sustainability

MDPI

Article

Application of Unsupervised Machine Learning Classification for the Analysis of Driver Behavior in Work Zones in the State of Qatar

Nour O. Khanfar [1], Huthaifa I. Ashqar [1,2], Mohammed Elhenawy [3], Qinaat Hussain [4], Ahmad Hasasneh [1] and Wael K. M. Alhajyaseen [4,5,*]

1 Natural, Engineering and Technology Sciences Department, Arab American University, 13 Zababde, Jenin P.O. Box 240, Palestine
2 Precision Systems, Inc., Washington, DC 20003, USA
3 CARRS-Q—Centre for Accident Research and Road Safety, Queensland University of Technology, Brisbane, QLD 4059, Australia
4 Qatar Transportation and Traffic Safety Centre, College of Engineering, Qatar University, Doha P.O. Box 2713, Qatar
5 Department of Civil & Architectural Engineering, College of Engineering, Qatar University, Doha P.O. Box 2713, Qatar
* Correspondence: wyaseen@qu.edu.qa

Abstract: Work zone areas are commonly known as crash-prone areas. Thus, they usually receive high priority by road operators as drivers and workers have higher chances of being involved in road crashes. The paper aims to investigate driving behavior in work zones using unsupervised machine learning and vehicle kinematic data. A dataset of 67 participants was gathered through an experiment using a driving simulator located at the Qatar Transportation and Traffic Safety Center (QTTSC). The study considered two different work zone scenarios where the leftmost lane was closed for maintenance. In the first scenario, drivers drove on the leftmost lane (Drive 1), while in the second, they drove on the second leftmost lane (Drive 2). The results show that the number of aggressive and conservative drivers was surprisingly more than normal drivers, as most participants either cautiously drove through or failed to drive without being aggressive. The results also show that drivers acted more aggressively in the leftmost lane rather than in the second leftmost lane. We also found that female drivers and drivers with relatively little driving experience were more likely to be aggressive as they drove through a work zone. The framework was found to be promising and can help policymakers take optimal safety countermeasures in work zones during construction.

Keywords: driving behavior; work zone; vehicle kinematics; classification; unsupervised machine learning

Citation: Khanfar, N.O.; Ashqar, H.I.; Elhenawy, M.; Hussain, Q.; Hasasneh, A.; Alhajyaseen, W.K.M. Application of Unsupervised Machine Learning Classification for the Analysis of Driver Behavior in Work Zones in the State of Qatar. *Sustainability* **2022**, *14*, 15184. https://doi.org/10.3390/su142215184

Academic Editor: Fausto Cavallaro

Received: 12 October 2022
Accepted: 14 November 2022
Published: 16 November 2022

1. Introduction

Every 5.4 min a work zone-related crash happens according to the Federal Highway Administration (FHWA) [1]; as such, work zone safety is considered a high priority for research studies and transportation agencies [1–5]. The topic is not only a concern for road travelers but also a serious concern for the workers in the work zones where they are usually at a high risk of being injured in crashes with vehicles. Therefore, studying work zone areas is important for researchers, policymakers, and transportation safety engineers [6–8].

Many studies have discussed different aspects of the work zone. For example, Bharadwaj et al. examined work zone events to identify the factors that affected crashes in work zones [9]. Hamdar et al. defined some variables that affected work zone conditions and situations [10]. Many other researchers observed different factors that affected the behavior of drivers in work zone areas [10–12]. These factors included work zone configurations, road and traffic conditions, environment characteristics, weather conditions, and driver behavior. In addition, the safety of work zones has been investigated in many studies [6,13] because providing a safe work zone is one of the essential requirements when starting a work zone [6].

A literature review revealed several generally studied aspects of the effect of driving behavior on safety in work zones. For example, a framework investigated longitudinal driving behavior, and the authors argued that the average driver speed was lower for short work zones than for extended ones [10]. On the other hand, another study discussed the relationship between the work zone activity and the average driver speed and that the work zone with long longitudinal buffers can potentially increase crash risks [11]. Another study also investigated if changing the traffic density may affect driving behavior [11]; the authors found that the work zone configuration, concentration, and geographical location impacted driving behavior. Reyes and Khan investigated the barrier impact on driving behavior [8] proving that drivers drove the fastest with concrete jersey barriers. Bharadwaj et al. considered three types of barriers that may affect driving behavior in work zones including using traditional cone pylons, metal barriers, or famous concrete jersey barriers [9]. Wang and Meng suggested that the gender of the drivers has a high impact on driving behavior. Middle-aged male drivers were found to be more likely to engage in risky behavior than middle-aged female [14]. Further, the same study indicated that rainy weather and poor light conditions on the roads were also associated with risky driving behavior.

Moreover, other studies analyzed the road conditions near work zones and their influence on driving behavior [6,10,15]. These characteristics included the total number of work zone lanes, type of closed lanes, number of left lanes open for drivers to proceed in work zones, and lane width. These factors are crucial for classifying driver behavior as drivers tend to skip the closed lane and try to force themselves at the head of the queue. This behavior leads to a high likelihood of a crash by creating turbulence in the lane of traffic. Studies also aimed to understand driver merging behavior and its relationship with the speed and traffic flow in work zone areas [16]. Moreover, a study built a model to assess the crash risk of rear-end collisions and their impact in work zone areas [17]. The results suggest that there was a higher risk of a rear-end collision in the lane closer to the work zone area. In addition to the work zone merging areas and configuration, Bharadwaj et al. argued that the level of inattention of driving through work zones might be higher than driving through other road infrastructures [9], which increases the chance of causing crashes in work zones. Almallah et al. proposed an improved work zone layout for highways using animated-based variable message signs (VMSs) [6]. They compared the driver behavior in the proposed setting with the conventional work zone layout using a driving simulator. They concluded that the VMS system yielded improved driver behavior in terms of speed reduction and lane-changing maneuvers [6].

Previous studies used various methods for understanding driving behavior in general. We can group these methods into two classes including statistical models and machine learning-based models. In the case of the statistical models, the analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) were both used to identify the impact of work zone configurations, road conditions, and traffic density on driving behavior in work zones [10]. In [14], the authors studied the impact of truck percentage and the capacity of road and lane configuration near work zones on driving behavior using a multivariate regression model. In the case of machine learning models, although their application in classifying driver behavior in work zones and in identifying risky behaviors and influencing factors can be very promising, we believe this is still a gap in the literature.

Before we go further, it is worth defining the different driving behaviors from a work zone perspective [18–21]. Conservative drivers were defined as those who drive at a lower speed than the average operating speed of other vehicles in work zones and who had a higher probability of making relatively slower lane-changing maneuvers if needed. Normal drivers were defined as drivers who proceeded within the speed limit near work zones and made the lane-changing maneuvers, if needed, without accelerating or stopping if they cannot complete the movement safely at their current speed. Aggressive drivers are the ones who proceeded at a higher speed than the average operating speed in work zones, making risky overtaking maneuvers and making relatively sudden lane-changing maneuvers when they should have slowed down/stopped because the available distance was not sufficiently safe to complete the lane-change maneuver; however, they decided to accelerate to shorten the required distance.

In this study, an unsupervised machine learning technique was used to comprehensively classify driving behavior in a work zone area. We classified the drivers into three behaviors including aggressive, conservative, and normal. The dataset used in building the model was the result of an experiment using the driving simulator located at the Qatar Transportation and Traffic Safety Center (QTTSC) at the College of Engineering, Qatar University. The driving simulator can collect various driving parameters [6]. This driving simulator was tested in terms of subjective and objective validity in previous studies [22]. It is worth mentioning that different aspects of driving behavior can be validated using this driving simulator including the relationship between driving speed perception and actual speed [23]; geometric field of view [22]; Advanced Driving Assistance Systems (ADASs) [24,25]. Moreover, most of the current studies that investigated the driving behavior in work zone areas used data collected by driving simulators [8,10,26]. Driving simulators are considered as an appropriate and safe tool to test work zone interventions to enhance road safety [8].

In general, several algorithms and techniques may be used to recognize driving behavior. The machine learning method is roughly divided into supervised and unsupervised learning approaches. For example, Ly et al. used a Support Vector Machine (SVM) as a supervised method to explore the possibility of using the labeled vehicle's inertial sensors from the Controller Area Network (CAN) of a bus to build a profile of the drivers [27]. Assigning unknown data into categories by mining the underlying sources of unlabeled data is called the Unsupervised Machine Learning method. For example, clustering and Principal Component Analysis (PCA) from exploratory statistics were used to identify and explain driver groupings according to their driving behavior [28]. However, the most used methods in driving style classification are the Support Vector Machine (SVM) [29], Artificial Neural Network (ANN) [30,31], Random Forest Decision [32], and K-means [33,34].

This study is organized as follows. After the introduction, the next section offers a description of the data collection process, information on the scenario design, and presents the adopted framework in this study. Next, the Analysis and Results section starts with the results of the clusters, presents the clustering results for the two drive scenarios, discusses the demographic characteristics of the drivers including education, gender, and number of years of experience, and compares driver behavior classification and work zones and intersections. Lastly, we conclude with the results and provide recommendations for future work.

2. Materials and Methods

We used a dataset that was collected by a driving simulator located at the QTTSC at Qatar University [6]. The study contained 67 participants who experienced two different drives (i.e., scenarios) on a road with a speed limit of 100 km/h. This will be explained further in the next sections.

2.1. Participants

Before inviting the participants, ethical approval was obtained from the Institutional Review Board (IRB) of Qatar University [6]. Participation announcements were spread within the Qatar University community (i.e., students and staff members) and on social media for the community outside the university. Less than 10% of the driver population of Qatar are Qatari nationals, while the rest are internationals with various background cultures and habits [35]. Furthermore, Qatar imposes heavy fines for speed violations, where speed cameras are installed densely along highways and at most intersections [36].

In total, sixty-nine individuals with valid Qatari drivers' licenses participated in the experiment voluntarily. Two participants could not complete the experiment due to simulation sickness and hence were removed from the analyses. Therefore, data were analyzed for 67 participants (46 males and 21 females). The age of the participants ranged from 18 to 68 years, averaging 24.84 years (SD: 8.3 years). Next, the driving experience of the participants ranged from 1 to 49 years (Mean: 5.58; SD: 7.1 years) [6].

2.2. Scenarios Design

The experiment contained two test drives, each of around 16 km long. The simulation environment of both test drives was designed to mimic the Doha Expressway in the state of Qatar, with a speed limit of 100 km/h [6]. Each test drive included two work zone locations where the temporary posted speed limit was dropped from 100 km/h to 80 km/h. Therefore, this experiment exposed each participant to four different work zone locations (i.e., two leftmost lanes and two second leftmost lanes). In the first scenario, drivers were asked to drive on the leftmost lane (Drive 1). They were instructed to drive on the leftmost lane until they felt a need to change lanes. In the second scenario, the drivers were asked to drive on the second leftmost lane (Drive 2). In this scenario, a preprogrammed vehicle was triggered in a way to drive on the leftmost lane, remain in front of the participant's car, and change lanes to the second leftmost lane when the participant's car approached [6]. Filler pieces were included in the middle of the work zones to reduce the learning effects [6].

2.3. Data Collection Procedure

Each participant was asked to sign an informed consent form by agreeing to transfer the right to use the data for research purposes. Next, they were asked to fill in a pretest questionnaire capturing the demographic characteristics followed by a short (5 to 10 min) warm-up drive in the simulator to get used to it [6]. Before the test drives, each participant was informed about their right to quit the experiment at any time and for any reason. Furthermore, they were instructed to drive as normally as they would in the real world. Afterward, each participant drove both test drives in a randomized order with a short break. During the test drives, the driving-related data, such as speed, acceleration/deceleration, lateral position, spacing with leading vehicles, and lane-changing maneuvers were collected using STISIM Drive® software integrated with CalPot32 [6].

2.4. Framework

In order to classify the driving behavior in work zones, we adopted a framework that we previously developed in [34]. The framework was used to classify the driving behavior at signalized intersection. In this study, we used the same main components of the framework but on a different road infrastructure (i.e., work zones). The first step of the framework was extracting features from the trajectory data of each driver using volatility measures, which have been used in many studies [34,37–39] as significant safety parameters to identify the behavior of drivers. It was also found that the higher the value of the volatility measures the more likely the driver is unstable and risky and thus more aggressive [34,37–40]. We used ten different volatility measures as shown in Table 1. Second, we used these extracted features as input for an unsupervised machine learning algorithm to cluster the behavior of each driver in the work zone. In this study, we used the K-means algorithm, which was used successfully in previous studies [34].

Table 1. Volatility Measures (where V: Speed; D_{long}: Longitudinal Deceleration; A_{long}: Longitudinal Acceleration; and AD_{long} : Longitudinal Deceleration or Acceleration).

Volatility Measure	Description	Equation
DV_1	Standard deviation of speed	$\sqrt{\frac{\sum_{i=1}^{N}\left(V_i - \overline{V}\right)^2}{N}}$
DV_2	Standard deviation of longitudinal deceleration or acceleration	$\sqrt{\frac{\sum_{i=1}^{N}\left(AD_{long_i} - \overline{AD}_{long}\right)^2}{N}}$
DV_3	Coefficient of variation of speed	$100 \times \frac{\sqrt{\frac{\sum_{i=1}^{N}(V_i - V)^2}{N}}}{\overline{V}}$
DV_4	Coefficient of variation of longitudinal acceleration	$100 \times \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(A_{long_i} - A_{long}\right)^2}{N}}}{\overline{A}_{long}}$
DV_5	Coefficient of variation of longitudinal deceleration	$100 \times \frac{\sqrt{\frac{\sum_{i=1}^{N}\left(D_{long_i} - D_{long}\right)^2}{N}}}{\overline{D}_{long}}$
DV_6	Mean absolute deviation of speed	$\frac{\sum_{i=1}^{N}\left\lvert V_i - \overline{V}\right\rvert}{N}$
DV_7	Mean absolute deviation of longitudinal acceleration	$\frac{\sum_{i=1}^{N}\left\lvert A_{long_i} - \overline{A}_{long}\right\rvert}{N}$
DV_8	Quantile coefficient of variation of normalized speed	$100 \times \frac{Q_{V_3} - Q_{V_1}}{Q_{V_3} + Q_{V_1}}$, where Q_1 and Q_3 are the sample 25th and 75th percentiles.
DV_9	Quantile coefficient of variation of longitudinal acceleration	$100 \times \frac{Q_{A_{long_3}} - Q_{A_{long_1}}}{Q_{A_{long_3}} + Q_{A_{long_1}}}$
DV_{10}	Quantile coefficient of variation of longitudinal deceleration	$100 \times \frac{Q_{D_{long_3}} - Q_{D_{long_1}}}{Q_{D_{long_3}} + Q_{D_{long_1}}}$

2.5. K-Means Algorithm

The K-means algorithm is an unsupervised machine learning technique. It aims to partition n observations into k clusters, in which each observation belongs to the cluster so that each group of observations clusters around the nearest centroid [40]. The outcome of the K-means algorithm is a group of clusters with corresponding centroids to minimize the following error function [41]:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} d(x, \mu(C_i))$$

where $C_1, C_2, \cdots, C_k$ are the number of clusters k; $\mu(C_i)$ is the centroid of the cluster C_i, and $d(x, \mu(C_i))$ is the distance between the observation n and the centroid $\mu(C_i)$. In this study, the Euclidean Distance was used. If $x = \{x_1, x_2, \cdots, x_n\}$ and $\mu = \{\mu_1, \mu_2, \cdots, \mu_n\}$ are the points and the clusters' centroids, respectively, the Euclidean Distance from x to μ can be computed as follows:

$$d = \sqrt{\sum_{k=1}^{n} (x_k - \mu_k)^2}$$

Assuming D as the dataset, the steps to implement K-means can be described as follows:

Step 1: Randomly determine initial values for centroids $(C_1, C_2, \cdots, C_k)$ from D.

Step 2: Use Euclidean Distance to assign observations to clusters that have nearest centroids (*min d*).

Step 3: Based on the outcome, recalculate the centroids of the new clusters based on the mean observations in each cluster.

Step 4: Reassign observations to the nearest new clusters based on the calculated centroids in Step 3.

Step 5: Repeat Steps 3 and 4 until no significant deviation in the error function.

Step 6: Report the outcomes.

The K-means algorithm offers many advantages that make it suitable for the proposed framework, including that it is relatively simple to implement and interpret its results, can be straightforwardly scaled to larger datasets, guarantees convergence as the iterations proceed, and can be easily generalized to different cluster datasets and situations.

3. Analysis and Results

Volatility measures for each of the 67 participants were found for each of the two drives. We applied standardization on the data as part of the clustering preparation. We clustered the data using K-means assuming three clusters $k = 3$, each referring to the three different styles: aggressive, conservative, and normal. The different styles were determined based on previous studies that focused on driving style classification [42–45]. Next, we ran the K-means algorithm and classified each driver using the three resulting clusters.

Each resulting cluster represented a different driving style. Python was used as a tool to build and execute the framework including using the K-means clustering algorithm to cluster the drivers into the three different driving styles. Each cluster was labelled as Cluster 1, 2, and 3, indicating conservative, aggressive, and normal driving styles, respectively. In order to identify the resulting clusters, we used the mean values of the classification features (i.e., centers). Higher driving volatility denoted higher driving instability and thus more aggressiveness [34,37–40]. The scaled cluster centers using the ten features are shown in Table 2, while Figure 1 shows the radar charts for those clusters. The axles that start from the center of the charts represent the ten volatility measures, where the length of the bar is equal to the value of the corresponding measure for each represented cluster. As shown in Figure 1, each cluster is represented as a region, where the larger the region the higher the values of the volatility measures. In that sense, we assigned the large region to aggressive driving behavior (Cluster 2), medium to normal (Cluster 3), and small to conservative (Cluster 1).

Table 2. Values of Cluster Centers.

	Cluster 1 (Conservative)	Cluster 2 (Aggressive)	Cluster 3 (Normal)
DV_1	−0.457	1.5604	−0.517
DV_2	−0.611	0.241	0.692
DV_3	−0.408	1.633	−0.257
DV_4	0.676	−0.371	−0.715
DV_5	−0.572	0.04	0.739
DV_6	−0.445	1.582	−1.84
DV_7	−0.355	−0.0605	0.501
DV_8	−0.315	1.243	−0.190
DV_9	0.531	−0.078	−0.666
DV_{10}	−0.151	−0.302	0.349

By applying the Principal Component Analysis (PCA), we represented the clusters in a two-dimensional space (see Figure 2). The results show that the clusters were reasonable, and drivers were grouped into aggressive, normal, and conservative. The aggressive (green) and conservative (purple) driving behaviors were more congested than the normal (red) driving behavior, which means most of the participants can be classified into these two clusters. This shows that most participants either cautiously drove through the work zones or failed to drive without taking aggressive maneuvers. We also performed a Silhouette Coefficient Analysis to study the separation distance between the resulting clusters. We found that the Silhouette coefficient was about 0.32. This means that the three clusters could be considered relatively compact within the cluster to which it belonged but not very far away from the other clusters.

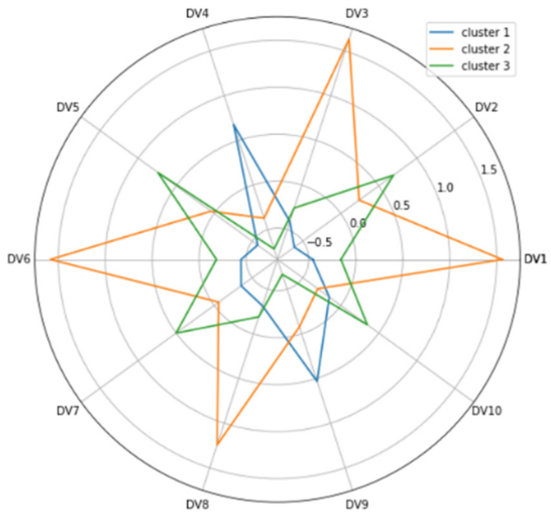

Figure 1. Radar charts, where aggressive is orange (cluster 2); normal is green (cluster 3), and conservative is blue (cluster 1).

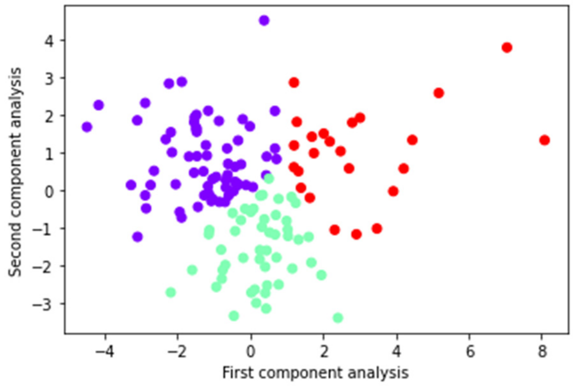

Figure 2. Representation of results in 2D space using PCA, where aggressive is green; normal is red, and conservative is purple.

3.1. Clustering Results of Drive Scenarios

Apart from profiling the behavior of drivers travelling through work zones, the study also investigated the difference in driving behavior in two traffic scenarios (i.e., drives), where in one scenario drivers used the leftmost lane (Drive 1) to traverse through the work zone, while in the second one, drivers used the second leftmost lane (Drive 2). We investigated the impact of driving through the two different drives or situations on driving behavior for the participants using K-means and assuming $k = 3$. Figure 3 shows the clusters of each driver for the two drives. It shows the number and percent of drivers whose behavior changed or stayed the same while driving in the leftmost lane (Drive 1) and the second leftmost lane (Drive 2). About 30% of the participants were classified as being conservative in Drive 1 and Drive 2, while about 22% of them were classified as being aggressive in both driving scenarios. However, only about 6% of the drivers kept their normal behavior as they drove through the two driving scenarios. Notably, about 13% of the aggressive drivers in Drive 1 were classified as conservative in Drive 2, and about 7% of the conservative drivers in Drive 1 changed their behavior to become aggressive in Drive 2.

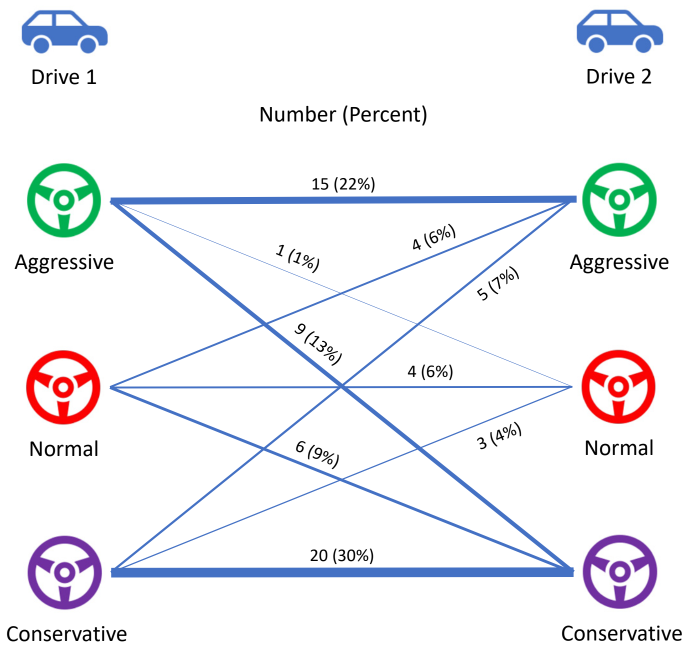

Figure 3. Results of classifying participants for the two drives.

Because work zones on roads and highways during the maintenance and construction phases create unusual driving environments by significantly decreasing highway capacity and increasing the crash risk, the results show that most drivers either approached the work zone cautiously or overreacted. The results show that there were slightly more aggressive (green) drivers and less conservative (purple) drivers in Drive 1 than in Drive 2, which means that the drivers acted more aggressively if they were in the leftmost lane rather than the second leftmost lane. In the leftmost lane, the drivers tended to brake harder and had to use an opportunity to change their lane. In the second leftmost lane, the drivers did not need to change their lane; however, they might have needed to slow down for the lane changers from the leftmost lane, which increased the number of conservative drivers in Drive 2.

3.2. Analysis of Driver Characteristics

Before the experiment, each participant was asked to fill in a pretest questionnaire capturing the demographic characteristics including education, gender, and number of years of experience as shown in Figures 4–6, respectively. We performed a *t*-test to compare the means of the two Drives and to determine whether they were different from one another. We found that the *p*-values in the three demographic characteristics including education, gender, and number of years of experience were less than 0.05, which means there were significant differences between Drive 1 and Drive 2. In terms of the participants' education, we found that the number of aggressive drivers with bachelor's degrees in Drive 1 (i.e., leftmost lane) was higher than Drive 2 (i.e., second leftmost lane). We found that in the second leftmost lane, most of the drivers tended to have conservative driving behavior. In contrast, the number of conservative drivers with a high school education in Drive 1 was more than in Drive 2, where most drivers tended to be aggressive.

In terms of gender, we found that most female drivers were grouped in the aggressive driving behavior category, while male drivers were grouped in the conservative driving behavior category. This means that a female is more likely to be aggressive while driving through a work zone than a male, who is more likely to have conservative driving behavior. Usually, males are more aggressive than females as shown in the literature [46,47]. However, in some of the previous studies, it was found that females tend to be more aggressive than males in the Gulf Arab Countries (GACs), such as the state of Qatar [48]. In the case of driving experience, we found that drivers with more driving experience tended to be less aggressive and have conservative driving behavior. Drivers with relatively little driving experience were more likely to be aggressive as they drove through a work zone [49].

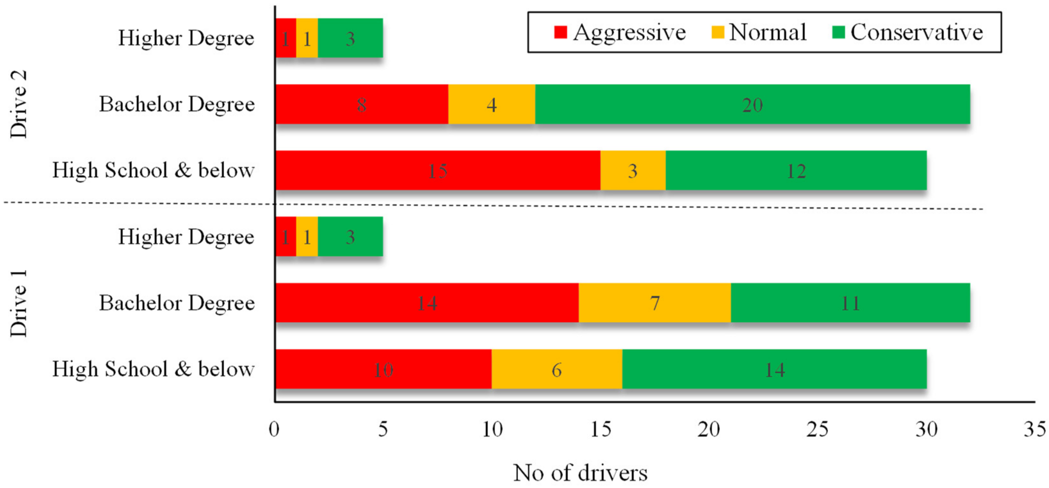

Figure 4. Education distribution between Drive 1 and Drive 2.

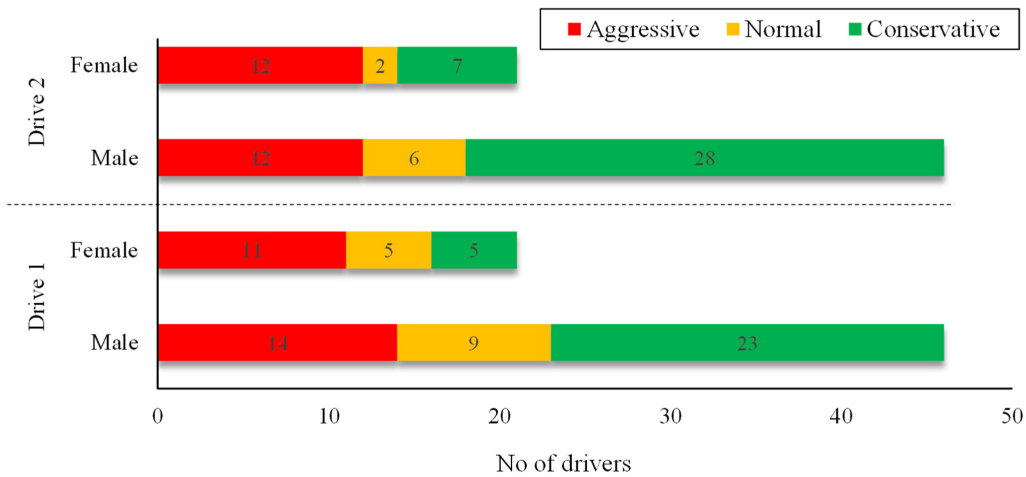

Figure 5. Gender distribution between Drive 1 and Drive 2.

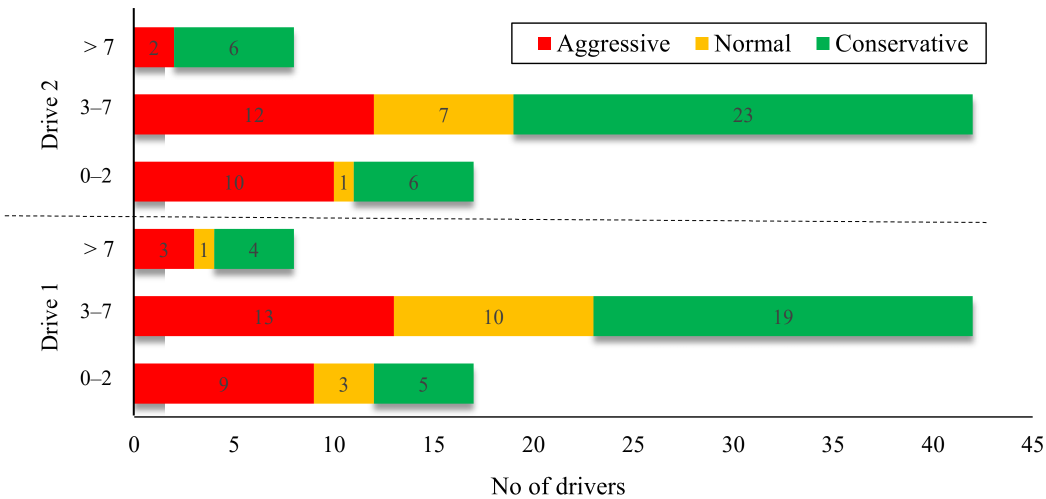

Figure 6. Number of years of experience distribution between Drive 1 and Drive 2.

To conclude the main findings, we found that a female is more likely to be more aggressive driving through a work zone than a male in Qatar. We also argue that drivers with relatively little driving experience are more likely to be aggressive as they drive through a work zone.

3.3. Driving Behavior Classification between Work Zones and Intersections

We found in this study that in work zones, drivers might be triggered to act abnormally due to the unusual configuration and environment of the road in work zone areas. This explains why in this study most of the participants were classified as either aggressive or conservative. Most participants either cautiously drove through the work zones or failed to drive without being aggressive. We believe that classifying driver behavior shows the impact of various road segments, configurations, environments, and infrastructures on the behavior of drivers. In [34], the drivers were classified based on their behavior at signalized intersections using data of 66 drivers extracted from the same simulator installed at Qatar University that was used in this study. Drivers experienced and tested two traffic signals conditions at six intersections. The conclusion was that the participants can be classified into the same three styles: conservative, normal, and aggressive driving behaviors. However, most of the participants driving through intersections were considered as normal drivers in terms of their behavior, in contrast to the findings of this study. At a typical road section in urban areas, which may include highways and intersections, there are usually not enough stimuli that might evoke abnormal behavior (i.e., aggressive or conservative) from drivers. Driver behavior and characteristics are influential contributing factors to traffic crashes. Incorporating these factors with infrastructure-related factors in the adopted traffic policies will decrease the crash severity and frequency.

4. Conclusions

This study applied a framework of unsupervised machine learning to classify the driving behavior in work zone areas into three styles, namely, aggressive, conservative, and normal using vehicle kinematic data. A dataset of 67 participants (46 males and 21 females) was collected by a driving simulator located at the QTTSC at Qatar University. The study considered two different scenarios, where drivers were asked to drive on the leftmost lane (Drive 1) or were asked to drive on the second leftmost lane (Drive 2). The framework was found to be promising in classifying driving behavior into the three styles. The results show that the clusters were reasonable, and the clusters for aggressive and conservative driving behavior were surprisingly more congested than for normal driving behavior. This shows that most of the participants either cautiously drove through the work zones or they failed to drive without taking aggressive maneuvers. The results also show that there were slightly more aggressive drivers and fewer conservative ones in Drive 1 than in Drive 2, which means that the drivers acted more aggressively if they were on the leftmost lane where the maintenance/construction activities were taking place rather than on the second leftmost lane. In the leftmost lane, the drivers tended to brake harder and had to use an opportunity to change their lane. In the second leftmost lane, the drivers did not need to change their lane; however, they might have needed to slow down for the lane changers from the leftmost lane, which increased the number of conservative drivers in Drive 2. In terms of the drivers' characteristics, we found that a female was more likely to be more aggressive when driving through a work zone than a male, who was more likely to be conservative. We also discovered that drivers with relatively little driving experience were more likely to be aggressive as they drove through a work zone. Finally, the results of this study and the insights will be very beneficial in helping policymakers and traffic engineers to take the optimal safety countermeasures in work zones during construction.

There are some considerations for future work. First, increasing the sample size of the drivers might enrich and strengthen the results of the study and lead to other insights. Second, this study was performed on a highway with a 100 km/h operating speed. Investigating the effect of work zones on driving behavior on lower-speed roads, such as arterials and collectors, might indicate other results and enhance the understanding of driving behavior in work zones. Finally, investigating other work zone configurations including the work zone on a different lane (e.g., the rightmost lane), will provide researchers and policymakers with the needed insights to determine the optimal safety countermeasures. Classifying driver behavior reveals the impact of different road infrastructures on the behavior of drivers, which can decrease crash severity and frequency if related policies were adopted.

Author Contributions: Conceptualization, All authors; funding acquisition, W.K.M.A.; software, N.O.K. and M.E.; validation, N.O.K., H.I.A. and M.E.; data curation, Q.H. and W.K.M.A.; methodology, N.O.K., H.I.A. and M.E.; writing—original draft preparation, N.O.K. and H.I.A.; writing—review and editing, All authors; visualization, N.O.K. All authors have read and agreed to the published version of the manuscript.

Funding: This publication was made possible by the NPRP award (NPRP 9-360-2-150) from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the author(s).

Institutional Review Board Statement: The ethical approval number is Qatar University (QU-IRB 609-EA/16).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon request and upon signing a consent.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shahin, F.; Elias, W.; Rosenfeld, Y.; Toledo, T. An Optimization Model for Highway Work Zones Considering Safety, Mobility, and Project Cost. *Sustainability* **2022**, *14*, 1442. [CrossRef]
2. Theofilatos, A.; Ziakopoulos, A.; Papadimitriou, E.; Yannis, G.; Diamandouros, K. Meta-analysis of the effect of road work zones on crash occurrence. *Accid. Anal. Prev.* **2017**, *108*, 1–8. [CrossRef] [PubMed]
3. Pal, R.; Sinha, K.C. Analysis of crash rates at interstate work zones in Indiana. *Transp. Res. Rec.* **1996**, *1529*, 43–53. [CrossRef]
4. Garber, N.J.; Zhao, M. *Final Report Crash Characteristics at Work Zones*; Rep. No. VTRC 02-R12; 2002. Available online: https://rosap.ntl.bts.gov/view/dot/20448 (accessed on 20 February 2022).
5. Ashqar, H.I.; Shaheen, Q.H.Q.; Ashur, S.A.; Rakha, H.A. Impact of risk factors on work zone crashes using logistic models and Random Forest. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; pp. 1815–1820.
6. Almallah, M.; Hussain, Q.; Alhajyaseen, W.K.; Pirdavani, A.; Brijs, K.; Dias, C.; Brijs, T. Improved traffic safety at work zones through animation-based variable message signs. *Accid. Anal. Prev.* **2021**, *159*, 106284. [CrossRef]
7. Edara, P.; Chatterjee, I. Multivariate regression for estimating driving behavior parameters in work zone simulation to replicate field capacities. *Transp. Lett.* **2010**, *2*, 175–186. [CrossRef]
8. Zhang, Y.; Guo, Z.; Sun, Z. Driving Simulator Validity of Driving Behavior in Work Zones. *J. Adv. Transp.* **2020**, *2020*, 4629132. [CrossRef]
9. Bharadwaj, N.; Edara, P.; Sun, C. Risk Factors in Work Zone Safety Events: A Naturalistic Driving Study Analysis. *Transp. Res. Rec. J. Transp. Res. Board* **2019**, *2673*, 379–387. [CrossRef]
10. Shakouri, M.; Ikuma, L.H.; Aghazadeh, F.; Punniaraj, K.; Ishak, S. Effects of work zone configurations and traffic density on performance variables and subjective workload. *Accid. Anal. Prev.* **2014**, *71*, 166–176. [CrossRef]
11. Hamdar, S.H.; Khoury, H.; Zehtabi, S. A simulator-based approach for modeling longitudinal driving behavior in construction work zones: Exploration and assessment. *Simulation* **2016**, *92*, 579–594. [CrossRef]
12. Weng, J.; Meng, Q. Effects of environment, vehicle and driver characteristics on risky driving behavior at work zones. *Saf. Sci.* **2012**, *50*, 1034–1042. [CrossRef]
13. Weng, J.; Meng, Q. Analysis of driver casualty risk for different work zone types. *Accid. Anal. Prev.* **2011**, *43*, 1811–1817. [CrossRef] [PubMed]
14. Osman, M.; Paleti, R.; Mishra, S.; Golias, M.M. Analysis of injury severity of large truck crashes in work zones. *Accid. Anal. Prev.* **2016**, *97*, 261–273. [CrossRef] [PubMed]
15. Khattak, A.J.; Khattak, A.J.; Council, F.M. Effects of work zone presence on injury and non-injury crashes. *Accid. Anal. Prev.* **2002**, *34*, 19–29. [CrossRef]
16. Chitturi, M.V.; Benekohal, R. Work zone queue length and delay methodology. *Transp. Lett.* **2010**, *2*, 273–283. [CrossRef]
17. Weng, J.; Meng, Q. Modeling speed-flow relationship and merging behavior in work zone merging areas. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 985–996. [CrossRef]

18. Vanlaar, W.; Simpson, H.; Mayhew, D.; Robertson, R. Aggressive driving: A survey of attitudes, opinions and behaviors. *J. Saf. Res.* **2008**, *39*, 375–381. [CrossRef]
19. Miles, D.E.; Johnson, G.L. Aggressive driving behaviors: Are there psychological and attitudinal predictors. *Transp. Res. Part F Traffic Psychol. Behav.* **2003**, *6*, 147–161. [CrossRef]
20. Tasca, L. *A Review of the Literature on Aggressive Driving Research*; Citeseer, 2000. Available online: https://www.stopandgo.org/research/aggressive/tasca.pdf (accessed on 21 February 2022).
21. Shinar, D. Aggressive driving: The contribution of the drivers and the situation. *Transp. Res. Part F Traffic Psychol. Behav.* **1998**, *1*, 137–160. [CrossRef]
22. Hussain, Q.; Almallah, M.; Alhajyaseen, W.K.; Dias, C. Impact of the geometric field of view on drivers' speed perception and lateral position in driving simulators. *Procedia Comput. Sci.* **2020**, *170*, 18–25. [CrossRef]
23. Hussain, Q.; Alhajyaseen, W.K.; Pirdavani, A.; Reinolsmann, N.; Brijs, K.; Brijs, T. Speed perception and actual speed in a driving simulator and real-world: A validation study. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *62*, 637–650. [CrossRef]
24. Bifulco, G.N.; Pariota, L.; Galante, F.; Fiorentino, A. Coupling instrumented vehicles and driving simulators: Opportunities from the DRIVE IN2 project. In Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, USA, 16–19 September 2012; pp. 1815–1820.
25. Brijs, T.; Mauriello, F.; Montella, A.; Galante, F.; Brijs, K.; Ross, V. Studying the effects of an advanced driver-assistance system to improve safety of cyclists overtaking. *Accid. Anal. Prev.* **2022**, *174*, 106763. [CrossRef] [PubMed]
26. Bai, Y.; Li, Y. *Determining the Major Causes of Highway Work Zone Accidents in Kansas*; Kansas Department of Transportation: Topeka, KS, USA, 2006.
27. Van Ly, M.; Martin, S.; Trivedi, M.M. Driver classification and driving style recognition using inertial sensors. In Proceedings of the 2013 IEEE Intelligent Vehicles Symposium (IV), Gold Coast, Australia, 23–26 June 2013; pp. 1040–1045.
28. Constantinescu, Z.; Marinoiu, C.; Vladoiu, M. Driving style analysis using data mining techniques. *Int. J. Comput. Commun. Control.* **2010**, *5*, 654–663. [CrossRef]
29. Wang, W.; Xi, J.; Chong, A.; Li, L. Driving Style Classification Using a Semisupervised Support Vector Machine. *IEEE Trans. Human-Mach. Syst.* **2017**, *47*, 650–660. [CrossRef]
30. MacAdam, C.; Bareket, Z.; Fancher, P.; Ervin, R. Using neural networks to identify driving style and headway control behavior of drivers. *Veh. Syst. Dyn.* **1998**, *29*, 143–160. [CrossRef]
31. Brombacher, P.; Masino, J.; Frey, M.; Gauterin, F. Driving event detection and driving style classification using artificial neural networks. In Proceedings of the 2017 IEEE International Conference on Industrial Technology (ICIT), Toronto, ON, Canada, 22–25 March 2017; pp. 997–1002. [CrossRef]
32. Xie, J.; Zhu, M. Maneuver-Based Driving Behavior Classification Based on Random Forest. *IEEE Sensors Lett.* **2019**, *3*, 1–4. [CrossRef]
33. Wang, W.; Xi, J. A rapid pattern-recognition method for driving styles using clustering-based support vector machines. In Proceedings of the 2016 American Control Conference (ACC), Boston, MA, USA, 6–8 July 2016; pp. 5270–5275.
34. Khanfar, N.O.; Elhenawy, M.; Ashqar, H.I.; Hussain, Q.; Alhajyaseen, W.K.M. Driving behavior classification at signalized intersections using vehicle kinematics: Application of unsupervised machine learning. *Int. J. Inj. Control Saf. Promot.* **2022**, 1–11. [CrossRef]
35. Timmermans, C.; Alhajyaseen, W.; Al Mamun, A.; Wakjira, T.; Qasem, M.; Almallah, M.; Younis, H. Analysis of road traffic crashes in the State of Qatar. *Int. J. Inj. Control Saf. Promot.* **2019**, *26*, 242–250. [CrossRef]
36. Toriumi, A.; Abu-Lebdeh, G.; Alhajyaseen, W.; Christie, N.; Gehlert, T.; Mehran, B.; Mussone, L.; Shawky, M.; Tang, K.; Nakamura, H. A multi-country survey for collecting and analyzing facts related to road traffic safety: Legislation, enforcement, and education for safer drivers. *IATSS Res.* **2022**, *46*, 14–25. [CrossRef]
37. Kamrani, M.; Arvin, R.; Khattak, A.J. Extracting Useful Information from Basic Safety Message Data: An Empirical Study of Driving Volatility Measures and Crash Frequency at Intersections. *Transp. Res. Rec. J. Transp. Res. Board* **2018**, *2672*, 290–301. [CrossRef]
38. Arvin, R.; Kamrani, M.; Khattak, A.J. The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. *Accid. Anal. Prev.* **2019**, *132*, 105226. [CrossRef]
39. Arvin, R.; Kamrani, M.; Khattak, A.J. How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data. *Accid. Anal. Prev.* **2019**, *127*, 118–133. [CrossRef] [PubMed]
40. Mohammadnazar, A.; Arvin, R.; Khattak, A.J. Classifying travelers driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning. *Transp. Res. Part C Emerg. Technol.* **2020**, *122*, 102917. [CrossRef]
41. Gan, G.; Ma, C.; Wu, J. *Data Clustering: Theory, Algorithms, and Applications*; SIAM, 2007. Available online: https://epubs.siam.org/doi/book/10.1137/1.9780898718348 (accessed on 25 February 2022).
42. Canale, M.; Malan, S. Analysis and Classification of Human Driving Behaviour in an Urban Environment. *Cogn. Technol. Work* **2002**, *4*, 197–206. [CrossRef]
43. Murphey, Y.L.; Milton, R.; Kiliaris, L. Driver's style classification using jerk analysis. In Proceedings of the 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, Nashville, TN, USA, 2 April 2009; pp. 23–28.

44. Feng, Y.; Pickering, S.; Chappell, E.; Iravani, P.; Brace, C. Driving style analysis by classifying real-world data with support vector clustering. In Proceedings of the 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE), Singapore, 3–5 September 2018; pp. 264–268.
45. Li, G.; Li, S.E.; Cheng, B.; Green, P. Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities. *Transp. Res. Part C Emerg. Technol.* **2017**, *74*, 113–125. [CrossRef]
46. Holland, C.; Geraghty, J.; Shah, K. Differential moderating effect of locus of control on effect of driving experience in young male and female drivers. *Pers. Individ. Differ.* **2010**, *48*, 821–826. [CrossRef]
47. Hennessy, D.A.; Wiesenthal, D.L. Aggression, violence, and vengeance among male and female drivers. *Transp. Q.* **2002**, *56*, 65–75.
48. Hussain, Q.; Alhajyaseen, W.K.; Brijs, K.; Pirdavani, A.; Brijs, T. Innovative countermeasures for red light running prevention at signalized intersections: A driving simulator study. *Accid. Anal. Prev.* **2019**, *134*, 105349. [CrossRef]
49. Li, F.; Yao, X.; Jiang, L.; Li, Y. Driving anger in China: Psychometric properties of the Driving Anger Scale (DAS) and its relationship with aggressive driving. *Pers. Individ. Differ.* **2014**, *68*, 130–135. [CrossRef]