Accident Analysis and Prevention 129 (2019) 250–262

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: An association rules mining approach

Anik Das*, Mohamed M. Ahmed, Ali Ghasemzadeh

University of Wyoming, Department of Civil & Architectural Engineering, 1000 E University Ave, Dept. 3295, Laramie, WY 82071, United States

ARTICLE INFO

Keywords:
Foggy weather conditions
Data mining techniques
Association rules mining
Lane-keeping
Naturalistic driving study
SHRP2
Limited visibility

ABSTRACT

The presence of fog has a significant adverse impact on driving. Reduced visibility due to fog obscures the driving environment and greatly affects driver behavior and performance. Lane-keeping ability is a lateral driver behavior that can be very crucial in run-off-road crashes under reduced visibility conditions. A number of data mining techniques have been adopted in previous studies to examine driver behavior including lane-keeping ability. This study adopted an association rules mining method, a promising data mining technique, to investigate driver lane-keeping ability in foggy weather conditions using big trajectory-level SHRP2 Naturalistic Driving Study (NDS) datasets. A total of 124 trips in fog with their corresponding 248 trips in clear weather (i.e., 2 clear trips: 1 foggy weather trip) were considered for the study. The results indicated that affected visibility was associated with poor lane-keeping performance in several rules. Furthermore, additional factors including male drivers, a higher number of lanes, the presence of horizontal curves, etc. were found to be significant factors for having a higher proportion of poor lane-keeping performance. Moreover, drivers with more miles driven last year were found to have better lane-keeping performance. The findings of this study could help transportation practitioners to select effective countermeasures for mitigating run-off-road crashes under limited visibility conditions.

1. Introduction

Foggy weather conditions have a significant negative influence on driving, including driver behavior and performance. According to the Federal Highway Administration, 1,259,000 vehicle crashes occur each year due to adverse weather conditions, where fog accounts for nearly 28,500 crashes (FHWA, 2018). In addition, previous studies revealed that foggy weather caused a significant number of crashes over different years (National Research Council, 2004; Wu et al., 2018). Fog can worsen visibility and create potential hazardous driving conditions (Elwood and Lyon, 2017). Reduced visibility due to fog affects safe driving behavior by obscuring the details of the environment and decreasing contrast (Hamilton et al., 2014). Due to the lower contrast, drivers are unable to perceive the necessary information from the roadway. Failure of recognizing this information can affect lateral control of driver (Saffarian et al., 2012). One of the lateral driver behaviors that can be highly associated with the run-off-road crashes is lane-keeping ability. A recent study revealed that run-off-road crashes contribute to an average of 57% of motor vehicle traffic fatalities occurred in each year, where a major portion of these crashes occurred at nighttime and inclement weather conditions (Jalayer et al., 2015). Therefore, it is worth investigating driver lane-keeping performance in inclement weather (foggy weather in this study) considering the contribution of poor lane keeping in run-off-road crashes.

Lane-keeping performance has been studied in the literature from driver distraction or reduction of visibility standpoints. A study conducted by Engström et al. examined the effect of visual and cognitive demand on lane-keeping performance and found that lane-keeping performance was decreased by visual demand (Engström et al., 2005). Another study investigated driver inattention on lane-keeping performance and concluded that driver inattention, eyes-off-road, significantly decreased lane-keeping ability (Peng et al., 2013). Furthermore, the study of Barham et al. noted that reduced visibility was the main reason for the inconsistent and uncertain lateral position of drivers, which adversely affected lane-keeping performance (Barham et al., 2000). Using naturalistic driving study data, some recent studies found that adverse weather significantly decreased driver lane-keeping ability (Das et al., 2019; Ghasemzadeh and Ahmed, 2018a, 2017).

While previous studies used traditional parametric model to examine driver behavior including lane-keeping ability (Ahmed and Ghasemzadeh, 2018; Das et al., 2019; Ghasemzadeh and Ahmed, 2017; Peng et al., 2013), numerous data mining techniques have been utilized

* Corresponding author.
E-mail addresses: adas4@uwyo.edu (A. Das), mahmed@uwyo.edu (M.M. Ahmed), aghasemz@uwyo.edu (A. Ghasemzadeh).

https://doi.org/10.1016/j.aap.2019.05.024
Received 15 November 2018; Received in revised form 29 April 2019; Accepted 26 May 2019
Available online 05 June 2019

in prior studies (e.g., Decision Tree, Random Forest (RF), Neural Network (ANN), Factor Analysis, Cluster Analysis, Multivariate Adaptive Regression Splines (MARS), etc.) because of their advantages over parametric models in many aspects (Ghasemzadeh et al., 2018b; Ghasemzadeh and Ahmed, 2018a; Mafi et al., 2018; Meseguer et al., 2013; Wu et al., 2016; Yamamoto et al., 2002). The association rules mining is one of the most popular and commonly used data mining techniques in transportation research; however, it was utilized mainly to identify contributing factors in crashes. Geurts et al. utilized association rule mining to identify accident patterns and characteristics in black spots (Geurts et al., 2005). Pande and Abdel-Aty used an association rules algorithm to discover indirect association in crash data (Pande and Abdel-Aty, 2008). Another study employed association rules in analyzing accident data of Iranian Railways to discover and explore hidden relationships and patterns among the data (Mirabadi and Sharifian, 2010). A study focused on identifying crash contributory factors at urban roundabouts utilized association rules to explore the interdependences between these factors (Montella, 2011). A study conducted by Das and Sun applied association rules to investigate the pattern of traffic crashes under rainy weather conditions (Das and Sun, 2014). In another study, association rules were used to discover patterns from vehicle-pedestrian crash database (Das et al., 2018a). Association rules technique has been attracted many attentions in recent years in different fields including market basket analysis, medical record analysis, product recommendation, and other fields to discover unknown patterns (Jeong et al., 2018; Montella and Ambrosio, 2011; Mousa et al., 2017).

Even though many studies utilized association rules from crash analysis perspective, there are no prior studies that utilized association rules to investigate the impact of adverse weather conditions on driver behavioral performance; to the extent of the authors' knowledge. One of the main advantages of using association rules is that it can uncover obscured relationships among variables within a large database. In fact, there are no predetermined assumptions in association rules mining. Additionally, the relationship between different variables can be clearly described by the rules (Montella et al., 2012). This study utilized association rules mining technique to investigate how driver lane-keeping performance is affected by foggy weather conditions in a naturalistic setting. The trajectory-level Naturalistic Driving Study (NDS) datasets (i.e., continuous time series driving data recording vehicles' trajectory information) utilized in this study were collected by the second Strategic Highway Research Program (SHRP2). The SHRP2 NDS provides a unique opportunity to researchers to analyze driver behavior across times and different spatial locations by providing numerous amounts of representative natural driving data, by compiling a wide variety of traffic, and environmental conditions (Hammit et al., 2019).

The main objective of this study was to examine the effect of foggy weather conditions on driver lane-keeping ability utilizing the association rules mining approach from the massive SHRP2 NDS and Roadway Information Database (RID) datasets. Using association rules mining technique can explore hidden associations among different contributing factors that might influence driver lane-keeping performance and help transportation practitioners to select effective countermeasures for mitigating run-off-road crashes. The remainder of the paper is organized as follows: Section 2 provides a brief description of the SHRP2 NDS and RID datasets used in this study; Section 3 describes the data querying, reduction, cleansing, and preparation processes; Section 4 provides details on the association rules mining methodology used in the analysis; Section 5 explains the descriptive statistics of selected variables; Section 6 discusses results of the association rules analyses; and finally, concluding remarks on the practical implications of the results and limitations of the study are provided.

2. Data description

The study utilized the SHRP2 NDS and RID datasets to achieve the study objective. The NDS compiled an unprecedented amount of data from more than 3400 drivers in six US states including Florida, Indiana, New York, North Carolina, Pennsylvania, and Washington between 2010 and 2013 (Hutton et al., 2014). During the 3-year period, a total of 2 petabytes of data were collected from in-vehicle Data Acquisition Systems (DAS). The DAS was developed by Virginia Tech Transportation Institute (VTTI), which includes vehicle network, forward radar, four video cameras, accelerometers, GPS, and several computer vision algorithms (Campbell, 2012). The RID dataset was developed by the Center for Transportation Research and Education (CTRE) of Iowa State University (Center for Transportation Research and Education (CTRE), 2015). The RID dataset gathered data from SHRP2 mobile data collection project and existing data from public and private sources, which included detailed highway inventory data of the six SHRP2 NDS states, and other supplementary data related to traffic, weather, work zones, crashes, construction, and active safety campaigns in the study sites (Center for Transportation Research and Education (CTRE), 2015; Hutton et al., 2014). This study utilized a subset of the SHRP2 NDS data that were acquired from VTTI with a main focus on freeways. The SHRP2 NDS and RID datasets were linked in this study to explore driver lane-keeping performance with different environmental, traffic and roadway characteristics.

3. Data reduction and preparation

It is worth mentioning that this study is part of the FHWA SHRP2 Implementation Assistance Program (IAP) in Wyoming. The University of Wyoming (UW) led the IAP for the Wyoming DOT (WYDOT) which three phases were awarded; 1) proof-of-concept, 2) comprehensive analysis, and 3) implementation of results. The University of Wyoming research team has developed a data acquisition and reduction protocol to query all adverse-related trips on freeways from the six SHRP2 states. VTTI was the subcontractor to extract the needed time-series and video SHRP2 data. Nevertheless, the University of Wyoming research team developed various methodologies to acquire, process, reduce, and verify the received data. The acquisition and identification of weather-related NDS trips was a challenging task considering several issues (e.g., high variability of weather conditions, and high percentage of old vehicles with errors/ missing data in the DAS recordings). To overcome these challenges, a unique methodology was utilized for extracting NDS trips that occurred in adverse weather conditions. In this method, weather data were extracted from the National Climatic Data Center (NCDC) database to identify weather events. In addition, weather-related crashes were used to pinpoint adverse weather conditions. The specific times and locations of these weather events were used to extract NDS trips occurred in the vicinity of the identified locations and times. More information on the method and data reduction protocol can be found in (Ahmed et al., 2018; Ahmed et al., 2017; Das et al., 2019; Ghasemzadeh et al., 2019; Khan et al., 2018). Note that previous weather-related studies based on SHRP2 NDS data utilized the similar data acquisition methodology to identify weather-related NDS trips (Ali et al., 2019; Das et al., 2018b; Ghasemzadeh et al., 2018a; Ghasemzadeh and Ahmed, 2019, 2018b, 2018c; Hammit et al., 2018; Khan and Ahmed, 2019). Utilizing the data acquisition methodology developed by the UW research team, fog-related NDS trips with their matched trips in clear conditions have been identified from the received data. Afterward, a manual video verification process was conducted in order to filter out the trips that occurred in clear weather conditions. The next step of data reduction procedure was to reduce the dimensionality of the NDS data by selecting the most relevant time-series variables of interest. Note that trajectory-level real-time weather and surface conditions from the driver's perspective could be identified from the NDS video data. Therefore, manual video verification of the trips was required to get accurate real-time trajectory-based information regarding weather, surface, visibility, and traffic conditions. Subsequently, the received NDS trips were segmented into 1-min time

Fig. 1. Heavy Fog. Few road markings are visible, roadway surroundings, ambient traffic is not clearly visible and the horizon cannot be seen (a, b). Signs are unreadable and horizon cannot be seen (c, d).

Fig. 2. Distant Fog. Road markings and route number signs are visible (a, b). Road markings, ambient traffic, roadway surroundings are discernable, horizon cannot be seen (a, b, c, d).

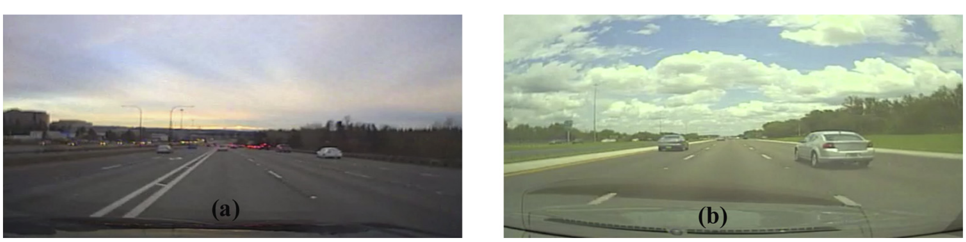


Fig. 3. Clear weather. Road markings and roadway surroundings are visible. Road signs and horizon are also apparent.

interval to preserve the consistent weather conditions within a single trip (Ahmed et al., 2018, 2012). This step also involved manual video observation and annotation of 1-min environmental and traffic conditions. To maintain consistency and minimize subjectivity in the manual video annotation process, video reviewers were trained comprehensively with several sample images and detailed written descriptions before the video observation and data reduction process began. In order to efficiently characterize driver responses (i.e., speed adaptation, lane wandering, etc.) during foggy weather, the UW research team developed a software called Wyoming NDS Visualization and Visibility Identification Tool. The video reviewers then utilized the tool to leverage the process of video observation and annotation of environmental and traffic conditions for each NDS trip. Moreover, post manual verification was conducted by external reviewers. Figs. 1–3 exhibits sample images of heavy fog, distant fog and clear weather that were provided during the manual annotation process.

In this study, 124 trips in foggy weather conditions with their corresponding 248 trips in matched clear weather (i.e., 2 clear trips: 1 foggy weather trip) were randomly selected and reduced for analyzing the lane-keeping behavior. The matching was developed as part of the design of experiment of this study to compare driver lane-keeping behavior in various weather conditions and controlling for other sundry factors. The selected NDS trips involved 61 drivers who drove the same vehicle and same routes in both fog and clear weather conditions on freeways. Table 1 exhibits the characteristics of the 61 drivers involved in the selected NDS trips. It can be seen from Table 1 that the participants are well balanced in categories with respect to gender, driving experience, and amount of yearly driving. However, specific characteristics of the involved SHRP2 participants (i.e., mostly young and educated) follow the same distribution as the entire SHRP2 database (Insight Website, 2019). These 61 drivers' ages ranged from 16 to 79 with a significant number of drivers in age group 30–34 and gender was

Table 1
Descriptive Statistics of NDS Drivers' Characteristics.

Demographics	Levels	Number of Drivers	Percentage (%)
Age	Young (< 25 years)	26	42.62
	Middle (25-55 years)	29	47.54
	Older (> 55 years)	6	9.84
Gender	Male	30	49.18
	Female	31	50.82
Education	Level = 1 (High school diploma or G.E.D.)	5	8.20
	Level = 2 (Some education beyond high school but no degree and College degree)	38	62.29
	Level = 3 (Some graduate or professional school, but no advanced degree (e.g., J.D.S., M.S. or Ph.D.) and Advanced degree (e.g., J.D.S., M.S. or Ph.D.))	18	29.51
Driving Experience	< = 10 years	31	50.82
	> 10 years	30	49.18
Driver Mileage Last Year Details	< 12,000 miles	26	42.62
	= > 12,000 miles	35	57.38

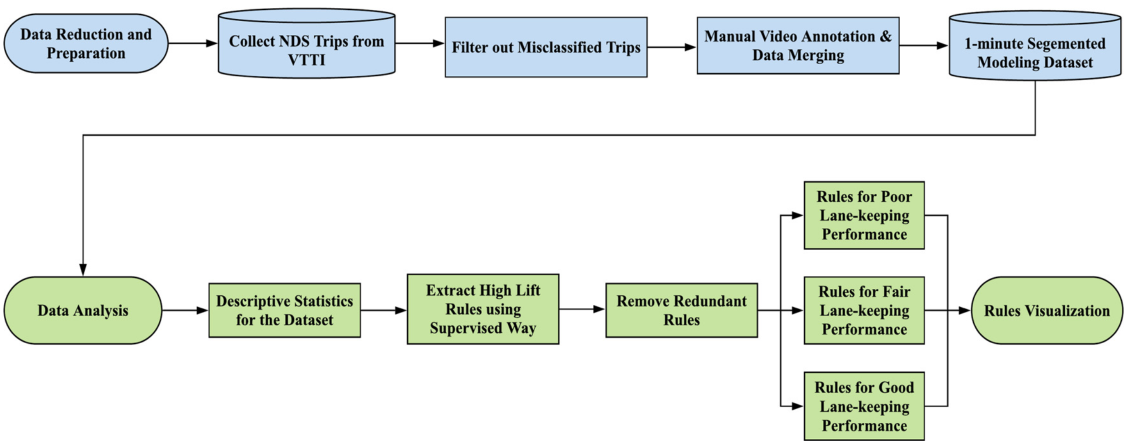


Fig. 4. Summary of Data Preparation and Analysis Process.

mainly balanced among age groups. In total, 7147 1-min segments were reduced from the selected NDS trips. Once the non-freeway segments were removed, 5584 1-min segments that are equivalent to nearly 93 h of driving time and around 8196 traveled km on freeways were considered for the final analysis. Subsequently, roadway characteristics provided in the RID database and driver demographics provided in the SHRP2 administrated survey questionnaires were linked with each 1-min segment to create a final modeling dataset. The overall process of data preparation and analysis is summarized in Fig. 4.

4. Methodology

As stated earlier, the main focus of this study was to investigate the impact of foggy weather conditions on driver lane-keeping performance by employing association rules mining approach. Therefore, this study developed three models of driver lane-keeping performance considering poor, fair, and good lane-keeping, which is discussed in the data analysis section. The details of the association rules mining methodology have been described in the following section.

4.1. Association rules mining

Data mining technique focuses on identifying valuable information from large datasets. The process involves machine learning, statistical techniques, and database management systems to discover this valuable information in the form of associations, interesting patterns, and significant structures. Data mining technique can be classified into two groups, descriptive and predictive. The descriptive data mining technique demonstrates the general properties of the data and illustrates the dataset in a compact way. On the contrary, the predictive approach attempts to anticipate the behavior of the new dataset (Exarchos et al., 2006).

Association rules is a popular descriptive data mining approach and a commonly used rule-based machine-learning method for discovering interesting relations of variables in large databases. The technique is intended to uncover obscured patterns in an itemset (i.e., a set of environmental, traffic, roadway geometry, and driver demographic factors in this study) that occur together or alone in a given event (i.e., a measure of lane-keeping performance in this study) using several algorithms. Among different algorithms, Apriori algorithm, introduced by (Agrawal et al., 1993), is one of the most widely used algorithms to mine association rules. The Apriori algorithm applies level-wise search for mining frequent itemsets. The study utilized Apriori algorithm of association rules to explore key association factors in driver lane-keeping performance.

Before explaining the method, a set of definitions needs to be provided, $I = \{i_1, i_2, \ldots, i_3\}$ be a set of items and $D = \{t_1, t_2, \ldots, t_n\}$ be a set of database lane-keeping performance information called transaction. Each lane-keeping performance information in D contains a subset of the items in I. An association rule can be defined as $A \rightarrow B$, where, $A, B \subseteq I$ and $A \bigcap B = \varnothing$ (Ding et al., 2014). Here, A is called the antecedent or left-hand-side (LHS), and B is consequent or right-hand-side (RHS) (Das et al., 2018a). It is worth mentioning that the inference made by an association rule does not suggest direct causation. Rather it implies the strong association between the antecedent and consequent of the rule.

4.2. Mining interesting rules

Various measures of significance and interest are used to select interesting rules. Among them, support, confidence, and lift are the most commonly used. The support of an association rule is defined as the percentage of lane-keeping performance (i.e., percentage of transaction) in the entire dataset covered by the rule (Hahsler, 2017). Eq. 1 represents the support of association rule (A → B)

$$\text{Support, s}(A \rightarrow B) = \frac{\sigma(A \cap B)}{N} \tag{1}$$

Where $\sigma(A \cap B)$ = Number of rules with particular lane-keeping performance (i.e., poor, fair, or good lane-keeping) where both A and B are present. N is the total number of lane-keeping performance.

The confidence of an association rule, c(A → B) can be measured as the percentage of lane-keeping performance (i.e., transaction) containing A which also contains B (Hahsler et al., 2005). The equation of confidence can be expressed as follows:

$$\text{Confidence, c}(A \rightarrow B) = \frac{s(A \rightarrow B)}{s(A)} \tag{2}$$

However, an association rule A → B needs to satisfy the following constrains:

$$\text{Support, s}(A \rightarrow B) \geq minsup \tag{3}$$

$$\text{Confidence, c}(A \rightarrow B) \geq mincon\,f \tag{4}$$

Where, *minsup* and *mincon f* are the minimum support and minimum confidence, respectively.

Considering minimum support is important to find out a particular significant item in the dataset. A more popular and practical measure to rank the found rules is lift (Brin et al., 1997). The lift can be expressed as a measure of the deviation of the support of the whole rule from the support expected under independence given the support of antecedent and consequent (Hahsler, 2017). In other words, the lift of a rule is the ratio of the confidence of the rule and its expected value. The lift of an association rule, l(A → B) can be calculated as:

$$\text{Lift, l}(A \rightarrow B) = \frac{s(A \rightarrow B)}{s(A) \times s(B)} \tag{5}$$

Where s(A) and s(B) denotes the support of an antecedent and a consequent. A lift value of 1 indicates the independence of the antecedent and consequent. A lift value greater than 1 suggests the positive independence (i.e., antecedent and consequent appear more often together than expected) between antecedent and consequent, whereas a lift value less than 1 suggests the negative independence (i.e., antecedent and consequent appear less often together than expected) (Montella et al., 2011; Geurts et al., 2005).

5. Descriptive statistics

As mentioned earlier, 5584 1-min segments from 372 trips (i.e., 124 trips in fog and 248 trips in matching clear weather) were selected for modeling lane-keeping behavior utilizing association rules mining technique. These 5584 1-min segments contain various number of variables; however, fourteen categorical variables were selected based on prior studies (Das et al., 2019; Ghasemzadeh and Ahmed, 2018a, 2017), as shown in Table 2. In order to investigate factors that might have an impact on driver lane-keeping ability, the association rules mining was utilized in a supervised way. From these fourteen variables, Standard Deviation of Lane Position Offset (SDLP) was considered as the consequents of the supervised association rules mining in this study. Lane Positon Offset is estimated from the distance to the left or right of the center of the lane and center of the vehicle based on machine vision techniques (Insight Website, 2019). SDLP is a widely used measure in assessing lane-keeping performance in existing literature (Ghasemzadeh and Ahmed, 2018a, 2017; Li et al., 2017; Peng et al., 2013; Zhou et al., 2008). A typical value of 20 cm SDLP for normal driving was utilized in this study following previous studies (Green et al., 2004; Zhou et al., 2008). Additionally, SDLP for driving on the curve roads was found to be around 30 cm in a previous study (Green, 1995). Therefore, this study utilized three level of SDLP; less than or equal 20 cm as good lane-keeping, greater than 20 to 30 cm as fair lane-keeping, and greater than 30 cm for poor lane-keeping ability.

The remaining thirteen variables were considered as the antecedents. These antecedents can be interpreted as the potential confounding factors that contribute to the driver lane-keeping ability under different weather and traffic conditions across time and various spatial locations. These variables include environmental, traffic, roadway geometry and driver characteristics. Table 2 summarizes the descriptive statistics of the selected variables.

A significant number of studeis was focused on drivers' lane-keeping behavior on curves in the literature. The study of Jeong and Liu related to the effects of road geometry (i.e., road curvature and curve direction) and lead vehicle on horizontal curve driving performance showed that road curvature affects driving performance, especially the variability of lane-keeping performance (Jeong and Liu, 2017). Hallmark et al. evaluated the influence of roadway geometries on drivers' lane-keeping behavior and indicated that lane position can be modeled as a function of roadway characteristics such as position within the curve (Hallmark et al., 2015). They also investigated drivers normal curve negotiation and suggested that drivers might be vulnerable to roadway departures at certain points during the curve negotiation process. Data were sampled for several upstream positions and along the curve from the time-series data. The sampling was conducted at 50 m increments and up to 300 m upstream of the curve (i.e., in the tangent section) and at equidistant points within the curve due to the variation of the curve lengths. A better understanding of drivers' curve negotiation process provides perspective into the characteristics of driver-roadway-environment that can potentially impact driving behavior. Therefore, knowing the number of drivers' normal lane deviation could interfere with roadway design or policy to tackle crashes due to roadway departures (Hallmark et al., 2015). Hence, an additional variable related to roadway alignment (i.e., curve radius) has been included in this study as an antecedent in addition to "presence of curve" variable to model lane-keeping performance. However, it is worth mentioning that no data were found in 'small/ sharp' curve radius category, considering that the focus of the study is on freeways which generally have a higher design standards compared to other roadway classifications. Therefore, three levels of curve radius (i.e., straight, moderate curve, and large curve) were considered while generating the rules related to three levels of lane-keeping performance. Since the segmentation of the data is temporal-based on 1-min chunks, "time spent driving along the curve" variable was not feasible to be characterized. A 1-min data point is considered on a curve if over 70% of the 1-min was driven on a curve. On the other hand, if more than 70% of the driving within the 1-min time segment occurred on a tangent, the whole epoch was considered a "tangent". It is also worth mentioning that either time or space could be used for segmenting NDS trips. However, time-chunking technique was selected in this study. This study did not consider specific geographic spaces, rather, it considered the NDS trips covering the whole roadway segments on freeways from start to end including straight and curve segments. Therefore, temporal-based segmentation was elected for data reduction, following previous studies which utilized SHRP2 data to investigate various driver behaviors (Ahmed et al., 2018; Das et al., 2019; Ghasemzadeh et al., 2019; Khan et al., 2018).

6. Data analysis

6.1. Results

To get meaningful and significant results, minimum support and confidence need to be defined. Although several optimization algorithms can be applied to identify the minimum support and confidence, this study utilized several trial and error processes to set the optimum values of support and confidence, as recommended by a previous study (Das et al., 2018a). Low minimum support can increase the number of

Table 2
Overview of Selected Variables.

Variable	Description	Levels	Frequency	Percentage (%)
Consequent (Response Variable)				
SDLP	Distance to the left or right of the center of the lane based on machine vision	Good Lane-keeping (≤20 cm)	2026	36.28
		Fair Lane-keeping (> 20-30 cm)	1174	21.02
		Poor Lane-keeping (> 30 cm)	2384	42.69
Antecedents (Explanatory Variables)				
Environmental Characteristics				
Weather Condition	Predominant weather condition in 1-min video observation	Clear	4795	85.87
		Distant Fog	579	10.37
		Heavy Fog	210	3.76
Visibility	Predominant visibility condition in 1-min video observation	Affected	664	11.89
		Not Affected	4920	88.11
Surface Condition	Predominant surface condition in 1-min video observation	Dry	5462	97.82
		Wet	122	2.18
Traffic Characteristics				
Traffic Condition	Predominant traffic condition in 1-min video observation	Free-flow (LOS A&B)	2514	45.02
		Non-Free-flow (LOS C-F)	3070	54.98
Roadway Characteristics				
Presence of Curve	Whether the participants drove curve or tangent in 1-min driving	Curve	1668	29.87
		Tangent	3916	70.13
Speed Limit	Predominant speed limit in 1-min video observation	< = 60 mph (median of Speed Limit)	3018	54.05
		> 60 mph	2566	45.95
Freeway Number of Lanes	Average number of lanes in 1-min segment on the traversed freeway	< = 2 lanes	1491	26.70
		> 2 lanes	4093	73.30
Curve Radius	Predominant curve radius in which participants drove during 1-min driving	No Curve	3916	70.13
		Small (< 700 ft)	0	0
		Moderate (> 700-1700 ft)	100	1.79
		Large (> 1700 ft)	1568	28.08
Driver Characteristics				
Age	The participants' age	Young (< 25 years)	2174	38.93
		Middle (25-55 years)	3116	55.80
		Older (> 55 years)	294	5.27
Gender	The participant's gender	Male	3141	56.25
		Female	2443	43.75
Education	The highest completed level of education of the participant	Level = 1 (High school diploma or G.E.D.)	445	7.97
		Level = 2 (Some education beyond high school but no degree and College degree)	2857	51.16
		Level = 3 (Some graduate or professional school, but no advanced degree (e.g., J.D.S., M.S. or Ph.D.) and Advanced degree (e.g., J.D.S., M.S. or Ph.D.))	2282	40.87
Driving Experience	Number of years driving experience	< = 10 years	2640	47.28
		> 10 years	2944	52.72
Driver Mileage Last Year Details	The approximate number of miles the participant drove last year	< 12,000 miles	1869	33.47
		= > 12,000 miles	3715	66.53

uninteresting rules, which may make it difficult to interpret. Conversely, setting minimum support too high could generate less number of rules, which could lead to a failure finding some interesting rules and the inherent relationship among different itemsets. Therefore, the process of trial and error was challenging to set the minimum threshold for support as well as confidence. In addition, the number of itemsets is another factor to interpret the results. A rule is defined as a 2-item rule if there are a single antecedent and single consequent in the rule. A 3-item rule indicates that there are two antecedents and one consequent or one antecedent and two consequents in the rule. In this study, the maximum length of association rules was set to 4 (i.e., 4-item rule) for easier interpretation of the results (Das and Sun, 2014).

The association rules were generated by using 'arules' package in R software (Hahsler, 2017; Hahsler et al., 2018). To generate association rules among the different characteristics (e.g., environmental, traffic, roadway geometry, and driver) in the database, Apriori algorithm was performed by keeping three levels of SDLP as consequents (i.e., SDLP = Poor, SDLP = Fair, and SDLP = Good). A higher value of lift in the rule indicates a stronger association between antecedent and consequent. As stated earlier, lift values greater than 1 indicate positive independence between antecedent and consequent. Therefore, rules were sorted according to the decreasing value of the lift and minimum threshold of the lift was considered as 1. Table 3 shows the summary statistics of the extracted rules (after removing the redundant rules) where all the rules had lift value higher than 1. From Table 3, it can be seen that the range of lift value corresponding to the poor lane-keeping performance rules varied between 1.29 to 1.77. The range of lift value related to the fair lane-keeping performance was observed from 1.43 to 1.93, with a mean of 1.55. However, the range of lift values from 1.24 to 1.73 was found in the rules corresponding to good lane-keeping performance.

It is worth noting that a lift increase threshold has been utilized in order to select rules with all the antecedents that significantly contribute to the strength of the rule (López et al., 2014; Montella et al., 2012). Once the sorted rules were found, rules with more antecedents were selected over simpler rules based on the following lift increase criteria,

$$\frac{l(A_{n+1} \rightarrow B)}{l(A_n \rightarrow B)} \geq 1.01$$

Where A_{n+1} is the antecedent of the rule having n + 1 items, A_n is the antecedent of the rule with n items, and B is the consequent of each rule (López et al., 2014; Montella et al., 2012).

6.1.1. Rules for poor lane-keeping performance

The association rules with poor lane-keeping performance as consequent were extracted from the generated rules. The minimum support and confidence were set at 1% and 55%, respectively. The number of rules obtained in this case was 105 and all of the rules had a lift value

Table 3
Summary Statistics of Association Rules in this Study.

Consequent Level	Number of Rules (Excluding Redundant Rules, Lift ≥ 1)	Support			Confidence			Lift		
		Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.
1. Poor Lane-keeping (SDLP = Poor)	105	0.045	0.01	0.133	0.611	0.550	0.756	1.432	1.292	1.770
2. Fair Lane-keeping (SDLP = Fair)	107	0.010	0.0025	0.046	0.326	0.300	0.406	1.552	1.427	1.932
3. Good Lane-keeping (SDLP = Good)	102	0.085	0.030	0.193	0.499	0.450	0.628	1.375	1.241	1.731

Table 4
Association Rules for Poor Lane-keeping Performance (using Lift Increase Criteria).

Rules	Antecedent	Consequent	S (%)	C (%)	L
2-item Rules					
1	{Radius = Moderate}	{SDLP = Poor}	1.07	60.00	1.405
3-item Rules					
1	{Education = Middle,Gender = Male}	{SDLP = Poor}	13.31	64.61	1.513
2	{Gender = Male,Visibility = Affected}	{SDLP = Poor}	3.38	60.00	1.405
3	{Age = young,Gender = Male}	{SDLP = Poor}	8.86	59.21	1.387
4	{Visibility = Affected,Weather = Distant Fog}	{SDLP = Poor}	1.70	58.64	1.374
5	{Age = middle,Education = Middle}	{SDLP = Poor}	8.35	57.67	1.351
6	{Driving_Experience < = 10 years,Gender = Male}	{SDLP = Poor}	9.49	57.55	1.348
7	{Lane > 2,Visibility = Affected}	{SDLP = Poor}	4.44	57.54	1.348
4-item Rules (First 40 Rules)					
1	{Driving_Experience > 10 years,Traffic_Condn. = Non-Free flow,Visibility = Affected}	{SDLP = Poor}	1.16	75.58	1.770
2	{Driving_Experience < = 10 years,Gender = Male,Traffic_Condn. = Non-Free-flow}	{SDLP = Poor}	5.57	73.35	1.718
3	{Lane > 2,Visibility = Affected,Weather = Clear}	{SDLP = Poor}	2.26	72.00	1.686
4	{Age = young,Gender = Male,Traffic_Condn. = Non-Free-flow}	{SDLP = Poor}	5.73	71.43	1.673
5	{Driving_Experience > 10 years,Lane > 2,Visibility = Affected}	{SDLP = Poor}	2.27	70.95	1.662
6	{Gender = Male,Traffic_Condn. = Non-Free-flow,Visibility = Affected}	{SDLP = Poor}	1.61	69.23	1.622
7	{Education = High,Lane > 2,Visibility = Affected}	{SDLP = Poor}	1.58	68.75	1.610
8	{Education = Middle,Gender = Male,Traffic_Condn. = Non-Free-flow}	{SDLP = Poor}	8.35	68.23	1.598
9	{Age = middle,Lane > 2,Visibility = Affected}	{SDLP = Poor}	2.49	67.15	1.573
10	{Driving_Experience > 10 years,Visibility = Affected,Weather = Distant Fog}	{SDLP = Poor}	1.02	67.06	1.571
11	{Age = young,Gender = Male,Weather = Distant Fog}	{SDLP = Poor}	1.02	67.06	1.571
12	{Lane > 2,Speed_Limit > 60 mph,Visibility = Affected}	{SDLP = Poor}	3.01	66.93	1.568
13	{Gender = Male,Lane > 2,Visibility = Affected}	{SDLP = Poor}	2.78	66.81	1.565
14	{Driving_Experience < = 10 years,Education = Middle,Gender = Male}	{SDLP = Poor}	7.97	66.32	1.553
15	{Driver_Mileage_Last_Year = > 12,000,Education = Middle,Gender = Male}	{SDLP = Poor}	9.56	66.09	1.548
16	{Driver_Mileage_Last_Year < 12,000,Gender = Male,Traffic_Condn. = Non-Free-flow}	{SDLP = Poor}	2.74	65.95	1.545
17	{Education = Middle,Gender = Male,Lane > 2}	{SDLP = Poor}	11.14	65.40	1.532
18	{Education = Middle,Gender = Male,Surface_Cond. = Dry}	{SDLP = Poor}	13.11	65.36	1.531
19	{Age = young,Education = Middle,Gender = Male}	{SDLP = Poor}	8.20	65.24	1.528
20	{Gender = Male,Visibility = Affected,Weather = Distant Fog}	{SDLP = Poor}	1.07	65.22	1.528
21	{Gender = Male,Radius = Large,Visibility = Affected}	{SDLP = Poor}	1.00	64.37	1.508
22	{Gender = Male,Presence_of_Curve = Curve,Visibility = Affected}	{SDLP = Poor}	1.00	64.37	1.508
23	{Age = middle,Traffic_Condn. = Non-Free-flow,Visibility = Affected}	{SDLP = Poor}	1.20	63.81	1.495
24	{Age = middle,Visibility = Affected,Weather = Distant Fog}	{SDLP = Poor}	1.07	63.16	1.479
25	{Age = middle,Education = Middle,Lane > 2}	{SDLP = Poor}	6.84	63.14	1.479
26	{Gender = Male,Visibility = Affected,Weather = Clear}	{SDLP = Poor}	1.79	62.89	1.473
27	{Age = middle,Education = Middle,Weather = Clear}	{SDLP = Poor}	7.25	62.60	1.466
28	{Driving_Experience < = 10 years,Gender = Male,Weather = Distant Fog}	{SDLP = Poor}	1.00	62.22	1.457
29	{Driving_Experience > 10 years,Gender = Male,Visibility = Affected}	{SDLP = Poor}	2.01	62.22	1.457
30	{Lane > 2,Visibility = Affected,Weather = Distant Fog}	{SDLP = Poor}	1.38	62.10	1.454
31	{Age = young,Gender = Male,Lane > 2}	{SDLP = Poor}	6.52	62.01	1.452
32	{Age = young,Gender = Male,Visibility = Affected}	{SDLP = Poor}	1.29	61.54	1.441
33	{Driver_Mileage_Last_Year = > 12,000,Education = Middle,Speed_Limit > 60 mph}	{SDLP = Poor}	8.43	61.41	1.438
34	{Lane > 2,Traffic_Condn. = Free-flow,Visibility = Affected}	{SDLP = Poor}	2.22	61.39	1.438
35	{Speed_Limit > 60 mph,Visibility = Affected,Weather = Distant Fog}	{SDLP = Poor}	1.09	61.00	1.429
36	{Age = young,Gender = Male,Speed_Limit < = 60 mph}	{SDLP = Poor}	4.03	60.98	1.428
37	{Speed_Limit > 60 mph,Traffic_Condn. = Non-Free-flow,Visibility = Affected}	{SDLP = Poor}	1.58	60.69	1.422
38	{Age = middle,Gender = Male,Visibility = Affected}	{SDLP = Poor}	2.04	60.64	1.420
39	{Age = young,Driving_Experience < = 10 years,Gender = Male}	{SDLP = Poor}	8.11	60.56	1.419
40	{Driver_Mileage_Last_Year = > 12,000,Education = High,Visibility = Affected}	{SDLP = Poor}	1.29	60.00	1.405

Note: S = Support, C = Confidence, L = Lift.

greater than 1. Table 4 shows the 2-item, 3-item, and 4-item rules (First 40 rules) rules by keeping the SDLP = Poor in consequent using the lift increase threshold. The rules were ordered according to the decreasing values of the lift for each item.

6.1.2. Rules for fair lane-keeping performance

The association rules with fair lane-keeping performance as consequent were extracted from the generated rules. The minimum support and confidence were set at 0.25% and 30%, respectively. The number of rules obtained in this case was 107. All the rules had a lift value greater than 1. Considering the lift increase threshold, the 3-item and 4-item rules (First 40 rules) listed in Table 5 were found by keeping the SDLP = Fair as consequent. All the rules were ordered according to the decreasing values of the lift for each item.

Table 5
Association Rules for Fair Lane-keeping Performance (using Lift Increase Criteria).

Rules	Antecedent	Consequent	S (%)	C (%)	L
3-item Rules					
1	{Age = old,Driver_Mileage_Last_Year= > 12,000}	{SDLP = Fair}	0.48	40.30	1.917
2	{Education = Low,Weather = Distant Fog}	{SDLP = Fair}	0.25	35.00	1.665
3	{Driving_Experience > 10 years,Education = Low}	{SDLP = Fair}	0.88	31.82	1.513
4	{Driver_Mileage_Last_Year= > 12,000,Surface_Cond. = Wet}	{SDLP = Fair}	0.34	31.67	1.506
5	{Education = High,Surface_Cond. = Wet}	{SDLP = Fair}	0.32	31.58	1.502
6	{Age = young,Education = Low}	{SDLP = Fair}	0.38	30.88	1.469
7	{Lane < = 2,Weather = Heavy Fog}	{SDLP = Fair}	0.27	30.61	1.456
8	{Education = High,Weather = Heavy Fog}	{SDLP = Fair}	0.32	30.51	1.451
9	{Gender = Male,Weather = Heavy Fog}	{SDLP = Fair}	0.43	30.38	1.445
10	{Driver_Mileage_Last_Year < 12,000,Speed_Limit > 60 mph}	{SDLP = Fair}	4.58	30.01	1.427
4-item Rules (First 40 Rules)					
1	{Gender = Male,Traffic_Condn. = Free-flow,Weather = Heavy Fog}	{SDLP = Fair}	0.29	40.00	1.903
2	{Driver_Mileage_Last_Year < 12,000,Lane < = 2,Speed_Limit > 60 mph}	{SDLP = Fair}	2.27	36.39	1.731
3	{Education = High,Traffic_Condn. = Free-flow,Weather = Heavy Fog}	{SDLP = Fair}	0.30	36.17	1.720
4	{Driver_Mileage_Last_Year= > 12,000,Driving_Experience > 10 years,Education = Low}	{SDLP = Fair}	0.39	36.07	1.715
5	{Driving_Experience > 10 years,Education = Low,Traffic_Condn. = Free-flow}	{SDLP = Fair}	0.50	35.90	1.707
6	{Age = middle,Driving_Experience < = 10 years,Gender = Male}	{SDLP = Fair}	1.11	35.84	1.705
7	{Age = middle,Education = High,Weather = Heavy Fog}	{SDLP = Fair}	0.27	35.71	1.699
8	{Driving_Experience > 10 years,Education = Low,Radius = No Curve}	{SDLP = Fair}	0.70	35.45	1.686
9	{Driving_Experience > 10 years,Education = Low,Presence_of_Curve = Tangent}	{SDLP = Fair}	0.70	35.45	1.686
10	{Driver_Mileage_Last_Year < 12,000,Education = High,Lane < = 2}	{SDLP = Fair}	0.84	35.34	1.681
11	{Education = High,Radius = No Curve,Surface_Cond. = Wet}	{SDLP = Fair}	0.25	35.00	1.665
12	{Education = High,Presence_of_Curve = Tangent,Surface_Cond. = Wet}	{SDLP = Fair}	0.25	35.00	1.665
13	{Driver_Mileage_Last_Year= > 12,000,Radius = No Curve,Surface_Cond. = Wet}	{SDLP = Fair}	0.27	34.88	1.659
14	{Driver_Mileage_Last_Year= > 12,000,Presence_of_Curve = Tangent,Surface_Cond. = Wet}	{SDLP = Fair}	0.27	34.88	1.659
15	{Driving_Experience > 10 years,Education = Low,Gender = Female}	{SDLP = Fair}	0.61	34.69	1.650
16	{Age = young,Education = Low,Visibility = Not Affected}	{SDLP = Fair}	0.38	34.43	1.637
17	{Age = middle,Driver_Mileage_Last_Year < 12,000,Lane < = 2}	{SDLP = Fair}	1.16	34.39	1.636
18	{Education = High,Lane < = 2,Traffic_Condn. = Non-Free-flow}	{SDLP = Fair}	0.81	34.35	1.634
19	{Education = High,Lane < = 2,Weather = Distant Fog}	{SDLP = Fair}	0.43	34.29	1.631
20	{Driving_Experience > 10 years,Gender = Male,Radius = Moderate}	{SDLP = Fair}	0.25	34.15	1.624
21	{Education = High,Radius = No Curve,Weather = Heavy Fog}	{SDLP = Fair}	0.25	34.15	1.624
22	{Education = High,Presence_of_Curve = Tangent,Weather = Heavy Fog}	{SDLP = Fair}	0.25	34.15	1.624
23	{Driving_Experience > 10 years,Education = High,Weather = Heavy Fog}	{SDLP = Fair}	0.27	34.09	1.621
24	{Education = High,Gender = Male,Weather = Heavy Fog}	{SDLP = Fair}	0.27	34.09	1.621
25	{Driving_Experience > 10 years,Gender = Male,Weather = Heavy Fog}	{SDLP = Fair}	0.30	34.00	1.617
26	{Age = middle,Gender = Male,Weather = Heavy Fog}	{SDLP = Fair}	0.30	34.00	1.617
27	{Education = Low,Lane < = 2,Speed_Limit < = 60 mph}	{SDLP = Fair}	0.30	34.00	1.617
28	{Age = middle,Traffic_Condn. = Free-flow,Weather = Heavy Fog}	{SDLP = Fair}	0.39	33.85	1.610
29	{Age = middle,Driving_Experience < = 10 years,Education = High}	{SDLP = Fair}	0.98	33.74	1.605
30	{Driving_Experience > 10 years,Education = Low,Speed_Limit < = 60 mph}	{SDLP = Fair}	0.73	33.61	1.598
31	{Age = middle,Gender = Male,Radius = Moderate}	{SDLP = Fair}	0.25	33.33	1.585
32	{Education = High,Surface_Cond. = Wet,Traffic_Condn. = Non-Free-flow}	{SDLP = Fair}	0.25	33.33	1.585
33	{Driving_Experience > 10 years,Traffic_Condn. = Free-flow,Weather = Heavy Fog}	{SDLP = Fair}	0.39	33.33	1.585
34	{Age = middle,Lane > 2,Weather = Heavy Fog}	{SDLP = Fair}	0.25	33.33	1.585
35	{Education = Low,Gender = Male,Radius = No Curve}	{SDLP = Fair}	0.29	33.33	1.585
36	{Education = Low,Gender = Male,Presence_of_Curve = Tangent}	{SDLP = Fair}	0.29	33.33	1.585
37	{Driver_Mileage_Last_Year < 12,000,Lane < = 2,Radius = Large}	{SDLP = Fair}	0.95	33.33	1.585
38	{Age = young,Driver_Mileage_Last_Year < 12,000,Radius = Large}	{SDLP = Fair}	1.47	33.20	1.579
39	{Driver_Mileage_Last_Year < 12,000,Gender = Male,Lane < = 2}	{SDLP = Fair}	0.97	33.13	1.576
40	{Age = middle,Driver_Mileage_Last_Year < 12,000,Education = High}	{SDLP = Fair}	1.67	32.98	1.569

Note: S = Support, C = Confidence, L = Lift.

6.1.3. Rules for good lane-keeping performance

The association rules with good lane-keeping performance as consequent were extracted from the generated rules. The minimum support and confidence were set at 3% and 45%, respectively. The number of rules obtained in this case was 102, where all the rules had lift value greater than 1. The rules with 3-item and 4-item (First 40 rules) listed in Table 6 were found using the lift increase threshold by keeping the SDLP = Good as consequent. Same as the previous two cases, all the rules were ordered according to the decreasing values of the lift for each item.

6.1.4. Visualization of extracted rules for lane-keeping performance

Manual examination of the many rules generated for different lane-keeping performances is not a feasible option. Therefore, a visual representation of the rules is needed. The results of the association rules can be visualized by using the 'arulesViz' package of R® software. Grouped balloon plots are used to visualize the relationship between the representative group of antecedents and consequents of all the rules. Fig. 5 illustrates grouped balloon plots of association rules generated for three levels of lane-keeping performance. The antecedent groups are represented in rows and consequent (i.e., particular level of lane-keeping performance) in column. The size of the balloon represents aggregated support values and the color indicates the aggregated lift values in the group with a particular lane-keeping performance as a consequent. The rules for poor, fair, and good lane-keeping performance can be easily identified based on high lift and high support groups in Fig. 5. The number of antecedents with the most important (frequent) items in the group are exhibited as the labels for the rows. The rows are plotted in such a way that the aggregated measure of interest (i.e., the lift in this study) is decreasing from top to down. In other words, the rows are sorted according to the decreasing values of lift. Therefore, the group of most interesting rules according to

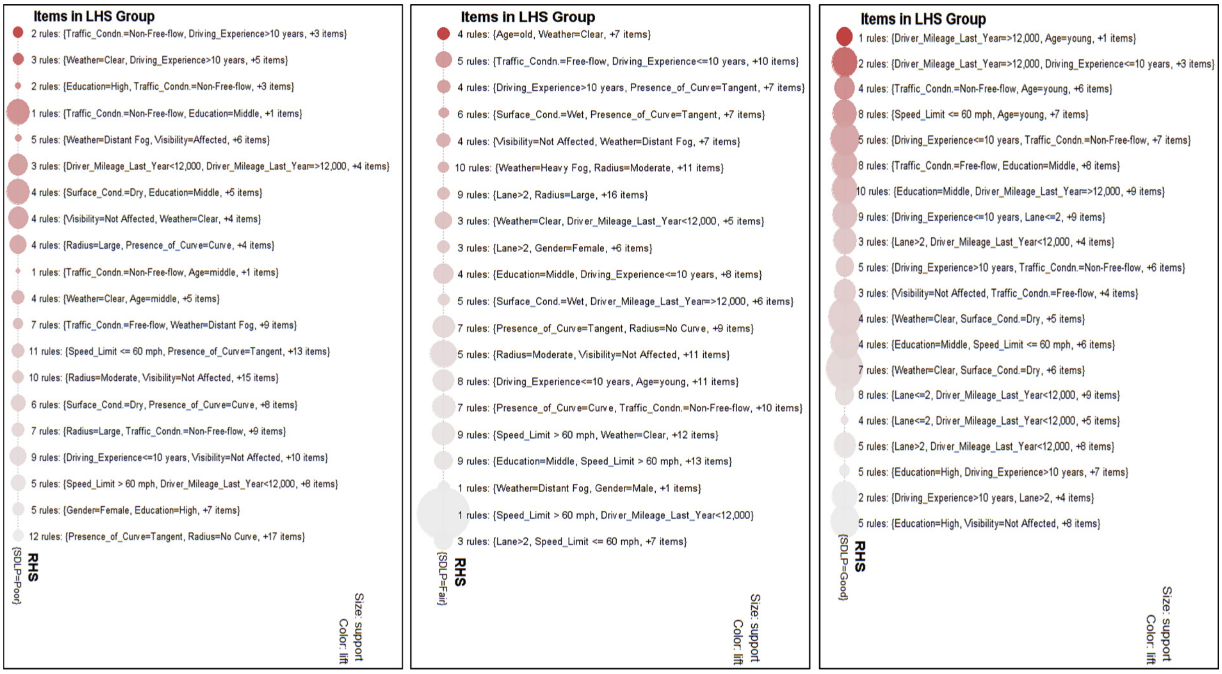
Table 6
Association Rules for Good Lane-keeping Performance (using Lift Increase Criteria).

Rules	Antecedent	Consequent	S (%)	C (%)	L
3-item Rules					
1	{Age = young,Gender = Female}	{SDLP = Good}	12.48	52.09	1.436
2	{Gender = Female,Speed_Limit < = 60 mph}	{SDLP = Good}	12.89	51.21	1.411
3	{Driving_Experience < = 10 years,Gender = Female}	{SDLP = Good}	15.51	50.38	1.389
4	{Education = Middle,Gender = Female}	{SDLP = Good}	14.79	48.39	1.334
5	{Driver_Mileage_Last_Year= > 12,000,Gender = Female}	{SDLP = Good}	10.76	47.85	1.319
6	{Age = young,Speed_Limit < = 60 mph}	{SDLP = Good}	10.44	47.67	1.314
7	{Gender = Female,Radius = No Curve}	{SDLP = Good}	14.63	47.50	1.309
8	{Gender = Female,Presence_of_Curve = Tangent}	{SDLP = Good}	14.63	47.50	1.309
9	{Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	8.29	46.91	1.293
10	{Gender = Female,Traffic_Condn.=Non-Free-flow}	{SDLP = Good}	10.78	46.70	1.287
11	{Driving_Experience < = 10 years,Speed_Limit < = 60 mph}	{SDLP = Good}	12.21	46.55	1.283
12	{Lane < = 2,Speed_Limit < = 60 mph}	{SDLP = Good}	3.30	46.12	1.271
13	{Gender = Female,Lane > 2}	{SDLP = Good}	14.31	45.12	1.243
14	{Gender = Female,Surface_Cond.=Dry}	{SDLP = Good}	19.27	45.06	1.242
4-item Rules (First 40 Rules)					
1	{Age = young,Driver_Mileage_Last_Year= > 12,000,Gender = Female}	{SDLP = Good}	6.29	62.79	1.731
2	{Age = young,Gender = Female,Speed_Limit < = 60 mph}	{SDLP = Good}	9.06	59.25	1.633
3	{Driver_Mileage_Last_Year= > 12,000,Driving_Experience < = 10 years,Gender = Female}	{SDLP = Good}	8.74	58.87	1.622
4	{Driver_Mileage_Last_Year= > 12,000,Education = Middle,Gender = Female}	{SDLP = Good}	6.84	56.43	1.555
5	{Age = young,Gender = Female,Traffic_Condn.=Non-Free-flow}	{SDLP = Good}	6.81	56.38	1.554
6	{Age = young,Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	5.14	56.16	1.548
7	{Driving_Experience < = 10 years,Gender = Female,Speed_Limit < = 60 mph}	{SDLP = Good}	10.83	55.76	1.537
8	{Age = young,Gender = Female,Radius = No Curve}	{SDLP = Good}	9.51	55.49	1.529
9	{Age = young,Gender = Female,Presence_of_Curve = Tangent}	{SDLP = Good}	9.51	55.49	1.529
10	{Education = Middle,Gender = Female,Speed_Limit < = 60 mph}	{SDLP = Good}	10.44	55.31	1.525
11	{Gender = Female,Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	5.18	55.15	1.520
12	{Driver_Mileage_Last_Year= > 12,000,Driving_Experience < = 10 years,Speed_Limit < = 60 mph}	{SDLP = Good}	8.09	54.99	1.516
13	{Gender = Female,Radius = No Curve,Speed_Limit < = 60 mph}	{SDLP = Good}	9.33	54.84	1.512
14	{Gender = Female,Presence_of_Curve = Tangent,Speed_Limit < = 60 mph}	{SDLP = Good}	9.33	54.84	1.512
15	{Age = young,Driver_Mileage_Last_Year= > 12,000,Speed_Limit < = 60 mph}	{SDLP = Good}	6.93	54.51	1.502
16	{Driver_Mileage_Last_Year= > 12,000,Gender = Female,Speed_Limit < = 60 mph}	{SDLP = Good}	8.04	53.45	1.473
17	{Driving_Experience < = 10 years,Gender = Female,Radius = No Curve}	{SDLP = Good}	11.57	53.43	1.473
18	{Driving_Experience < = 10 years,Gender = Female,Presence_of_Curve = Tangent}	{SDLP = Good}	11.57	53.43	1.473
19	{Driving_Experience < = 10 years,Gender = Female,Traffic_Condn.=Non-Free-flow}	{SDLP = Good}	8.51	53.37	1.471
20	{Age = young,Gender = Female,Lane > 2}	{SDLP = Good}	9.29	53.34	1.470
21	{Education = Middle,Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	5.28	52.87	1.457
22	{Driving_Experience < = 10 years,Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	5.62	52.86	1.457
23	{Radius = No Curve,Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	6.12	52.62	1.450
24	{Presence_of_Curve = Tangent,Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	6.12	52.62	1.450
25	{Driver_Mileage_Last_Year= > 12,000,Gender = Female,Traffic_Condn.=Free-flow}	{SDLP = Good}	5.68	51.80	1.428
26	{Driver_Mileage_Last_Year= > 12,000,Driving_Experience < = 10 years,Traffic_Condn.=Free-flow}	{SDLP = Good}	6.30	51.16	1.410
27	{Education = Middle,Gender = Female,Traffic_Condn.=Non-Free-flow}	{SDLP = Good}	8.09	50.96	1.404
28	{Driving_Experience < = 10 years,Radius = No Curve,Speed_Limit < = 60 mph}	{SDLP = Good}	8.65	50.26	1.385
29	{Driving_Experience < = 10 years,Presence_of_Curve = Tangent,Speed_Limit < = 60 mph}	{SDLP = Good}	8.65	50.26	1.385
30	{Driver_Mileage_Last_Year= > 12,000,Speed_Limit < = 60 mph,Traffic_Condn.=Free-flow}	{SDLP = Good}	5.05	50.18	1.383
31	{Gender = Female,Lane < = 2,Radius = No Curve}	{SDLP = Good}	4.37	50.10	1.381
32	{Gender = Female,Lane < = 2,Presence_of_Curve = Tangent}	{SDLP = Good}	4.37	50.10	1.381
33	{Education = Middle,Gender = Female,Lane > 2}	{SDLP = Good}	10.46	49.49	1.364
34	{Driver_Mileage_Last_Year= > 12,000,Gender = Female,Lane > 2}	{SDLP = Good}	8.67	49.44	1.363
35	{Driver_Mileage_Last_Year < 12,000,Gender = Female,Traffic_Condn.=Non-Free-flow}	{SDLP = Good}	5.69	49.30	1.359
36	{Driver_Mileage_Last_Year= > 12,000,Gender = Female,Radius = No Curve}	{SDLP = Good}	7.43	48.94	1.349
37	{Driver_Mileage_Last_Year= > 12,000,Gender = Female,Presence_of_Curve = Tangent}	{SDLP = Good}	7.43	48.94	1.349
38	{Gender = Female,Radius = No Curve,Traffic_Condn.=Non-Free-flow}	{SDLP = Good}	7.90	48.89	1.348
39	{Gender = Female,Presence_of_Curve = Tangent,Traffic_Condn.=Non-Free-flow}	{SDLP = Good}	7.90	48.89	1.348
40	{Speed_Limit < = 60 mph,Surface_Cond.=Dry,Traffic_Condn.=Free-flow}	{SDLP = Good}	8.17	47.45	1.308

Note: S = Support, C = Confidence, L = Lift.

the lift are displayed in the first row of each plot (Hahsler, 2017). For instance, the highest interest group consists of 2 rules which contain “Traffic_Condn.=Non-Free-flow” (i.e., non-free-flow traffic) and “Driving_Experience > 10” (i.e., driving experience greater than 10 years) as well as up to 3 other items in the antecedent when the consequent is “SDLP = Poor” (i.e., poor lane-keeping).

In addition to the grouped balloon plots, scatter plots were generated to visualize the relationship among lift, support, and confidence in the rules generated from poor, fair, and good lane-keeping performance. Fig. 6 shows the scatter plots of the rules for poor, fair, and good lane-keeping performance with support and confidence values on the x-axis and y-axis, respectively. One scatter point denotes an association rule. The color of scatter points (light red to dark red) represents the lift value of each rule, where dark red color indicates the rules with high lift values. It can be seen from Fig. 6 that most of the rules for poor lane-keeping performance are distributed in between the support value of 0.01 to 0.12 and up to 65% confidence, although several high lift rules are located above 65% confidence. On the contrary, most of the rules for fair lane keeping performance are located close to the minimum support threshold and up to a confidence value of 35% indicating that majority of the rules are interesting. However, a number of rules are located in-between the support value of 0.03 to 0.15 with a confidence value of up to 55% for good lane-keeping performance. Additionally, the most interesting rules (i.e., support-confidence-optimal rules) related to poor, fair, and good lane-keeping performance can be easily identified from the support/confidence border (Bayardo and Agrawal,



(a) (b) (c)

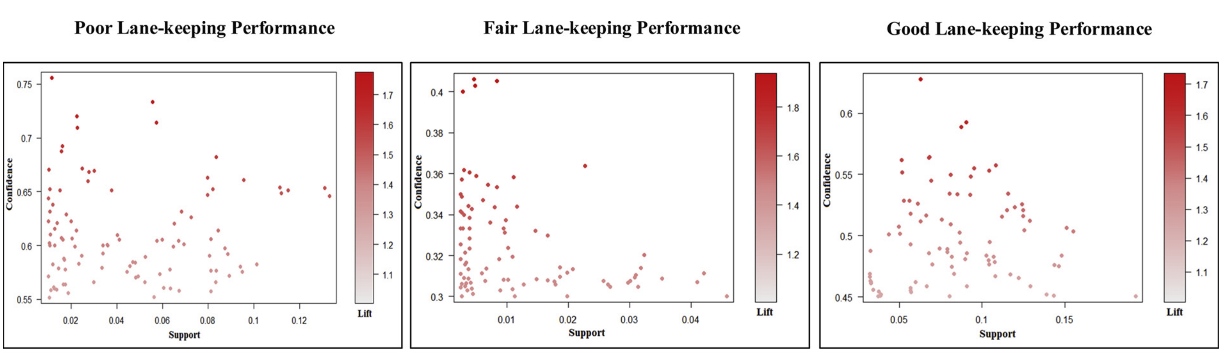
Fig. 5. Grouped Balloon Plots of the Generated Association Rules (a) Poor Lane-keeping Performance, (b) Fair Lane-keeping Performance, and (c) Good Lane-keeping Performance; LHS is the antecedent, RHS is the consequent. The rows are sorted according to the decreasing lift values. The size and color of the balloon indicate the aggregated support and lift values, respectively in the group with a specific lane-keeping performance level as consequent.

1999). In general, Fig. 6 gives an overview of the distribution of support and confidence in the extracted rules set for each of the lane-keeping performance. As the majority of the rules are located close to the support/confidence border, the distribution of the rules can be considered as acceptable to achieve the study objectives regarding their support and confidence values.

6.2. Discussions

From the generated rules for poor lane-keeping performance using the lift increase threshold (Table 4), several 2-item, 3-item, and 4-item rules were found. It was observed that drivers with more than 10 years of driving experience in non-free-flow traffic conditions under affected visibility (4 items, Rule 1: Driving_Experience > 10 years, Traffic_Condn. = Non-Free-flow, Visibility = Affected = > SDLP = Poor) produced highest lift value and were highly associated with poor lane-keeping performance (Support = 1.16%, Confidence = 75.58%, Lift = 1.770). In addition, the effect of distant fog on driver poor lane-keeping performance was found in one rule with 3-items and several rules with 4-items. For example, the rule indicating that young male drivers drove in distant fog conditions had higher propensity to have poor lane-keeping performance (4-items, Rule 11: Age = young, Gender = Male, Weather = Distant Fog = > SDLP = Poor).

According to Table 4, it was also perceived that affected visibility was associated with poor lane-keeping performance in several rules (Rule 2, Rule 4, and Rule 7 in 3-items; Rule 1, Rule 3, Rules 5–7, Rules 9–10, Rules 12–13, Rules 20–24, Rule 26, Rules 29–30, Rule 32, Rules 34–35, Rules 37–38, and Rule 40 in 4-items). This finding is in line with existing literature (Brooks et al., 2011; Das et al., 2019; Ghasemzadeh and Ahmed, 2018a, 2017). For instance, the combined effect of experienced driver, affected visibility, and distant fog on driver poor lane-keeping performance was found in Rule 10 under the 4-item rules category. (Driving_Experience > 10 years, Visibility = Affected, Weather = Distant Fog = > SDLP = Poor). The rule can be expressed



(a) (b) (c)

Fig. 6. Scatter Plots of the Generated Association Rules (a) Poor Lane-keeping Performance, (b) Fair Lane-keeping Performance, and (c) Good Lane-keeping Performance. Each point (light red to dark red) indicates an association rule and position of the rules are determined by the combination of support and confidence of the rule. The plots provide the overall distribution of support and confidence in the extracted rules. The distribution of the rules looks acceptable to achieve the study goals with respect to their support and confidence values as the majority of the rules are located close to the support/confidence border.

as (a) 1.02% of lane-keeping performances in the dataset occurred by drivers with more than 10 years of driving experience in affected visibility and distant fog conditions and produced poor lane-keeping, (b) Out of all lane-keeping performances in the dataset that occurred by drivers with more than 10 years of driving experience in affected visibility and distant fog conditions, 67% were poor lane-keeping, (c) The proportion of poor lane-keeping performances that occurred by drivers with more than 10 years of driving experience in affected visibility and distant fog conditions was 1.57 times the proportion of poor lane-keeping performances in the overall dataset.

It is worth noting that male drivers were dominant in most of the rules for having poor lane-keeping performance. For example, the rule specifies that a proportion of poor lane-keeping performances involving male drivers and affected visibility under distant fog conditions is almost 1.53 times the proportion of poor lane-keeping performances in the overall dataset (4-items, Rule 20: Gender = Male, Visibility = Affected, Weather = Distant Fog = > SDLP = Poor). However, the generated rules indicated that male drivers had worse lane-keeping ability than female drivers. Additionally, the proportion of poor lane-keeping performance was found to be higher for more than two lanes in some of the rules. For instance, the rule implies that drivers drove on freeways with more than two lanes and affected visibility under distant fog condition were associated with poor lane-keeping performance (4-items, Rule 30: Lane > 2, Visibility = Affected, Weather = Distant Fog = > SDLP = Poor). Intuitively, the presence of curve was found to be an important factor for having a higher proportion of poor lane-keeping performance. The finding is consistent with the existing literature (Das et al., 2019; Ghasemzadeh and Ahmed, 2018a). As expected, large radius was found to be a contributing factor for having a higher proportion of poor lane-keeping performance. A study explored that drivers usually begin reacting to the curve sooner for curves with a larger radius than for curves with a smaller radius, which might be the reason for a higher proportion of poor lane-keeping performance on curves with large radiuses (Hallmark et al., 2015).

Several 3-item, and 4-item rules were found in the generated rules for fair lane-keeping performance using lift increase threshold (Table 5). It can be seen from Table 5 that, both fog conditions (i.e., heavy and distant fog) have a higher association in several 3-items rules. The highest lift value was found for a 3-item rule (Rule 1: Age = old, Driver_Mileage_Last_Year = > 12,000 = > SDLP = Poor), which indicates that older drivers with mileage equal or greater than 12,000 miles per year were highly associated with fair lane-keeping performance (Support = 0.48%, Confidence = 40.30%, Lift = 1.917) amongst all the observed rules. However, heavy fog was found to be a contributing factor in several 4-item rules (Rule 1, Rule 3, Rule 7, Rules 21–26, Rule 28, and Rules 33–34). For instance, the rule in 4-item indicates that the proportion of fair lane-keeping performances involving male drivers in free-flow-traffic under heavy fog conditions is almost 1.9 times the proportion of fair lane-keeping performances in the overall dataset (4-items, Rule 1: Gender = Male, Traffic_Condn. = Free-flow, Weather = Heavy Fog = > SDLP = Fair). The factors like less than or equal two lanes, wet surface conditions, the presence of tangent, and education of the participants were dominant in several fair lane-keeping performance rules. Among the three education levels, higher and lower level of education showed significant associations with drivers' fair lane-keeping performance.

Considering the lift increase threshold, several 3-item and 4-item rules were found from the generated rules for good lane-keeping performance in Table 6. As can be seen in Table 6, young female drivers who drove equal or more than 12,000 miles last year (4-items, Rule 1: Age = young, Driver_Mileage_Last_Year = > 12,000, Gender = Female = > SDLP = Good), produced highest lift value and were highly associated with good lane-keeping performance (Support = 6.29%, Confidence = 62.79%, Lift = 1.731). The rule can be expressed as (a) 6.29% of lane-keeping performances in the dataset occurred by young female drivers who drove equal or more than 12,000 miles last year and produced good lane-keeping, (b) Out of all lane-keeping performances in the dataset that occurred by young female drivers who drove equal or more than 12,000 miles last year, 62.79% were good lane-keeping, (c) The proportion of good lane-keeping performances that occurred by young female drivers who drove equal or more than 12,000 miles last year was 1.731 times the proportion of good lane-keeping performances in the complete dataset.

In addition, female drivers were found to be dominant in the antecedent part of the good lane-keeping performance. The findings are expected in comparison with the findings found in the rules for poor lane-keeping performance. However, the combined effect of female drivers who drove on a tangent/straight segment was found to be dominant in some of the rules (Rule 8 in 3-items; Rule 9, Rule 14, Rule 18, Rule 37, and Rule 39 in 4-items). Among different age groups, younger drivers were found to be dominant in having good lane-keeping performance. The finding is consistent with a previous study (Ghasemzadeh and Ahmed, 2018a). Moreover, drivers who drove equal or more than 12,000 miles last year were found to be a significant factor for having a higher proportion of good lane-keeping performance compared to drivers who drove less than 12,000 miles last year. Furthermore, drivers had better lane-keeping ability on roadway segments in which there was no curve (i.e., on tangent segments).

7. Conclusions

The framework presented in this study provided a thorough investigation on how association rules mining approach can be utilized to identify contributing factors that affect driver lane-keeping performance under foggy weather conditions from enriched trajectory-level datasets of the SHRP2 NDS and RID. In this study, 124 trips in fog and additional 248 trips in matching clear weather were extracted representing 61 drivers and ages ranging from 16 to 79 years. Using a 1-min time chunking technique, these trips were reduced to 5584 1-min segments and considered for the final lane-keeping analysis.

Association rules mining were performed to analyze driver lane-keeping performance under foggy weather conditions. Keeping SDLP as a consequent, some interesting findings were observed by mining association rules among a set of environmental, roadway geometry, traffic and driver demographics. It was found that affected visibility was associated with poor lane-keeping performance in several rules. Male drivers were found to be dominant in the rules for having poor lane-keeping performance compared to female drivers. In addition, a higher number of lanes, diving on a large radius of curve, and presence of curve were found to be significant factors for having higher proportions of poor lane-keeping performance. On the other hand, rules from the good lane-keeping performance indicated that visibility had no/little effect on good lane-keeping performance. However, drivers who drove equal or more than 12,000 miles were found to have better lane keeping. Considering the extracted rules from fair lane-keeping performance, it was observed that participants' higher and lower level of education, wet surface conditions, presence of tangent and lower number of freeway lanes were found to be some of the contributing factors for having higher proportions of fair lane-keeping performance on the dataset.

The implementation of association rules mining to naturalistic driving data, with a focus on lane-keeping behavior in fog is relatively new in the area of driver behavior research under inclement weather conditions. Few studies exist in the literature to compare the findings with. Therefore, the framework was developed in such a way that could serve most of the expectations related to probable contributing factors in fog that affect lane-keeping performance. Furthermore, most of the findings of this study were consistent with the results of some earlier studies, where different analytical methods, including statistical parametric and non-parametric techniques, were applied. As stated earlier, several contributing factors including affected visibility, presence of curve, larger radius, higher number of lanes, male drivers, etc., have been found to significantly produce poor lane-keeping performance, which is consistent with the findings from previous studies (Brooks et al., 2011; Das et al., 2019; Ghasemzadeh and Ahmed, 2018a, 2017; Hallmark et al., 2015). Additionally, it was expected that good visibility should be a contributing factor to good lane-keeping ability. However, it was observed that factors like female and younger drivers, the presence of tangent, and drivers with more miles driven last year were found to have better lane-keeping ability, as consistent with some prior studies (Ghasemzadeh and Ahmed, 2018a).

Understanding the impact of foggy weather conditions on driver lane-keeping ability is important to help identifying safety-related issues and countermeasures. On this pivotal point, one of the major contributions is made by determining more domain-specific patterns associated with several factors, including geometric characteristics; driver demographics; as well as environmental and traffic characteristics on drivers' specific lane-keeping performance level using association rules mining approach under different foggy weather events. Form the generated rules found for each lane-keeping performance level, it was noticed that fog conditions had a significant effect on fair and poor lane-keeping performance. Fog was associated with other contributing factors like male drivers, number of lanes, higher and lower education levels, free-flow traffic, tangent/no curve segments, as well as experienced and middle-aged drivers in rules related to fair lane-keeping performance. Additionally, fog was significantly associated with factors including affected visibility, driving experience, male participants, middle and young-aged drivers, higher number of lanes, and higher speed limits in rules corresponding to poor lane-keeping performance. Valuable insights for determining potential countermeasures to enhance the safety in foggy weather conditions could be accomplished by considering the stated numerous factors with fog conditions in each lane-keeping performance.

The investigation on the association rules mining in this study could provide a better understanding of several risk factors associated with driver lane-keeping performance, where visibility is a main issue. As affected visibility was found to be a contributing factor in several rules for occurring poor lane keeping, therefore, it would be recommended to implement safety countermeasures which proven to work effectively in limited visibility. These countermeasures may include Changeable Message Signs (CMSs) to disseminate safety messages at roadway segments with limited visibility due to fog. Safety messages could include restriction of vehicles to one lane for effectively mitigating run-off-road crashes in foggy conditions (Shepard, 1996). In addition, having a curve in a segment in most of the generated rules for poor lane-keeping performance suggests implementing potential countermeasures, which include shoulder rumble strips, paved shoulders, high friction treatment, reflective barrier delineation, chevron, improved lighting, etc., in order to ameliorate drivers' lane keeping through the curve (Hallmark et al., 2015). However, male drivers showed greater propensity towards poor lane-keeping, therefore necessary trainings, educational programs, and campaigns are needed to improve their lane-keeping ability. Identifying fog prone locations with affected visibility, and implement retroreflective 'All-Weather Paint' (AWP) with standard pavement markings may help in increasing the visibility of roadway cues and improve drivers' lane-keeping ability (Cunningham et al., 2013; Das et al., 2019).

Analyzing driver behavior at a trajectory level provided unprecedented opportunity to overcome the unpredictability of driver behavioral trends, especially in different weather and traffic conditions. Therefore, utilizing Big trajectory-level driving Data, the findings from this study may also help in calibrating and advancing drive behavior modeling needed for Advanced Driving Assistance Systems (ADAS) and Connected and Autonomous Vehicle (CAV) under limited visibility conditions.

The findings of this study provided valuable insights regarding the hidden associations among several contributing factors that affect driver lane-keeping performance. Identifying the most suitable support and confidence for each lane-keeping performance was a daunting task, which may need further investigation. Another limitation of this study was to explore the issues specific to affected visibility conditions caused by other inclement weather conditions (i.e., rain and snow), which could be considered in future studies. Therefore, future study is needed to develop more in-depth associations among the contributing factors based on a range of support and confidence values, which might provide broader view of driver lane-keeping performance under reduced visibility due to other weather conditions such as snow and heavy rain.

Acknowledgments

This work was conducted under the second Strategic Highway Research Program (SHRP2), which is administrated by the Transportation Research Board of the National Academies of Sciences, Engineering, and Medicine, and it was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials (AASHTO) and the Wyoming Department of Transportation (WYDOT).

References

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association in large databases. Proc. ACM SIGMOD Int. Conf. Manag. Data-SIGMOD' 93 207–216. https://doi.org/10.1145/170036.170072.

Ahmed, M.M., Abdel-Aty, M., Yu, R., 2012. Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. Transp. Res. Rec. J. Transp. Res. Board 2280 (06), 51–59. https://doi.org/10.3141/2280-06.

Ahmed, M., Young, R., Ghasemzadeh, A., Hammit, B., Ali, E., Khan, N., Das, A., Eldeeb, H., 2017. Implementation of SHRP2 Results within the Wyoming Connected Vehicle Variable Speed Limit System: Phase 2 Early Findings Report and Phase 3 Proposal, Wyoming Department of Transportation (WYDOT).

Ahmed, M.M., Ghasemzadeh, A., 2018. The impacts of heavy rain on speed and headway behaviors: an investigation using the SHRP2 naturalistic driving study data. Transp. Res. Part C Emerg. Technol. 91 (February), 371–384. https://doi.org/10.1016/j.trc.2018.04.012.

Ahmed, M.M., Ghasemzadeh, A., Hammit, B., Khan, N., Das, A., Ali, E., Young, R.K., Eldeeb, H., 2018. Driver Performance and Behavior in Adverse Weather Conditions: An Investigation Using the SHRP2 Naturalistic Driving Study Data-Phase 2. Final Rep. WY-18/05F, Wyoming Dep. Transp.Driver Performance and Behavior in Adverse Weather Conditions: An Investigation Using the SHRP2 Naturalistic Driving Study Data-Phase 2. Final Rep. WY-18/05F, Wyoming Dep. Transp.

Ali, E.M., Ahmed, M.M., Wulff, S.S., 2019. Detection of critical safety events on freeways in clear and rainy weather using SHRP2 naturalistic driving data: parametric and non-parametric techniques. Saf. Sci. 1–9. https://doi.org/10.1016/j.ssci.2019.01.007. November 2018.

Barham, P., Andreone, L., Zhang, X.H., Vache, M., 2000. The "Darwin" driver vision support system - its potential impact on driving behaviour and road safety. Conditions of Reduced Visibility, in: European Transport Conference. Association for European Transport 77–88.

Bayardo, R.J., Agrawal, R., 1999. Mining the Most Interesting Rules. pp. 145–154. https://doi.org/10.1145/312129.312219.

Brin, S., Motwani, R., Ullman, J.D., Tsur, S., 1997. Dynamic itemset counting and implication rules for market basket data. ACM SIGMOD Rec. 26 (2), 255–264. https://doi.org/10.1145/253262.253325.

Brooks, J.O., Crisler, M.C., Klein, N., Goodenough, R., Beeco, R.W., Guirl, C., Tyler, P.J., Hilpert, A., Miller, Y., Grygier, J., Burroughs, B., Martin, A., Ray, R., Palmer, C., Beck, C., 2011. Speed choice and driving performance in simulated foggy conditions. Accid. Anal. Prev. 43 (3), 698–705. https://doi.org/10.1016/j.aap.2010.10.014.

Campbell, K.L., 2012. The SHRP 2 Naturalistic Driving Study. Transp. Res. News.

Center for Transportation Research and Education (CTRE), 2015. SHRP2-Roadway Information Database [WWW Document]. URL http://www.ctre.iastate.edu/shrp2-rid/ (Accessed 26 March 2018). .

Cunningham, C.M., Hummer, J.E., Chang, J., Katz, D., Vaughan, C.L., Findley, D.J., Schroeder, B.J., McAvoy, D., 2013. All-Weather Pavement Marking for Work Zones : Field Evaluation in North Carolina and Ohio. Publ. No. FHWA-HIF-13-004. U.S. Dep. Transp. Fed. Highw. Adm. May.

Das, S., Sun, X., 2014. Investigating the pattern of traffic crashes under rainy weather by association rules in data mining. Proc. 93rd Transp. Res. Board Annu. Meet. pp. 1–19.

Ding, J., Wang, J., Liu, C., Lu, M., Li, K., 2014. A driver steering behavior model based on lane-keeping characteristics analysis. 2014 17th IEEE Int. Conf. Intell. Transp. Syst. ITSC 2014 623–628. https://doi.org/10.1109/ITSC.2014.6957759.

Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., Jalayer, M., 2018a. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. Int. J. Urban Sci. 0 (0), 1–19. https://doi.org/10.1080/12265934.2018.1431146.

Das, A., Ghasemzadeh, A., Ahmed, M.M., 2018b. A comprehensive analysis of driver lane-keeping performance in fog weather conditions using the SHRP2 naturalistic driving study data. Proc. 97th Transp. Res. Board Annu. Meet.

Das, A., Ghasemzadeh, A., Ahmed, M.M., 2019. Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data. J. Saf. Res. 68, 71–80. https://doi.org/10.1016/j.jsr.2018.12.015.

Elwood, J., Lyon, R., 2017. Surface Weather Observations and Reports, Federal Meteorological Handbook No. 1. US Department of Commerce/National Oceanic and Atmospheric Administration.

Engström, J., Johansson, E., Östlund, J., 2005. Effects of visual and cognitive load in real and simulated motorway driving. Transp. Res. Part F Traffic Psychol. Behav. 8 (2), 97–120. https://doi.org/10.1016/j.trf.2005.04.012. SPEC. ISS.

Exarchos, T.P., Papaloukas, C., Fotiadis, D.I., Michalis, L.K., 2006. An association rule mining-based methodology for automated detection of ischemic ECG beats. IEEE Trans. Biomed. Eng. 53 (8), 1531–1540. https://doi.org/10.1109/TBME.2006.873753.

FHWA, 2018. How Do Weather Events Impact Roads? [WWW Document]. URL https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm (Accessed 1 March 2018). .

Geurts, K., Thomas, I., Wets, G., 2005. Understanding spatial concentrations of road accidents using frequent item sets. Accid. Anal. Prev. 37 (4), 787–799. https://doi.org/10.1016/j.aap.2005.03.023.

Ghasemzadeh, A., Ahmed, M.M., 2017. Drivers' lane-keeping ability in heavy rain: preliminary investigation using SHRP2 naturalistic driving study data. Transp. Res. Rec. J. Transp. Res. Board 2663, 99–108. https://doi.org/10.3141/2663-13.

Ghasemzadeh, A., Ahmed, M.M., 2018a. Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: non-parametric MARS and parametric logistic regression modeling approaches. Transp. Res. Part C Emerg. Technol. 90 (March), 379–392. https://doi.org/10.1016/j.trc.2018.03.018.

Ghasemzadeh, A., Ahmed, M.M., 2018b. Exploring factors contributing to injury severity at work zones considering adverse weather conditions. IATSS Res. https://doi.org/10.1016/j.iatssr.2018.11.002.

Ghasemzadeh, A., Ahmed, M.M., 2018c. A framework to enhance the transferability of the SHRP2 naturalistic driving study by considering heterogeneity of driver behavior using spatial-temporal factors in a trajectory level. Proc. 2018 21st Int. Conf. Intell. Transp. Syst. 1826–1833. https://doi.org/10.1109/ITSC.2018.8569927.

Ghasemzadeh, A., Ahmed, M.M., Gaweesh, S.M., 2018a. Multivariate adaptive regression splines and logistic regression models to identify the impact of rainy weather on driver lane-keeping performance considering driver demographics and roadway characteristics using SHRP2 naturalistic driving data. Proc. 97th Transp. Res. Board Annu. Meet.

Ghasemzadeh, A., Hammit, B.E., Ahmed, M.M., Young, R.K., 2018b. Parametric ordinal logistic regression and non-parametric decision tree approaches for assessing the impact of weather conditions on driver speed selection using naturalistic driving data. Transp. Res. Rec. J. Transp. Res. Board 2672 (12), 137–147. https://doi.org/10.1177/0361198118758035.

Ghasemzadeh, A., Ahmed, M.M., 2019. Complementary parametric probit regression and nonparametric classification tree modeling approaches to analyze factors affecting severity of work zone weather-related crashes. J. Mod. Transp. 27 (2), 129–140. https://doi.org/10.1007/s40534-018-0178-6.

Ghasemzadeh, A., Hammit, B.E., Ahmed, M.M., Eldeeb, H., 2019. Unlocking new potentials of SHRP2 naturalistic driving study data: complementary methodologies to supplement real-time weather conditions and effective frameworks for data reduction and preparation. Saf. Sci. 1–8. https://doi.org/10.1016/j.ssci.2019.01.006.

Green, P., 1995. Suggested Procedures and Acceptance Limits for Assessing the Safety and Ease of Use of Driver Information Systems, FHWARD-94-089, Federal Highway Administration. US Department of Transportation, Virginia.

Green, P., Cullinane, B., Zylstra, B., Smith, D., 2004. Typical Values for Driving Performance With Emphasis on the Standard Deviation of Lane Position: a Summary of the Literature Prepared, A Report on Safety Vehicles Using Adaptive Interface Technology (Task 3A). The University of Michigan Transportation Research, Ann Arbor, MI.

Hahsler, M., 2017. arulesViz: Interactive Visualization of Association Rules With R. R J. 9/2 December. pp. 1–13.

Hahsler, M., Buchta, C., Gruen, B., Michael, A., Buchta, C., Gruen, B., Hornik, K., Johnson, I., 2018. Arules: Mining Association Rules and Frequent Itemsets. [WWW Document]. .

Hahsler, M., Grün, B., Hornik, K., 2005. Introduction to arules – a computational environment for mining association rules and frequent item sets. J. Stat. Softw. 14 (15), 1–25. https://doi.org/10.1007/978-1-84800-201-2.

Hallmark, S.L., Oneyear, N., Tyner, S., Wang, B., 2015. Analysis of Naturalistic Driving Study Data: Roadway Departures on Rural Two-Lane Curves. SHRP2 Rep. S2-S08D-RW-1 91.

Hammit, B.E., Ghasemzadeh, A., James, R.M., Ahmed, M.M., Young, R.K., 2018. Evaluation of weather-related freeway car-following behavior using the SHRP2 naturalistic driving study database. Transp. Res. Part F Traffic Psychol. Behav. 59, 244–259. https://doi.org/10.1016/j.trf.2018.08.023.

Hamilton, B., Tefft, B., Arnold, L., Grabowski, J., 2014. Hidden Highways: Fog and Traffic Crashes on America's Roads. AAA Found. Traffic Saf. November.

Hammit, B.E., James, R., Ahmed, M., Young, R., 2019. Towards the development of weather-dependent microsimulation models. Transp. Res. Rec. J. Transp. Res. Board 1–24.

Hutton, J.M., Bauer, K.M., Fees, C.A., Smiley, A., 2014. Analysis of Naturalistic Driving Study Data: Offset Left-Turn Lanes, SHRP2 REPORT S2-S08B-RW-1 Analysis. https://doi.org/10.17226/22315.

Insight Website, 2019. SHRP2 NDS Data Access [WWW Document]. URL https://insight.shrp2nds.us/ (Accessed 15 April 2019). .

Jalayer, M., Zhou, H., Satterfield, C., 2015. Overview of safety countermeasures for roadway departure crashes. Proc. 95th Transp. Res. Board Annu. Meet.

Jeong, H., Gan, G., Valdez, E., 2018. Association rules for understanding policyholder lapses. Risks 6 (3), 69. https://doi.org/10.3390/risks6030069.

Jeong, H., Liu, Y., 2017. Horizontal curve driving performance and safety affected by road geometry and lead vehicle. Proc. Hum. Factors Ergon. Soc. 2017 Annu. Meet. 33, 1629–1633. https://doi.org/10.1177/1541931213601893.

Khan, M.N., Ghasemzadeh, A., Ahmed, M.M., 2018. Investigating the impact of fog on freeway speed selection using the SHRP2 naturalistic driving study data. Transp. Res. Rec. J. Transp. Res. Board 2672 (16), 93–104.

Khan, M.N., Ahmed, M.M., 2019. Snow detection using in-vehicle video camera with texture-based image features utilizing k-nearest neighbor, support vector machine, and random. Forest. Transp. Res. Rec. J. Transp. Res. https://doi.org/10.1177/0361198119842105. Board 036119811984210.

Li, P., Merat, N., Zheng, Z., Markkula, G., Li, Y., Wang, Y., 2017. Does cognitive distraction improve or degrade lane keeping performance? Analysis of time-to-line crossing safety margins. Transp. Res. Part F Traffic Psychol. Behav. https://doi.org/10.1016/j.trf.2017.10.002. Press 2014.

López, G., Abellán, J., Montella, A., Oña De, J., 2014. Patterns of single-vehicle crashes on two-lane rural highways in Granada Province, Spain in-depth analysis through decision rules. Transp. Res. Rec. J. Transp. Res. Board 2432, 133–141. https://doi.org/10.3141/2432-16.

Mafi, S., Abdelrazig, Y., Doczy, R., 2018. Analysis of gap acceptance behavior for unprotected right and left turning maneuvers at signalized intersections using data mining methods: a driving simulation approach. Transp. Res. Rec. J. Transp. Res. Board 1–18. https://doi.org/10.1177/0361198118783111.

Meseguer, J.E., Calafate, C.T., Cano, J.C., Manzoni, P., 2013. Driving styles: a smartphone application to assess driver behavior. Proc. - Int. Symp. Comput. Commun. 535–540. https://doi.org/10.1109/ISCC.2013.6755001.

Mirabadi, A., Sharifian, S., 2010. Application of association rules in Iranian Railways (RAI) accident data analysis. Saf. Sci. 48 (10), 1427–1435. https://doi.org/10.1016/j.ssci.2010.06.006.

Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid. Anal. Prev. 43 (4), 1451–1463. https://doi.org/10.1016/j.aap.2011.02.023.

Montella, A., Ambrosio, A.D., 2011. Classification trees and association rules for exploratory analysis of powered two-wheeler crashes. Proc. 90th Transp. Res. Board Annu. Meet.

Montella, A., Aria, M., Ambrosio, A.D., Mauriello, F., 2012. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accid. Anal. Prev. 49, 58–72. https://doi.org/10.1016/j.aap.2011.04.025.

Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2011. Data-mining techniques for exploratory analysis of pedestrian crashes. Transp. Res. Rec. J. Transp. Res. Board 2237, 107–116. https://doi.org/10.3141/2237-12.

Mousa, S., Ishak, S., Osman, O.A., 2017. Extracting association rules from the SHRP 2 naturalistic driving data: a market basket analysis. 10th SHRP 2 Saf. Data Symp. From Anal. to Results October. pp. 11–12.

National Research Council, 2004. Where the Weather Meets the Road: A Research Agenda for Improving Road Weather Services, National Academy of Sciences. The National Academies Press, Washington, D.C., D.C. https://doi.org/10.17226/10893.

Pande, A., Abdel-Aty, M., 2008. Discovering indirect associations in crash data through probe attributes. Transp. Res. Rec. J. Transp. Res. Board 2083, 170–179. https://doi.org/10.3141/2083-20.

Peng, Y., Boyle, L.N., Hallmark, S.L., 2013. Driver's lane keeping ability with eyes off road: insights from a naturalistic study. Accid. Anal. Prev. 50, 628–634. https://doi.org/10.1016/j.aap.2012.06.013.

Saffarian, M., Happee, R., de Winter, J.C.F., 2012. Why do drivers maintain short headways in fog? A driving-simulator study evaluating feeling of risk and lateral control during automated and manual car following. Ergonomics 55 (9), 971–985. https://doi.org/10.1080/00140139.2012.691993.

Shepard, F.D., 1996. Reduced visibility due to fog on the highway. Natl. Coop. Highw. Res. Progr. 228.

Wu, C., Sun, C., Chu, D., Huang, Z., Ma, J., Li, H., 2016. Clustering of several typical behavioral characteristics of commercial vehicle drivers based on GPS data mining. Transp. Res. Rec. J. Transp. Res. Board 2581, 154–163. https://doi.org/10.3141/2581-18.

Wu, Y., Abdel-Aty, M., Cai, Q., Lee, J., Park, J., 2018. Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data. Transp. Res. Part C Emerg. Technol. 87 (February), 11–25. https://doi.org/10.1016/j.trc.2017.12.012.

Yamamoto, T., Kitamura, R., Fujii, J., 2002. Drivers' route choice behavior: analysis by data mining algorithms. Transp. Res. Rec. J. Transp. Res. Board 1807 (02), 59–66. https://doi.org/10.3141/1807-08.

Zhou, J., Peng, H., Gordon, T.J., 2008. Characterization of the lateral control performance by human drivers on highways. SAE Int. J. Passeng. Cars- Mech. Syst 2008-01–0561, 1–9.