this topic based on a driving simulation study.

III. METHODOLOGY

A. OVERALL FRAMEWORK

After improving the overall local traffic state prediction framework presented by [23], this paper outlines the overall framework in Fig. 1. It includes the main methodological components along with the information flow. Generally, each observation may hold multiple attributes, such as speed, headway and lateral location.

The methodology includes training and application steps. During the training step, archived surveillance data are used to (i) determine the optimal time interval to aggregate the high-frequency surveillance data; (ii) find the optimal number of clusters, presenting the ideal number of driving behavior safety zones or levels; (iii) identify the various driving behavior safety levels through clustering the available observations; and (iv) estimate the transition processes between these regimes. Finally, the information is stored into a knowledge base and further supports the application of the framework. During the application step, the appropriate classification model was selected to evaluate the driving safety levels with the input of the real-time aggregated surveillance data. The driving behavior safety levels in this paper is defined as the dangerous levels of drivers' behavior with different probabilities of the accident occurrence, ranging from "normal" to "dangerous" driving. It should be noted that there are three obvious improvements compared to [23]. The first improvement is that this paper adds a step about determining aggregation time interval to deal with the high frequency issue of the data collection. The second improvement is that this paper adds a step about searching the optimal number of clusters, which is an important point. The third improvement is that this paper does not include the step about predicting speed based on the predicted traffic state, which is the final target in [23], since the purpose of the framework in this paper is to prediction / evaluate the driving behavior safety levels.

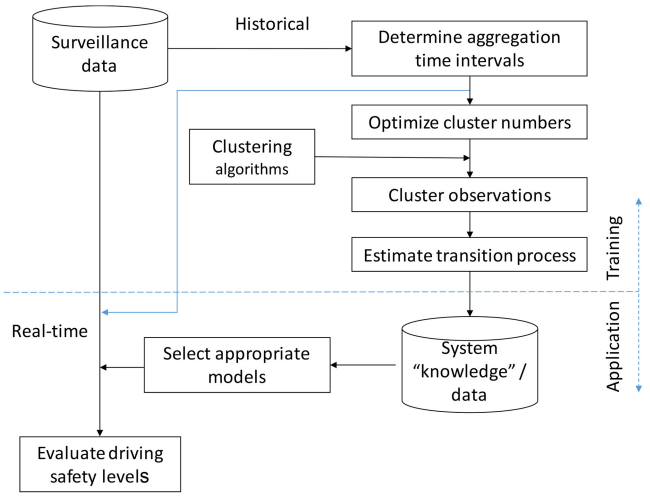

FIGURE 1. Overall framework of driving behavior safety level classifications and estimations.

In this study, three clustering algorithms including k-means clustering, hierarchical clustering and a model-based clustering, were employed to find the optimal number of clusters. These three algorithms were then used to cluster the available observations. Finally, support vector machines (SVM), decision trees (DT) and naïve Bayes (NB) classifiers were developed with the input of the labeled datasets based on three clustering algorithms to evaluate driving behavior safety levels and further to test the performance of developed models and clustering algorithms.

B. CLUSTERING ALGORITHMS

1) K-MEANS CLUSTERING

The standard k-means clustering is equivalent to known procedures for approximately maximizing the multivariate

normal classification likelihood when the covariance matrix is the same for each component and proportional to the identity matrix [23], [24]. One of the key parameters for the input of k-means clustering is to first specify the number of clusters. The average silhouette method is the most popular for determining it [25].

The basic idea in the k-means clustering consists of defining clusters so that the total intra-cluster variation (known as total within-cluster variation) is minimized. The total within-cluster variation is popularly defined as the sum of squared Euclidean distances between items [26], and can be formulated as follows.

$$Total.D = \sum_{k=1}^{k} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \qquad (1)$$

where x_i is a driving behavior data observation belonging to the cluster C_k, and μ_k is the mean value of the observations assigned to the cluster C_k.

The total within-cluster variation measures the compactness (i.e., goodness) of the clustering and it should be as small as possible. Each observation x_i is assigned to the closest cluster based on the Euclidean distance between the object and the centroid so that k-means clustering iteratively minimizes the total within-cluster variation (Eq. (1)).

2) HIERARCHICAL CLUSTERING

Hierarchical clustering can create a hierarchy of clusters, and presents the hierarchy in a dendrogram to cluster multidimensional data sets, by evaluating dissimilarities of objects in the variables space, or similarities of variables in the objects space [27]. Some studies, such as [5], [27], describe in detail the hierarchical clustering methods.

Hierarchical clustering adopts either an agglomerative technique, which is proceeded by a series of fusions of the n objects into groups, or a divisive technique, which separates n objects successively into finer groups, to build a hierarchy of clusters. Since agglomerative techniques are more commonly used [28], they are used in this paper. Agglomerative hierarchical clustering methods are characterized by: *the distance metric* and *the linkage method.*

The distance metric presents the similarity between each cluster. Euclidean distance whose equation is $d = \sum_{x_i \in C_k} (x_i - \mu_k)^2$ [28], is used in this paper. The linkage methods determine how to define the distance between two clusters. The common linkage methods include single linkage, complete linkage, and ward linkage. After comparing these methods and conducting initial analysis, we found the complete linkage could best fit our dataset. Therefore, it was used in our final analysis for hierarchical clustering. Complete linkage refers to the longest distance between two observations in each cluster, and its equation is $D_{12} = max_{ij}(X_i, Y_j)$ where X_i *and* Y_j are two observations. The distance between two clusters is the maximum distance between an observation in one cluster and an observation in the other cluster.

3) MODEL-BASED CLUSTERING

Unlike k-means clustering and hierarchical clustering, a model-based clustering assumes a data model and applies an expectation-maximization (EM) algorithm to find the most likely model components and the number of clusters. As a parametric method that uses the Gaussian distribution, the Gaussian mixture model (GMM) is a widely used model-based clustering [6], [29]. Each component probability distribution in GMM corresponds to a cluster. The problems of determining the number of clusters and of choosing an appropriate clustering method can be recast as statistical model choice problems and outliers are dealt with by adding one or more components representing a different distribution for outlying data [23].

GMM attempts to optimize the fit between the observed data and some mathematical model using a probabilistic approach. First, a specific-form mixture of Gaussians is assumed, and the density of the Gaussian mixture model [6] is:

$$f(x \mid \theta) = \sum_{m=1}^{M} \pi_m \varphi(x|\rho_m, \Sigma_m)$$

where $\varphi(x|\rho_m, \Sigma_m)$ is the density of a multivariate Gaussian random variable X with mean ρ_m and covariance matrix Σ_m, and $\theta = (\pi_1, \ldots, \pi_M, \rho_1, \ldots, \rho_M, \Sigma_1, \ldots, \Sigma_M)$.

Second, the parameters of this model are estimated with the use of the Expectation Maximization (EM) algorithm. EM starts off with a random or heuristic initialization and then iteratively uses two steps to resolve the circularity in computation: (i) E-Step, which determines the expected probability of assignment of data points to clusters with the use of current model parameters. (2) M-Step, which determines the optimum model parameters of each mixture by using the assignment probabilities as weights [30].

Note that Gaussian distribution, sometimes known as the normal distribution, has two parameters (i.e., the mean and the standard deviation), that we need to learn in order to fit this equation to our data.

C. EVALUATION MODELS FOR DRIVING BEHAVIOR SAFETY LEVELS

When the new observation comes in real time, it is difficult to use the clustering algorithm to identify the driving behavior safety levels since the clustering models are unsupervised machine learning models. Therefore, the evaluation models based on supervised machine learning models for driving behavior safety levels should be developed based on the historical dataset with the clustered labels. And then the developed evaluation models can evaluate the driving behavior safety levels for the new observation in real time. For this purpose, a support vector machine (SVM), a decision tree (DT) and a naïve Bayes classifier are used in this paper as well as the parameter fine-tune of developed models.

1) SUPPORT VECTOR MACHINE

Support vector machine (SVM) was originally designed based on statistical learning theory and structural risk minimization (e.g., [31], [32]); it will be used in this paper to classify driving behavior safety states. The SVM models [32], [33] were developed in RStudio®1.4.1717., using Package '*e1071*' [34].

Given training vectors $\boldsymbol{x}_i \in R^n$, $i = 1, \ldots, l$, in two classes and an indicator vector $\boldsymbol{y} \in R^l$ such that $y_i \in \{1, -1\}$, assuming that for the crash cases $y_i = 1$ and $y_i = -1$ for the non-crash cases. C-SVM [31] solves the following primal optimization problem.

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{b},\boldsymbol{\xi}} \quad & \frac{1}{2} w^T \boldsymbol{w} + C \sum_{i=1}^{l} \xi_i \\ s.t \quad & y_i\Big(\boldsymbol{w}^T \varnothing(\boldsymbol{x}_i) + b\Big) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, \ldots, l \end{aligned} \tag{2}$$

where $\varnothing(\boldsymbol{x}_i)$ maps $\boldsymbol{x}_i$ into a higher-dimensional space and $C > 0$ is the regularization parameter. Due to the possible high dimensionality of the vector variable $\boldsymbol{w}$, the following dual problem is solved [32].

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{Q} \boldsymbol{\alpha} - \boldsymbol{e}^T \boldsymbol{\alpha} \\ s.t \quad & \boldsymbol{y}^T \boldsymbol{\alpha} = 0, \\ & 0 \leq \alpha_i \leq \mathrm{C}, \quad i = 1, \ldots, l, \end{aligned} \tag{3}$$

where $\boldsymbol{e} = [1, \ldots, 1]^T$ is the vector of all ones, $\boldsymbol{Q}$ is an l by l positive semi definite matrix, $Q_{ij} \equiv y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and $K(\boldsymbol{x}_i, \boldsymbol{x}_j) \equiv \varnothing(\boldsymbol{x}_i)^T \varnothing(\boldsymbol{x}_j)$ is the kernel function. The optimal $\boldsymbol{w}$ satisfies $\boldsymbol{w} = \sum_{i=1}^{l} y_i \alpha_i \varnothing(\boldsymbol{x}_i)$ and the decision function is

$$\operatorname{sgn}\Big(\boldsymbol{w}^T \varnothing(\boldsymbol{x}) + b\Big) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}) + \mathrm{b}\right)$$

In this paper, the two kernel functions were considered since they provide two assumptions, i.e., non-linear relationship and linear relationship, respectively.

a) Radial Kernel: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma |\boldsymbol{x}_i - \boldsymbol{x}_j|^2)$
b) Linear Kernel: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T \boldsymbol{x}_j$.

2) DECISION TREE

A decision tree (DT) is a decision support algorithm that represents nodes of the tree and helps take decisions depending upon the inputs of a node. It uses a tree-like model of decisions and their possible consequences to perform both classifications and regressions. A decision tree consists of nodes (i.e., root, decision, and leaf), and branches. The nodes and branches are composed of each tree or sub-tree. Each root node and decision node represents features in a category to be classified and each leaf node contains a class label. Fig. 2 illustrates a structure of DT. The detailed knowledge of DT is introduced in previous papers, such as [35]. In this paper, the package 'rpart' [36] was used to develop the DT in RStudio®1.4.1717.

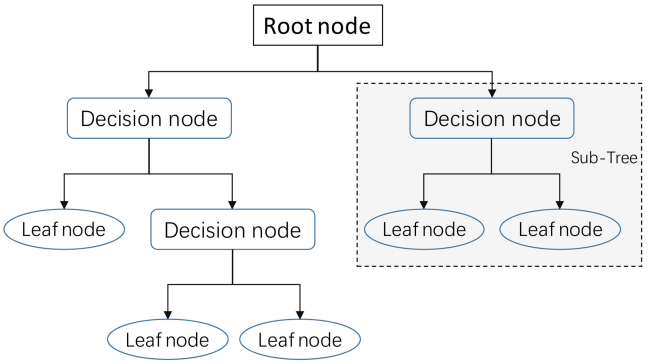

FIGURE 2. An example of the decision tree.

3) NAIVE BAYES CLASSIFIER

Naive Bayes classifiers [37] are a throng of classification algorithms based on Bayes' Theorem. Now, suppose there are n predictors, denoted by X_i, $i = 1, \ldots, n$. And the outcome variable is y, coming from one of the k classes, denoted by C_j, $j = 1, \ldots, k$. Therefore, $P(C_j|X_1, \ldots, X_i, \ldots, X_n)$ is the conditional probability of the observation coming from any of the k classes. According to Bayes formula, we can know

$$P\big(C_j|X_1, \ldots, X_i, \ldots, X_n\big) = \frac{P\big(X_1, \ldots, X_i, \ldots, X_n|C_j\big)}{P(X_1, \ldots, X_i, \ldots, X_n)} \tag{4}$$

We assumed that all the predictors $X_1, \ldots, X_i, \ldots, X_n$ are independent conditioned on class C_j. Therefore, we have

$$P\big(C_j|X_1, \ldots, X_i, \ldots, X_n\big) = \frac{P\big(C_j\big) \prod_{i=1}^{n} P\big(X_i|C_j\big)}{P(X_1, \ldots, X_i, \ldots, X_n)} \tag{5}$$

where $P(C_j)$ and $\prod_{i=1}^{n} P(X_i|C_j)$ are known as the prior and the likelihood respectively. The prior can be estimated as $P(C_j) = n_j / \sum n_j$. For the likelihood, the conditional probability distributions, $f(X_i|C_j)$ for $i = 1, \ldots, n$, is needed. When X_i is continuous, the normality is a common assumption. When X_i is discrete, a common assumption is the mulnomial distribution or non-parametric distribution.

In this paper, the package '*naivebayes*', [38] in RStudio®1.4.1717 is used to develop the naive Bayes classifier. And the kernel based density is used for the continuous predictors.

D. EXPERIMENTAL DESIGN

The driving simulator experiment was conducted in the Department of Transportation Planning and Engineering of the School of Civil Engineering of the National Technical University of Athens (NTUA), where the FOERST Driving Simulator FPF is located (see Fig. 3). The Foerst GmbH is a DIN ISO 9001-certified company and this specific simulator has been manufactured by the FOERST Company in order to serve research purposes. The driving simulator consists of 3 LCD wide screens 40"(fullHD), a total angle view of 170 degrees, a driving position, and a support base. The dimensions at full development are 230cm x 180cm, while the base width is 78cm.

FIGURE 3. FOERST Driving Simulator FPF.

FIGURE 4. Simulated rural area with the unexpected incident - donkey crossing lanes.

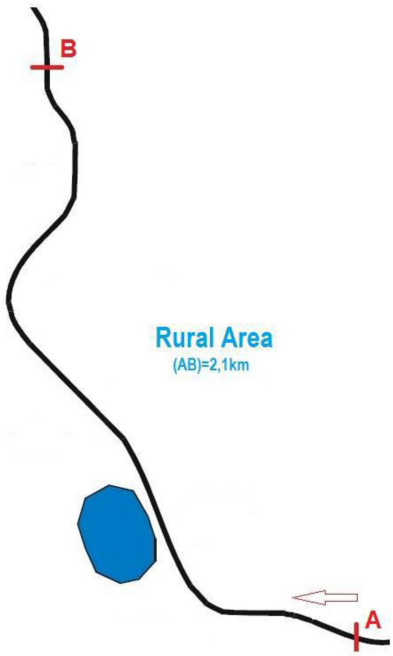

FIGURE 5. The simulated rural route.

The simulated road environment is an undivided two-lane rural road (see Fig. 4), which is a single carriageway with length 2,1 km, width 3m, with zero gradient, and mild horizontal curves (see Fig. 5). In the driving simulations, two traffic scenarios (i.e., moderate traffic conditions and high traffic conditions) and three distraction conditions (i.e., no distraction, cell-phone conversation, and conversation with passenger) were examined in a full factorial within-subject design. It should be noted that in moderate traffic conditions, ambient vehicles' arrivals are drawn from a Gamma distribution with mean = 12s, and variance = 6s, corresponding to an average traffic volume Q = 300 vehicles/hour. In high traffic conditions –ambient vehicles' arrivals are drawn from a Gamma distribution with mean = 6s, and variance = 3s, corresponding to an average traffic volume of Q = 600 vehicles/hour. The trials that demand conversation as a distractor are covered by the following topics: family, origin, accommodation, travelling, geography, interests, hobbies, everyday life, news, business.

During each trial of the experiment, two unexpected incidents that are the sudden appearance of an animal (deer or donkey) on the roadway were scheduled to occur at approximately fixed points along the drive. The driving simulator provides a "Free Driving" scenario that familiarizes the participants with the demands of an everyday drive. After a familiarization drive and a necessary short brake, each participant has only one chance to drive approximately 12.6 km within about 20min in total. A sample of 140 participants with a pathological condition were examined during approximately two years. A similar control group of another 120 participants with no known pathological conditions, of the same age groups was found to be sufficient. Finally, the sample of participants is a total of 260 individuals.

E. DATA AGGREGATION

The simulator records data at intervals of 33 to 50 milliseconds, including at first, 33 variables in each session. In order to explore driving behavior safety level classification and estimation, 36 variables were further aggregated and collected, during the whole driving scenario. The variables are listed in TABLE 1. It is noted that the crash could only happen at unexpected incidents. And an average of 0.65 crashes occurred for each driver during the driving simulation.

For the purpose of determining the optimal aggregation time interval, this paper proposed an improved cross-validation mean square error model based on driver behavior vectors according to [39], [40], which can estimate data fluctuations of driving behavior at different aggregation time intervals. The cross-validated mean square error (MSE) seeks to determine the minimal sufficient statistics necessary to capture the full information contained within a driving behavior parameter distribution. Only considering driving speed cannot characterize actual driving behavior correctly enough. Therefore, this proposed cross-validated mean square error (MSE) based on driver behavior vectors includes all the parameters (i.e., 35 variables in TABLE 1) related to driver characteristics and driver behavior. It is defined as

$$MSE_{ij} = \sum_{k=1}^{K} \left(x_{ij}^{k} - \overline{x_{ij}^{k}} \right)^2 \tag{6}$$

TABLE 1. Variables description and summary statistics for analysis.

No	Variables	Description	Mean	Std.dev	Min	Max
1	age	Driver's age	57.61	17.75	21.00	90.00
2	LateralPosition	Average distance to the right road board (m)	0.80	0.33	-3.88	7.33
3	StdevLateralPos	Standard deviation of distance to the right road board (m)	0.04	0.04	0	2.27
4	AverageSpeed	Average speed (km/h)	36.48	18.17	0	118.83
5	StdevSpeed	Standard deviation of speed (km/h)	0.86	1.30	0	41.93
6	AverageRspur	Average track of the vehicle from the middle of the road (m)	1.53	0.33	-5.48	6.00
7	StdRspur	Standard deviation of track of the vehicle from the middle of the road (m)	0.04	0.04	0	2.06
8	AverageRalpha	Average direction of the vehicle compared to the road direction in degrees	3.35	2.85	0	6.28
9	StdRalpha	Standard deviation of direction of the vehicle compared to the road direction in degrees	0.61	1.13	0	3.63
10	AverageBrake	Average brake pedal position (%)	3.93	17.41	0	100.00
11	StdBrake	Standard deviation of brake pedal position (%)	1.50	6.99	0	57.74
12	AverageGear	Average chosen gear (0 = idle, 6 = reverse)	2.64	1.20	0	5.00
13	StdGear	Standard deviation of chosen gear (0 = idle, 6 = reverse)	0.07	0.28	0	2.61
14	AverageRpm	Average motor revolutions in 1/min	2418.93	930.32	0	7057.17
15	StdRpm	Standard deviation of motor revolutions (1/min)	84.43	120.28	0	4073.64
16	AverageHWay	Average headway, distance to the ahead driving vehicle (m)	342.00	248.54	-0.70	1000.00
17	StdHWay	Standard deviation of headway, distance to the ahead driving vehicle (m)	1.93	18.81	0	501.48
18	AverageDleft	Average distance to the left road board (m)	0.71	0.33	-7.00	5.20
19	StdDleft	Standard deviation of distance to the left road board (m)	0.04	0.04	0	2.71
20	AverageWheel	Average steering wheel position in degrees	-2.67	17.21	-487.00	540.00
21	StdWheel	Standard deviation of steering wheel position in degrees	3.07	4.36	0	170.65
22	AverageThead	Average time to headway, i.e. to collision with the ahead driving vehicle (s)	60.84	124.22	-0.04	999.96
23	StdThead	Standard deviation of time to headway, i.e. to collision with the ahead driving vehicle (s)	7.44	46.22	0	507.99
24	AverageTTL	Average time to line crossing, time until the road border line is exceeded (s)	16.91	21.35	0.02	99.99
25	StdTTL	Standard deviation of time to line crossing, time until the road border line is exceeded (s)	47.80	95.00	0	534.48
26	AverageTTC	Average time to collision (all obstacles) (s)	35.41	30.26	0	98.83
27	StdTTC	Standard deviation of time to collision (s)	10.48	16.90	0	66.19
28	AverageClutch	Average clutch pedal position (%)	85.56	29.53	0	100.00
29	StdClutch	Standard deviation of clutch pedal position (%)	3.45	8.86	0	50.27
30	AverageAcc	Average gas pedal position in percent.	27.56	21.18	0	100.00
31	StdAcc	Standard deviation of gas pedal position (%)	8.80	8.09	0	50.85
32	AverageAccLat	Average acceleration lateral, in m/s2	0	0.01	-1.24	1.00
33	StdAccLat	Standard deviation of lateral acceleration (m/s2)	0.01	0.02	0	0.86
34	AverageAccLon	Average longitudinal acceleration (m/s2)	0.44	3.13	-13.99	14.00
35	StdAccLon	Standard deviation of longitudinal acceleration (m/s2)	1.28	2.29	0	15.12

where MSE_{ij} is the cross-validated mean square error of the of the i-th observation in the j-th aggregated group. x_{ij}^{k} are the parameters (i.e., 35 variables in TABLE 1), e.g., speed, accelerations and etc., of the i-th observation in the j-th aggregated group. And the $K = 35$ in Eq. (6). $\overline{x_{ij}^{k}}$ are the average values of the corresponding parameters, e.g., speed, accelerations and etc., in the j-th aggregated group without the i-th observation.

For the cross-validated mean square error in the j-th aggregated group, $MSE_j = \sum_i MSE_{ij} / \sum i$. And the cross-validated mean square error in the driving behavior data for each driver at the T aggregation time interval, $MSE^T = \sum_j MSE_j / \sum j$. It should be noted that the parameters should be standardized firstly. In this paper, the mean standardization is applied, which will not lose variation information of observations. It means that the observed value of each parameter is divided by the average value in the whole driving for each driver.

In the preliminary analysis, we chose randomly the data of five drivers. 21 aggregation time intervals (i.e., 1s, 2s, 3s, 4s, 5s, 6s, 7s, 8s, 9s, 10s, 20s, 30s, 40s, 50s, 60s, 70s, 80s, 90s, 100s, 110s, 120s) are applied. The results are shown in Fig. 6. It is found that the cross-validated mean square error of the dataset gradually increases with the growth of the aggregation time intervals, and that the improvement of cross-validated mean square error tends slowly to be gentle. In theory, the aggregation time interval corresponding to the minimum cross-validated mean square error is the optimal.

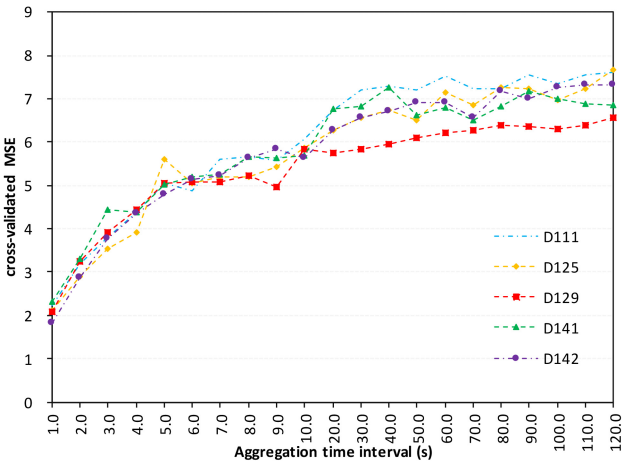

FIGURE 6. The cross-validated mean square error for different aggregation time intervals: D111, D125, D129, D141 and D142 are the driver ID.

Considering the application, 1s is determined as the aggregation time interval in this paper, like [41]. The summary statistics pf variables in the dataset are listed in TABLE 1. There are 193,453 observations after aggregations in the dataset. For each crash, the three observations prior to the moment when the driver starts to take measures (i.e., braking actions) to avoid crashes are considered as crash cases, which means that these three observations are labeled as crashes. The other observations are considered as non-crash cases.

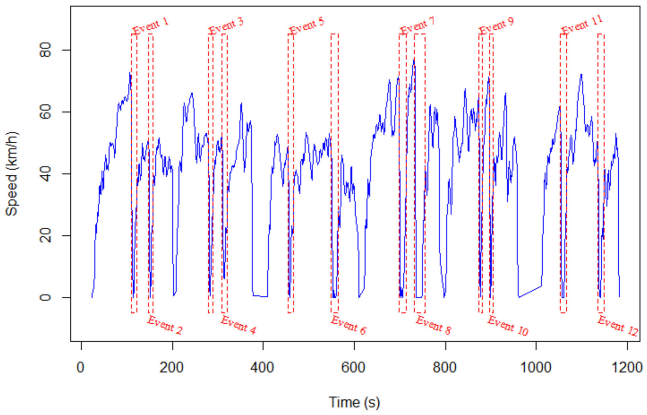

FIGURE 7. The speed of a driver (i.e., D111) during the whole driving.

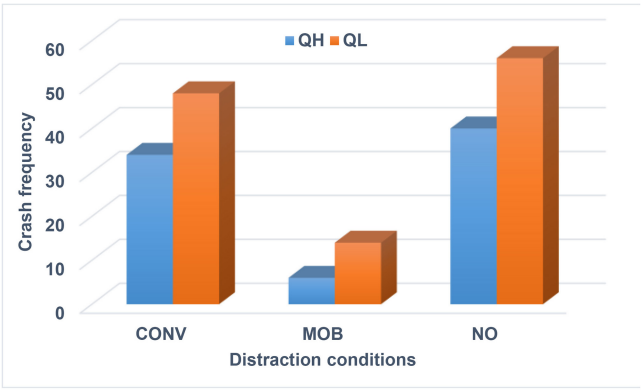

FIGURE 8. Crash frequency distribution at different conditions.

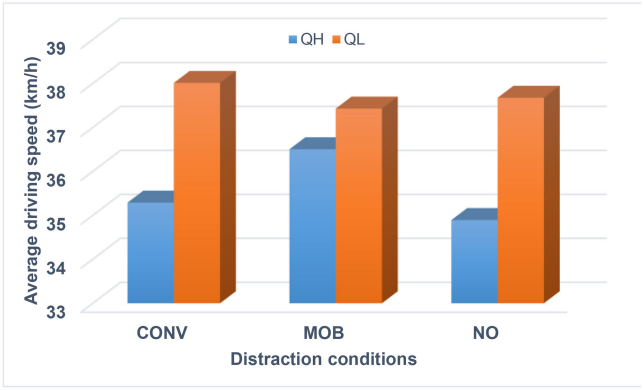

FIGURE 9. Average driving speed at different conditions.

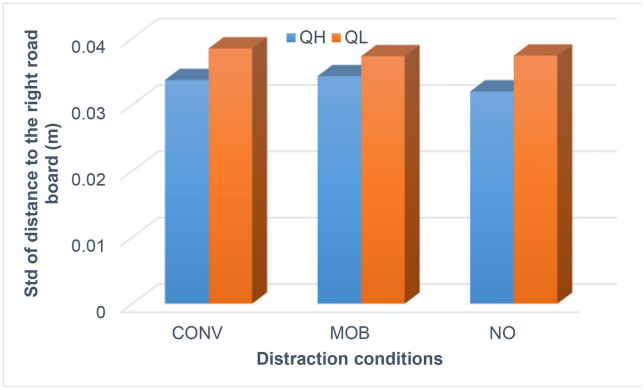

FIGURE 10. Average Std of distance to the right road board at different conditions.

It is noted that a non-crash sample were sampled randomly for the further analysis including searching the optimizing driving behavior safety levels, clustering and classification, and driving behavior safety level evaluations. Finally, there are 5423 observations in the sampling dataset, including 423 crashes and 5000 non-crashes.

Fig. 7 illustrates the speed of a driver (i.e., D111) during the whole driving as an example. During each event, the driver speed reduces due to the sudden appearance of the animal (deer or donkey).

IV. RESULTS AND ANALYSES

A. PRIMARY ANALYSIS

Fig. 8 shows the crash frequency distribution at different conditions. We can find that the crash frequency at high traffic conditions (QH) is lower than that at moderate traffic conditions (QL). The crash frequency at no distraction conditions (NO) is higher than that at the conversation with passenger conditions (CONV), and both of them are significantly higher than that at cell-phone conversation (MOB). These results are not consistent with our intuitive knowledge that the traffic safety at moderate traffic conditions and no distraction conditions are higher. However, here the crash frequency presents the traffic safety level of the driving behavior. Generally, it is easier for drivers to be slack and careless when the driving environment is good. Therefore, these results are reasonable since the traffic safety level of the driving behavior may be higher when the driving environment is bad (e.g., higher traffic flow, distractions). Additionally, under the distraction condition, negative impacts of the conversation with passenger conditions (CONV) on the traffic safety level of the driving behavior are significantly more than that of cell-phone conversation (MOB), which is consistent with our intuitive knowledge and existing studies.

Fig. 9 shows the average driving speed at different conditions. The average driving speed at high traffic conditions (QH) is lower than that at moderate traffic conditions (QL). The result of the paired T-test based on the control of drivers and distraction conditions is $t = -8.7989$ (p-value <0.0001), which means that the difference is significant and the mean of the differences is -2.55 km/h. The average driving speed at cell-phone conversation (MOB) and conversation with passenger conditions (CONV) are higher than that at no distraction conditions (NO), which is reasonable since the driver's ability to perceive speed decreases when distracted or drunk, which further will make drivers unconsciously increase the driving speed. The paired T-test shows that the speed difference between MOB and NO (i.e., -3.8km/h) is statistically significant ($t = -6.4668$, p-value <0.0001), and the speed difference between CONV

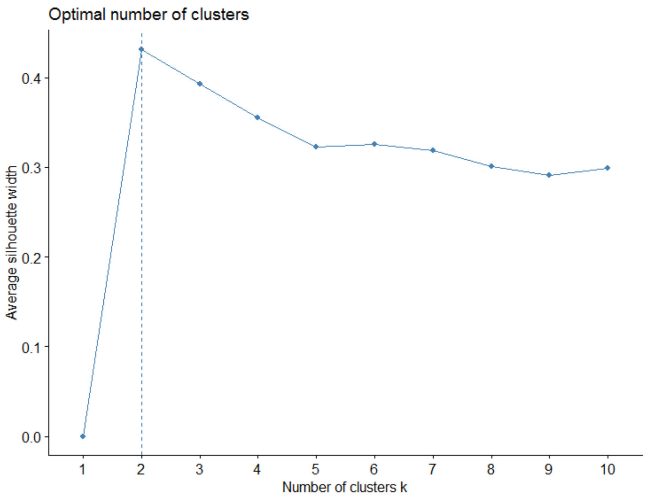

FIGURE 11. Optimal number of clusters based on k-means clustering.

and NO (i.e., −0.9km/h) is also statistically significant (t = −2.8909, p-value = 0.004).

Fig. 10 shows the average standard deviation of distance to the right road board at different conditions, which means the amplitude fluctuation of the vehicle swinging from side to side. When drivers are distracted, the amplitude fluctuation is slightly bigger than no distraction conditions (NO) since the average standard deviation of distance to the right road board at distraction conditions (CONV, MOB) are bigger than that at no distraction conditions (NO). However, the paired T-test results show that the difference is not quite statistically significant (CONV and NO: t = 1.5679, p-value = 0.1181; MOB and NO: t = −2.0172, p-value = 0.046). Additionally, when the traffic flow is higher, the amplitude fluctuation of the vehicle swinging is lower (paired T-test results: t = −12.274, p-value < 0.0001) since drivers are more careful.

B. OPTIMIZING DRIVING BEHAVIOR SAFETY LEVELS

K-means clustering, hierarchical clustering and a model-based approach were used to identify the optimal levels of driving behavior safety for the non-crash cases. Fig. 11 illustrates the optimal number of clusters based on the average silhouette method. The bigger the average silhouette method is, the better the number of clusters is. Therefore, the optimal number of driving behavior safety levels for non-crash cases is two. Totally, the optimal number is three levels including the crash level.

Fig. 12 shows the cluster dendrogram of hierarchical cluster analysis. The Ward's minimum variance method to perform agglomerative clustering. In the dendrogram, each leaf corresponds to one observation, and we can see the hierarchy of clusters. As we move up the tree, observations that are similar to each other are combined into branches. However, we can determine the number of clusters within the dendrogram and cut the dendrogram at a certain tree height to separate the data into different groups. The optimal number

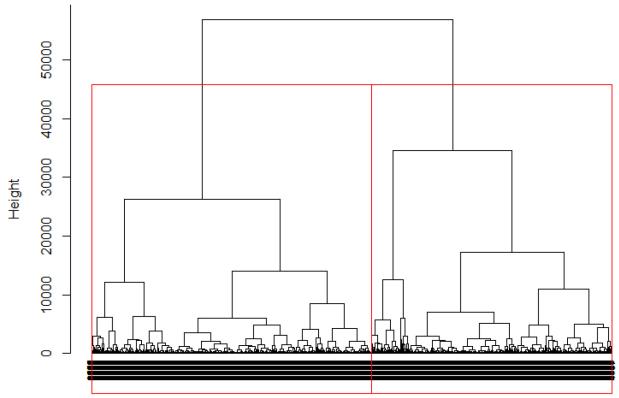

FIGURE 12. Cluster dendrogram of hierarchical cluster analysis.

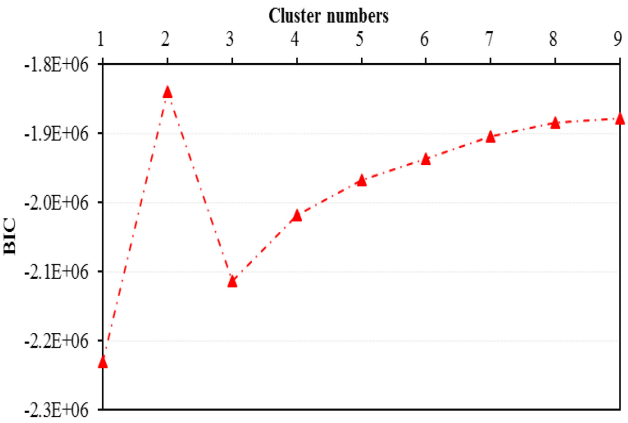

FIGURE 13. Optimal number of clusters based on GMM.

of state levels of driving behavior safety for non-crash cases is found to be two. The red rectangle borders show the three clusters in Fig. 12. Similarly, the optimal number is three levels including the crash level.

Bayesian Information Criterion (BIC) [23] is an important index to find the number of clusters by selecting the best clustering model and it uses the likelihood and a penalty term to guard against overfitting. The bigger the BIC is, the better the number of clusters is. Fig. 13 shows the optimal number of clusters based on GMM. Therefore, the optimal number of driving behavior safety levels for non-crash cases is also two based on GMM, where the BIC, which is −1840350, is biggest.

C. CLUSTERING AND CLASSIFICATION

After defining the optimal driving behavior safety levels, the non-crash observations in the dataset were further clustered and classified into different clusters. K-means clustering, hierarchical clustering and a model-based approach were used for this purpose, as presented in Fig. 14, Fig. 15, and Fig. 16. Three variables, namely average speed, average headway and average TTC, were selected as examples. It is interesting to note that the resulting sets of clusters based on these three clustering algorithms have similar geometries.

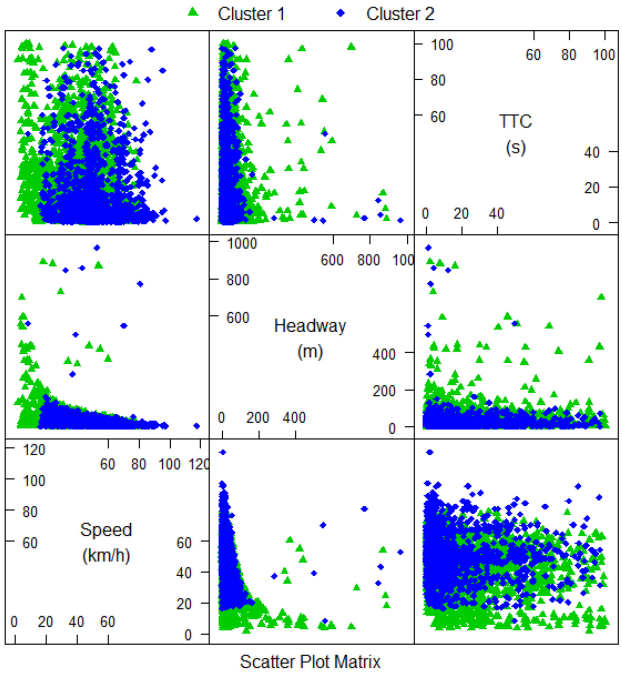

FIGURE 14. Different clustering scenarios based on k-means clustering.

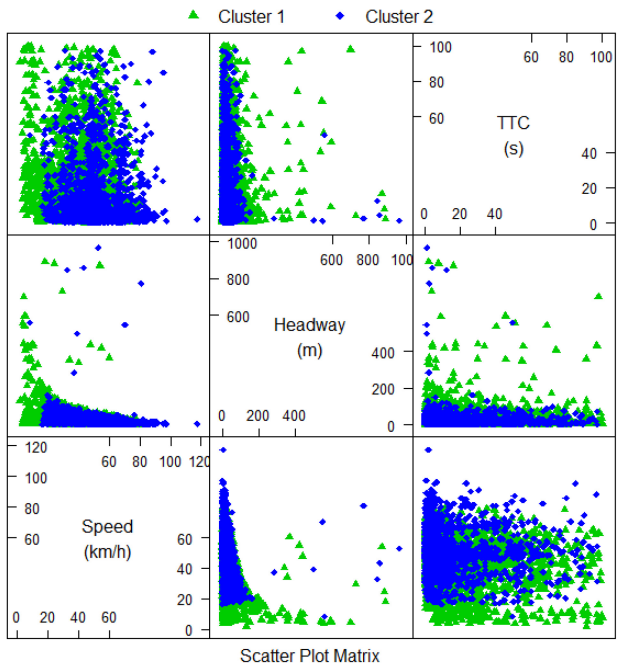

FIGURE 16. Different clustering scenarios based on GMM.

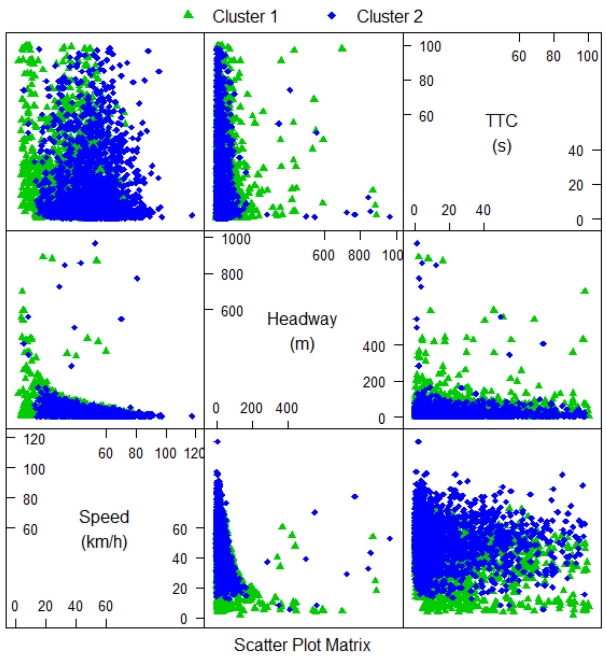

FIGURE 15. Different clustering scenarios based on hierarchical clustering.

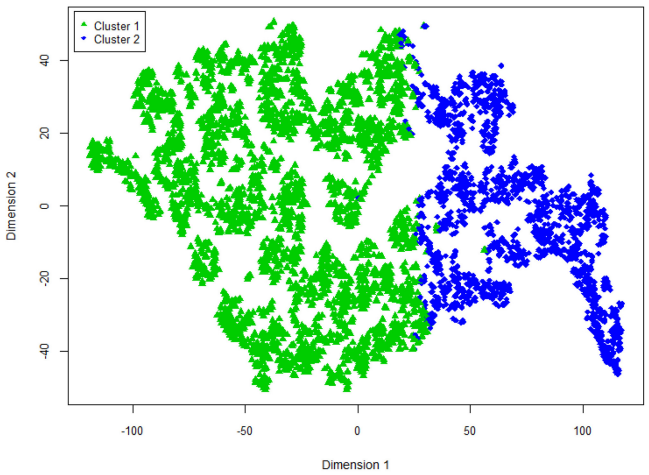

FIGURE 17. Visualization of clustering results based on k-means: elbow method.

T-SNE [7] was used to visualize the clustering algorithms. It is extremely useful for visualizing high-dimensional data [42], and it has a dimensionality reduction method to visualize data embedded in a lower number of dimensions, to see patterns and trends in the data. It can deal with more complex patterns of Gaussian clusters in multidimensional space compared to Principal Component Analysis. T-SNE results are shown in Fig. 17, Fig. 18 and Fig. 19. These visualizations show that driving behavior is well clustered into several levels.

In order to identify the best clustering algorithm, four widely used indices, i.e., the within clusters sum of squares, the average silhouette width, Dunn index and Calinski-Harabasz index, were used. The within clusters sum of squares is a measurement showing how closely related objects are in a cluster. The smaller the value, the more closely related objects are within the cluster. The average silhouette width is a measurement considering how closely related objects are within the cluster and how clusters are separated from each other. The silhouette value ranges from 0 to 1, and a value closer to 1 suggests that the data is better clustered [43]. The Dunn index [44] is an internal evaluation

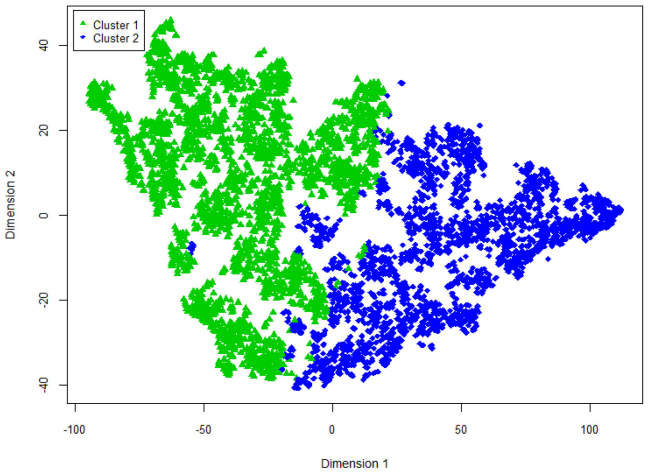

FIGURE 18. Visualization of clustering results based on hierarchical clustering.

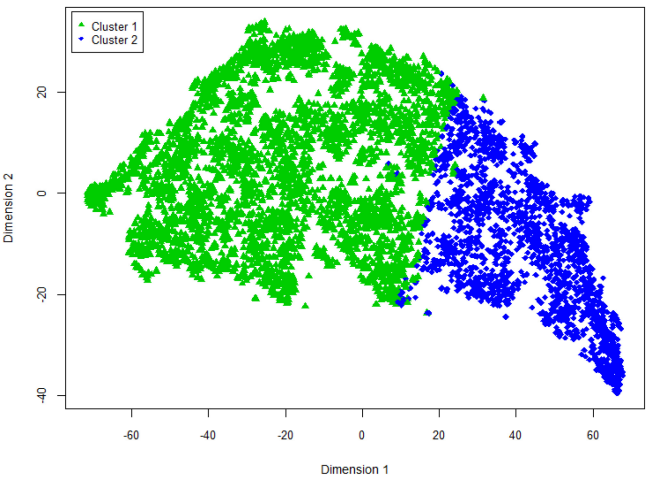

FIGURE 19. Visualization of clustering results based on GMM.

TABLE 2. Comparing three clustering algorithms.

Index	K-means	Hierarchical cluster	GMM
The within clusters sum of squares	1.73E+09	3.06E+09	1.73E+09
The average silhouette width	0.4486	0.6464092	0.4488
Calinski-Harabasz index	5099.560	719.981	5099.539
Dunn index	1.7222	3.148955	1.7228

scheme, where the result is the ratio of minimum separation and maximum diameter for all clusters based on the clustered data itself. The higher the Dunn index value is, the better the clustering is. The Calinski-Harabasz index [45] also known as the Variance Ratio Criterion, is the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters. The higher the Calinski-Harabasz index is, the better the clustering performance is. The results are listed in TABLE 2.

The indices show that the k-means algorithm is the best since its within clusters sum of squares is the smallest and its Calinski-Harabasz index is the biggest. The Hierarchical

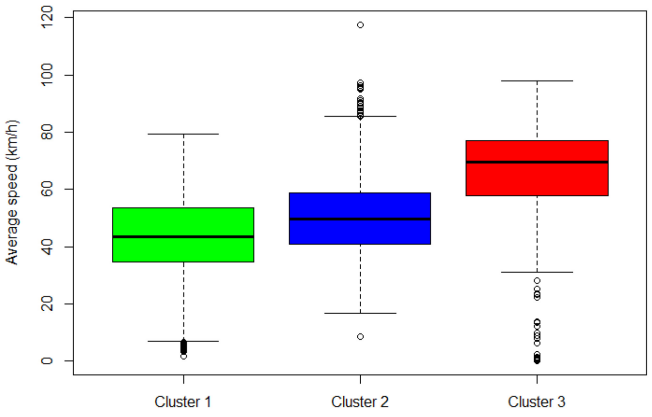

FIGURE 20. Boxplot of average speeds.

cluster is the best since its average silhouette width and Dunn index are the biggest among the three. Additionally, the difference of indictor values of k-means algorithm and GMM are not significant, respectivelly.

D. DRIVING BEHAVIOR SAFETY LEVEL EVALUATIONS

After clustering the driving behavior safety levels for non-crash observations, the crash observations are added into the dataset and labelled as "Cluster 3". Since most of crashes in the experiment are to collide the suddenly appeared animal (deer or donkey), driving at the higher speed is more difficult to stop before the animal to avoid accidents, which is more dangerous. Therefore, we use the driving speed of the clusters to identify their order of driving behavior safety levels. The higher the driving speed is, the more dangerous the safety level is. Fig. 20 shows the boxplot of the average speed of the three clusters and cluster 3 is the crash group. Therefore, we can label them as "Cluster 1", "Cluster 2" and "Cluster 3" to present "normal" driving, "low-risk" driving, and "high-risk" driving, respectively. It is reasonable since their frequencies are 3353, 1647, and 423, respectively. Importantly, classification methods were developed to evaluate the crash risk of driving behavior for the new observations and further identify the safety levels in real time. For this purpose, the widely used support vector machine (SVM), decision tree (DT) and naïve Bayes (NB) classifier were used.

The original dataset with clustered labels was divided randomly with the help of the stratified sampling technique into training data and test data, with 4338 observations (i.e., 80.0%) and 1085 observations (i.e., 20.0%), respectively. It means that about 80.0% of observations were used to train the SVM models, DT models and NB classifiers whereas the other 20.0% of observations were employed to test these models. Firstly, 80 SVM models, with different key parameters (the kernel function, the gamma and the cost), were developed to identify the best SVM model. Eight different gammas (i.e., 0.001, 0.01, 0.1, 0.5, 1, 2, 5, 10) and five different costs (i.e., 0.01, 0.01, 1, 10, and 100) were considered for each of two kernel functions (i.e., radial and linear).

TABLE 3. Results of SVM models.

Parameters	K-means clustering scenario	Hierarchical clustering scenario	GMM scenario
Kernel	Radial	Radial	Radial
Gamma	0.01	0.01	0.001
Cost	10	10	10
Number of Support Vectors	989	1118	1349
Best performance	0.2328	0.2141	0.2244
Total Accuracy	94.6%	93.6%	93.8%

TABLE 4. The evaluated levels of three models in K-means clustering scenario.

Model	Clustered	*Predicted*			
		1	*2*	*3*	*Accuracy*
SVM	1	659	11	1	98.2%
	2	8	304	2	96.8%
	3	12	17	71	71.0%
	Total				95.3%
Decision tree	1	671	0	0	100.0%
	2	0	314	0	100.0%
	3	4	0	96	96.0%
	Total				99.6%
Naïve Bayes	1	549	111	11	81.8%
	2	10	299	5	95.2%
	3	8	25	67	67.0%
	Total				84.3%

TABLE 5. The evaluated levels of three models in hierarchical clustering scenario.

Model	Clustered	Predicted			
		1	*2*	*3*	*Accuracy*
SVM	1	533	13	2	97.3%
	2	19	416	2	95.2%
	3	9	23	68	68.0%
	Total				93.7%
Decision tree	1	535	13	0	97.6%
	2	19	418	0	95.7%
	3	4	0	96	96.0%
	Total				96.7%
Naïve Bayes	1	378	162	8	69.0%
	2	7	423	7	96.8%
	3	4	33	63	63.0%
	Total				79.6%

Finally, the best model was identified for different clustering algorithms. TABLE 3. lists the results of SVM models. The total accuracy of the three best SVM models are quite high (i.e., $> 93.0\%$). It means that the developed SVM models can well identify the driving behavior safety levels in the data in the scenarios based on the three clustering methods.

The test data was further used to test the developed SVM models and decision trees. The results are listed in TABLE 4, TABLE 5, and TABLE 6. The total accuracy of SVM model in the k-means clustering scenario is $(659 + 304 + 71)/1085 = 95.3\%$, and the percentages of true predictions for each traffic safety levels are higher than 70.0%. Similarly, the total accuracy of SVM models in hierarchical clustering and GMM scenario are 93.7% and 95.2%, respectively. The total accuracy of decision trees in k-means clustering scenario, hierarchical clustering and GMM scenario are 99.6%, 96.7% and 99.6%, respectively. Besides, the total accuracy of naïve Bayes classifiers in k-means clustering scenario, hierarchical clustering and GMM scenario

TABLE 6. The evaluated levels of three models tree GMM scenario.

Model	Clustered	Predicted			
		1	*2*	*3*	*Accuracy*
SVM	1	656	10	5	97.8%
	2	4	309	1	98.4%
	3	10	22	68	68.0%
	Total				95.2%
Decision tree	1	671	0	0	100.0%
	2	0	314	0	100.0%
	3	4	0	96	96.0%
	Total				99.6%
Naïve Bayes	1	550	110	11	82.0%
	2	10	299	5	95.2%
	3	7	25	68	68.0%
	Total				84.5%

are 84.3%, 79.6% and 84.5%, respectively. Therefore, it can be found that the decision trees perform the best in these three clustering scenarios and the SVM models perform the second best. For each model, there are no significant differences between the accuracy from the training data and the test data in the three clustering algorithms. For instance, the accuracy from the training data and the test data are 94.6% and 95.3% (in k-means clustering scenario), 93.6% and 93.7% (in Hierarchical clustering scenario), 93.8% and 95.2% (in GMM scenario), respectively. This indicates that the developed SVM model, DT models and NB classifiers are reasonable and well developed. Besides, the safety levels of driving behaviors are all well identified. Importantly, the total accuracy in k-means clustering scenario is the highest among the three scenarios and it is slightly higher than that in GMM scenario. By ignoring the performance difference between the developed models, we can conclude that the k-means clustering can slightly improve the clustering performance of the safety level of driving behaviors. This can also reflect that the optimal safety level / cluster is three. Specially, the combination of k-means clustering and decision trees is the best.

E. DRIVING BEHAVIOR SAFETY LEVEL ANALYSIS

Based on the clustered results, we can further analysis the factors with the help of the boxplot. Fig. 20 shows the boxplot of the speed for each cluster. The speed of cluster 3 (i.e., "high-risk" driving) is the highest and the speeds of cluster 1 (i.e., "normal" driving) is the lowest. Therefore, it is important to drive at the reasonable speed in rural area since it is difficult to stop when sudden actions (e.g., animal, unexpected obstacles, walkers, cyclers) appear.

Fig. 21 and Fig. 22 show the boxplot of standard deviation of direction angle of the vehicle and the boxplot of standard deviation of steering wheel position, and they present the swing scope of vehicle head. These standard deviation values are not big since they are calculated based on the observations in 1s. We can find that the swing scopes of cluster 2 and cluster 3 are bigger than the cluster 1 (i.e., "normal" driving). It is reasonable since the greater swing scopes of vehicle head when moving is easy to cause sideslips, drifts,

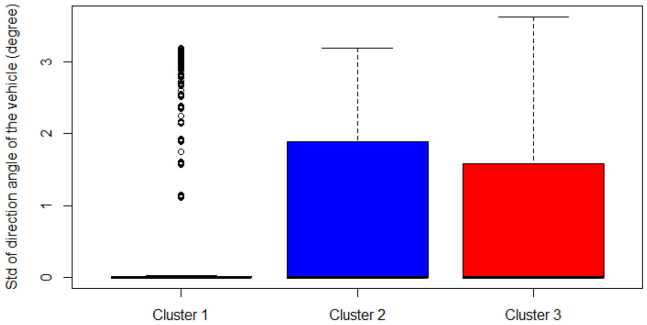

FIGURE 21. Boxplot of std of direction angle of the vehicle.

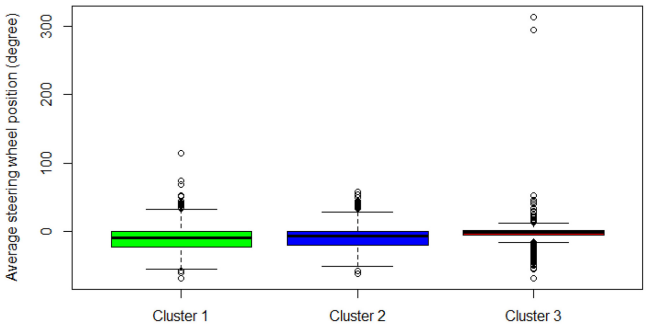

FIGURE 22. Boxplot of std of steering wheel position.

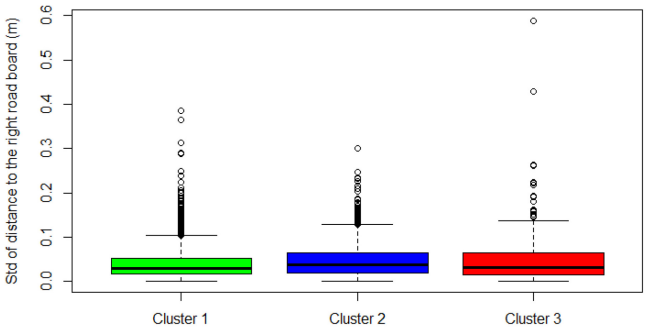

FIGURE 23. Boxplot of std of distance to the right road board.

and even rollover accidents and makes drivers and passengers uncomfortable. Fig. 23 and Fig. 24 show the boxplot of standard deviation of distance to the right road board and boxplot of standard deviation of lateral accelerations, respectively. Similarly, these two factors present the fluctuations of lateral positions and accelerations, and their values of cluster 2 and cluster 3 are slightly bigger than the cluster 1 (i.e., "normal" driving). It is also reasonable since a safe driving behavior should have small fluctuations at the lateral direction.

F. IMPACT OF UNEXPECTED INCIDENTS

In order to explore the impact of unexpected incidents in the driving simulation experiments on drivers, the driver behavior data during no events and events are extracted. The paired T-test is used to further test the difference of

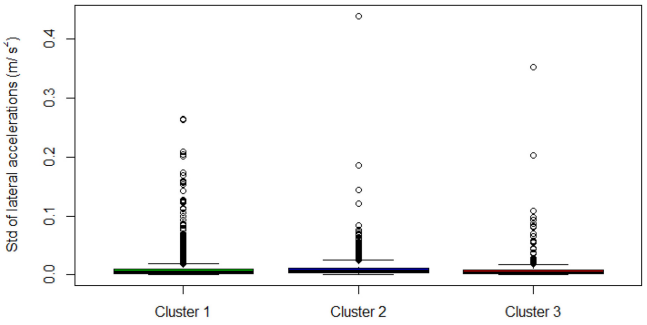

FIGURE 24. Boxplot of std of lateral accelerations.

TABLE 7. Paired T-test results of the driver behavior between with and without events.

Variables	T value	Df	P.value	Mean of the differences
AverageRspur	-1.04	892	0.298	-0.01
StdRspur	-65.92	892	< 0.0001	-0.17
AverageRalpha	8.02	892	< 0.0001	0.38
StdRalpha	-41.48	892	< 0.0001	-1.22
AverageSpeed	-29.66	892	< 0.0001	-7.48
StdevSpeed	42.71	892	< 0.0001	5.63
AverageBrake	77.60	892	< 0.0001	37.57
StdBrake	130.37	892	< 0.0001	35.15
AverageAcc	-91.85	892	< 0.0001	-19.67
StdAcc	-23.76	892	< 0.0001	-6.14
AverageClutch	-47.91	892	< 0.0001	-33.77
StdClutch	34.02	892	< 0.0001	13.47
AverageGear	2.97	892	0.003	0.06
StdGear	-10.04	892	< 0.0001	-0.20
AverageRpm	-55.20	892	< 0.0001	-631.44
StdRpm	11.98	892	< 0.0001	81.73
AverageHWay	17.29	768	< 0.0001	88.53
StdHWay	-59.29	767	< 0.0001	-141.99
AverageDleft	-0.95	892	0.344	-0.01
StdDleft	-65.83	892	< 0.0001	-0.17
LateralPosition	1.26	892	0.209	0.01
StdevLateralPos	-64.96	892	< 0.0001	-0.16
AverageWheel	-5.29	892	< 0.0001	-1.53
StdWheel	-77.46	892	< 0.0001	-10.62
AverageThead	28.85	791	< 0.0001	44.99
StdThead	38.84	790	< 0.0001	69.29
AverageTTL	29.10	892	< 0.0001	26.21
StdTTL	30.67	892	< 0.0001	39.48
AverageTTC	2.02	876	0.044	0.21
StdTTC	-32.77	875	< 0.0001	-1.94
AverageAccLat	4.32	892	< 0.0001	0.01
StdAccLat	30.63	892	< 0.0001	0.04
AverageAccLon	-27.54	892	< 0.0001	-1.63
StdAccLon	21.53	892	< 0.0001	1.05

the driver behavior between with and without events and the result is listed in TABLE 7.

Initially, when the unexpected incidents that are the sudden appearance of an animal (deer or donkey) happens, the driver will taking measures, e.g., releasing gas pedals (i.e., mean of the AverageAcc differences $= -19.67$), increasing braking actions (i.e., mean of the AverageBrake differences $= 37.57$), reducing motor revolutions (i.e., mean of the AverageRpm differences $= -631.44$) and reducing driving speed (i.e., mean of the AverageSpeed differences $= -7.48$ km/h) which increases the standard deviation of speeds (i.e., mean of the StdevSpeed differences $= 5.63$ km/h) and longitudinal acceleration. Besides, the (time and distance) headway increases (i.e., mean of the differences: AverageHWay,

88.53s; AverageThead, 44.99s; AverageTTL, 26.21s) since the speed drop increases the distance to the followed vehicle. Additionally, compared with no event, the lateral acceleration during the events has a slightly improvement and the longitudinal acceleration during the events has a decrease. With the help of paired T-test, we can also find that most of the driving behavior variable have the statistically significant change (i.e., P.value < 0.0001). However, the change of the lateral position (AverageDleft, LateralPosition, AverageRspur) is not significant. The reason may be that drivers can more easily control the vehicle at the lateral orientation when the speed is lower.

V. CONCLUSION

Driving simulators and naturalistic driving studies are often used to understand driving behavior characteristics. It is essential to evaluate the traffic safety of driving behavior in real time, which is helpful to trigger interventions of Advanced Driver Assistance Systems (ADAS) to ensure the driving safety. There are four contributions in this paper. Firstly, this paper proposed a framework of driving behavior safety level classification and evaluation in real time, which is helpful for the further research in the driver behave safety study. Secondly, this paper proposed an improved cross-validation mean square error model based on driver behavior vectors to find the optimal aggregation time interval, which is 1s. Thirdly, the findings of this paper proved that driving safety could be clustered into several levels: ideally three. They can be labelled as "normal" driving, "low-risk" driving, and "high-risk" driving. Fourthly, after comparing k-means clustering, hierarchical clustering, and a model-based clustering (i.e., GMM), k-means clustering and hierarchical clustering gave the optimal number of clusters, and the combination of developed decision trees and k-means clustering outperformed the other combined algorithms. This further supports the hypothesis that the driving data is well clustered in various levels, and that models could be developed for safety level classifications. Additionally, this paper further analyzed and compared the driver behavior in different traffic flow scenarios and distraction scenarios. The factors were also analyzed in the driving behavior safety level and the impact of the unexpected incidents in the driving simulation experiment was also discussed.

In the future, the finding of this paper can help to design Advanced Driver Assistance Systems (ADAS) and active traffic management systems. Once the driver behavior is identified as the "high-risk" driving, some interventions will be triggered to warn the driver. Besides, three safety levels should be applied in these systems. Additionally, this paper proposed the overall framework to conduct the study of the real-time driving behavior safety level classification and evaluation.

Still, this research does not come without limitations. For instance, the dataset did not include existing variables on drivers' demographics and attitudes and perceptions. These will be included in the next stage, to further improve the clustering and classification. Future work should also consider more different driving scenarios.

REFERENCES

[1] K. Yang, C. Al Haddad, G. Yannis, and C. Antoniou, "Driving behavior safety levels: Classification and evaluation," in *Proc. 7th Int. IEEE Conf. Models Technol. Intell. Transp. Syst.*, 2021, pp. 1–6.

[2] S. Kallas, *White Paper on Transport: Roadmap to a Single European Transport Area—Towards a Competitive and Resource-Efficient Transport System*, Office Off. Publ. Eur. Communities, Brussels, Belgium, 2011

[3] S. Singh, *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*, document DOT HS 812 115, Nat. Highway Traffic Safety Admin., Washington, DC, USA, 2015.

[4] P. Bholowalia and A. Kumar, "EBK-means: A clustering technique based on elbow method and k-means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014.

[5] V. Cohen-Addad, V. Kanade, F. Mallmann-Trenn, and C. Mathieu, "Hierarchical clustering: Objective functions and algorithms," *J. ACM*, vol. 66, no. 4, pp. 1–42, 2019.

[6] P. D. McNicholas and T. B. Murphy, "Model-based clustering of microarray expression data via latent Gaussian mixture models," *Bioinformatics*, vol. 26, no. 21, pp. 2705–2712, 2010.

[7] G. C. Linderman, M. Rachh, J. G. Hoskins, S. Steinerberger, and Y. Kluger, "Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data," *Nat. Methods*, vol. 16, no. 3, pp. 243–245, 2019.

[8] D. Pollock, S. Bayarri, and E. Vicente, "A historical perspective of the use of driving simulators in road safety research," in *Progress in System and Robot Analysis and Control Design*. London, U.K.: Springer, 1999, pp. 309–320.

[9] R. G. Mortimer, "Effect of low blood-alcohol concentrations in simulated day and night driving," *Percept. Motor Skills*, vol. 17, no. 2, pp. 399–408, 1963.

[10] S. Noth, "A multi-user driving simulator for studying human driving," Ph.D. dissertation, Fakultät Elektrotechnik Informationstechnik, Ruhr-Universität Bochum, Bochum, Germany, 2016

[11] M. T. Sarwar, P. C. Anastasopoulos, N. Golshani, and K. F. Hulme, "Grouped random parameters bivariate probit analysis of perceived and observed aggressive driving behavior: A driving simulation study," *Anal. Methods Accid. Res.*, vol. 13, pp. 52–64, Mar. 2017.

[12] X. Zhao, Y. Ju, H. Li, C. Zhang, and J. Ma, "Safety of raised pavement markers in freeway tunnels based on driving behavior," *Accid. Anal. Prevent.*, vol. 145, Sep. 2020, Art. no. 105708.

[13] A. Benedetto, A. Calvi, and F. D'Amico, "Effects of mobile telephone tasks on driving performance: A driving simulator study," *Adv. Transp. Stud.*, vol. 26, no. 26, pp. 29–44, 2012.

[14] F. Meng, S. C. Wong, W. Yan, Y. C. Li, and L. Yang, "Temporal patterns of driving fatigue and driving performance among male taxi drivers in Hong Kong: A driving simulator approach," *Accid. Anal. Prevent.*, vol. 125, pp. 7–13, Apr. 2019.

[15] A. H. Jamson and N. Merat, "Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving," *Transp. Res. F Traffic Psychol. Behav.*, vol. 8, no. 2, pp. 79–96, 2005.

[16] A. Mehmood and S. M. Easa, "Modeling reaction time in car-following behaviour based on human factors," *Int. J. Appl. Sci. Eng. Technol.*, vol. 5, no. 14, pp. 93–101, 2009.

[17] 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.

[18] J. Wang, M. Lu, and K. Li, "Characterization of longitudinal driving behavior by measurable parameters," *Transp. Res. Rec.*, no. 1, pp. 15–23, 2010.

[19] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 724–736, Jun. 2012.

[20] L. Eboli, G. Mazzulla, and G. Pungillo, "How to define the accident risk level of car drivers by combining objective and subjective measures of driving style," *Transp. Res. F Traffic Psychol. Behav.*, vol. 49, pp. 29–38, Aug. 2017.

[21] Y. Zheng, J. Wang, X. Li, C. Yu, K. Kodaka, and K. Li, "Driving risk assessment using cluster analysis based on naturalistic driving data," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, 2014, pp. 2584–2589
[22] Y. Yan, Y. Dai, X. Li, J. Tang, and Z. Guo, "Driving risk assessment using driving behavior data under continuous tunnel environment," *Traffic Injury Prevent.*, vol. 20, no. 8, pp. 807–812, 2019.
[23] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, "Dynamic data-driven local traffic state estimation and prediction," *Transp. Res. C Emerg. Technol.*, vol. 34, pp. 89–107, Sep. 2013.
[24] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *J. Amer. Stat. Assoc.*, vol. 97, no. 458, pp. 611–631, 2002.
[25] A. Et-Taleby, M. Boussetta, and M. Benslimane, "Faults detection for photovoltaic field based on K-means, elbow, and average silhouette techniques through the segmentation of a thermal image," *Int. J. Photoenergy*, vol. 2020, Dec. 2020, Art. no. 6617597, doi: 10.1155/2020/6617597.
[26] J. Zhang, W. Chen, M. Gao, and G. Shen, "K-means-clustering-based fiber nonlinearity equalization techniques for 64-QAM coherent optical communication system," *Opt. Exp.*, vol. 25, no. 22, pp. 27570–27580, 2017.
[27] A. Smoliński, B. Walczak, and J. W. Einax, "Hierarchical clustering extended with visual complements of environmental data set," *Chemometr. Intell. Lab. Syst.*, vol. 64, no. 1, pp. 45–54, 2002.
[28] M. F. Balcan, Y. Liang, and P. Gupta, "Robust hierarchical clustering," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3831–3871, 2014.
[29] P. D. McNicholas and T. B. Murphy, "Model-based clustering of longitudinal data," *Can. J. Stat.*, vol. 38, no. 1, pp. 153–168, 2010.
[30] "Model-Based Clustering and Gaussian Mixture Model in R." [Online]. Available: https://en.proft.me/2017/02/1/model-based-clustering-r/ (accessed Jan. 2, 2017).
[31] R. Yu and M. Abdel-Aty, "Utilizing support vector machine in real-time crash risk evaluation," *Accid. Anal. Prevent.*, vol. 51, pp. 252–259, Mar. 2013.
[32] K. Yang, X. Wang, and R. Yu, "A Bayesian dynamic updating approach for urban expressway real-time crash risk evaluation," *Transp. Res. C Emerg. Technol.*, vol. 96, pp. 192–207, Nov. 2018.
[33] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 10–27, 2011.
[34] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *Misc Functions of the Department of Statistics, V R Package Version 1.6-8*, Probability Theory Group, TU Wien, Vienna, Austria, 2017.
[35] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991.
[36] T. Therneau, B. Atkinson, B. Ripley, and M. B. Ripley. "Package 'rpart.'" 2015. [Online]. Available: https://cran.r-project.org/web/packages/rpart/rpart.pdf
[37] K. P. Murphy, *Naive Bayes Classifiers*, vol. 18, Univ. British Columbia, Vancouver, BC, Canada, 2006, pp. 1–8.
[38] M. Majka. "Package 'NaiveBayes.'" 2020. [Online]. Available: https://cran.r-project.org/web/packages/naivebayes/naivebayes.pdf
[39] D. Park, L. R. Rilett, B. J. Gajewski, C. H. Spiegelman, and C. Choi, "Identifying optimal data aggregation interval sizes for link and corridor travel time estimation and forecasting," *Transportation*, vol. 36, no. 1, pp. 77–95, 2009.
[40] Z. Lu, J. Xia, T. Jiao, X. Shi, and W. Huang, "Analysis of optimal temporal aggregation interval of traffic flow data for urban road traffic monitoring," *J. Southeast Univ.*, vol. 42, no. 5, pp. 1000–1005, 2012.
[41] Q. Cai, "Investigation of driver behavior during crash and near-crash events using naturalistic driving data," M.S. thesis, Dept. Civil Constr. Environ. Eng., Iowa State Univ., Ames, IA, USA, 2018. [Online]. Available: https://core.ac.uk/download/pdf/212843876.pdf
[42] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[43] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-SNE effectively," *Distill*, vol. 1, no. 10, p. e2, 2016.
[44] C. Hennig. "FPC: Flexible Procedures for Clustering, R Package Version 2.2-5." 2020. [Online]. Available: https://CRAN.R-project.org/package=fpc
[45] A. B. Garay, G. P. Contreras, and R. P. Escarcina, "A GH-SOM optimization with SOM labelling and dunn index," in *Proc. 11th Int. Conf. Hybrid Intell. Syst. (HIS)*, 2011, pp. 572–577.
[46] S. Łukasik, P. A. Kowalski, M. Charytanowicz, and P. Kulczycki, "Clustering using flower pollination algorithm and Calinski-Harabasz index," in*Proc. IEEE Congr. Evol. Comput. (CEC)*, 2016, pp. 2724–2728.

KUI YANG received the B.Sc. degree in transportation engineering and the Ph.D. degree in transportation engineering from Tongji University, China, in 2014 and 2019, respectively. He has been a Postdoctoral Research Associate with the Technical University of Munich, Germany, since December 2019. His key qualifications and research interests are traffic safety analysis, traffic data mining, machine learning, intelligent transportation system, connected and autonomous vehicle, driving behavior, and big data.

CHRISTELLE AL HADDAD received the bachelor's degree in civil engineering from the American University of Beirut, Lebanon, and the M.Sc. degree in transportation systems from the Technical University of Munich in 2018, where she is currently pursuing the Ph.D. degree. Her research interests include human factors, road safety, and modeling emerging mobility systems.

GEORGE YANNIS is a Professor of Traffic and Safety Engineering with particular focus on data management and analysis with the Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens. For more than 30 years, he has contributed extensively in more than 285 research and engineering projects and studies and in several scientific committees of the European Commission and other International Organizations. He has published more than 675 scientific papers (201 in scientific journals) widely cited worldwide.

CONSTANTINOS ANTONIOU received the Diploma degree in civil engineering from National Technical University of Athens in 1995, and the M.S. degree in transportation and the Ph.D. degree in transportation systems from Massachusetts Institute of Technology in 1997 and 2004, respectively. He is currently a Full Professor with the Chair of Transportation Systems Engineering, Technical University of Munich, Germany. His research interests focus on data analytics, modeling and simulation of transportation systems, intelligent transport systems, calibration and optimization applications, road safety, and sustainable transport systems.