

## Research Article

# Predicting the Collisions of Heavy Vehicle Drivers in Iran by Investigating the Effective Human Factors

**Hossein Naderi** , **Habibollah Nassiri** , and **Farnaz Zahedieh**

*Civil Engineering Department, Sharif University of Technology, Tehran, Iran*

Correspondence should be addressed to Habibollah Nassiri; [nassiri@sharif.edu](mailto:nassiri@sharif.edu)

Received 20 January 2021; Revised 8 June 2021; Accepted 13 September 2021; Published 6 October 2021

Academic Editor: Haneen Farah

Copyright © 2021 Hossein Naderi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traffic collisions are one of the most important challenges threatening the general health of the world. Iran's crash statistics demonstrate that approximately 16,500 people lose their lives every year due to road collisions. According to the traffic police of Iran, heavy vehicles (including trailers, trucks, and panel trucks) contributed to 20.5% of the fatal road traffic collisions in the year 2013. This highlights the need for devoting special attention to heavy vehicle drivers to further explore their driving characteristics. In this research, the effect of heavy vehicle drivers' behavior on at-fault collisions over three years has been investigated with an innovative approach of structural equation modeling (SEM) and Bayesian Network (BN). The database utilized in this research was collected using a questionnaire. For this purpose, 474 heavy vehicle drivers have been questioned in the Parviz Khan Border Market, located on the border of Iran and Iraq. The response rate of the survey was 80%. The participants answered the questions on Driver Behavior Questionnaire (DBQ) and a sleep assessing questionnaire named Global Dissatisfaction with Sleep (GSD). In this research, human factors affecting at-fault collisions of heavy vehicles were identified and their relationships with other variables were determined using the SEM approach. Then the descriptive model constructed by the SEM method was used as the basis of the BN, and the conditional probabilities of each node in the BN were calculated by the database collected by the field survey. SEM indicates that other attributes including GSD, mobile usage, daily fatigue, exposure, and education level have an indirect relation with heavy vehicle drivers' at-fault collisions. According to the BN, if there is no information about the characteristics of a heavy vehicle driver, the driver will likely have at least one collision during the next three years with the probability of 0.17. Also, it was indicated that the minimum probability of the at-fault collision occurrence for a heavy vehicle is 0.08.

## 1. Introduction

Traffic collisions have become one of the most important challenges threatening the general health of the world [1]. According to the World Health Organization's (WHO) report, about 1.24 million people were killed on the roads' collision in the year 2010, which is approximately 3,400 people per day [2]. In a global status report on road safety 2018, the World Health Organization has predicted that 16,426 people had died due to traffic collisions in Iran in the year 2016 [3]. On the other hand, heavy vehicles (such as trailers, trucks, and panel trucks) constitute 8.3% of the total number of vehicles in Iran [4], while these vehicles were present in 20.5% of the fatal road traffic collisions in 2013.

Given the high contribution of heavy vehicles to fatal road traffic collisions and the much smaller number of these vehicles compared to the total number of vehicles in Iran, the collision risk of this type of vehicle is way higher than that of cars that has led to a much higher drop in this category of cars than on riding cars, indicating the necessity of conducting a separate study on this category of drivers.

One of the first steps to study road traffic collisions is the recognition of factors affecting the occurrence of these collisions. Researchers divide the factors affecting the collisions into four general categories:

- (1) Human factors
- (2) Environmental factors

(3) Vehicle factors

(4) Road factors

Studies have demonstrated that human factors were effective in the occurrence of 93% of collisions, while environmental factors and vehicle factors were found to be effective in 34% and 13%, respectively [1]. This highlights the importance of the role of the human factors in the occurrence of traffic collisions and the need for special attention to this factor.

This research identifies the effective human factors in at-fault collisions of heavy vehicle drivers with an innovative approach of the structural equation modeling (SEM) and Bayesian Network (BN). For this purpose, SEM was used to explore the relationships between the effective human factors on collisions.

The innovations of this study can be divided into the following categories: the construction of the BN requires cause and effect relationships among the variables effective in the study. Although in most studies conducted on the traffic collisions using the BN, the BN was usually based on the researcher's expert opinion [5]; in this study, first, the variables affecting the incidence of traffic collisions and the way they influence the traffic collision were identified using the SEM method, then this model was employed to define the BN.

The SEM model is only a descriptive model and has no prediction capability, but with the integration of SEM with the BN, this method gains the prediction capability. Considering the widespread application of SEM in behavioral sciences, the integration of these two methods would be extremely appealing for the prediction of human behavior.

Finally, the goal of this research can be divided into the following parts: first, with the help of the SEM, human factors affecting the incidence of at-fault collisions of heavy vehicle drivers were identified and then determined how they affect the collision occurrence. In the next step, the relationships discovered by the SEM were used in the construction of the BN and then the probabilities used in the BN were calculated according to the collected data.

## 2. Literature Review

*2.1. Literature Review on Human Factors Affecting the Incidence of Traffic Collisions.* Researchers have demonstrated that driving behavior is one of the most important factors affecting the crash risk factor. So far, numerous questionnaires have been developed to study driving behavior, but the most popular one is the Driver Behavior Questionnaire (DBQ). This questionnaire is a tool for identifying and categorizing aberrant driving behavior. Aberrant driving behavior is divided into four general categories: driving slips (lapses), errors, ordinary violations, and aggressive violations [1]. In this research, we will identify which of these factors is effective in the incidence of at-fault collisions.

Researchers have indicated that inappropriate driving behavior is one of the most important factors influencing the Crash Risk Factor [1]. Inappropriate driving behavior can be subdivided into three general categories which are slips

(lapses), errors, and driving violations. Driving slips are generally memory-related problems, insufficiencies, and neglects, for instance when a driver enters an intersection or the fourth lane incorrectly. Errors in driving are mistakes that result from a misjudgment and an inability to see, such as when a driver is turning right and suddenly encounters a cyclist or a motorcyclist with a high risk of impact. Driving violations include failing to comply with the necessary laws that are the absolute requirements of safe driving.

In the studies that have been carried out on the variable of inappropriate driving so far, the effect of this variable on the number of road collisions has provided different results. For example, in a study that focused on elderly drivers, only the driving errors factor had a meaningful correlation with the total number of collisions [6], but in another study that focused on bus drivers, the driving violations variable had a meaningful correlation with the total number of crashes [7]. Also, other studies showed differences among work-related variables between urban bus drivers and bus rapid transit (BRT) drivers that illuminate the importance of studying each segment of drivers separately [8]. In some studies, only one of the items on the questionnaire was related to the total number of collisions. For example, in a study on the driving behavior of truck drivers, it was indicated that the increase in the number of high driving speed violations increased the chance of traffic collisions [9]. In a study conducted on the general Iranian drivers' population, it was noted that out of the three inappropriate driving factors including the slips, errors, and driving violations, only the driving offenses factor was associated with the total number of driving incidents [10]. In another study, which introduced a new classification of violations including risky violations and highway violations, there was a direct relationship between highway violations and driving errors and the total number of traffic incidents [11]. Tabibi's study on predicting drivers' inappropriate behaviors concluded that the factors of driving errors and driving offenses are directly related to the number of driving collisions reported by the interviewees. Also, the driving error factor has a much stronger relationship with the total number of collisions compared to the driving violation factor [12]. But in most studies conducted in countries other than Iran, the variable of driving offenses is more associated with the total number of collisions compared to the driving error variable [13]. The reason behind this difference may be the high tendency of Iranian drivers towards high-risk driving, such as chasing other cars with very small spacing in between, and in these situations, driving errors can be the most important cause of traffic collisions [12].

Other important human factors affecting the occurrence of collisions are drivers' sleep quality and daily fatigue. These variables have great importance for drivers of heavy vehicles owing to their continuous driving. However, the official statistics related to car collisions that are caused by drowsiness are often not collected [14]. The few statistics that consider drowsiness as a factor that affects the total number of driving collisions show a very low rate (1% to 3% of the total collisions), while self-proclaimed reports by drivers often indicate a very high percentage of drowsy

driving and collisions caused by driver drowsiness [15]. The results of a study that focused on the effect of drowsiness while driving on traffic collisions in the United Kingdom showed that 29% of male drivers had the experience of a nap while driving during the past 12 months. Also, 7% of drivers who have had a driving collision within the past three years indicated tiredness as the main reason behind the incident [16].

The Case Studies of the National Transportation Safety Board on collisions found that collisions related to drowsiness and sleeps are extremely widespread among truck drivers. In 1995, the National Transportation Safety Board conducted a study on single-vehicle collisions (in which the drivers survived) and found that 58% of the collisions were caused by fatigue and 19 out of the 107 drivers (17%) interviewed in the study reported that they had been asleep at the time of the collision [17]. Therefore, this factor has also been considered as one of the important factors in this study.

Drivers' use of mobile phones is becoming more common [18]. So far, many studies have been carried out to identify the effect of mobile phone use on safe driving. Brown et al. found that using mobile phones while driving reduces speed and increases the number of judgmental errors made by drivers [19]. Other studies have indicated that using mobile phones while driving reduces drivers' performance [20]. Therefore, as an important and effective factor, the use of mobile phones has also been considered in this research.

Socioeconomic characteristics of drivers such as gender, age, urbanization, education, occupation, working hours, marital status, and nationality have been measured almost in all of the human factor studies. The age variable had a decreasing effect in all the studies that assessed the relationship of this variable with the total number of collisions in a certain period [9, 10, 21, 22]. In another study conducted on bus drivers in Iran, the increase of the age indicator had no significant relationship with drivers' collisions [11]. In one study, three reasons for the increased likelihood of a collision involving younger drivers have been presented. First, young drivers drive more miles per year. Second, they generally drive faster than older drivers, and finally, those young drivers commit more hostile driving violations than older drivers [9, 23].

Also, in some of the conducted studies, the variable of educational status is considered as a feature that determines a person's social stance [10, 22]. The results of the study conducted by Moghaddam and Ayati in 2014 showed that the higher a driver's education level is, the lower the number of their collisions becomes [10]. But the results of Factor's study showed no correlation between the educational level and the total number of a driver's collisions [22].

In summary, six effective variables have been identified by studying human factors influencing traffic collisions. These variables include driving behavior, daily fatigue, sleep quality and health, mobile phone usage, socioeconomic characteristics, and driving experience. In this study, these variables have been used to explore the relationship between human factor variables and traffic collisions.

*2.2. Literature Review on Structural Equation Modeling (SEM).* Utilizing structural equation modeling in the field of transportation safety does not have a prolonged history. This section reviews some of the studies carried out in this field. In a study conducted on elderly drivers, Lucidi et al. used structural equation modeling. They assessed the relationship between personal characteristics variables, habits related to traffic safety, and dangerous driving (violations, errors, slips, and the total number of driving tickets received in the past 12 months). Structural equation modeling results showed that a driver's characteristics can, directly and indirectly, predict the dangerous driving variable [6].

In another study that was conducted using the structural equation modeling approach, the relationship between the risk of collisions and the inappropriate driving habits of 301 Italian bus drivers was investigated. This study showed that there is only a relationship between the variable of traffic offenses and the risk of collisions, meaning that the more traffic offenses bus drivers have, the greater their risk of collisions [7].

Eboli and Mazzulla also used the structural equation modeling to analyze traffic collisions in Italy [24]. They considered the severity of a traffic collision as a hidden variable, measured by two variables that are the number of casualties and the total number of vehicles involved in the collision. In the model developed in this study, the dependent crash severity variable was affected by three independent variables that are the environmental condition's factor, the path specifications factor, and the drivers' specifications factor. In the final model, the total number of casualties is defined as the most important indicator of collision severity. Also, it was found that the road classification variable is the most effective variable representing path characteristics, and the climate conditions variable is the most effective environmental factor that indirectly affects the severity of traffic collisions [25].

*2.3. Literature Review on Bayesian Network (BN).* Numerous scholars have investigated the modeling of drivers' behavior. These studies have devoted most of their emphasis on the typical driving conditions. There are two modeling routines to follow: the performance model and the cognitive model [26]. The latter approach represents and defines the different inherent processes of driving behavior. The techniques applied in this approach are the Adaptive Control of Thought—Rational (ACT-R) and the multiagent systems. However, the first approach concentrates on the driver's actions reproduction for a special task during a certain condition [26]. The major techniques applied for the same objective include neural networks, transfer systems, fuzzy logic, Markov chains, etc. [26]. The outputs of the model are categorized as samples, and the target is the anticipation of the most suitable class output for each of the drivers. The technique satisfying these requirements is Bayesian networks in architecture with an augmented naïve nature.

Chen et al. in a comprehensive study developed a hybrid approach for combining BN approaches and multinomial

logit models to analyze the severities of drivers' injuries in rear-end crashes based on crash data compiled in New Mexico for the 2010-2011 period. To identify and investigate the major factors contributing to the severities of rear-end crash driver injury, a multinomial logit model is developed [27]. The identified major factors are then used to develop a BN for formulating statistical relations between injury severity implications and explanatory features, such as demographic specifications, driver's behavior, environmental and geometric features, and vehicle factors.

Both efficiency and safety are typically deemed as two major performance indices of transport systems. The planning of road networks has concentrated on transportation efficiency and road capacity; however, a road network safety level has not received much attention in the planning phase. Another research presents a joint model of Bayesian hierarchical nature for evaluation of road network safety to assist planners to take the safety of traffic into account at the time of road network planning [28]. The presented model develops relations between the risk of the road network and microlevel variables associated with traffic volume and road entities, and also trip generation, network density, and socioeconomic variables commonly utilized for long-term transportation schemes [29].

As the application of in-vehicle information systems (IVISs), including navigation systems and mobile phones, is continually increased, driver distraction has become a major safety concern. An approach that allows individuals to benefit from in-vehicle information systems is creating adaptive in-vehicle information systems adjusting their operations per driver and roadway state [29]. A crucial element in adaptive in-vehicle information systems is the real-time monitoring of driver distraction. Such a monitoring function makes it possible to reduce that distraction. Bayesian networks were used in this study to establish a real-time procedure to detect cognitive distractions through driving performance and drivers' eye movements [29]. In another study, a Bayesian regression procedure has been taken for developing travel time prediction equations for central region streets through intuitive contributory variables [30].

### 3. Materials and Methods

In this part of the report, the overall trend of the modeling has been studied. First, general information about the database is provided. In the next step, the method of SEM construction is described and the quality of the SEM is evaluated. Then, the methods of BN construction and probability tables' preparation are discussed.

**3.1. Data Collection.** In this study, standardized questionnaires such as Driver Behavior Questionnaire (DBQ) and Global Dissatisfaction with Sleep (GSD) were used to collect data [1, 31]. First, the English versions of these questionnaires were translated into Persian. Then, a pilot survey was conducted involving a few heavy vehicle drivers, and the ambiguities of the questionnaires were modified. Eventually,

the final questionnaire language was checked out by a language specialist for resolving possible inconsistencies between the English and translated versions. The English version of the questionnaire used in this study could be found in the Supplementary Materials (available here).

Data gathering and interviewing were carried out in Parviz Khan Border Market, Qasr-e Shirin County, Kermanshah Province in Iran. The Parviz Khan Border Market is located on the zero-point border between Iran and Iraq. A total of 474 truck drivers participated in this survey. All the participants were male (there is less than 5 female heavy vehicle drivers in Iran because of cultural issues). The average age of the participants was 44.1 (SD = 9.8). About 20% of the participants had a college education and 88% of them were married. Data gathering took over 20 days, from March 12, 2016, to April 8, 2016. It is worth noting that the data collection was stopped from March 17, 2016, until March 24, 2016, because of the closure of the border market. Truck drivers completed the questionnaire in the parking lot of the border market from 10 A.M. to 5 P.M. The truck drivers were informed that participating in this survey is completely voluntary and would remain anonymous. Before making the modeling for constructing a database, all data were preprocessed.

**3.2. Structural Equation Modeling (SEM).** SEM as a generalized statistical method has had a wide range of applications in behavioral sciences, especially in sociology and economics since the 1970s. After that, the application of this method has expanded to other sciences including psychology, political science, and educational sciences. All of the multivariate methods, including multiple regression and principal components, provide researchers with tools to examine a wide range of theoretical problems. But all of these methods have a common weakness: each technique at a time can only examine a single correlation relationship. Despite other multivariate analysis methods, SEM simultaneously investigates a series of interdependent relationships that can be considered as a set of multiple regression equations [32].

An SEM model is composed of two structural and measurement patterns. The structural pattern, a set of dependency relationships, interconnects structures that exist in the pattern, and the measurement pattern, a part of the overall model, identifies the measurement indicators of each latent variable. A latent variable cannot be measured directly, but rather must be measured by several other variables (indicators) [32]. For instance, the number of errors in the driving of a person cannot be verified, but rather there must be a variety of questions to identify the various aspects of driving errors and measured the theory using them. Figure 1 depicts the structural and measurement patterns. In Figure 1, circles represent latent variables and rectangles are questionnaire questions.

**3.3. Bayesian Network.** BN has been introduced for more than three decades. BN is a directed acyclic graph (DAG) that shows a set of random variables and their independent

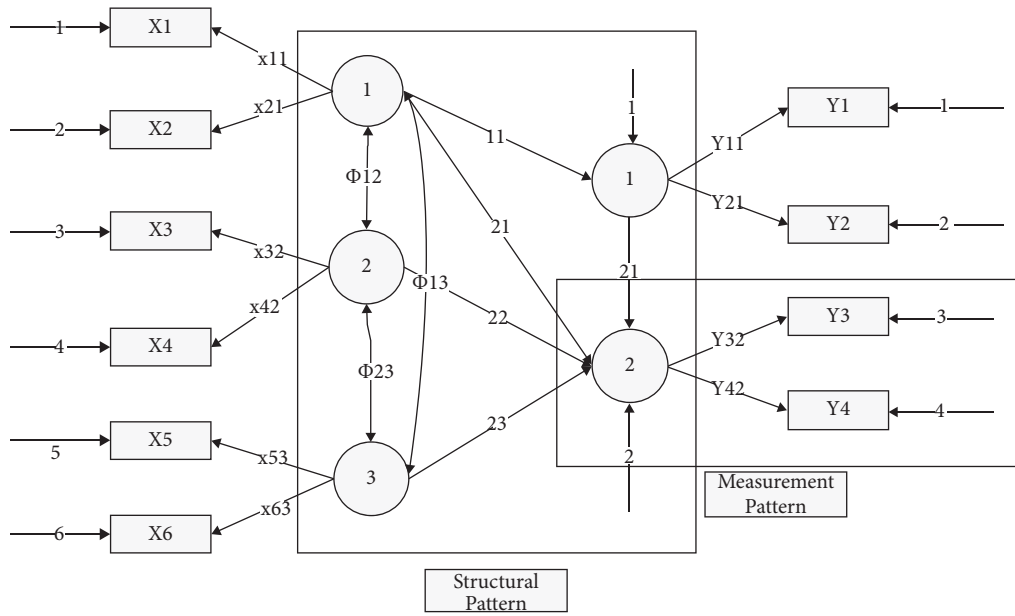


FIGURE 1: Structural and measurement patterns in SEM.

connections. BN is formed by a set of variables  $V = \{x_1 \cdot x_2 \cdot \dots \cdot x_9\}$  provided that  $\vartheta > 1$ .

The DAG is used to represent meaningful relationships between variables. A set of probability tables in the form of  $B_p = \{p(x_i | pr(x_i)) \cdot x_i \in V\}$  exists to show the occurrence probability of each node. In this relationship,  $pr(x_i)$  is the set of parent variables of  $x_i$  in  $B_p$ . The probability of each of the variables displayed in the nodes is calculated using equations (1) and (2) [27]:

$$P(Y|X) = \frac{P(Y \cdot X)}{P(X)} = \frac{P(V)}{P(X)}, \quad (1)$$

$$P(Y|X) = \prod_{x_i \in V} p(x_i | pr(x_i)). \quad (2)$$

To construct the BN, the specified relationships between data must first be determined, and then the probability table for each node is defined. For this purpose, Rapidminer software was employed to calculate the number of observations of each variable. This software is extremely useful in database and data mining studies. GeNIe 2.1 software was used for BN calculations. GeNIe software has a user-friendly graphical interface developed by the University of Pittsburgh. This software is available to students for free.

In response to questionnaires, people tend to choose some options of the questionnaires more, and the answers are somewhat biased. For example, over 85% of the answers to the questions in the error section of the driving behavior questionnaire are among the three options of never, rarely, and occasionally. One of the advantages of the BN approach is to consider this topic for future predictions. In the next section, a more detailed explanation will be provided.

To calculate the conditional probability on the BN, since each variable and indicator are asked by a few questions, the mean of the indicators is calculated and rounded up to the

higher integer. For example, for the error variable, if the mean of responses is equal to 1.1, the driving error value is considered on a scale of 2.

## 4. Results

**4.1. Results from Structural Equation Modeling (SEM).** For SEM, SmartPLS (v 3.2.7) software was employed [33]. Over the past two years, more than 1,000 scientific articles have been published with the help of this software. The pro version of the software is simply available for free for 30 days. The constructed SEM model is shown in Figure 2. In the constructed model, among driving behavior variables, only the variable error had a meaningful relationship with the number of traffic collisions, and the remaining cases were not identified as a variable that could describe traffic collisions, and, as shown in Figure 2, they indirectly affected at-fault collisions. The study of the quality and accuracy of the SEM model has many and varied indicators evaluating each part of the model. For instance, by studying the path coefficients and their statistical significance ( $t$  values), as shown in Figure 2, the significance of the structural pattern is examined.

To examine the quality of the SEM model, more than 7 indicators that examine the reliability of the indicators, the reliability of the model's integrity, the model's convergence validity, the model's segmental validity, and so on have been examined.

Figure 2 illustrates the ultimate result of structural equation modeling. In this figure, only the statistically meaningful paths have been drawn. There are two numbers written on each path. The first represents the influence coefficient of the related endogenous variable that is used for the prediction of the exogenous variable which the path leads to it. The second is equal to the  $t$  value. Therefore, how larger the first number is means that the alteration of its related

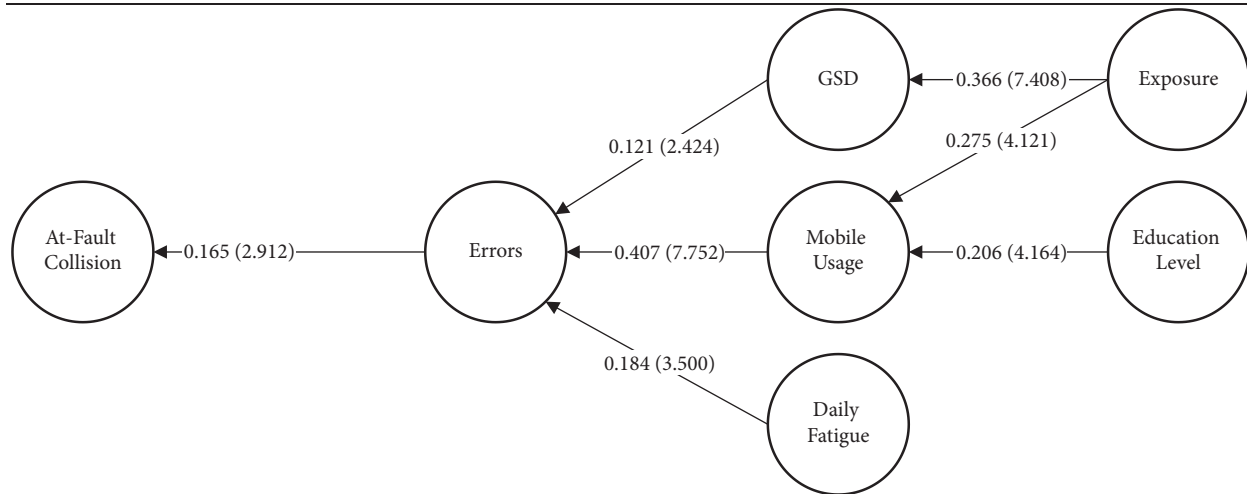


FIGURE 2: SEM model for the relationship between driver characteristics and at-fault collisions over 3 years; the numbers shown are path coefficients and the values in parentheses are their statistical significance ( $t$  value).

endogenous variable has more influence on the exogenous variable of the end of the path. At the significance level of 0.01, all path coefficients are meaningful (for more details refer to Figure 3).

The SEM results showed that only the “error” factor, among the other factors of DBQ, has a statistically meaningful relationship with at-fault collision directly.

The results related to factor loadings of indexes are shown in Table 1. Fifty-eight percent of items have a factor loading of more than 0.7, which is the preferred value for the final survey of indexes. Moreover, the factor loadings of more than 42% of items are more than 0.4, which is the acceptable level for the final survey of indexes. Since education and the item that represents the level variable of GSD is one question (The GSD question is, “How do you rate your sleep in general?” Respondents are to choose the answer from these items: “I sleep well,” “Occasionally I do not sleep well, but I am generally satisfied with my sleep,” “My sleep has already caused me problems,” “I think I have a problem with my sleep,” “I sleep badly,” “I sleep very badly” [34]), the value of its factor loading is equal to one that is demonstrated in Table 1. Therefore, according to these results, the reliability of each of the observable variables was confirmed.

To confirm the reliability of the reflective measurement model, in addition to confirming the reliability of each of the observable variables, the composite reliability (CR) and average variance extracted (AVE) were extracted. Table 1 shows the calculated CR and AVE of each construct. The minimum acceptable values for the CR and AVE are 0.7 and 0.5 [34–37].

Then, the Fornell-Larker test was used to assess the discriminant validity of the measurement model. The matrix shown in Table 2 is the correlation matrix of the latent variables of the model, except that the numbers on its diameter have been replaced by the square root of the AVE of the corresponding latent variable. According to the Fornell-Larker test, discriminant validity is confirmed if the numbers on the diameter are greater than the correlation values of

that column [38]. Results show that this criterion is also met in this model.

In structural equation modeling through PLS, there are four main criteria for structural model testing that include 1—coefficient of determination ( $R^2$ ), 2—the significance of path coefficients, 3— $Q^2$  value, and 4—Cohen effect size criterion ( $f^2$ ) [39].

The values of  $R^2$  equal to 0.19, 0.33, and 0.67 are described as weak, medium, and significant, respectively. Significance of path coefficients means confirming the assumptions of the structural equation model.  $Q^2$ -criteria above zero indicate that the observed values are well reconstructed.  $Q^2$  equals 0.02, 0.15, and 0.35 are the weak, medium, and strong values for this index, respectively [36]. The Cohen criterion also shows the intensity of the relationship between the hidden variables of the model. Cohen introduced three values of 0.02, 0.15, and 0.35 for weak, medium, and strong effects, respectively [40].

According to the results of SEM in Tables 3 and 4, all the criteria for testing the overall fitness of the model are met.

**4.2. Results from Bayesian Network (BN).** In this research, BN was used to answer the question of how likely heavy car drivers in Iran were likely to be involved in a collision. The problem we were faced within the first step was how to determine the relationships between the variables affecting the incidence of collisions. The relationship was determined using SEM in the previous section. In the next step, the probability of occurrence of each variable for the conditional probability associated with that variable with its parent node was obtained using

$$P(Y|X) = \frac{P(Y \cdot X)}{P(X)} = \frac{n(Y \cdot X)}{n(X)}. \quad (3)$$

According to equation (3), the probability of  $Y$  occurrence given that the event  $X$  has already occurred is equal to the number of observations that both  $X$  and  $Y$  variables

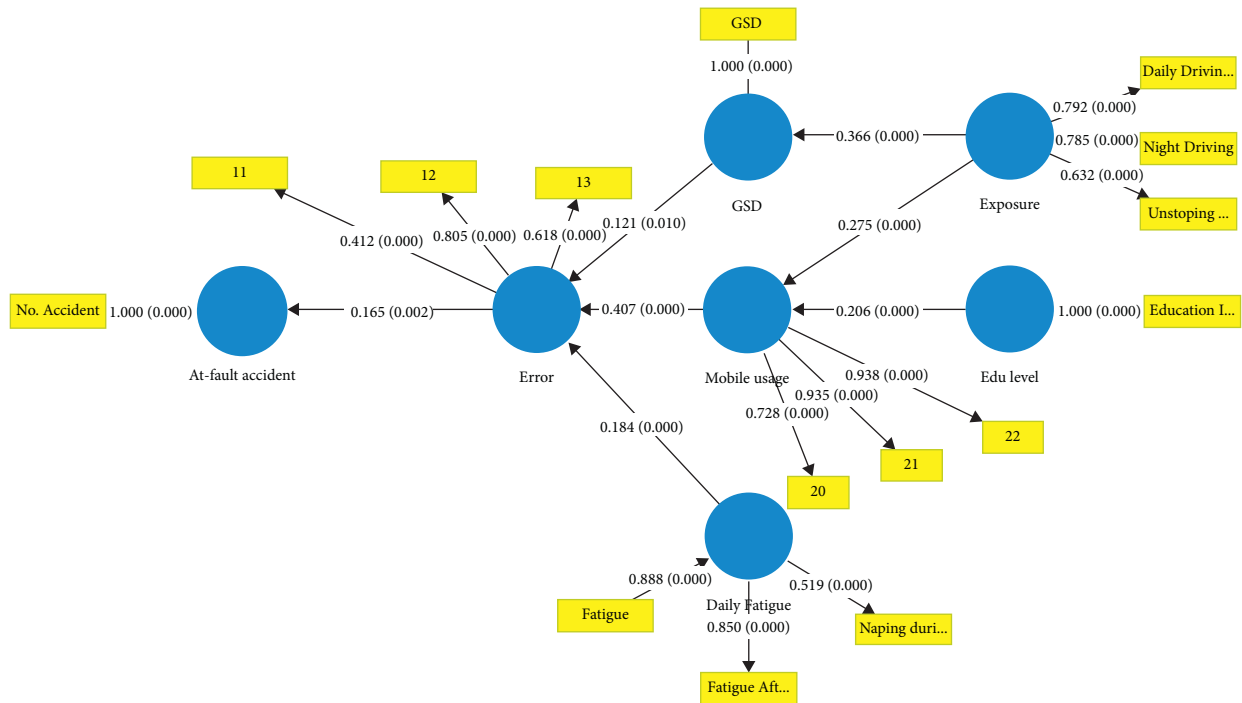


FIGURE 3: Full diagram with path coefficients of SEM.

TABLE 1: The factor loadings of latent variable indexes and their statistical significance level.

Latent variable	Index	Factor loading	t value	P value	CR	AVE
At-fault collision	In the last three years, in how many collisions have you been found guilty as a driver?	1.000	—	—	1.000	1.000
GSD	How do you rate your sleep in general?	1.000	—	—	0.807	1.000
Education level	What is your level of education?	1.000	—	—	1.000	1.000
Mobile usage	Do you talk through the cellular phone while driving?	0.728	18.901	0.001	0.904	0.761
	Do you read SMS while driving?	0.935	82.270	0.001		
	Do you send SMS while driving?	0.938	97.037	0.001		
Exposure	Driving hours in a day	0.792	17.609	0.001	0.782	0.547
	Driving duration from midnight to 6 am	0.785	17.852	0.001		
	Hours of restless nonstop driving	0.632	6.940	0.001		
Daily fatigue	Level of fatigue in a day	0.888	32.275	0.001	0.807	0.594
	Level of boredom after awakening	0.850	25.108	0.001		
	Level of doze while driving	0.519	4.656	0.035		
Errors	Attempt to overtake and had not noticed signaling left	0.412	3.346	0.003	0.652 ~ 0.7	0.400
	When turning right have nearly hit a cyclist on the inside	0.805	12.114	0.001		
	Missed give way signs and avoided colliding with traffic	0.618	8.800	0.001		

TABLE 2: Fornell-Larker test results.

	At-fault accident	Daily fatigue	Education level	Error	Exposure	GSD	Mobile usage
At-fault accident	1.000	*	*	*	*	*	*
Daily fatigue	0.035	0.770	*	*	*	*	*
Education level	-0.056	0.076	1.000	*	*	*	*
Error	0.165	0.369	0.103	0.632	*	*	*
Exposure	0.031	0.346	-0.136	0.221	0.74	*	*
GSD	0.033	0.415	-0.032	0.289	0.366	1.000	*
Mobile usage	0.126	0.33	0.168	0.495	0.247	0.225	0.873

TABLE 3:  $R^2$  and  $Q^2$  results.

	$R^2$	$Q^2$
At-fault accident	0.027	0.019
Error	0.304	0.107
GSD	0.134	0.123
Mobile usage	0.102	0.069

TABLE 4:  $f^2$  results and path coefficients.

	$f^2$	Path coefficient	t value ( $p$ value)
Daily fatigue $\rightarrow$ error	0.037	0.184	3.500 (0.000)
Education level $\rightarrow$ mobile usage	0.046	0.206	4.164 (0.000)
Error $\rightarrow$ at-fault accident	0.028	0.165	2.912 (0.002)
Exposure $\rightarrow$ GSD	0.155	0.366	7.408 (0.000)
Exposure $\rightarrow$ mobile usage	0.083	0.275	4.121 (0.000)
GSD $\rightarrow$ error	0.017	0.121	2.424 (0.012)
Mobile usage $\rightarrow$ error	0.210	0.407	7.752 (0.000)

occur together, divided by the number of observations of  $X$ . For example,

$$\begin{aligned}
 &P(\text{Accident} = \text{"Involved"} | \text{Error} = \text{"Never"}) \\
 &= \frac{n(\text{Accident} = \text{"Involved"} \cdot \text{Error} = \text{"Never"})}{n(\text{Error} = \text{"Never"})} \quad (4) \\
 &= 0.082.
 \end{aligned}$$

The Rapidminer software was used to calculate the number of these cases and then they were introduced to the GeNIe software as the probability tables of the BN. For example, in Table 5, the probability of the collision occurrence for a heavy vehicle driver over the past three years has been set for different categories of driving errors. For all the nodes shown in Figure 2, conditional probability was introduced to the software.

In the previous, the BN construction process was fully described. In this section, the results obtained by the BN are discussed.

Figure 4 demonstrates the results of the database on the BN for the time when we do not have any information about the driver's characteristics. According to Figure 4, if there is no information on the driver's characteristics, there is a probability of 17% that any heavy vehicle driver may have had an at-fault collision in the past three years, which can be used to predict the future. Meaning that in the next three years, there is a probability of 17% that any heavy vehicle driver may have an at-fault collision. The probability distributions of other variables are also depicted in Figure 4 for the case that there is no other information.

Now, if any information is somehow obtained about any of these variables affecting the incidence of a collision by the SEM model of Figure 2, and the uncertainty of that variable is eliminated, the probability pertinent to the occurrence of the at-fault collision and other variables can be updated. For example, suppose a driver was asked the three questions regarding the amount of mobile phone usage, and the "often" level was specified. In this case, the probabilities of

the BN were updated, and the probability of an at-fault collision of this driver over three years became 24%.

By examining different probability states and identifying different variables as the observed variable, the minimum and maximum probabilities pertinent to an at-fault collision of a driver were calculated according to the BN. The minimum probability of the at-fault collision occurrence for a heavy vehicle driver is when the result of the variable "error" of the driving behavior test is "Never". In this case, the probability of an at-fault collision for a heavy vehicle driver over 3 years is 8%. Also, the maximum probability of the at-fault collision occurrence for a heavy vehicle driver is when the result of the variable "error" of the driving behavior test is identified as "Often"; in this case, the probability of an at-fault collision for a driver is equal to 60% over three years. Table 6 presents the probabilities of the at-fault collision occurrence for heavy vehicle drivers for three years considering all variables except one.

One of the disadvantages of the variable "error" in driving behavior test is that, depending on the nature of the questions, drivers may not answer these questions with full integrity; however, variables such as exposure, education level, mobile phone usage, daily fatigue, and sleep satisfaction criterion can be verified and are easier to evaluate. Furthermore, drivers are less inclined to express values such as "Always" and "Repeatedly" for some offenses, and thus, having several variables of this type and the probabilities related to them can help to predict probabilities pertinent to the at-fault collision occurrence of a driver more accurately.

## 5. Discussion

The results of the SEM have been compared to some similar studies across the world in Table 7. Although plenty of researches showed that driving violations have a direct effect on collision numbers [7, 8, 10–12], but studies such as this research and Lucidi et al. showed that there is no significant relationship between driving violations and collision numbers [6]. However, research by Lucidi et al. showed that



TABLE 5: The probability of an at-fault collision for heavy vehicle drivers under the condition of error.

At-fault Collisions	Errors					
	never	rarely	Occasionally	often	repeatedly	Always
Involved	8.20%	9.80%	20.80%	60.00%	50.00%	50.00%
not involved	91.80%	90.20%	79.20%	40.00%	50.00%	50.00%

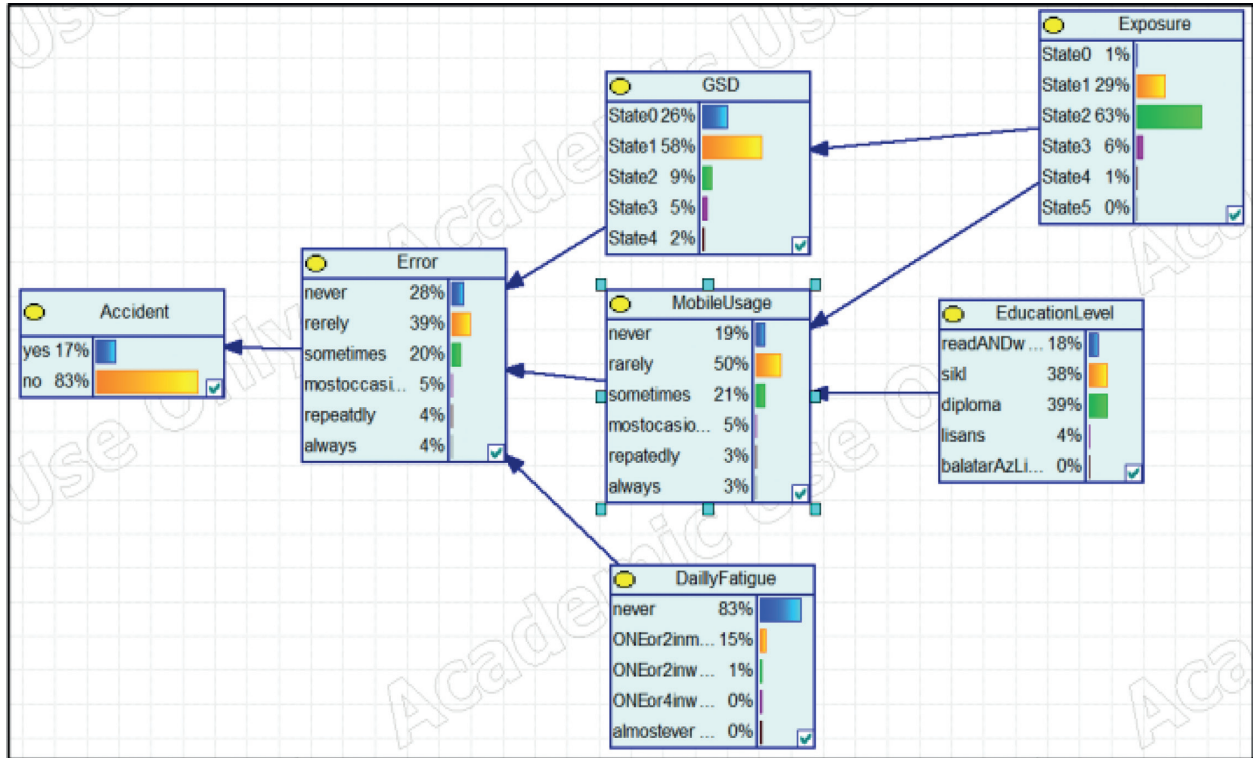


FIGURE 4: The distribution of conditional probabilities of the BN for the database studied in this research.

TABLE 6: The probability of the at-fault collision occurrence of heavy vehicle drivers for each variable.

Variable	Probability of At-Fault Collision Occurrence (Percentage)					
	Never	1-2 times per month	3-4 times per month	1-2 times per week	3-4 times per week	Always
Daily Fatigue	17%	20%	33%	33%	33%	33%
Mobile Usage While Driving	Never	Rarely	Occasionally	Often	Repeatedly	Always
	16%	16%	20%	24%	18%	20%
Exposure	Stage Zero (low)	Stage One	Stage Two	Stage Three	Stage Four	Stage Five
	20%	18%	16%	25%	25%	22%
Sleep Dissatisfaction	Stage Zero (I am satisfied with my sleep)		Stage One	Stage Two	stage Three	Stage Four
	25%		11%	23%	32%	30%

TABLE 7: Comparison of the SEM result with previous researches [6–8, 10–12, 15, 21, 23, 41, 42].

Researcher	Study territory	Number of participants	Type of drivers	Mobile usage	Sleep dissatisfaction	Exposure	Error	Slips	Driving violations
Lucidi et al. 2014	Italy	485	Old drivers	—	—	—	↑	○	○
Mallia et al. 2014	Italy	301	Bus drivers	—	—	—	○	○	↑
Sullman et al. 2002	New Zealand	378	Truck drivers	—	—	○	○	○	↑
Moghaddam and Ayati 2014	Iran	1667	General drivers	—	—	—	○	○	↑
Varmazyar et al. 2013	Iran	628	Bus drivers	—	—	↑	↑	○	↑
Tabibi et al. 2015	Iran	107	General drivers	—	—	○	↑	○	↑
Qu et al. 2015	China	317	General drivers	—	—	—	—	—	—
Radun et al. 2015	Finland	1121	General drivers	—	↑	—	—	—	—
Chen 2007	Taiwan	194	General drivers	↑	—	—	—	—	—
Cantor et al. 2010	USA	560695	Truck drivers	—	—	↑	—	—	—
Shams et al. 2020	Iran	780	Truck drivers	—	↑	○	—	—	—
<b>Current study</b>	<b>Iran</b>	<b>474</b>	<b>Truck drivers</b>	<b>▲</b>	<b>▲</b>	<b>▲</b>	<b>↑</b>	<b>○</b>	<b>○</b>

The “—” sign shows that the variable had not been studied. The “↑” sign means that this variable had an incremental effect on collision numbers and the “▲” means this variable had an incremental effect on collision numbers through other variables. The “↓” sign means that this variable had a decreasing effect on collision numbers and finally the “○” means that this variable has no meaningful effect on collision numbers.

the variable of violations is strongly related to the number of driving fines received by drivers.

The results of most studies on the slip variable showed that this variable has a statistically significant relationship with the collision number [6–8, 10–12]. The result obtained in this study indicates that the structural equation model provides results consistent with the results of previous studies.

The effect of the error on the number of traffic collisions is one of the most controversial subjects in previous studies. Research results in this field can be divided into two categories. In the first category, research studies such as Mallia et al., Sullman et al., and Moghaddam and Ayati showed that this variable did not have a significant effect on the number of drivers’ collisions [7, 8, 10]. But in the second category, studies such as Lucidi et al., Varmazyar et al., and Tabibi et al., showed that this variable has a positive and increasing effect on the number of drivers’ collisions [6, 11, 12]. The results of the present study confirmed the second category studies results. The results of the structural equation model showed that among all the variables studied in this research, only the error variable is directly related to the rate of traffic collisions.

According to the results of some studies that have examined the effect of exposure, the more the drivers are exposed, the higher the number of their collisions [11, 41]. Other studies show that there is no significant relationship between exposure and the number of collisions [8, 12, 42]. The results of the present study showed that the increase in exposure is not directly related to the rate of collisions, but it

leads to an increase in collisions through mediators such as mobile usage, GSD, and error.

According to Radun et al. and Shams et al. studies [15, 42], drivers’ dissatisfaction with their sleep and its poor quality have led to an increase in the number of accidents. Unlike the results of these studies, no significant direct relationship between sleep quality and health with collision was discovered in this research. But the results of SEM showed that the GSD was directly related to the error. Therefore, it can be said that the more the driver is dissatisfied with his sleep, the more the error increases and the higher the probability of an accident.

Also, the results of the current study showed that the more a driver uses a cell phone while driving, the more likely he/she is to crash. This result is compatible with other studies such as Chen study [23].

The results of this research show that among all the variables studied in this study, only the error variable directly affects the number of traffic collisions. On the other hand, the error variable is influenced by the latent variables including mobile usage, the amount of fatigue and boredom in the day (daily fatigue), and sleep dissatisfaction (GSD). This means that the more the driver uses the mobile phone while driving, the more tired and drowsy he feels, the more he will make a driving error and be exposed to a collision.

The variable of mobile usage is affected by two latent variables of education level and exposure, and the higher the value of these two variables, the more mobile phone uses by the driver. It is natural that the higher the level of literacy of

the driver, the more it is possible for him/her to use mobile applications, send text messages and read text messages. Also, the longer the driver is driving, the more exposed he/she is to answering incoming calls or making calls, sending and reading text messages.

## 6. Conclusions

In this study, to construct the BN of the relationship between the characteristics of heavy vehicle drivers and their at-fault collisions, first, a model depicting the relationships between human factors affecting at-fault collisions was required. For this purpose, the main structure of the BN was developed using the SEM. Then, the conditional probability table for the developed BN was calculated with the help of the Rapidminer software and introduced to the GeNIe software used for Bayesian calculations. The following results were obtained with the help of the BN:

- (i) In this research, an innovative method of using the SEM in the construction of BN was employed to study the behavioral sciences of humans.
- (ii) If there is no information about the characteristics of a heavy vehicle driver, the probability of an at-fault collision over the next 3 years for this driver is 17%.
- (iii) According to the SEM model developed in this study, only the “driving error” factor directly affects the incidence of collisions involving heavy vehicles.
- (iv) Factors “slip,” “ordinary violations,” and “aggressive violations” were not recognized effective in the incidence of at-fault collisions of heavy vehicle drivers.

Sleep dissatisfaction is one of the most important factors in the collision occurrence of heavy vehicles, which can increase the chance of a collision occurrence to 32%. A way to reduce the at-fault collision occurrence of heavy vehicle drivers is to increase the sleep satisfaction of drivers by setting daily driving hours limits for heavy vehicle drivers.

Eventually, with the help of the obtained network, the likelihood of the at-fault collision occurrence involving a heavy vehicle driver can be calculated using a certain amount of information, including sleep quality, daily fatigue, and the amount of mobile phone usage.

It is expected that this model will help insurance companies in the calculation of annual third-party insurance rates, traffic police in the renewal of licenses, freight shipping companies in hiring drivers, and legislators in macro policies to enforce working hour's rules or driving experience to make better decisions.

The BN developed in this study can help freight shipping companies to make decisions on hiring drivers, insurance companies to determine the annual third-party insurance rates in proportion to the risk of at-fault collision occurrence for each driver, and traffic police to decide on the renewal of heavy vehicle drivers' licenses or to require them to take part in retraining courses or to impose heavier fines for mobile phone usage while driving.

## Appendix

Full diagram with path coefficients of SEM:

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors received no financial support for the research, authorship, and/or publication of this article.

## Supplementary Materials

The questionnaire used in this study has been uploaded as supplementary. (*Supplementary Materials*)

## References

- [1] R. Rowe, G. D. Roman, F. P. McKenna, E. Barker, and D. Poulter, “Measuring errors and violations on the road: a bifactor modeling approach to the driver behavior questionnaire,” *Accident Analysis & Prevention*, vol. 74, pp. 118–125, 2015.
- [2] World Health Organization, “Supporting a decade of action,” *Global Status Report on Road Safety*, World Health Organization, Geneva, Switzerland, 2013.
- [3] World Health Organization, *Global Status Report on Road Safety 2018*, World Health Organization, Geneva, Switzerland, 2018.
- [4] H. Naderi, H. Nassiri, and S. Sahebi, “Assessing the relationship between heavy vehicle driver sleep problems and confirmed driver behavior measurement tools in Iran,” *Transportation Research Part F*, vol. 59, pp. 57–66, 2018.
- [5] J. de Oña, R. O. Mujalli, and F. J. Calvo, “Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks,” *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 402–411, 2011.
- [6] F. Lucidi, L. Mallia, L. Lazuras, and C. Violani, “Personality and attitudes as predictors of risky driving among older drivers,” *Accident Analysis & Prevention*, vol. 72, pp. 318–324, 2014.
- [7] L. Mallia, L. Lazuras, C. Violani, and F. Lucidi, “Crash risk and aberrant driving behaviors among bus drivers: the role of personality and attitudes towards traffic safety,” *Accident Analysis & Prevention*, vol. 79, pp. 145–151, 2014.
- [8] M. J. M. Sullman, M. L. Meadows, and K. B. Pajo, “Aberrant driving behaviours amongst New Zealand truck drivers,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 5, no. 3, pp. 217–232, 2002.
- [9] S. Useche, B. Cendales, F. Alonso, and A. Serge, “Comparing job stress, burnout, health and traffic crashes of urban bus and BRT drivers,” *American Journal of Applied Psychology*, vol. 5, no. 1, pp. 25–32, 2017.

- [10] A. M. Moghaddam and E. Ayati, "Introducing a risk estimation index for drivers: a case of Iran," *Safety Science*, vol. 62, pp. 90–97, 2014.
- [11] S. Varmazyar, S. B. Mortazavi, S. Arghami, and E. Hajizadeh, "Factor analysis of driver behavior questionnaire (DBQ) in public transportation bus company: investigation of the relationship between DBQ factors and crashes," *Scientific Journal of Review*, vol. 3, pp. 155–165, 2014.
- [12] Z. Tabibi, H. H. Borzabadi, D. Stavrinou, and A. Mashhadi, "Predicting aberrant driving behaviour: the role of executive function," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 34, pp. 18–28, 2015.
- [13] J. C. F. De Winter and D. Dodou, "The driver behaviour questionnaire as a predictor of accidents: a meta-analysis," *Journal of Safety Research*, vol. 41, no. 6, pp. 463–470, 2010.
- [14] J. Horne and L. Reyner, "Vehicle accidents related to sleep: a review," *Occupational and Environmental Medicine*, vol. 56, no. 5, pp. 289–294, 1999.
- [15] I. Radun, J. Radun, M. Wahde, C. N. Watling, and G. Kecklund, "Self-reported circumstances and consequences of driving while sleepy," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 32, pp. 91–100, 2015.
- [16] G. Maycock, "Sleepiness and driving: the experience of U.K. car drivers," *Accident Analysis & Prevention*, vol. 29, no. 4, pp. 453–462, 1997.
- [17] National Transportation Safety Board, *Factors that Affect Fatigue in Heavy Truck Accidents*, Vol. 1-2, National Transportation Safety Board, Washington, DC, USA, 1995.
- [18] D. Haigney and S. J. Westerman, "Mobile (cellular) phone use and driving: a critical review of research methodology," *Ergonomics*, vol. 44, no. 2, pp. 132–143, 2001.
- [19] I. D. Brown, A. H. Tickner, and D. C. Simmonds, "Interference between concurrent tasks of driving and telephoning," *Journal of Applied Psychology*, vol. 53, no. 5, pp. 419–424, 1969.
- [20] K. A. Brookhuis, G. de Vries, and D. de Waard, "The effects of mobile telephoning on driving performance," *Accident Analysis & Prevention*, vol. 23, no. 4, pp. 309–316, 1991.
- [21] W. Qu, Y. Ge, Q. Zhang, W. Zhao, and K. Zhang, "Assessing dangerous driving behavior during driving inattention: psychometric adaptation and validation of the attention-related driving errors scale in China," *Accident Analysis & Prevention*, vol. 80, pp. 172–177, 2015.
- [22] R. Factor, "The effect of traffic tickets on road traffic crashes," *Accident Analysis & Prevention*, vol. 64, pp. 86–91, 2014.
- [23] D. E. Cantor, T. M. Corsi, C. M. Grimm, and K. Özpolat, "A driver focused truck crash prediction model," *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 5, pp. 683–692, 2010.
- [24] L. Eboli and G. Mazzulla, *A Structural Equation Model for Road Accident Analysis*, 4th International SHV Congress, Palermo, Italy, 2007.
- [25] M.-Y. Sohrabi and Siamak, "A study on gender differences of affecting factors on intensity of driving accidents," Master of Science Thesis, Sharif University of Technology, Tehran, Iran, 2011.
- [26] W. Bouslimi, M. Kasugai, D. Lourdeaux, and P. Fuchs, "Augmented naive Bayesian network for driver behavior modeling," in *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium*, Las Vegas, NV, USA, June 2005.
- [27] C. Chen, G. Zhang, R. Tarefder, J. Ma, H. Wei, and H. Guan, "A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes," *Accident Analysis & Prevention*, vol. 80, pp. 76–88, 2015.
- [28] J. Wang and H. Huang, "Road network safety evaluation using Bayesian hierarchical joint model," *Accident Analysis & Prevention*, vol. 90, pp. 152–158, 2016.
- [29] Y. Liang, J. D. Lee, and M. L. Reyes, "Nonintrusive detection of driver cognitive distraction in real-time using Bayesian networks," *Journal of the Transportation Research Board*, vol. 2018, pp. 1–8, 2018.
- [30] L. A. Frechette and A. M. Khan, "Bayesian regression-based urban traffic models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1644, no. 1, pp. 157–165, 1998.
- [31] M. M. Ohayon and J. Zulley, "Correlates of global sleep dissatisfaction in the German population," *Sleep*, vol. 24, no. 7, pp. 780–787, 2001.
- [32] R. O. Mujalli and J. de Oña, "A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks," *Journal of Safety Research*, vol. 42, no. 5, pp. 317–326, 2011.
- [33] C. M. Ringle, S. Wende, and J. M. Becker, *SmartPLS 3: Boenningstede*, SmartPLS GmbH, Kiel, Germany, 2015.
- [34] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.
- [35] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: indeed a silver bullet," *Journal of Marketing Theory and Practice*, vol. 19, no. 2, pp. 139–152, 2011.
- [36] J. Henseler, C. M. Ringle, and R. R. Sinkovics, "The use of partial least squares path modeling in international marketing," *Advances in International Marketing*, vol. 20, pp. 227–320, 2009.
- [37] Y. Ye, S. C. Wong, F. Meng, and P. Xu, "Right-looking habit and maladaptation of pedestrians in areas with unfamiliar driving rules," *Accident Analysis & Prevention*, vol. 150, Article ID 105921, 2021.
- [38] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research*, vol. 18, no. 1, pp. 39–50, 1981.
- [39] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," *European Business Review*, vol. 31, no. 1, 2019.
- [40] J. Cohen, "A power primer," *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992.
- [41] Y.-L. Chen, "Driver personality characteristics related to self-reported accident involvement and mobile phone use while driving," *Safety Science*, vol. 45, no. 8, pp. 823–831, 2007.
- [42] Z. Shams, H. Naderi, and H. Nassiri, "Assessing the effect of inattention-related error and anger in driving on road accidents among Iranian heavy vehicle drivers," *IATSS Research*, vol. 45, no. 2, pp. 210–217, 2020.