Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 4857013, 29 pages
https://doi.org/10.1155/2022/4857013

Research Article

Presentation of Machine Learning Approaches for Predicting the Severity of Accidents to Propose the Safety Solutions on Rural Roads

Mohammad Habibzadeh, Mahmoud Ameri, Hassan Ziari, Neda Kamboozia, and Seyede Mojde Sadat Haghighi

School of Civil Engineering, Iran University of Science & Technology (IUST), Tehran, Iran

Correspondence should be addressed to Mohammad Habibzadeh; habibzadeh1991@gmail.com

Received 31 March 2022; Revised 7 August 2022; Accepted 21 August 2022; Published 14 September 2022

Academic Editor: Yanyong Guo

The aim of the current research was to develop models to predict the severity of accidents on rural roads in Tehran province, Iran. In this regard, using accident data from 2017 to 2020, the machine learning algorithms, including multiple logistic regression, multilayer perceptron neural network (MLPNN), and radial basis function neural network (RBFNN) models, as well as statistical methods, including Kolmogorov–Smirnov test, Friedman test, and factor analysis, were implemented to determine the contributory factors in the severity of accidents. Thus, nine variables affecting the severity of accidents were considered in modeling, and then the effect of each variable was calculated. By comparing the results of artificial neural network (ANN) models and the Friedman test, it was indicated that the human factor had a remarkable effect on accident severity. In addition, both machine learning and statistical methods can be served as guidance for safety authorities to provide safety solutions, thereby leading to reducing accidents. Finally, the performances of ANN models were analyzed by other mathematical models built by MATLAB programming.

1. Introduction

Road traffic accidents have been considered one of the most important issues among other topics in today's lives and are regarded as the 8th leading cause of death in the world [1]. Based on the statistics, every year, 1.35 million citizens are killed in highway accidents, and 20–50 million are injured around the world due to traffic accidents [2]. The safety of rural roads is related to the road users, which is the result of each person's thinking and approach. Therefore, there is an urgent need to identify and analyze these issues that can be led to the understanding of traffic behaviors [3].

The traffic system is considered one of the most complex and dangerous systems that people should be faced with it in society, and today, the danger of traffic accidents made by humans threatens the lives of people more than other factors such as pathogenicity [4–8]. Due to the high number of traffic accidents and imposing costs on society, it is important to evaluate and understand the variables, which affect the accident occurrence, helping governments and safety authorities to take preventive and corrective measures [9–11]. With respect to the high number of accidents on the rural roads of Tehran province, Iran, it is necessary to evaluate the contributory factors to the severity of accidents, including property damage only (PDO) and fatal/injury, which can lead to finding solutions in reducing accidents and increasing safety.

Modeling accident severity in terms of their significant variables makes it possible to predict accident occurrence that requires relief equipment. In this regard, different statistical analyses and machine learning methods can be employed. Moreover, using these methods, the influence of each variable in intensifying traffic accidents can be evaluated. This approach will lead to the possibility of formulating traffic safety plans for traffic engineers, and they may also have a better understanding of the parameters having a negative or positive effect on accident severity. This study examined the risk factors in the severity of vehicle accidents on Tehran rural roads by the use of various methods to detect more precisely the factors that affect the severity of accidents. Results provide insight with respect to the relationship between the various risk factors and the severity of accidents for improving traffic safety on these roads. The original idea of this study was to develop optimum models that have a robust and accurate capability to calculate the influence of each independent variable affecting the severity of accidents.

1.1. Literature Review. A lot of efforts have been made using machine learning methods to identify different contributing factors in accident occurrence. In this section, some of the research carried out in some countries are explained.

In a study conducted by Abdelwahab and Abdel-Aty, the performances of multilayer perceptron neural network (MLPNN) and radial basis function neural network (RBFNN) models were compared with calibrated logit models by the use of road accident data. The results of this research indicated that the RBFNN model was more robust in analyzing the severity of driver injuries. In addition, based on the RBFNN model, the likelihood of older drivers involved in injury accidents was more than younger drivers, and female drivers accounted for a higher percentage of injured accidents compared to males [12]. Xie et al. found out that the ANN model is more strong in predicting travel mode compared to other alternatives such as decision tree (DT) and multinomial logit (MNL) models [13]. Kim et al. developed the MNL model to predict the variables affecting motor vehicle accidents. They focused on four categories of collisions, which included fatal, incapacitating, nonincapacitating, possible, and no injury. The study represented that the variables of the truck vehicle, consuming alcohol by driver or cyclist, high speed, inclement weather, the age group related over 55 years old for the cyclist, and head-on collisions increased the probability of injury accidents, leading to fatal [14].

Boodlal et al. using the two MNL models in predicting the probability of total and fatal/injury accidents evaluated the combination effects of lane and shoulder widths on rural roads in Illinois and Minnesota, USA. The results of this research indicated that the geometry features had no effect on the severity of accidents and only affected their occurrence [15]. Omrani utilized four types of machine learning methods, which included RBFNN, MLPNN, support vector machines (SVM), and MNL models, to assess the prediction of the individuals' travel mode in the city of Luxembourg. The results illustrated that the performance of ANN models was better [16]. Huang et al. aimed at investigating the prediction of accidents and identifying risk parameters important to accident occurrence by the use of the RBFNN model. By comparing the results of backpropagation neural network (BPNN), RBFNN, and negative binomial (NB) models, it was indicated that the RBFNN model had better performance [17]. Alkedher et al. applied the MLPNN model to predict the severity of accidents divided into three classes using the k-means algorithm to improve the prediction accuracy. The clustering of the accidents data led to a substantial enhancement in the accuracy of predicting accidents [18].

Behbahani et al. compared four types of artificial neural network (ANN) methods, including MLPNN, RBFNN, probabilistic neural network (PNN), and extreme learning machine (ELM). The results of this study showed that the ELM technique was recognized to be the fastest algorithm, and also, this method was very accurate in predicting accidents. In addition, the RBFNN algorithm identified the most effective factors in the prediction of accidents [19]. Rezapour et al. developed and then compared a single-layer perceptron neural network (SLPNN) with a recurrent neural network to predict the frequency and severity of motorbike accidents. The SLPNN was employed to investigate accident research as a distance-based pattern matching technique to detect the correct road segment [20]. Najafi Moghaddam Gilani et al. (2021) provided a study to identify the influential variables on vehicle accident occurrence. They designed machine learning methods such as ANN and logistic regression models to recognize the variables affecting the severity of accidents in two types of injury/fatal and PDO in Rasht city, Iran. The results indicated that the variables of accidents time (18:00 to 24:00) and KIA Pride vehicle were found as the most important variables in accident occurrence. Moreover, the performance of the ANN model was more capable of predicting the severity of accidents in comparison with logistic regression in terms of accuracy and efficacy [21].

Thus, the results of these previous studies have indicated that using machine learning methods can be a proper approach for understanding the contributory factors and provide better predictive accuracy for accident occurrence. So, in this study, the Kolmogorov–Smirnov test was used to examine the normality of data. Then, the Friedman test and factor analysis were applied to prioritize the variables and detect the underlying variables, respectively. The machine learning approach was then employed to recognize the contributory factors in accident occurrence. In this method, the multiple logistic regression, multilayer perceptron neural network (MLPNN), and radial basis function neural network (RBFNN) models were built to detect the importance of each variable in the severity of accidents. By comparing these techniques, the most effective variables in the severity of accidents were identified, and the best model with a higher accuracy was chosen for accidents occurring in Tehran province rural roads, Iran. While various studies have used machine learning approaches to predict the severity of accidents, limited studies have studied to measure the effectiveness of education rate, type of collision, type of vehicles, and other variables. To address this gap, in this study, we considered 9 variables split into 64 variables, which have been rarely considered in previous studies, providing safety authorities with practical safety approaches. In this study, we try to make an effort to answer these research questions: what are the most influential variables that affect accident severity on rural roads in Iran? What are the optimum models to predict accident severity? What is the percentage of importance of each independent variable in predicting accident severity? This helps recognize gaps and enhance public safety toward promoting the overall highway safety situation of rural roads in Iran.

2. Methodology

In the current study, various statistical methods have been examined and then compared to determine the factors that affect accident severity, each of which has its own unique characteristics and can be applied based on its functions. Comparing the results of the different methods can enable the researcher to adopt suitable measures to reduce accident occurrence. In the first step, the K-S test was utilized to examine the data distribution normality and select the appropriate tests. After that, the Friedman test and factor analysis were used to determine the priority of factors and the underlying variables, respectively. In the last section, machine learning approaches such as logit regression, MLPNN, and RBFNN models were considered in order to recognize and measure the importance of independent variables in accident severity. Finally, comparing the statistical methods and the machine learning approach provides suitable solutions for reducing accidents and increasing safety on Tehran province rural roads.

2.1. Data Collection. The current research was conducted in the Tehran metropolis, which is the capital of the Tehran province, Iran. Tehran is the most populous metropolis in the country and western Asia and includes 22 districts spread across 730 square kilometers and is located at 51°17′ to 51°33′ E and 35°36′ to 35°44′ N and 900 to 1,800 meters above sea level. Based on the official census in 2016, the population of this city was 8,693,706, of which 4,369,551 were female, and the population density of females was recorded as 5,986 per square. The influential parameters on increasing the likelihood of accidents might be different from city to city; due to high traffic interferences with other flows in several parts of this city and the congested traffic flow on ring roads [22], a separate research is needed to be done in the Tehran metropolis. Because of provided information and numerous traffic interferences and a very dense urban texture between passing vehicles and nonmotorized users, there is a clear need to recognize the most effective parameters for the severity of accidents and create a model with higher accuracy for future accidents. As it is mentioned earlier, many efforts have aimed at evaluating the number and severity of accidents, which considered the particular condition of accidents, which included accidents with unknown severity, high severity accidents, or the severity of accidents that may be expected to happen sometime in the future. In this research, however, due to a balanced number of datasets in terms of three classes of fatal, injury, and damage accidents, we analyzed and developed prediction models using machine learning methods, including logit, ANN models, and statistical methods. To conduct the research, on road safety improvements, accident data, including parameters such as accident time, environmental characteristics, human characteristics, and type of collision should be precisely gathered. The information related to 36-month was obtained from visiting Tehran Traffic Police Statistic Center. The data used in this study were gathered over three periods of 2017 to 2020, which included accident severity, accident time, accident day, reason of accident, gender of the driver, collision type, driver age, vehicle type, education rate, and weather condition. In total, 2,585 accidents data were collected from police accident reports of Tehran province, of which 2,013 cases (77.9%) were damage accidents, and 572 cases (22.1%) were fatal and injury ones. The dependent variable in this research was accident severity, divided into three classes: damage, fatal, and injury accidents. Because the amount of fatal accidents was small compared to whole accidents, the goodness of fit and the significance of the models cannot be satisfied regarding the three types of dependent variables; injury accidents were combined with fatal accidents; and the dependent variable was classified into two groups of damage and fatal/injury accidents.

2.1.1. Data Description. Table 1 represents the variables affecting accident severity in Tehran province as well as the suitable coding set for the variables in developing models.

2.2. Statistical Methods

2.2.1. Kolmogorov–Smirnov (K-S) Test. The Kolmogorov–Smirnov test is utilized to recognize the normality of samples [23]. The comparison of the result of the K-S test called sig. with a critical value of 0.05 determines the normal distribution of the data. If the sig. amount (p value) is greater than the critical value (0.05) for a fixed significance level, the null hypothesis of normality rejects, indicating that the normality assumption of the desired distribution is rejected [24].

2.2.2. Friedman Test. The Friedman test is a nonparametric test utilized to detect the difference between related data. This test is equal to the parametric two-factor analysis of variance. The null hypothesis of Friedman's two-way analysis of variance based on ranks describes that the K repeated measures or matched categories came from populations with the same median or the same population. The Friedman test is an analysis of variance based on ranks, which means that rank values or observed rank values are obtained by numerical outcomes or ordering ordinal, and is applied when there is no motivation to provide robust distributional assumptions [25].

2.2.3. Factor Analysis (FA). The FA is an important statistical technique applied in the modeling process the covariation between a set of data. When there are many number variables and there are no established relationships among them, the FA is utilized to recognize the contributory

Table 1: Description and coding of each variable used in the study.

Variable	Variable levels
Accident severity	(1) Damage (2) Injury/fatal
Reason of accident	(1) Not paying attention to the front (2) Failure to observe the longitudinal distance (3) Failure to observe the transverse distance (4) Fatigue and drowsiness (5) Exceeding the lawful speed (6) Failure to yield the right-of-way (7) Violating the left lane law (8) Deviating the left lane (9) Traveling in opposite directions (10) Defections of pavement (11) Turning in a forbidden place (12) Changing lanes dramatically (13) Inability to control the vehicle (14) Not paying attention to travel with reverse gear (15) Technical defect in the vehicle
Gender of the driver	(1) Male (2) Female
Type of collision	(1) Single vehicle (2) Vehicle-vehicle (3) Vehicle-pedestrian (4) Vehicle-motorcycle (5) Motorcycle-motorcycle (6) Motorcycle-pedestrian (7) Overturning vehicle (8) Overturning motorcycle (9) Vehicle-object
Time of accident	(1) 00:00 to 06:00 (2) 06:00 to 12:00 (3) 12:00 to 18:00 (4) 18:00 to 24:00
Day of accident	(1) Start of the week (2) Middle of the week (3) Weekend
Age	(1) Less or than equal to 18 (2) 18 to 30 (3) 30 to 45 (4) 45 to 60 (5) 60 and over
Vehicle	(1) Motorcycle (2) KIA pride (3) Samand (4) Peugeot (5) Paykan (6) Taxi (7) Pickup (8) Truck (9) Trailer (10) Bus (11) Minibus (12) Other vehicles

Table 1: Continued.

Variable	Variable levels
Education rate	(1) Illiterate (2) Elementary (3) High school (4) Diploma (5) Associate degree (6) Bachelor's (7) Master's (8) PhD
Weather conditions	(1) Clear (2) Cloudy (3) Rainy (4) Snowy (5) Frost (6) Stormy

factors. Using the FA leads to recognizing the most effective parameters in forming the phenomena and declining the dimensions of factors observed. In general, the FA is split into two types: confirmatory factor analysis (CFA) and exploratory factor analysis (EFA). The EFA is utilized to recognize the hidden structures once the structure of the relationships between the parameters is unknown and to develop hypotheses about their possible structures, and in the CFA method, the determination of parameters is performed by the use of dimensions [26]. In the current research, EFA was used.

2.3. Machine Learning Approach. Machine learning methods such as ANN and logistic regression models have the ability to solve traffic accident problems, including the nonlinear relationships between input variables as well as presenting reliable and fast techniques for designing models. In addition, in developing these models, there is no need to consider the statistical distribution of the data and knowing knowledge about the relationships between the variables applied in the modeling.

2.3.1. Multiple Logistic Regression. The multiple logistic regression model is one of the most applicable and common models used to find the relationships between contributory factors and the dependent variable. This model can represent a closed form in suggesting the possibility of selecting choices interpreted. In order to create this model, two fundamental assumptions must be provided. The first assumption is that each independent variable is unique, which means that a unit value is for each variable. Second, the dependent variable is not entirely predictable from the independent variables. The independent assumption of variables is not required in this model [27, 28]. Logistic regression is another type of logit model employed for dichotomous dependent variables. This model is applied to predict accidents based on continuous and/or categorical independent variables. Since the occurrence of the dependent variable is binary (0 and 1), the logistic binary can be applied. The chances are calculated as follows [29]:

$$\text{odds} = \frac{P_i}{1 - P_i}, \tag{1}$$

where P_i is the probability of an event and $1 - P_i$ is the probability of the lack of an event.

Logarithm or logit related to success's chance is achieved by performing the logarithm of equation (2) as follows:

$$\text{Logit}\left(P_i\right) = \text{Log}\frac{P_i}{1 - P_i}. \tag{2}$$

The reverse transfer function, called antilogistics, is applied in computing likelihood in terms of logistics from [29] as follows:

$$\text{Logit}\left(Z_i\right)^{-1} = \frac{e^{z_i}}{1 - e^{z_i}}. \tag{3}$$

2.3.2. Neural Network. ANN models have been recognized as one of the most common and applicable techniques of artificial intelligence (AI) in predicting the number and severity of accidents [30, 31]. Based on the recent studies carried out by some researchers on the safety of roads, neural network models had an acceptable level of accuracy in predicting road accidents over other techniques such as zero-inflated negative binomial, zero-inflated Poisson, negative binomial, and Poisson regression [32]. They are promising tools for evaluating complicated data and computing complex nonlinear problems. Because the internal structure of ANN models that do not give insights into the causal relationships between inputs and outputs, they are regarded as a black-box method [33–35]. Because of their accuracy and efficiency in recognizing the importance and amount of factors that affect accident severity, ANN models were considered a reliable tool in modeling techniques. In the present study, developing and comparing two ANN models, which include multilayer perceptron neural network (MLPNN) and radial basis function neural network (RBFNN), were considered.

(1) MLPNN. The MLPNN has been one of the most common types of feedforward backpropagation neural networks that each neuron in one layer has directed connections to the

neurons of the subsequent layer and the gradient of the loss function is computed with respect to the weights of the network for a single input-output example [36, 37], which have been used to predict the occurrence of traffic accidents. The framework of the MLPNN is as same as the single-layer perceptron neural network (SLPNN) with more hidden layers. At least three-node layers that include input, hidden, and output layers have usually been considered in the MLP model. In general, a supervised learning method called backpropagation is applied to train the model. In order for the model validation, the softmax activation function (AF) at the output layer and cross-entropy as the error function were used. The outputs are computed at the output layer as follows [38]:

$$y(x) = W_0 + \sum_{j=1}^{n} W_j f\left(W_{j0} + \sum_{i=1}^{1} \lambda_{ij} X_i \right), \tag{4}$$

where $W = (W_{0,\ldots,}W_n)$ and $\lambda = (\lambda_{10}, \ldots, \lambda_{nI})$ illustrate the weights in the second and first layers, respectively, and f is the AF. The hyperbolic tangent was used as an AF in the hidden layer as follows [38]:

$$f\left(x_j\right) = \tanh\left(x_j\right) = \frac{e^{-x_j} - e^{-x_j}}{e^{-x_j} + e^{x_j}}. \tag{5}$$

Softmax was also applied as an AF in the output layer as follows:

$$f\left(x_j\right) = \frac{e^{x_j}}{\sum e^{x_k}}. \tag{6}$$

(2) RBFNN. The structure of the RBFNN model is similar to the structure of the MLPNN; however, the major difference between them is that the hidden layer of the RBFNN model includes nodes called RBF units. Two main parameters in the RBFNN models illustrate the position of its width or deviation and the function center. The hidden layer of the RBFNN model calculates the distance between the center of its RBF and an input data vector. The RBF has its peak by the zero distance between the input data vector and the RBF center and reducing progressively by increasing the distance; the output of the network is built as follows [39]:

$$y_j(x) = \sum_{i=1}^{m} w_{ji} f_i(x) + w_0, \tag{7}$$

where m is the number of hidden nodes, $x = (x_1 = \text{vehicle type}, x_2 = \text{driver age}, \ldots, x_m = \text{collision type})$ is the input data, w_{ji} illustrates the weight related to the i_{th} hidden node to the j_{th} output node, w_0 is the bias value, and f_i is the basis function of unit j. The Gaussian is one of the most common types of RBFNN [39].

$$f_i(x) = e^{\left(-\left|x - c_i\right|^2/2\sigma_i\right)}, \tag{8}$$

where x is the input data, c_j is the center for neuron j, and r is the spread of the basis function. A normalized RBF (NRBF) network is one of the Gaussian RBFNN types, using the softmax function; accordingly, the activations are normalized in the hidden units until the sum of them gets the value of one. An NRBF network with equal heights and unequal widths is written as follows [39]:

$$f_i(x)(\text{softmax}) = \frac{e^{h_i}}{\sum_{i=1}^{m} e^{h_i}}, \tag{9}$$

$$h_i = \left(-\sum_{k=1}^{2} \frac{\left(x_k - c_{ik}\right)^2}{2\sigma_i^2} \right), \tag{10}$$

where x_k is the input data (damage, and injury/fatal), c_{ik} is the center of the i-th hidden node related to the k_{th} $(k = 1, 2)$ input data, and softmax function (h_i) is the output vector of the i-th hidden node.

2.4. Performance Evaluation. To compare the predictive performance of the developed models, five criteria of root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and average absolute deviation (MAD) were employed. The RMSE, MSE, MAE, MAPE, and MAD indicate the differences between predicted and actual values. In comparing, the lower values of RMSE, MSE, MAE, MAPE, and MAD, the better performance of the model. Equations (11)–(15) define the RMSE, MSE, MAE, MAPE, and MAD, respectively. To evaluate the performance of models with different architectures, we calculated each criterion for models to determine which one reaches the greatest accuracy [40–48].

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(P_i - A_i\right)^2}{N}}, \tag{11}$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{n} \left(P_i - A_i\right)^2, \tag{12}$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{n} \left|P_i - A_i\right|, \tag{13}$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{n} \frac{\left|\left(A_i - P_i\right)^2\right|}{A_i}, \tag{14}$$

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^{n} \left|P_i - A_i\right|, \tag{15}$$

where N is the amount of data for the training set and P, A, and $\overline{A_i}$ are actual values, predicted values, and the average of actual values, respectively.

3. Results

3.1. Frequency Analysis

3.1.1. The Evaluation of Accident Year. As shown in Figure 1(a), vehicle accidents had an overall upward trend. The highest percentage of accidents happened in 2019–2020

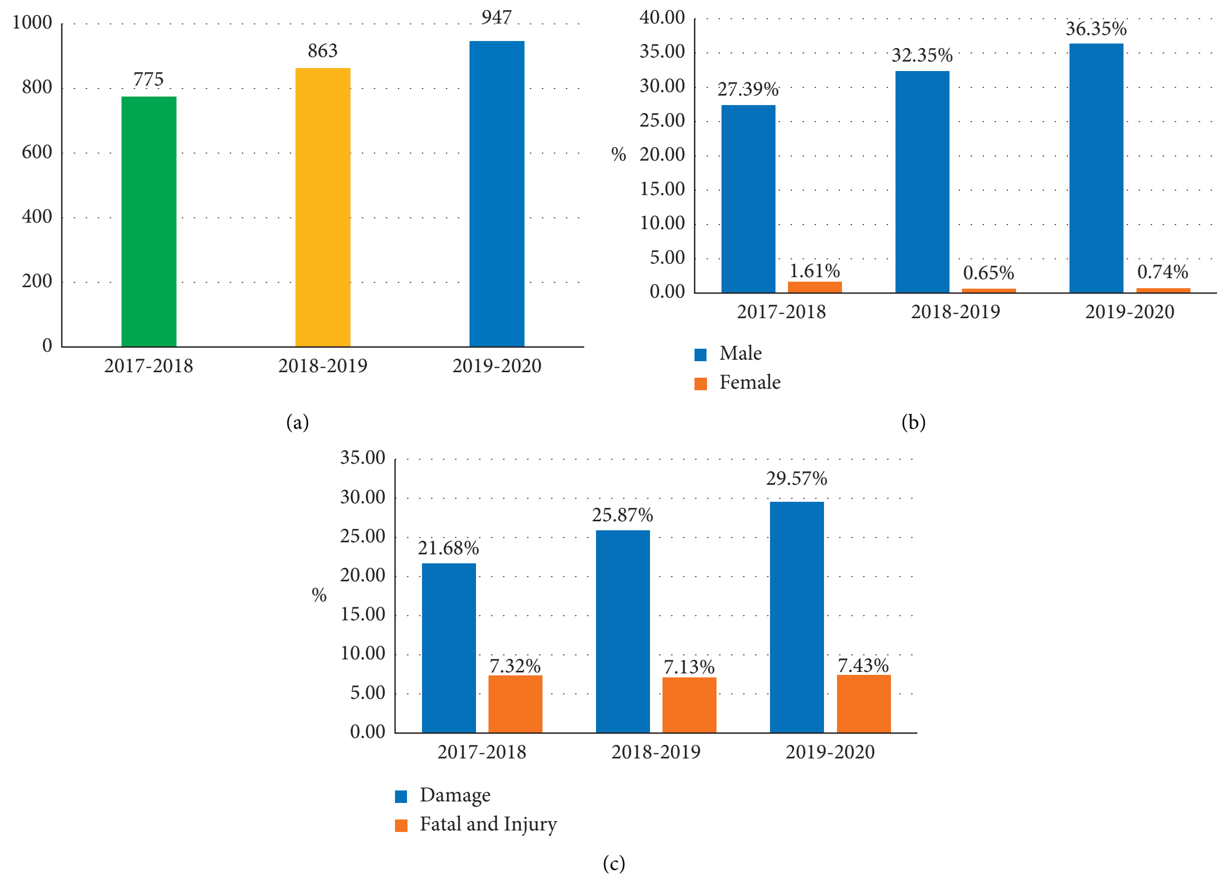


Figure 1: Statistics of vehicle accidents based on: (a) accident year, (b) severity of accidents and accident year, and (c) gender of the driver and accident year.

($n = 947$), and the lowest rate of accidents occurred in 2017–2018 ($n = 775$). Overall, from 2017 to 2020, 2,585 vehicle accidents occurred, of which 77.9% (2013) were damaged, and 22.1% (571) were injury/dead accidents. As shown in Figure 1(b), the greatest rate of male and female accidents occurred in 2019–2020 (36.35%) and 2017–2018 (1.61%), respectively. Figure 1(c) represents that the rural accidents in Tehran often resulted in damage. The greatest percentage of damage (29.57%) and fatal/injury (7.43%) accidents happened in 2019–2020.

3.1.2. The Evaluation of Accident Time. As shown in Figure 2(a), the greatest percentage of traffic accidents for both males and females (41.6%) happened from 12:00 to 18:00, of which 40.36% were males and the rest (1.24%) were females. Given Figure 2(b), 28.84% of vehicle accidents as the largest percentage resulted in damages, and the rest (12.8%) resulted in fatalities and injuries between 12:00 and 18:00.

3.1.3. Evaluation of Collision Type. As shown in Figure 3(a), 62.19% of traffic accidents were related to single vehicle as the largest percentage, of which 60.30% were males and 1.89% were females, and the minimum percentages of vehicle accidents were related to overturning motorcycle. Based on Figure 3(b), the largest and the least percentage of fatal/injury was related to single-vehicle (13.04%) and overturning motorcycle (0.18%), respectively.

3.1.4. Evaluation of Accident Day. Figure 4(a) shows that the largest percentage of accidents (48.9%) occurred in the middle of the week, of which 47.2% were male drivers and 1.69% were female drivers. The least accident occurred on weekends (22.3%), of which 21.8% and 0.5% were males and females, respectively. According to Figure 4(b), from these days, 38.88% of accidents resulted in damage accidents, and 10.02% were fatal/injury.

3.1.5. Evaluation of Accident Reason. As shown in Figure 5(a), the highest percentage in males and females (27.3%) was related to not paying attention to the front (males and females accounted for 26.98% and 0.31%, respectively). The least percentage in males and females was related to fatigue and drowsiness, which was 0.43% and 0.07%, respectively. Also, Figure 5(b) indicates fatigue and drowsiness had the least percentage of damage (0.43%) and fatal/injury (0.07%) accidents.

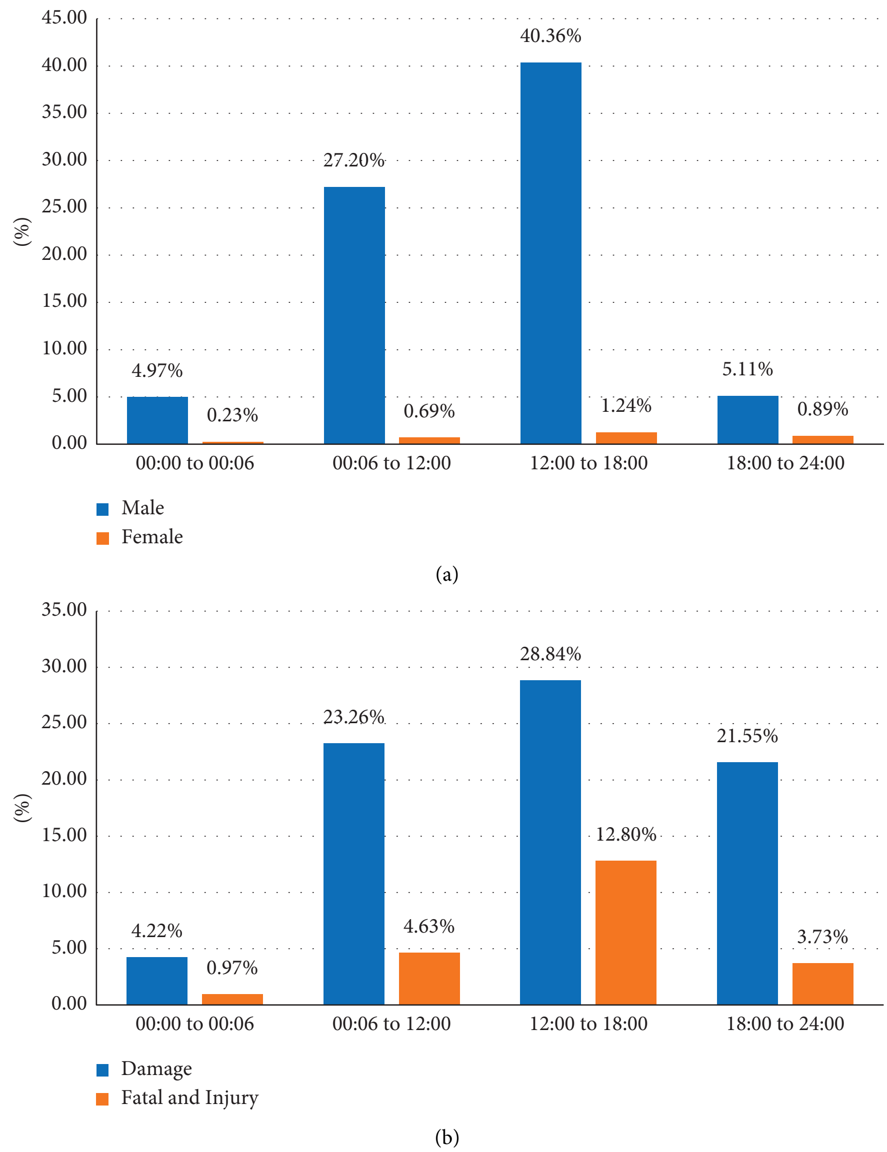


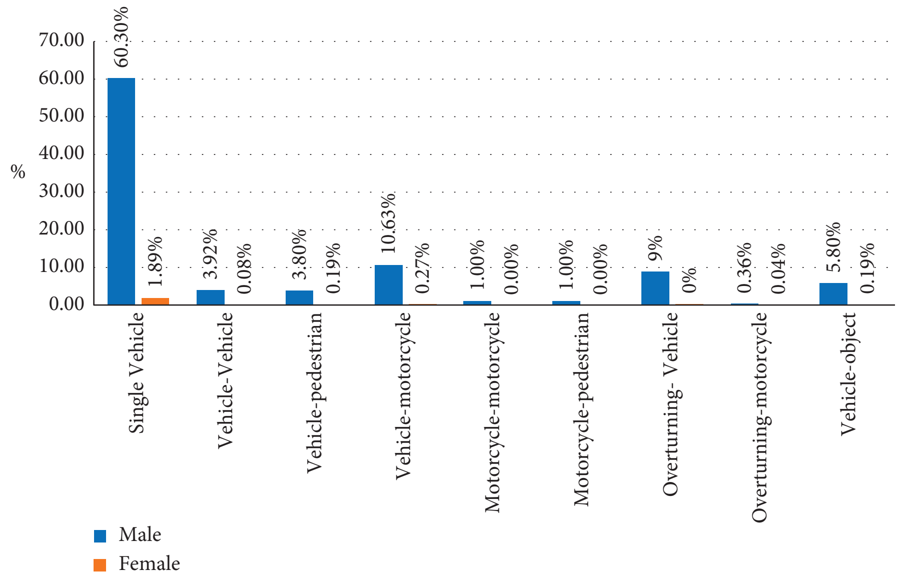
Figure 2: Statistics of vehicle accidents based on accident time: (a) gender of the driver and (b) severity of accidents.

3.1.6. Evaluation of Age. As shown in Figure 6(a), the highest accident rates of males and females were in the age group of 30 to 45 (46.68% and 1.31%, respectively). The lowest accident rate was 2.3%, with the age group less than or equal to 18, of which 2.07% were males, and the rest (0.23%) were females. According to Figure 6(b), the age group 30–45 years had the highest percentage of damage (38.44%) and fatal/injury (9.56%).

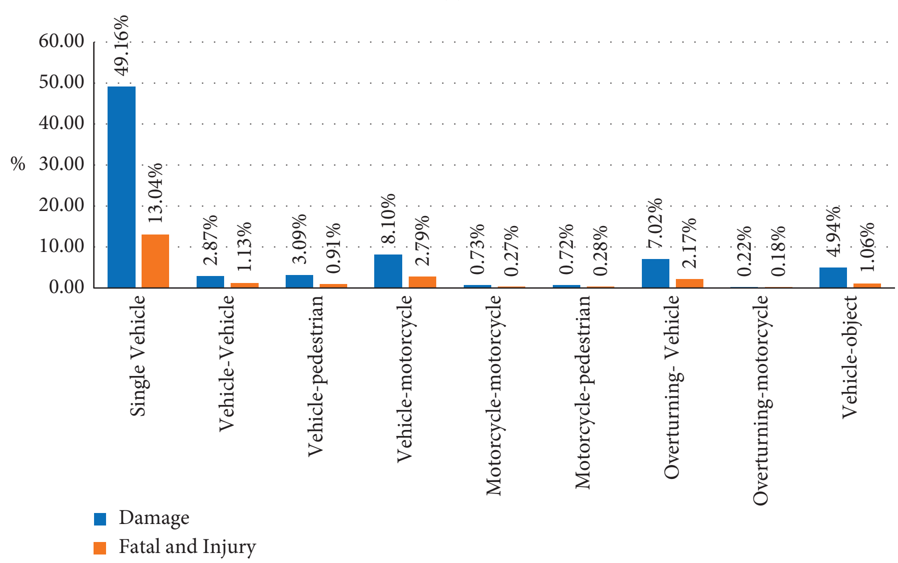
3.1.7. Evaluation of Type of Vehicle Accidents. As shown in Figure 7(a), truck vehicles had the largest percentage among other types of vehicles involved in road accidents, of which 21.83% of accidents were males, and only 0.57% of accidents were females. The results of Figure 7(b) show that the largest portion of both fatal/injury and damage (22.39%) accidents was related to truck vehicles (18.07% for damage and 4.32% for fatal/injury).

3.1.8. Evaluation of Education Rate. According to Figure 8(a), most of the accidents (43.9%) occurred by people having diploma degrees, of which 42.78% were males and 1.12% were females. As shown in Figure 8(b), the largest percentage of traffic accidents that resulted in damage and fatal/injury accidents was also related to people having diploma degrees (33.22%, and 10.6%, respectively).

3.1.9. Evaluation of Weather Conditions. According to Figure 9(a), the greatest accident rates for males and females were recorded in clear weather (81.22% for males and 1.27% for females), and stormy weather had the least accident rate for both males and females (0.63% and 0.07%, respectively). Figure 9(b) shows that the largest percentage (82.5%) of accidents was related to clear weather (63.66% for damage and 18.83% for fatal/injury).

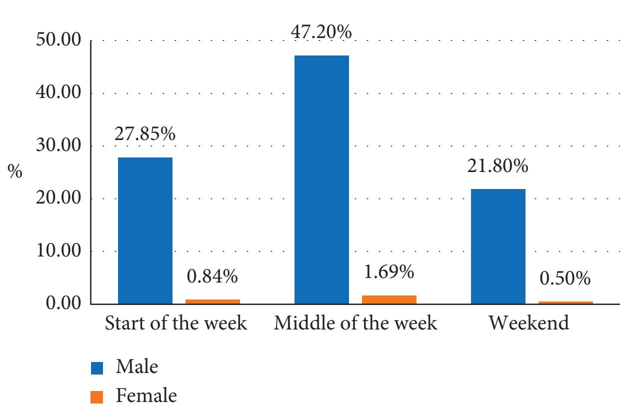


(a)

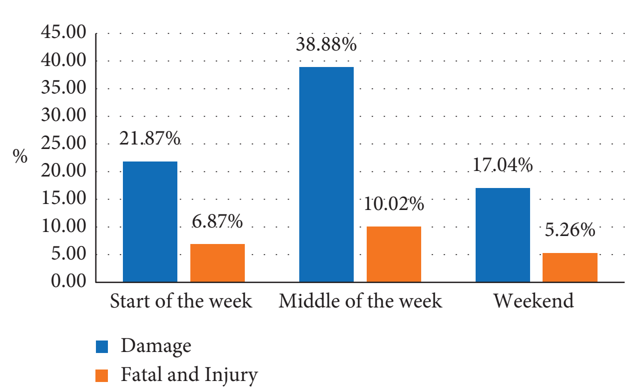


(b)

Figure 3: Statistics of vehicle accidents based on collision type: (a) gender of the driver and (b) severity of accidents.

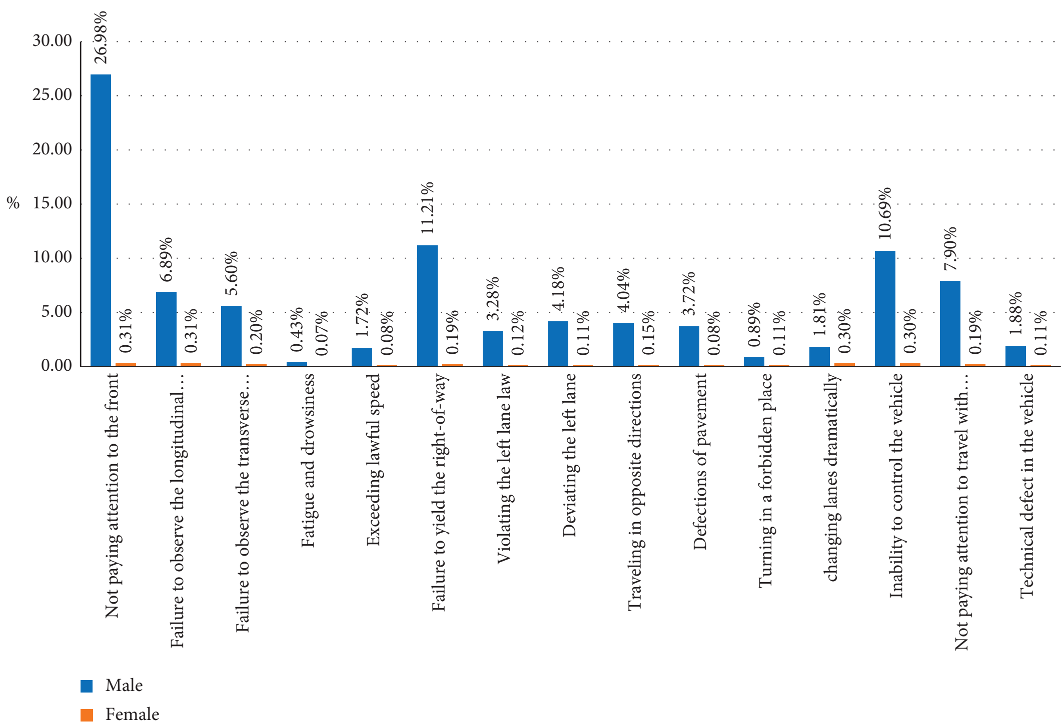


(a)

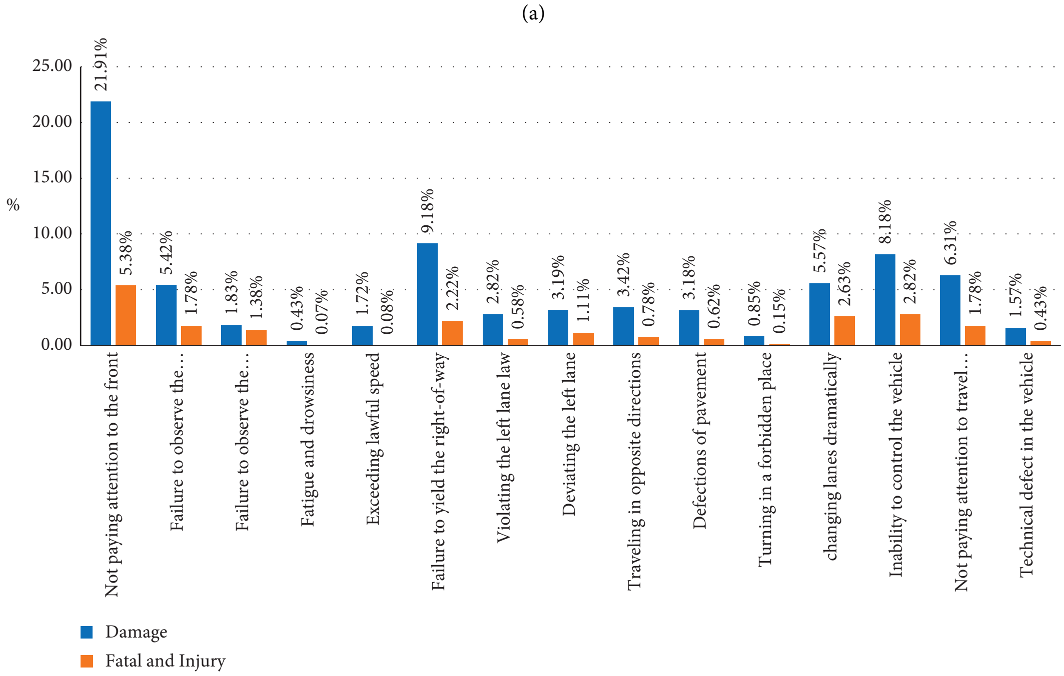


(b)

Figure 4: Statistics of vehicle accidents based on accident day: (a) gender of the driver and (b) severity of accidents.



(a)



(b)

FIGURE 5: Statistics of vehicle accidents based on the reason of accident: (a) gender of the driver and (b) severity of accidents.

3.2. Kolmogorov–Smirnov Test. A related statistical test is needed to evaluate the data normality. So the K-S test was utilized to check the normality of the data. The results of the K-S test are indicated in Table 2. Based on Table 2, all significance levels in the K-S test were less than 0.05, and regarding the 5% error, the null hypothesis (H_0) was rejected, indicating that the natural distribution of variables and H1 was accepted. Therefore, because of the lack of data normality, the nonparametric tests could have been applied.

3.3. Friedman Test. In the presented paper, the Friedman test was used to analyze the rank of each variable. The FT was applied to examine the rank equality related to parameter levels. The level of significance, chi-square value, degrees of freedom, and statistical significance of the statistical sample volume, and α that is illustrated by sig. are reported in Table 3.

As shown in Table 3, the significance level of less than 5% is indicative of rejecting $H_{0,}$ and equal rank claims for these four named parameters were not accepted. Thus, ratings

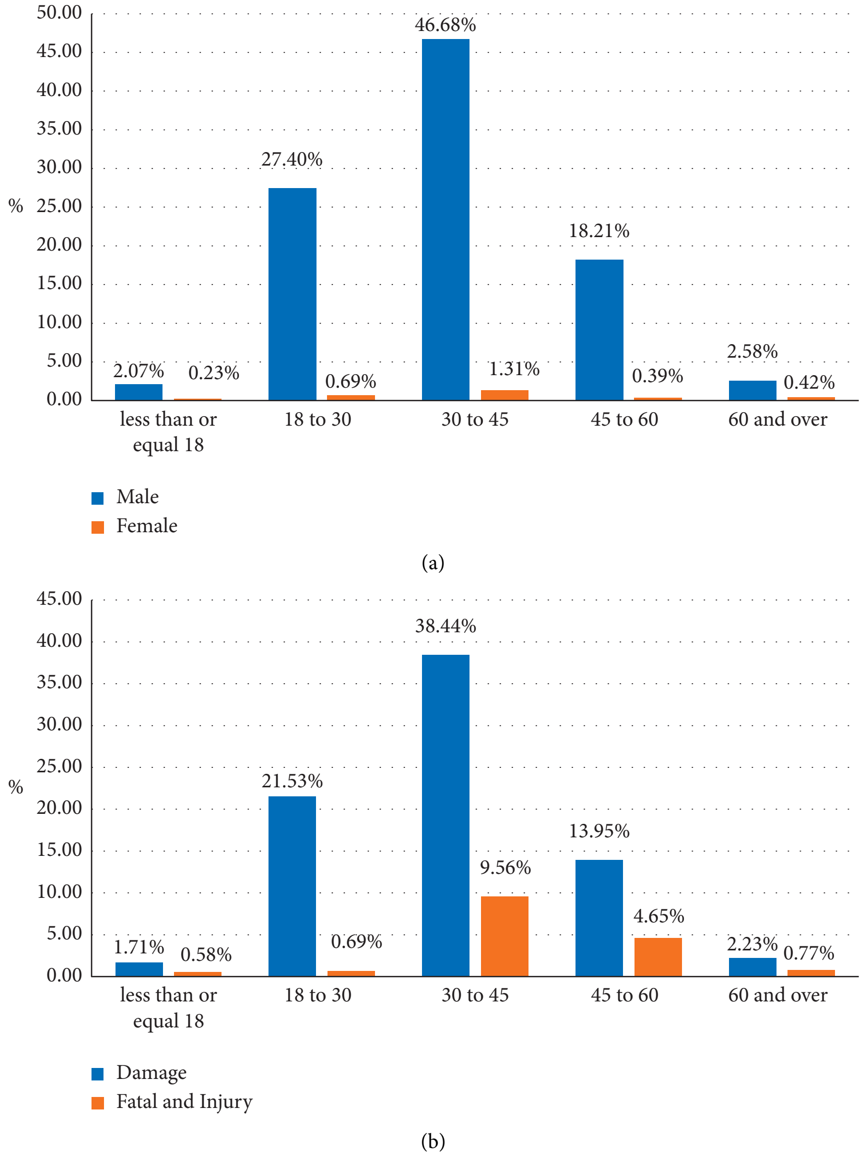


Figure 6: Statistics of vehicle accidents based on age: (a) gender of the driver and (b) severity of accidents.

were inconsistent. Table 4 reports the ranking condition of each variable, representing the average score for independent variables: the lower the average rating, the more influential the variable.

Table 4 indicates that the parameters of gender of the driver, weather condition, collision type, and accident day were considered the most significant parameters on accident severity, respectively, with values of 2.17, 2.71, and 4.12, respectively. On the other hand, the variables of education rate, accident reason, and vehicle type had the least influence on accident occurrence. It can be concluded that gender of the driver had a significant influence on accident severity, which means that the role of humans should be considered as an important factor, and the condition of weather was recognized as the second most factor that increased accident rate. The collision type and accident day parameters were also detected as the third and fourth factors, which affected accident severity.

3.4. Factor Analysis. Table 5 reports the results of the Kaiser–Meyer–Olkin (KMO) index and Bartlett's test in the factor analysis. The closer the value of the KMO index to 1, the better the result of factor analysis. Based on Table 5, the value of the KMO index was 0.524, which shows that the factor analysis is appropriate. Bartlett's test confirms that there is no correlation between variables achieved from the significant level of chi-square. The chi-square value was much greater than 5. Also, the significant level was lower than 0.05; in other words, it means that the alternative hypothesis was affirmed, and a significant correlation existed among the parameters. Thus, all factors in the study affected accident occurrence.

Table 6 reports the variance and eigenvalues corresponding to components. In order for the remaining components in the model, it is required for each component to have values that are greater than 1. As shown in Table 6, components 1 to 4 had values larger than 1. The first block

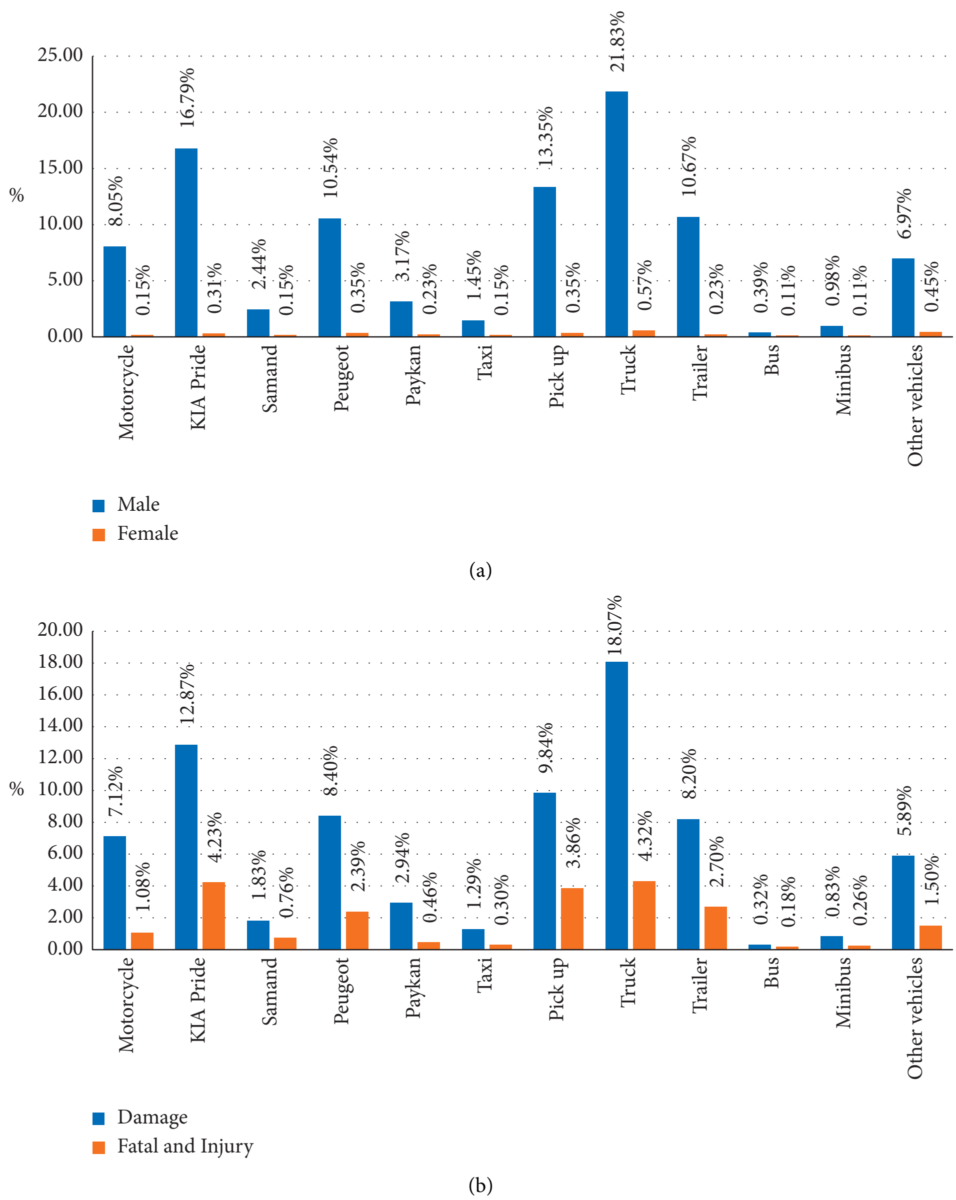
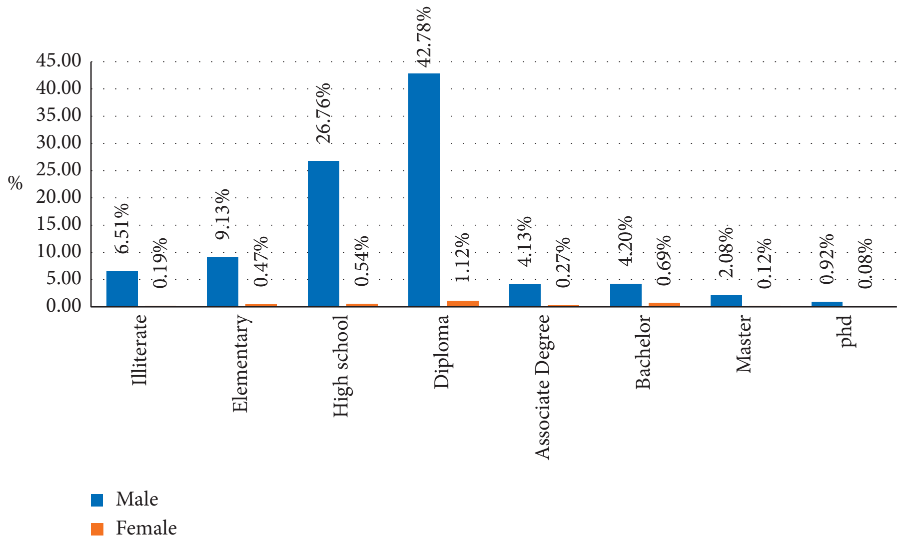


Figure 7: Statistics of vehicle accidents based on vehicle type: (a) gender of the driver and (b) severity of accidents.

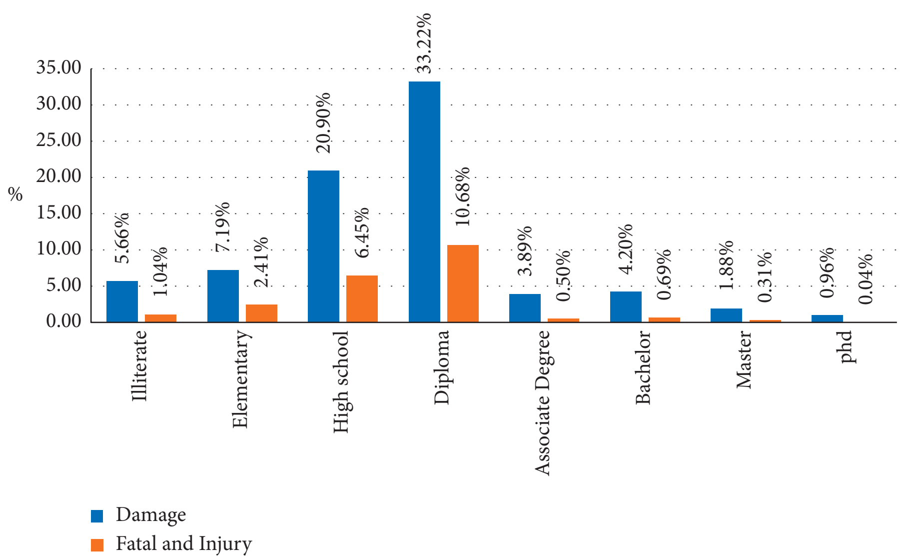
included three columns with labels of initial eigenvalues related to the eigenvalues of the correlation matrix. Eigenvalues are the values estimated by special factors, and the total variance for each test is equal to 100%. The eigenvalue for the first factor was 1.335. Other eigenvalues for later factors are reported in Table 6. The four components that had values greater than 1 could estimate the variability of variables and variances containing about 53% of the variance. The second block included the three columns with labels of extraction total of squared loadings, and rotation total of squared loadings are not represented. Table 6 also displays the rotation matrix of the remaining components. The principal component analysis is the extraction technique used for both matrices, and the method of rotation is Equamax using Kaiser normalization. The values of the correlation degree of the proper parameters related to components are represented in Table 6.

Table 7 represents the components matrix, which includes nine variables in eight factors extracted. Once these factors have no correlation with each other, these factors are the same correlation coefficients between variables in factors. Therefore, the closer the absolute value is to 1, the greater the role of the desired factor in the total variance of numbers. In order to decline the difficulty of interpreting nonrotating component loads, the component matrix was rotated. The aim of the rotated component matrix was to create a new condition for factors to make a better interpretation. The effect of variables on the severity of accidents is explained in the factor analysis, and the higher the absolute value of variables, the more influential the relevant component in the total changes of variables.

The results of Table 8 indicated that driver age, vehicle type, and collision type variables were under the first component, and the coefficients of significance between the first component



(a)



(b)

Figure 8: Statistics of vehicle accidents based on education rate: (a) gender of the driver and (b) severity of the accident.

and each parameter were 0.719, 0.637, and 0.528 positives, respectively. Thus, driver age, vehicle type, and collision type variables were recognized as the most influential factors, which affect accident severity on Tehran province rural roads. In addition, accident time, accident day, and reason of accident variables with positive values of 0.755, 0.641, and 0.508, respectively, under the second most component had a significant impact on increasing the severity of accidents. Similarly, the variables of weather condition and gender of the driver with scores of 0.756, and 0.755, respectively, were under the third component, and then education rate with a coefficient of 0.874 was considered the most influential parameter in increasing the severity of accidents on rural roads of Tehran province.

3.5. Logit Model. Multiple logistic regression was applied in the presented research to predict the severity of accidents. In designing the logit model, 64 independent variables and 2 dependent variables were set and then modeled. Generally, there are three techniques to enter the variables in developing the model. The common method used for modeling is the entering method; however, it has its own disadvantages, in which the model cannot process the data correctly and recognize the effective variables. So, because of this limitation, the forward and backward stepwise methods are commonly utilized in processing the data. Afterward, two criteria of R^2 and the correct percentage were considered to choose the superior method. As shown in Table 9, the values of the correct percentage and the goodness of fit of the backward were greater than the forward. Due to its accuracy in predicting accident severity, the backward method with the correct percentage of 78.5% and the R^2 value of 0.324 was chosen as a superior method to create the logit model on rural roads of Tehran province.

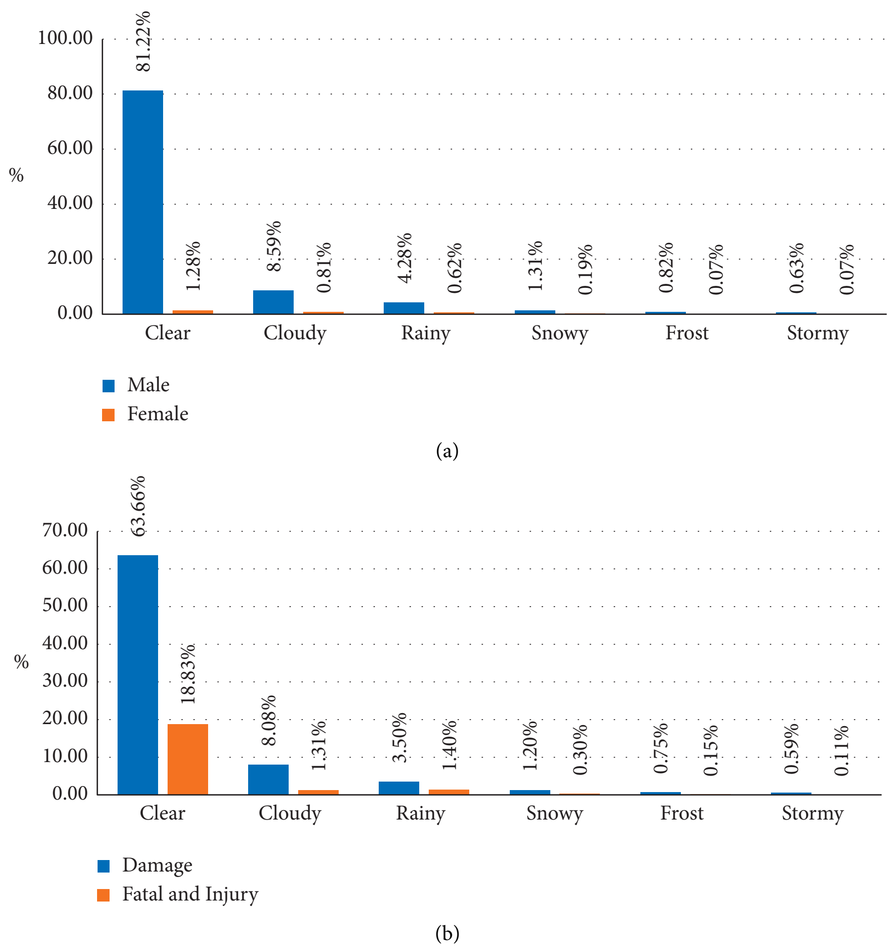


Figure 9: Statistics of vehicle accidents based on the condition of weather: (a) gender of the driver and (b) severity of the accident.

Table 2: The Kolmogorov–Smirnov test results.

Number	Variables	Most extreme differences Absolute	Positive	Negative	Test statistic	Asymp. sig. (2-tailed)
1	Reason of accident	0.173	0.173	−0.151	8.815	0.0
2	Gender of the driver	0.539	0.539	−0.431	27.381	0.0
3	Time of accident	0.230	0.186	−0.230	11.674	0.0
4	Type of collision	0.362	0.362	−0.260	18.391	0.0
5	Day of accident	0.249	0.241	−0.249	12.635	0.0
6	Age	0.245	0.245	−0.235	12.450	0.0
7	Vehicle type	0.181	0.145	−0.181	9.129	0.0
8	Education rate	0.249	0.249	−0.191	12.649	0.0
9	Weather condition	0.474	0.474	−0.351	24.110	0.0

Table 3: Results of the Friedman test.

Number of data	Chi-square	Degrees of freedom	Asymp. sig.
2,585	9,735.064	8	0.0

Table 10 shows that out of 2013 damage accidents, 1,996 accidents were correctly predicted, and out of 572 fatal and injury accidents, only 33 cases were properly predicted by the model. The predictive values of the logit model related to damage and fatal/injury accidents were measured to be 99.2% and 5.8%, respectively. Accordingly, the model's capability was better to classify and separate the damage accidents in comparison to fatal/injuries ones. Also, the overall percentage of the model in recognizing accident severity was 78.5%. The poor ability of the model to predict fatal and injury accidents does not mean that this model is inaccurate.

For modeling the severity of accidents, 64 independent variables were fitted into the model by the use of the backward method and then reduced to include 12 independent variables found as the significant variables in predicting accident severity. As shown in Table 11, the

Table 4: Mean rank in the Friedman test.

Variables	Mean	Rank
Gender of the driver	2.17	1
Weather condition	2.71	2
Collision type	4.12	3
Accident day	4.12	4
Accident time	5.74	5
Driver age	5.81	6
Education rate	6.48	7
Accident reason	6.62	8
Vehicle type	7.24	9

Table 5: The KMO and Bartlett's test.

KMO measure of sampling adequacy		0.524
	Approx. chi-square	479.993
Approx. chi-square	df	36
	Sig.	0.00

Table 6: Total variance explained.

	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
Component	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	1.335	14.834	14.834	1.335	14.834	14.834	1.308	14.528	14.528
2	1.287	14.295	29.129	1.287	14.295	14.295	1.271	14.123	28.652
3	1.149	12.768	41.897	1.149	12.768	41.897	1.179	13.102	41.754
4	1.036	11.506	53.403	1.036	11.506	53.403	1.048	11.650	53.403
5	0.995	10.609	64.012						
6	0.899	9.987	73.999						
7	0.822	9.134	83.133						
8	0.777	8.636	91.769						
9	0.741	8.231	100.00						

Table 7: Component matrix before rotation.

	Component			
Variables	1	2	3	4
Reason of accident	0.122	0.582	0.150	0.359
Gender of the driver	0.121	0.264	0.755	0.055
Collision type	0.411	−0.347	−0.085	−0.087
Accident time	0.500	0.454	−0.013	0.075
Accident day	0.475	0.284	−0.045	−0.300
Driver age	0.629	−0.318	0.058	−0.072
Vehicle type	0.512	−0.282	0.043	0.199
Education rate	−0.015	0.070	−0.038	0.874
Weather condition	0.043	0.533	0.756	−0.061

Table 8: Rotated component matrix.

	Component			
Variables	1	2	3	4
Reason of accident	−0.224	0.508	0.150	0.359
Gender of the driver	0.131	−0.101	0.755	0.055
Collision type	0.528	−0.010	−0.085	−0.087
Accident time	0.074	0.755	−0.013	0.075
Accident day	0.114	0.641	−0.045	−0.300
Driver age	0.719	0.076	0.058	−0.072
Vehicle type	0.637	0.009	0.043	0.199
Education rate	0.048	−0.029	−0.038	0.874
Weather condition	−0.136	0.120	0.756	−0.061

statistics of each variable indicate that the variables of exceeding lawful speed, vehicle-motorcycle, overturning-motorcycle, 12:00 to 18:00, KIA Pride vehicle, Samand vehicle, pickup vehicle, trailer vehicle, bus vehicle, and education rate of elementary, high school, and diploma had the greatest effect on the severity of accidents leading to damage and injury/fatal accidents in Tehran. Moreover, the most significant factors that increase the severity of accidents were 12:00 to 18:00 time, pickup vehicles, KIA Pride vehicles, and education rate of elementary, diploma, and high school. On the other hand, the most influential variable reducing accident occurrence was exceeding lawful speed. The result of the logit model represented that the impact of 12:00 to 18:00 time variable had the highest on increasing the likelihood of vehicle accidents and then pickup vehicles had the most influence on increasing the accident rate. Therefore, regarding the high rate of accidents at 12:00 to 18:00 time, the presence and controlling police at this time may have a suitable impact on reducing the occurrence of accidents. In addition, installing strong lighting along the road at the time

Table 9: Summary of the forward and backward methods of accident severity.

Number	Logit regression methods	Correct percentage	Goodness of fit (R^2)
1	Forward stepwise	78.1	0.213
2	Backward stepwise	78.5	0.324

Table 10: Classification table in the logit model.

Observed		Predicted		
		Accident severity		Percentage correct
		Damage	Fatal and injury	
Accident severity	Damage	1,996	17	99.2
	Fatal and injury	539	33	5.8
Overall percentage		78.5		

Table 11: The results of logit analysis in the first step.

Variables	β	S.E.	Wald	Sig	Exp (β)
Exceeding lawful speed	−1.618	0.708	5.215	0.022	0.198
Vehicle motorcycle	0.461	0.268	2.955	0.086	1.585
Overturning motorcycle	1.156	0.673	2.946	0.086	3.177
12:00 to 18:00	0.972	0.136	51.220	0.000	2.644
KIA Pride	0.486	0.227	4.584	0.032	1.625
Samand	0.635	0.340	3.491	0.062	1.887
Pickup	0.679	0.230	8.681	0.003	1.972
Trailer	0.421	0.241	3.049	0.081	1.524
Bus	1.183	0.631	3.519	0.061	3.265
Elementary	2.133	1.041	4.199	0.040	8.442
High school	2.048	1.033	3.930	0.047	7.754
Diploma	2.106	1.032	4.164	0.041	8.214
Constant	−4.560	1.335	11.665	0.001	0.010

of 12:00 to 18:00, especially in the winter (due to the darkness of the air at 17:00), can be considered an effective way to reduce the severity of the accidents. Considering the vehicle manufacturers to design the vehicles with high and safe quality as well as address their technical problems will have a direct impact on declining the severity of these accidents.

Table 12 reports the values of significance (sig.), degree of freedom (df), and chi-square of the backward method in the first step of the modeling process. The chi-square value of the logit model related to step 1 was 217.333 with a significant value of less than 5%, which means that the predictive power of the model for accidents was affirmed.

Table 12: The backward stepwise model results.

		Chi-square	Df	Sig.
Step 1	Step	217.333	56	0.00
	Block	217.333	56	0.00
	Model	217.333	56	0.00

3.6. ANN Models

3.6.1. MLPNN. In the current research, the number of accidents leading to damage, and fatal injury was developed by the MLPNN model. In designing the model, 70% of the data were selected to train the network, and the remaining 30% were considered for testing. The number of input variables entered in the MLPNN model was 9, each of which was divided into different categories. The numbers of neurons in the input and output layers were 64 and 2, respectively. The automatic architecture was applied to develop the model, which computed six neurons in the hidden layer and selected one hidden layer for the model. Table 13 represents the correct percentage of the MLPNN model for each category. Table 13 reports that among 1,426 cases related to damage accidents, 1,409 cases were accurately predicted by the MLPNN model, and out of 403 cases related to fatal-injury accidents, only 37 cases were accurately classified by the model. Overall, in the training sample, 98.8% of cases were predicted, and 1.2% of cases related to damage accidents were not accurately predicted. Moreover, the cross-entropy error was considered a criterion to evaluate the MLPNN model in predicting accidents. The results showed that the values of training and testing samples were 860, 985, and 383.342, respectively. According to Figure 10, the receiver operating characteristic (ROC) curve is a plot of the sensitivity versus specificity of a diagnostic test utilized to visualize and organize the classification of the model in forecasting each category of accidents. The value of the area under the curve (AUC) can be different from 0 to 1; the closer the value to 1, the more precise classification of the model (27). In addition, the value of AUC should not be less

Table 13: Classification in the MLPNN model.

Sample	Observed	Predicted Damage	Predicted Fatal and injury	Predicted Correct percentage (%)
Training	Damage	1,409	17	98.8
	Fatal and injury	366	37	9.2
	Overall percentage (%)	97.0	3.0	79.1
Testing	Damage	577	10	98.3
	Fatal and injury	157	12	7.1
	Overall percentage (%)	97.1	2.9	77.9

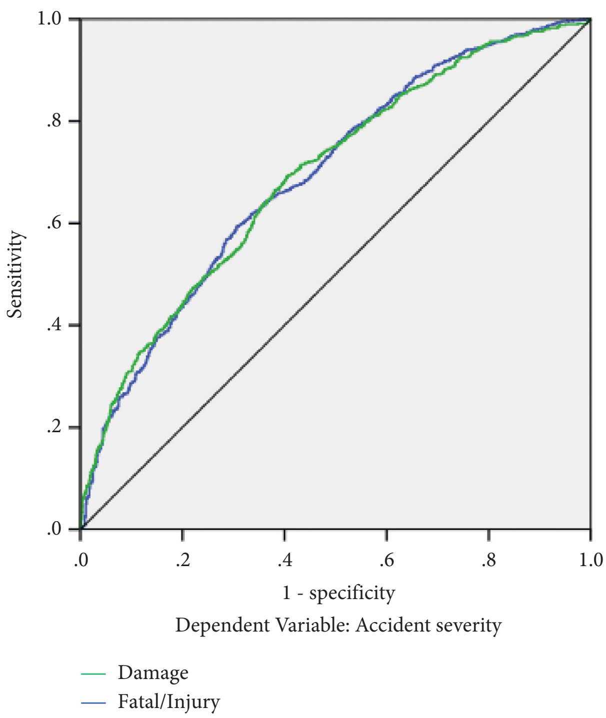


Figure 10: The ROC curve for the MLPNN model.

than 0.5. According to Table 14, in this model, the likelihood of accidents that led to damage and fatal/injury was calculated to be 0.693, larger than 0.5, meaning that the classification of the model was in a good range. Figure 11 also represents the relative importance of independent variables for the MLPNN model, which indicates the extent of the effect of the independent variables in predicting accidents. Normalization of this significance was achieved by dividing the importance values by their highest value and defined as a percentage. As shown in Figure 11, the variables of gender of the driver, accident time, weather condition, driver age, and education rate were found as the most significant variables of the severity of accidents, respectively. Moreover, Figure 11 shows that gender of the driver had the greatest indication of all parameters, which affect accident severity.

3.6.2. RBFNN. At first, in designing the RBFNN model, the data set was divided into 70% and 30% for training and testing, respectively. The numbers of neurons in input and output layers were measured to be 64 and 2, respectively. In

Table 14: The area under the curve for the MLPNN model.

		Area
Accident severity	Damage	0.693
	Fatal/injury	0.693

order to develop the RBFNN model, the automatic architecture computed 9 neurons in the hidden layer. Table 15 indicates the classifying of the prediction of damage and fatal/injury accidents. The RBFNN model correctly predicted 78.8% of cases. Among 1,402 cases related to damage accidents, all cases were correctly predicted with the correct percentage of 100, but among 379 cases of fatal/injury accidents, only 2 cases were correctly predicted. Therefore, the model was capable of predicting damage accidents, but it was not able to predict accidents leading to fatal/injury. In addition, the values of cross-entropy error for training and testing were found to be 296.281 and 126.994, respectively. Figure 12 represents the ROC curve for the damage and fatal/injury accidents. As shown in Table 16, the values of AUC for both damage and fatal/injury accidents were measured to be 0.635, which is larger than 0.5, meaning that the response of the RBFNN model is positive. The independent variable importance, which affects accident severity, is shown in Figure 13.

3.7. Validation of the Machine Learning Models

3.7.1. Development of Mathematical Models. In order to prove the proposed models, several ANN models using MATLAB programming with a different learning method and active functions were developed and then trained on the data to evaluate the performance of the machine learning methods. In this section, 40 multilayer perceptron network (MLP) models with different structures were built. In order to have an effective comparison between the machine learning methods and the mathematical models (MLP) used for validation, the partitioning data set, as same as the machine learning methods, was considered to design MLP models. Thus, 70% of the data were utilized for training, and 30% were used for validation. The criterion in dividing the data set was to determine enough samples for the ANN training and the rest for cross-validation. Selecting the right architecture, transfer functions, training algorithm, and the number of neurons in the hidden layer is regarded as the critical factors in creating models. The best method for training is that once making the network convergent and

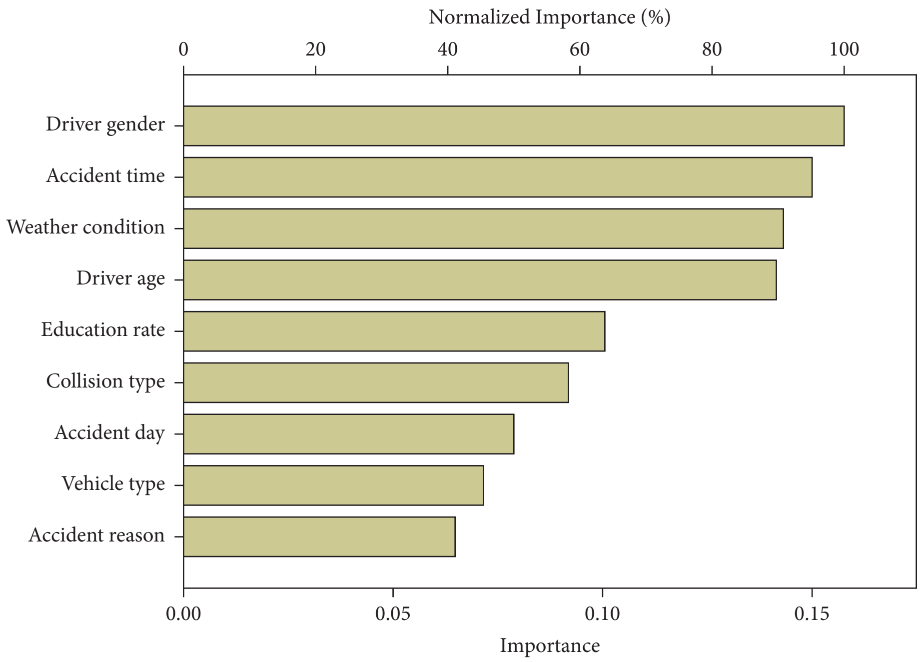


Figure 11: Independent variable importance chart in the MLPNN model.

Table 15: Classification in the RBFNN model.

		Predicted		
Sample	Observed	Damage	Fatal and injury	Correct percentage (%)
Training	Damage	1,402	0	100.0
	Fatal and injury	377	2	0.5
	Overall percentage (%)	99.9	0.1	78.8
Testing	Damage	611	0	100.0
	Fatal and injury	192	1	0.5
	Overall percentage (%)	99.9	0.1	76.1

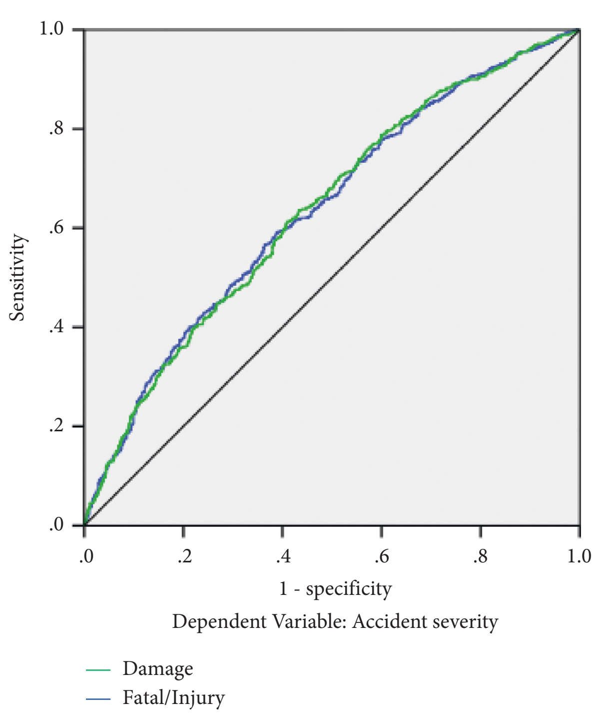


Figure 12: The ROC curve for the RBFNN model.

Table 16: The area under the curve for the RBFNN model.

		Area
Accident severity	Damage	0.635
	Fatal/injury	0.635

having minimum, the error function of the training has the modeling capability by providing new data. Among the several methods for training, the Levenberg–Marquardt (LM) algorithm has the best acceleration to train neurons in comparison to other kinds of methods (Bayesian regularization-scaled conjugate gradient). This method, which is different from the method of machine learning approaches, can be used to train the network with a large number of variables. Its most prominent feature is ensuring the model's convergence and convergence speed. Thus, the Levenberg–Marquardt (LM) algorithm was applied to train the neural networks.

3.7.2. Model Details. Log-sigmoid function was used as the activation function. Initially, the developed models were trained using the training set and afterward, they should be validated using the validation set. Since during the training process, the initial weights were randomly chosen, and there

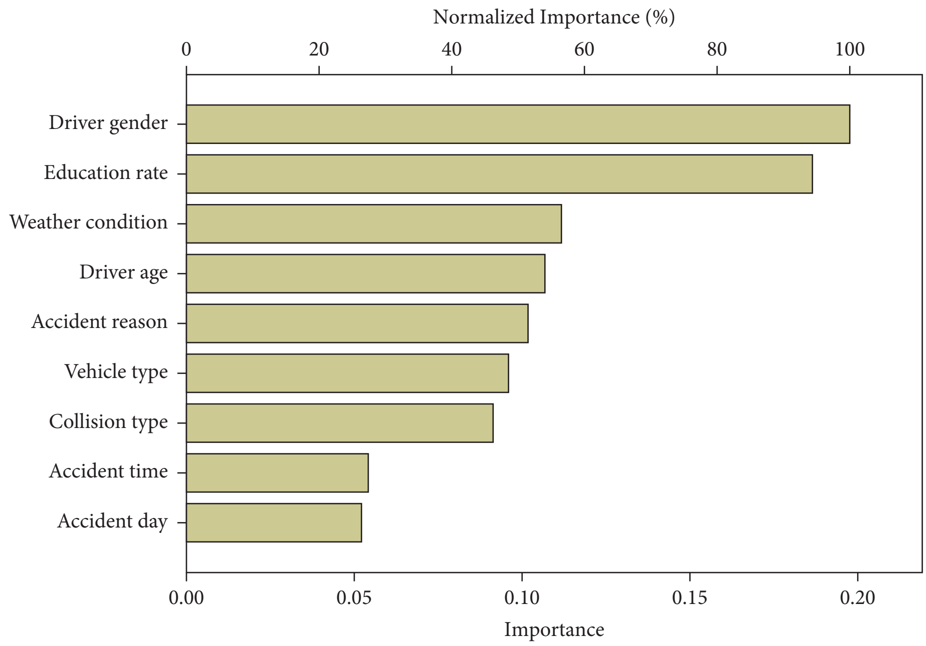


Figure 13: Independent variable importance chart in the RBF model.

was a probability the network fell into a local minimum; the networks were initially trained using one hidden layer and various numbers of neurons. Each model was developed for 14 iterations, and the results were recorded. A similar process was then repeated for models with 2 hidden layers. The 40 different network structures as well as the accuracy of their prediction are shown in Table 17. Afterward, the optimum network structure was chosen by comparing the results of different structures.

3.7.3. Determination of the Optimized Model. Forty models with different structures have been developed to train on the data set. The specification of 40 models with various structures is represented in Table 17. For both stages of training and validation, two comparison criteria have been utilized in order to compare these diverse neural network architectures, which included RMSE. An important question may arise as to how we can determine the best network structure associated with the least amount of variance. The most ideal range for selecting an optimum model is in which the training and validation error graphs start converging with each other. Thus, this research used the values of RMSE as the error criterion. In other words, the criteria for selecting the optimized model are based on the values of training RMSE and validation RMSE. According to Figure 14, the optimized model is achieved in which both the values of training RMSE and validation RMSE are at the minimum level and have a little difference (corresponding to each other) [49]. As shown in Figure 14, the horizontal axis is indicative of the number of neurons in the hidden layer, and the vertical axis is indicative of the training RMSE and validation RMSE values. Initially, the number of neurons is small, and the values of the training RMSE and validation RMSE are very close to each other; however, by moving toward the right side of the plot, meaning that the number of neurons is increased, it is indicated the values of the training RMSE and validation RMSE will be going to be far away from each other (high variances). By comparing 40 models (both single layer and double layers), it is clearly shown that the best architecture for the network is the neural network with 3 neurons in 1 hidden layer (M1), an optimal selection. Model M1 has the least amount of difference between the values of the training RMSE and validation RMSE compared to other models. The results of each network training are represented in Figure 15. As can be seen in Figure 15, in prediction model M1, the training procedure has reached the best value for validation error after 14 repeats. It can be seen in Figure 15 that by increasing the number of epochs, the amount of MSE is reduced, improving the performance of the model both in the training and validation phase. Figure 15 represents the prediction model M1 had the minimum MSE compared to other networks.

3.8. Performance Comparison of Mathematical Models with the Machine Learning Methods. In this section, since the aim was to analyze the performance of machine learning methods, the comparison was made between the ANN (MLP and RBF) models developed by the scaled conjugate gradient (SCG) method for training neurons to six kinds of ANN (M1, M4, M13, M20, M22, and M30) models trained by Levenberg–Marquardt (LM) algorithm. In addition, the active function used for machine learning methods (MLP and RBF) was the hyperbolic tangent, and the active function utilized for MLP models was log-sigmoid. In order to prove the machine learning methods, after the training set was done, the performance of the trained model was analyzed; for this purpose, the error criteria of root mean square error (RMSE), mean square error (MSE), mean absolute error

Table 17: Different ANN structures for the prediction of accident severity.

Model	Hidden layer	Hidden neurons	RMSE Training	RMSE Validation
M1	1	3	0.3956	0.3985
M2	1	6	0.3731	0.3986
M3	1	10	0.3727	0.3993
M4	1	15	0.3794	0.3885
M5	1	20	0.3688	0.3952
M6	1	23	0.3599	0.3905
M7	1	25	0.3558	0.4129
M8	1	28	0.3665	0.3864
M9	1	30	0.3716	0.4027
M10	1	33	0.3665	0.4041
M11	1	35	0.3607	0.3963
M12	1	40	0.3662	0.3983
M13	1	43	0.3874	0.3775
M14	1	45	0.3629	0.4034
M15	1	48	0.3488	0.4011
M16	1	50	0.3436	0.4092
M17	1	52	0.3629	0.3930
M18	1	55	0.3393	0.3980
M19	1	58	0.3458	0.4029
M20	1	60	0.3758	0.3921
M21	2	4–7	0.4174	0.4026
M22	2	4–6	0.3806	0.3914
M23	2	5–10	0.3739	0.3965
M24	2	8–14	0.3741	0.4018
M25	2	9–17	0.3615	0.3855
M26	2	12–20	0.3645	0.3951
M27	2	14–23	0.3439	0.4073
M28	2	16–27	0.3534	0.4068
M29	2	18–31	0.3461	0.4092
M30	2	21–34	0.3848	0.4003
M31	2	24–37	0.3732	0.4012
M32	2	26–39	0.3536	0.3817
M33	2	29–41	0.3654	0.3944
M34	2	31–44	0.3586	0.3854
M35	2	32–47	0.3364	0.4058
M36	2	34–51	0.3449	0.4263
M37	2	38–54	0.3396	0.3951
M38	2	41–56	0.3146	0.4060
M39	2	44–57	0.3034	0.4042
M40	2	45–60	0.3428	0.3950

(MAE), mean absolute percentage error (MAPE), and average absolute deviation (MAD) were considered to compare the performance of the mathematical models with machine learning methods. These approaches used for evaluating the performances of machine learning models have been applied in earlier studies [50, 51]. In this study, the SPSS and MATLAB software programs were utilized to assess the obtained results. Furthermore, the optimum model (M1), as well as the models with suitable performances, were determined to analyze the efficiency of machine learning methods (as shown in Table 18). In the stage of training, the comparison of the obtained results indicated that with respect to the values of RMSE (the machine learning method) equal to 0.4679 and 0.4683, there is a slight difference between the results of the machine learning method and other ANN models (with the values of RMSE = 0.3956, 0.3794, 0.3874, 0.3758, 0.3806, and 0.3848). A similar conclusion can be found by comparing the other criteria such as MSE, MAE, MAPE, and MAD. As a result of comparing the mentioned approaches, it can be concluded that the difference between the developed models used for validation and the machine learning methods was a little (as shown in Table 18). Therefore, the performance of the machine learning methods can be proved in terms of accuracy and efficiency, showing that these methods can be considered robust and accurate models for predicting the severity of accidents.

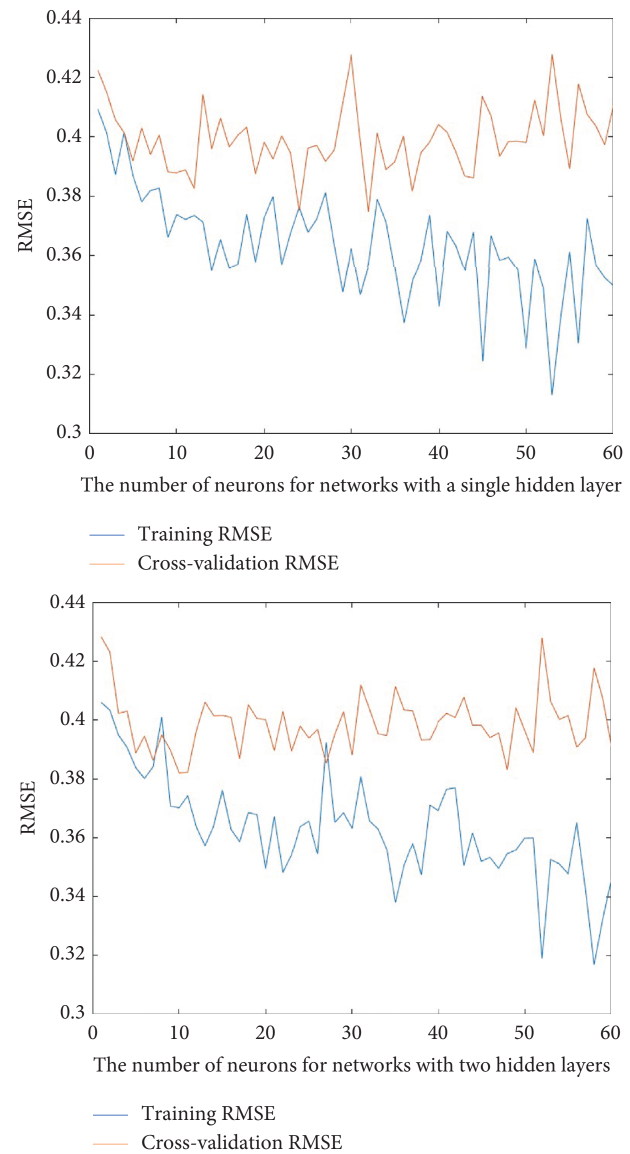


Figure 14: The number of neurons sets based on the values of RMSE for the developed models.

4. Comparison and Discussion

The current research examined several effective variables using different methods on accident occurrence on rural roads of Tehran, Iran. Each method had its own unique

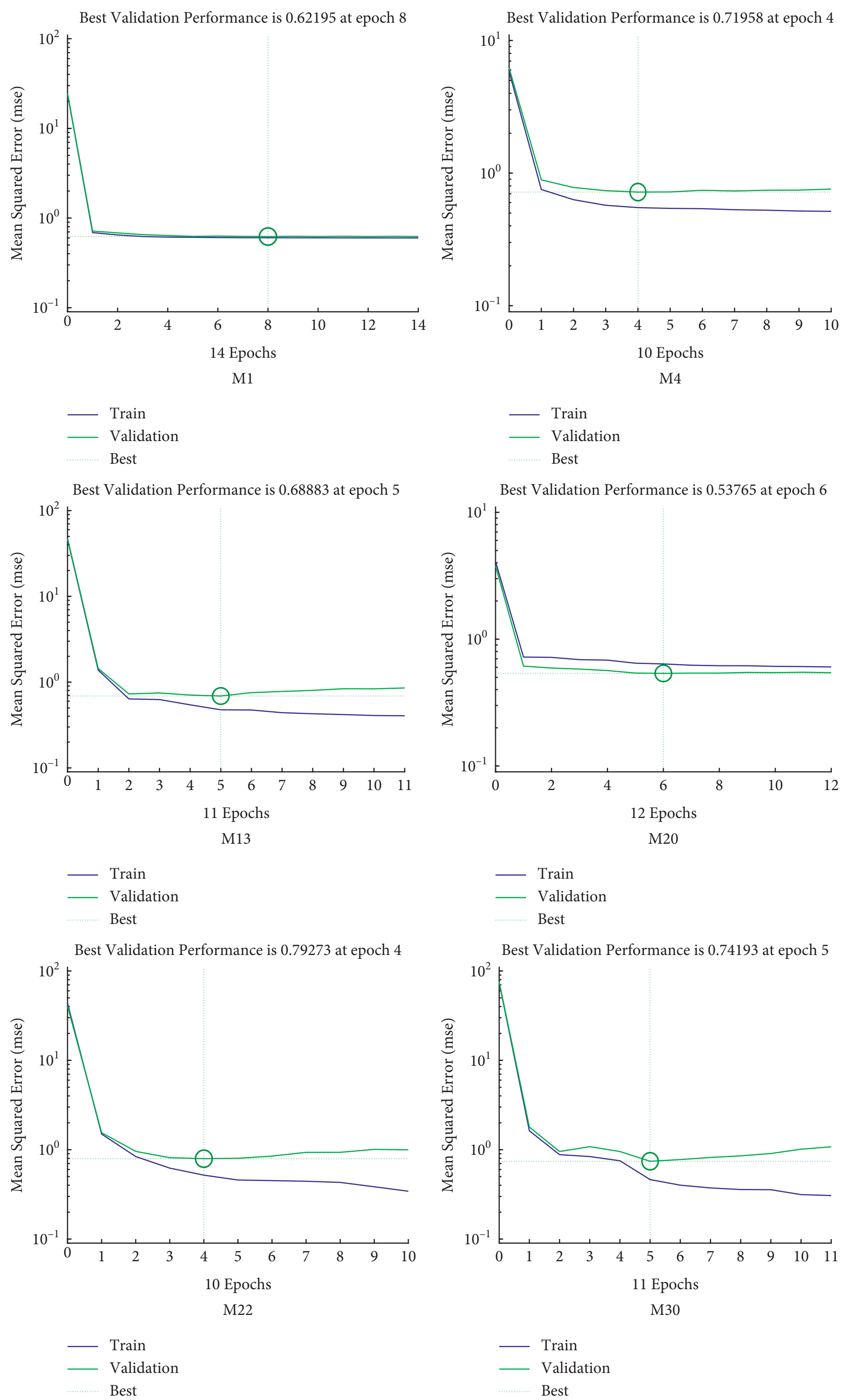


FIGURE 15: MSE versus epoch curve for best six models.

results, and there was no conflict between them, so each technique should be assessed based on its own result. Comparison of these methods enables for summarizing the results. The K-S test was applied to check the normality of the distribution, the results of which indicated that the test was significant; thus, the data did not have a normal distribution. Therefore, nonparametric tests were utilized. The Friedman test was utilized in this study to detect the priority

Table 18: Performance comparison between the mathematical models and machine learning models.

Model	RMSE	MSE	MAE	MAPE (%)	MAD	Learning method	Active function	Hidden layer	Hidden neurons
MLPNN	0.4679	0.2189	0.2212	11.1218	0.2212	SCG	Tanh	1	6
RBFNN	0.4683	0.2193	0.2192	11.0831	0.2176	SCG	Tanh	1	9
M1	0.3956	0.1565	0.0046	12.579	0.0045	LM	Sigmoid	1	3
M4	0.3794	0.1439	0.0147	13.210	0.0156	LM	Sigmoid	1	15
M13	0.3874	0.1501	0.0159	13.550	0.0675	LM	Sigmoid	1	43
M20	0.3758	0.1412	0.0781	13.430	0.0092	LM	Sigmoid	1	60
M22	0.3806	0.1448	0.0877	13.960	0.0597	LM	Sigmoid	2	4–6
M30	0.3848	0.1481	0.0517	13.210	0.0041	LM	Sigmoid	2	21–34

of parameters in accident severity, in which the variables of gender of the driver, condition of weather, collision type, and accident day were recognized as the first to fourth rankings leading to vehicle accidents. On the other hand, the factor analysis was performed to recognize the underlying parameters, in which driver age, vehicle type, and collision type were considered as the first influential factors and time of accident, accident day, and reason of accident were distinguished as the second significant parameters in vehicle accidents. The machine learning approach was then used to recognize the contributory parameters of accident occurrence in Tehran province rural roads. In this method, the logit, MLPNN, and RBFNN models were built to detect the significance of each parameter in the severity of accidents. The criteria for selecting optimized machine learning methods are based on the smallest cross-entropy error, the correct classification rate, and the area under the ROC curve [52]. That means that the model that has a minimum cross-entropy, the best percentage of the correct classification rate, and the maximum value of the area under the ROC curve is selected as the best model. By the comparison of the correct prediction of the logit and ANN models, it was shown that the MLPNN model provided better performance and had a greater prediction percentage than other models. The prediction accuracy of the logit, MLPNN, and RBFNN models was 78.5%, 79.1%, and 78.8%, respectively. In addition, the prediction accuracy of the ANN models was greater than the logit model. Moreover, the logit model indicated the variables of exceeding lawful speed, vehicle-motorcycle, overturning-motorcycle, 12:00 to 18:00, KIA Pride vehicle, Samand vehicle, pickup vehicle, trailer vehicle, bus vehicle, and education rate of elementary, high school, and diploma had the highest effect on increasing the severity of accidents, while in the MLPNN model, due to better prediction, gender of the driver, accident time, weather condition, driver age, and education rate were the most significant variables in accident occurrence. Also, in the RBFNN model, the variables of gender of the driver, education rate, condition of weather, and driver age were the most significant parameters. Consequently, gender of the driver was recognized as the most influential variable in accident severity, indicating the role of the human factor in accident occurrence on Tehran province rural roads.

The results of this research, in comparison to the previous research, demonstrated that based on the results of MLP, RBF models, and the Friedman test (FT), the most effective variable on the severity of accidents is gender of the driver. It means that the role of humans is recognized as a precursory or primary cause of vehicle accidents, in line with other studies [53–56]. To be more specific, the results of factor analysis showed that the driver males (96.9%) accounted for a larger portion of accidents in comparison to driver females (3.1%), consistent with these studies [26, 57–59]. Also, in some previous studies, similar to the results of the MLP model, a significant effect of accident time has been emphasized [60, 61], in contradiction with the results of the RBF model in this study, indicating the low impact of this factor on accidents. Moreover, the results of frequency analysis and logistic regression model revealed that the time of 12:00 to 18:00 has a significant impact on accidents, similar to this study [26]. A significant effect of weather conditions has been found in the previous research [62], which is in contradiction with the results of the MLP and RBF models, showing that this factor is not a high priority in reducing accidents. In addition, in the frequency analysis, the greatest accident happened once the weather was clear, in line with the results of other studies [26, 58]. The significant effect of age in reducing accidents has been emphasized in some studies, contrary to the result of this research, indicating that there is less effect of this factor [51, 63]. In addition, the results of frequency analysis indicated that most accidents occurred by the driver between the age of 30 and 45, similar to these studies [26, 58]. Some researchers have shown a significant effect on accidents [64, 65], contrary to the results of these studies, indicating a low effect of this factor. The collision type in both MLP and RBF models has been considered a low effect of this factor on accidents, contrary to this study [66]. Accident day in declining accidents has been recognized as low impact, contrary to some studies [67, 68]. The results of MLP and RBF models indicated low importance for the variable of vehicle type, contrary to this study [21]. The variable of exceeding lawful speed in the results of the logit model has a significant impact on accidents, similar to the study [69]; however, in ANN models, accident reason has a low impact.

5. Safety Solutions

In order to reduce the number and severity of accidents in Tehran province rural roads, it is necessary to consider the effective variables in the occurrence of accidents and propose safety solutions according to their importance. Based on the results of the analysis, safety solutions that can be utilized to reduce rural road accident risks are as follows.

The likelihood of single-vehicle accidents, mostly trucks, at 12:00 to 18:00 accident time (the time of returning people to their home) once drivers did not have enough attention to the front in the middle of the week, especially in the clear weather, was significant. Moreover, the male drivers of single-vehicle accidents, mainly young (between age 30 and 45) with diploma degrees, were more likely to be involved in accident occurrence. Thus, it is recommended that drivers pay more attention to the type of movement (not paying attention to the front) due to poor visibility in driving [70]. Special attention should also be paid to disabled people [71]. Since the air is going to be dark in the winter sooner, hence, installing powerful lights along roads can help reduce the severity of accidents after 17:00. The role of education may also have an impact effect on reducing the occurrence of accidents. Accordingly, it is recommended the official authorities allocate enough budget for educating young male drivers. In addition, allocating separate lines for trucks, leading to an increase in the level of service, especially on two-lane two-way (TLTW) rural highways, is recommended to control trucks on roads. Pavement damages [72–74] also have a great impact on lane changing of drivers and, as a result, driving safety, so the necessary measures such as various additives as well as nanomaterials [75–95] and improvement of friction [96–100] should be taken into account. It is recommended that motorcycle riders keep their distance from other vehicles on roads to prevent accidents of overturning motorcycle and vehicle motorcycle, so allocating separate lines can be an effective way to promote safety [101]. Improving the brake system and checking its frequency and using high-quality tires in motorcycles can be helpful in reducing overturning motorcycle accidents, especially in wet conditions and curves. Since the drivers of motorcycles are considered vulnerable road users, promoting the skills of drivers by providing related education is another type of solution to increase safety. Young male drivers will have the possibility to be prepared with last education annually, which can be included various educational activities and instructions. Passing strict rules such as physical or emotional tests or preventing people who have diplomas or lower degrees from getting a motorcycle license can enhance safety (it is just recommended). Since pickup and KIA Pride vehicles are the most common types of vehicles involved in the occurrence of accidents, it is necessary for vehicle designers to reconsider the design of these vehicles and address their defects. Giving suitable warnings to drivers using intelligent warning signs or other types of warnings can be a useful way for drivers to be concentrated on controlling their speed and increasing safety [102].

6. Conclusion

In the current paper, machine learning methods were developed to detect the variable that affects accident severity on rural roads of Tehran province. The study examined the different techniques to recognize the most influential variables in the occurrence of traffic accidents. Each of them has its own characteristics and can be helpful in providing practical results in reducing accident severity. The machine learning methods indicated the following meaningful findings:

(1) The results of the frequency analysis indicated that most of the accidents (damage and fatal/injury) occurred at the hours of 12:00–18:00 in the middle of the week, once the weather was clear and drivers did not pay attention to the front. Moreover, the truck vehicle was the most common type of vehicle involved in accidents (damage and fatal/injury). The drivers who were in the age group of 30–45 accounted for the highest percentage of the occurrence of accidents, and the males who had a diploma degree had the greatest rate of accidents. Among other types of accidents, single-vehicle was the most common type of accidents, having the greatest rate. In addition, the number of male drivers involved in accidents was much greater compared to female drivers.

(2) According to the results of the K-S test, the test was significant, which means a lack of normality distribution existed in the variables that affected vehicle accidents. Thus, nonparametric tests were employed.

(3) The Friedman test represented the variables of gender of the driver, weather condition, collision type, and accident day were detected as the most effective variables to predict the severity of accidents. Gender of the driver and type of collision variables were considered the first and the third factors, enlightening a human factor. In addition, accident day was identified as the second factor that was considered an environmental factor, which affected accident severity on the rural roads of Tehran. With respect to the results of the Friedman test, gender of the driver accounted for the largest portion of accident occurrences.

(4) The results in factor analysis illustrated that four parameters were recognized as the most influential factors in the accident rate, which means that out of nine variables, four factors were detected in the FA. The analysis indicated that the variables of driver age, vehicle type, and collision type were the first factor that influenced the severity of accidents, respectively. Therefore, the human factor (as the first factor) was found to be associated with accident occurrence. Accident time, accident day, and reason of accident variables were the second factors; weather condition and gender of the driver variables were the third factors; and education rate was the fourth effective factor in the occurrence of accidents on the Tehran province rural roads.

(5) The results of the logit model indicated that between the two kinds of methods considered for building the logit model in predicting accident severity on the Tehran province rural roads, based on the two criteria of the prediction accuracy of 78.5% and the goodness of fit (R^2) of 0.324, the backward method was selected as an effective way in designing the logit

model. Moreover, the results in the logit model represented that the variables of accident time (12:00 to 18:00), the type of vehicle (pickup and KIA Pride), education rate (elementary, diploma, and high school), the type of vehicle (bus, Samand, and trailer), and the type of collision (vehicle motorcycle and overturning motorcycle) have been identified as the most influential parameters with positive coefficients in increasing accident severity.

(6) In addition, the role of pickup vehicles in accident occurrence, especially at 12:00 to 18:00 accident time (the time people return from the workplace) was significant. Also, the reason for the accident (exceeding safe speed) declined the likelihood of vehicle accidents due to its negative coefficient. Therefore, pickup and KIA Pride vehicles have been identified as the most common types of vehicles that were involved in the occurrence of accidents. So reconsidering the design of these vehicles as well as addressing their technical problems is necessary. Moreover, by giving suitable warnings to drivers, such as intelligent warning signs or other types of warnings, it is possible to provide an effective way for drivers in controlling their speed and improving safety.

(7) Based on the results of ANN models, by comparing the predictive accuracy of the MLPNN and the RBFNN models, it was found that the MLPNN model provided better performance in predicting accident severity in terms of accuracy and efficiency. Moreover, the comparison of the ROC curves of the RBFNN and MLPNN models indicated the prediction accuracy of the MLPNN model in the severity of accidents was more. The result of the MLPNN model illustrated the variables of gender of the driver, accident time, weather condition, driver age, and education rate had the most influential effect on predicting accidents; however, in the RBFNN model, the variables of gender of the driver, education rate, condition of weather, and driver age were found to be the most effective variables in accident severity. By comparing the MLPNN and RBFNN models, it was indicated that gender of the driver was the most significant variable in accident severity, which affirms the role of the human factor in accident occurrence. Finally, the performance of MLP and RBF models were evaluated by other kinds of ANN models developed by MATLAB programming.

There are, however, some limitations. MLP and RBF models as well as statistical methods do not provide the sufficient ability to consider all the required details of the problem. Thus, it is recommended for future studies to utilize deep learning methods such as the recurrent neural network (RNN), the convolutional neural network (CNN), and so on, recently successfully employed to design models with higher accuracy and efficiency in predicting accident severity [103–105]. Moreover, since the transportation sector was the second largest contributor to CO_2 emissions, a study can be done to investigate the impact of pollutants on accidents in the continuation of this study [106]. By understanding the perception of road users concerning the facility improvements, these approaches, along with a survey analysis, can also improve work zone safety [107]. Various data collection methods can be incorporated into the proposed approach to obtain more variables [108, 109]. In addition, for future studies, it is recommended to utilize other kinds of cross-validation techniques [110, 111], machine learning methods [112–119], and optimization algorithms [120–124] to make a better decision about the interference of nonmotorized users and vehicle flow within a city and to develop models with high prediction accuracy in taking preventive measures for decreasing pedestrian accidents in urban environments. Also, various crash modeling and before-after safety evaluation methods can be incorporated into the proposed approaches to examine risk factors on crash rates [125–128].

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 they have no conflicts of interest.

Acknowledgments

The authors would like to express their gratitude to the traffic police for their cooperation in granting them access to accidents data for the purpose of the present research.

References

[1] N. Kamboozia, M. Ameri, and S. M. Hosseinian, "Investigation of effective factors in the severity of rural road accidents in guilan to determine the most effective factors and provide safety solutions," *Road*, vol. 29, no. 106, pp. 115–128, 2021.

[2] World Health Organization, *Global Status Report on Road Safety 2018*, China, 2018.

[3] S. M. Hosseinian, N. Kamboozia, and M. Ameri, "Prioritization of effective factors in rural road accidents of guilan province based on exploratory factor Analysis and logistic regression model," *Journal of Transportation Engineering*, vol. 13, no. 2, pp. 1489–1509, 2022.

[4] H. Mohammadpour, S. M. Hosseinian, and V. Najafi Moghaddam Gilani, "Analysis and investigation of pedestrian accidents on suburban roads of Rasht and providing safety solutions to reduce traffic accidents," *Quarterly Journal of Gilan Police Science*, vol. 10, no. 39, pp. 91–108, 2021.

[5] A. Sharf Aldeen, A. Abdi Kordani, S. M. Hosseinian, and M. Mehdi Nazar, "Evaluation of the application of maximum radius in horizontal curves using vehicle dynamic simulation," *Advances in Civil Engineering*, vol. 2022, pp. 1–19, Article ID 5237541, 2022.

[6] M. Heidarian, A. Abdi Kordani, A. Sarkar, and S. M. Hosseinian, "Location of emergency escape ramps on

two-lane rural highways based on the fuzzy multicriteria decision-making method," *Journal of Engineering*, vol. 2022, pp. 1–20, Article ID 7699870, 2022.

[7] S. Abolfathi, M. Saffarzadeh, H. Mirzahossein, and S. M. Hosseinian, "Signalized Intersection Delay Analysis Using Transit Signal Priority (Tsp) and Dedicated Short-Range Communication (Dsrc) System for Bus Rapid Transit (Brt)," *Innovative Infrastructure Solutions*, vol. 7, 2022.

[8] H. Yuan and B. Yang, "System dynamics approach for evaluating the interconnection performance of cross-border transport infrastructure," *Journal of Management in Engineering*, vol. 38, no. 3, Article ID 04022008, 2022.

[9] H. Mirzahossein, M. Sashurpour, S. M. Hosseinian, and V. N. M. Gilani, "Presentation of machine learning methods to determine the most important factors affecting road traffic accidents on rural roads," *Frontiers of Structural and Civil Engineering*, vol. 12, pp. 1–10, 2022.

[10] S. M. Hosseinian, V. Najafi Moghaddam Gilani, B. Mirbaha, and A. Abdi Kordani, "Statistical analysis for study of the effect of dark clothing color of female pedestrians on the severity of accident using machine learning methods," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5567638, 21 pages, 2021.

[11] M. B. Movahhed, J. Ayoubinejad, F. N. Asl, and M. Feizbahr, "The effect of rain on pedestrians crossing speed," *Computational Research Progress in Applied Science and Engineering*, vol. 6, no. 3, 2020.

[12] H. T. Abdelwahab and M. A. Abdel-Aty, "Artificial neural networks and logit models for traffic safety analysis of toll plazas," *Transportation Research Record*, vol. 1784, no. 1, pp. 115–125, 2002.

[13] C. Xie, J. Lu, and E. Parkany, "Work travel mode choice modeling with data mining: decision trees and neural networks," *Transportation Research Record*, vol. 1854, no. 1, pp. 50–61, 2003.

[14] J.-K. Kim, S. Kim, G. F. Ulfarsson, and L. A. Porrello, "Bicyclist injury severities in bicycle–motor vehicle accidents," *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 238–251, 2007.

[15] L. Boodlal, E. T. Donnell, R. J. Porter et al., *Factors Influencing Operating Speeds and Safety on Rural and Suburban Roads*, NY China, 2015.

[16] H. Omrani, "Predicting travel mode of individuals by machine learning," *Transportation Research Procedia*, vol. 10, pp. 840–849, 2015.

[17] H. Huang, Q. Zeng, X. Pei, S. C. Wong, and P. Xu, "Predicting crash frequency using an optimised radial basis function neural network model," *Transportmetrica: Transport Science*, vol. 12, no. 4, pp. 330–345, 2016.

[18] S. Alkheder, M. Taamneh, and S. Taamneh, "Severity prediction of traffic accident using an artificial neural network," *Journal of Forecasting*, vol. 36, no. 1, pp. 100–108, 2017.

[19] H. Behbahani, A. M. Amiri, R. Imaninasab, and M. Alizamir, "Forecasting accident frequency of an urban road network: a comparison of four artificial neural network techniques," *Journal of Forecasting*, vol. 37, no. 7, pp. 767–780, 2018.

[20] M. Rezapour, S. Nazneen, and K. Ksaibati, "Application of deep learning techniques in predicting motorcycle crash severity," *Engineering Reports*, vol. 2, no. 7, Article ID e12175, 2020.

[21] V. Najafi Moghaddam Gilani, S. M. Hosseinian, M. Ghasedi, and M. Nikookar, "Data-driven urban traffic accident analysis and prediction using logit and machine learning-based pattern recognition models," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9974219, 2111 pages, 2021.

[22] G. Luo, H. Zhang, Q. Yuan, J. Li, and F. Y. Wang, "ESTNet: embedded spatial-temporal network for modeling traffic flow dynamics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 2022, pp. 1–12, Article ID 3167019, 2022.

[23] I. Bargegol, V. Najafi Moghaddam Gilani, S. M. Hosseinian, and M. Habibzadeh, "Pedestrians crossing and walking speeds analysis in urban areas under the influence of rain and personality characteristics," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–13, Article ID 7768160, 2022.

[24] V. Najafi Moghaddam Gilani, M. R. Ghanbari Tamrin, S. M. Hosseinian, M. Nikookar, D. Safari, and S. YektaParast, "Investigation of bus special lane performance using statistical analysis and optimization of the signalized intersection delay by machine learning methods," *Journal of Optimization*, vol. 2022, pp. 1–24, Article ID 2984803, 2022.

[25] H. Hariri Asli, "Investigation of the factors affecting pedestrian accidents in urban roundabouts," *Computational Research Progress in Applied Science and Engineering*, vol. 8, no. 1, 4 pages, 2022.

[26] S. M. Hosseinian, V. Najafi Moghaddam Gilani, H. Tahmasbi Amoli, M. Nikookar, and A. Orouei, "Presentation of analytical methods for better decision making about the most important factor influencing rural accidents," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5564269, 2116 pages, 2021.

[27] I. Bargegol, S. M. Hosseinian, V. Najafi Moghaddam Gilani, M. Nikookar, and A. Orouei, "Presentation of regression analysis, GP and GMDH models to predict the pedestrian density in various urban facilities," *Frontiers of Structural and Civil Engineering*, vol. 2, pp. 250–265, 2022.

[28] Z. Said, M. Jamei, L. Syam Sundar, A. Pandey, A. Allouhi, and C. Li, "Thermophysical properties of water, water and ethylene glycol mixture-based nanodiamond+ Fe3O4 hybrid nanofluids: an experimental assessment and application of data-driven approaches," *Journal of Molecular Liquids*, vol. 347, Article ID 117944, 2022.

[29] V. Najafi Moghaddam Gilani, M. Sashurpour, S. Hassanjani, and S. M. Hosseinian, "Analysis of the effect of the speed factor on highway safety using the machine learning method," *Slovak Journal of Civil Engineering*, vol. 29, no. 4, pp. 19–28, 2021.

[30] Z. Said, N. K. Cakmak, P. Sharma et al., "Synthesis, stability, density, viscosity of ethylene glycol-based ternary hybrid nanofluids: experimental investigations and model-prediction using modern machine learning techniques," *Powder Technology*, vol. 400, Article ID 117190, 2022.

[31] B. Alizadeh and A. H. Behzadan, "Crowdsourced-based deep convolutional networks for urban flood depth mapping," *European Conference on Computing in Construction*, Ixia, Rhodes, Greece, 2022.

[32] L.-Y. Chang, "Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network," *Safety Science*, vol. 43, no. 8, pp. 541–557, 2005.

[33] L. Zhou, Q. Fan, X. Huang, and Y. Liu, "Weak and strong convergence analysis of Elman neural networks via weight decay regularization," *Optimization*, vol. 2022, pp. 1–23, Article ID 2057852, 2022.

[34] V. Najafi Moghaddam Gilani, S. M. Hosseinian, and M. Nikookar, "Presentation of a new deicer with the least moisture and fatigue failures in asphalt mixtures," *Arabian*

Journal for Science and Engineering, vol. 46, no. 11, pp. 10457–10471, 2021.

[35] E. Eslami and H.-B. Yun, "Attention-based multi-scale convolutional neural network (A+ MCNN) for multi-class classification in road images," *Sensors*, vol. 21, no. 15, p. 5137, 2021.

[36] Q. Fan, Z. Zhang, and X. Huang, "Parameter conjugate gradient with secant equation based elman neural network and its convergence analysis," *Advanced Theory and Simulations*, Article ID 2200047, 2022.

[37] J. Li, K. Xu, S. Chaudhuri, E. Yumer, H. Zhang, and L. Guibas, "Grass: generative recursive autoencoders for shape structures," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.

[38] C. Eswaran, C. Eswaran, R. Logeswaran, and A. R. A. Rahman, "Prediction models for early risk detection of cardiovascular event," *Journal of Medical Systems*, vol. 36, no. 2, pp. 521–531, 2012.

[39] J. Wang, J. Tian, X. Zhang et al., "Control of time delay force feedback teleoperation system with finite time convergence," *Frontiers in Neurorobotics*, vol. 16, Article ID 877069, 2022.

[40] C. Liu, D. Wu, Y. Li, and Y. Du, "Large-scale pavement roughness measurements with vehicle crowdsourced data using semi-supervised learning," *Transportation Research Part C: Emerging Technologies*, vol. 125, Article ID 103048, 2021.

[41] Z. Zhang et al., "Residual stress of MoS2 nano-lubricant grinding cemented carbide," *International Journal of Advanced Manufacturing Technology*, vol. 1, pp. 1–15, 2021.

[42] M. Yang, C. Li, Z. Said et al., "Semiempirical heat flux model of hard-brittle bone material in ductile microgrinding," *Journal of Manufacturing Processes*, vol. 71, pp. 501–514, 2021.

[43] X. Wang, C. Li, Y. Zhang et al., "Influence of texture shape and arrangement on nanofluid minimum quantity lubrication turning," *International Journal of Advanced Manufacturing Technology*, vol. 119, no. 1-2, pp. 631–646, 2022.

[44] H. Li, Y. Zhang, C. Li et al., "Cutting fluid corrosion inhibitors from inorganic to organic: progress and applications," *Korean Journal of Chemical Engineering*, vol. 39, no. 5, pp. 1107–1134, 2022.

[45] Z. Said, S. Arora, S. Farooq, L. S. Sundar, C. Li, and A. Allouhi, "Recent advances on improved optical, thermal, and radiative characteristics of plasmonic nanofluids: academic insights and perspectives," *Solar Energy Materials and Solar Cells*, vol. 236, Article ID 111504, 2022.

[46] M. Yang, C. Li, Y. Zhang et al., "Effect of friction coefficient on chip thickness models in ductile-regime grinding of zirconia ceramics," *International Journal of Advanced Manufacturing Technology*, vol. 102, no. 5-8, pp. 2617–2632, 2019.

[47] D. Jia, L. Zhang-he, Z. Yan-bin et al., "Grinding performance and surface morphology evaluation of titanium alloy using electric traction bio micro lubricant," *Journal of Mechanical Engineering*, vol. 58, no. 5, pp. 198–211, 2022.

[48] A. E. Anqi, C. Li, H. A. Dhahad et al., "Effect of combined air cooling and nano enhanced phase change materials on thermal management of lithium-ion batteries," *Journal of Energy Storage*, vol. 52, Article ID 104906, 2022.

[49] H. Ziari, A. Amini, A. Goli, and D. Mirzaiyan, "Predicting rutting performance of carbon nano tube (CNT) asphalt binders using regression models and neural networks," *Construction and Building Materials*, vol. 160, pp. 415–426, 2018.

[50] B. Sharma, S. Kumar, P. Tiwari, P. Yadav, and M. I. Nezhurina, "ANN based short-term traffic flow forecasting in undivided two lane highway," *Journal of Big Data*, vol. 5, no. 1, pp. 48–16, 2018.

[51] A. M. Amiri, A. Sadri, N. Nadimi, and M. Shams, "A comparison between artificial neural network and hybrid intelligent genetic algorithm in predicting the severity of fixed object crashes among elderly drivers," *Accident Analysis & Prevention*, vol. 138, Article ID 105468, 2020.

[52] M. Habibzadeh, M. Ameri, S. M. Sadat Haghighi, and H. Ziari, "Application of artificial neural network approaches for predicting accident severity on rural roads (case study: tehran-qom and tehran-saveh rural roads)," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–14, Article ID 5214703, 2022.

[53] F. Paolo, F. Gianfranco, F. Luca et al., "Investigating the role of the human element in maritime accidents using semi-supervised hierarchical methods," *Transportation Research Procedia*, vol. 52, pp. 252–259, 2021.

[54] A. Zaib, J. Yin, and R. U. Khan, "Determining role of human factors in maritime transportation accidents by fuzzy fault tree analysis (FFTA)," *Journal of Marine Science and Engineering*, vol. 10, no. 3, p. 381, 2022.

[55] P. Cullen, H. Moller, M. Woodward et al., "Are there sex differences in crash and crash-related injury between men and women? A 13-year cohort study of young drivers in Australia," *SSM-population health*, vol. 14, Article ID 100816, 2021.

[56] Y. Liu, Y. Liu, X. X. Ma, and W. L. Qiao, "A comprehensive model for human factor risk assessment: hfacs-fft-ann," in *Proceedings of the in 5th Annual International Conference on Management, Economics and Social Development (ICMESD 2019)*, Atlantis Press, Amsterdam, Netherlands, September 2019.

[57] J. P. Bernstein and M. Calamia, "Dimensions of driving-related emotions and behaviors: an exploratory factor analysis of common self-report measures," *Accident Analysis & Prevention*, vol. 124, pp. 85–91, 2019.

[58] N. Casado-Sanz, B. Guirao, and D. Gálvez-Pérez, "Population ageing and rural road accidents: analysis of accident severity in traffic crashes with older pedestrians on Spanish crosstown roads," *Research in Transportation Business & Management*, vol. 30, Article ID 100377, 2019.

[59] C. A. Butts, R. Gonzalez, L. Nguyen et al., "Twelve-year review of urban versus rural off-road vehicle injuries at a level 1 trauma center," *Journal of Surgical Research*, vol. 233, pp. 331–334, 2019.

[60] N. M. Pawar, R. K. Khanuja, P. Choudhary, and N. R. Velaga, "Modelling braking behaviour and accident probability of drivers under increasing time pressure conditions," *Accident Analysis & Prevention*, vol. 136, Article ID 105401, 2020.

[61] B. N. Husain, M. S. Vyawahare, and R. Pate, "Correlation of injury severity score with survival time in fatal road traffic accidents in central Indian population," *Prof.(Dr) RK Sharma*, vol. 20, no. 2, p. 237, 2020.

[62] F. Malin, I. Norros, and S. Innamaa, "Accident risk of road and weather conditions on different road types," *Accident Analysis & Prevention*, vol. 122, pp. 181–188, 2019.

[63] K. Gomes-Franco, M. Rivera-Izquierdo, L. M. Martin-delosReyes, E. Jimenez-Mejias, and V. Martinez-Ruiz, "Explaining the association between driver's age and the risk of causing a road crash through mediation analysis,"

International Journal of Environmental Research and Public Health, vol. 17, no. 23, p. 9041, 2020.
[64] D. Topolšek, D. Babić, and M. Fiolić, "The effect of road safety education on the relationship between Driver's errors, violations and accidents: slovenian case study," *European transport research review*, vol. 11, no. 1, pp. 18–8, 2019.
[65] M. Haghighi, F. Bakhtari, H. Sadeghi-Bazargani, and H. Nadrian, "Strategies to promote pedestrian safety from the viewpoints of traffic and transport stakeholders in a developing country: a mixed-method study," *Journal of Transport & Health*, vol. 22, Article ID 101125, 2021.
[66] M. J. Giummarra, B. Beck, and B. J. Gabbe, "Classification of road traffic injury collision characteristics using text mining analysis: implications for road injury prevention," *PLoS One*, vol. 16, no. 1, Article ID e0245636, 2021.
[67] I. Ashraf, S. Hur, M. Shafiq, and Y. Park, "Catastrophic factors involved in road accidents: underlying causes and descriptive analysis," *PLoS One*, vol. 14, no. 10, Article ID e0223473, 2019.
[68] N. Kamboozia, M. Ameri, and S. M. Hosseinian, "Statistical analysis and accident prediction models leading to pedestrian injuries and deaths on rural roads in Iran," *International Journal of Injury Control and Safety Promotion*, vol. 27, no. 4, pp. 493–509, 2020.
[69] M. Ghasedi, M. Sarfjoo, and I. Bargegol, "Prediction and analysis of the severity and number of suburban accidents using logit model, factor analysis and machine learning: a case study in a developing country," *SN Applied Sciences*, vol. 3, no. 1, pp. 13–16, 2021.
[70] C. Zhao, F. Liao, X. Li, and Y. Du, "Macroscopic modeling and dynamic control of on-street cruising-for-parking of autonomous vehicles in a multi-region urban road network," *Transportation Research Part C: Emerging Technologies*, vol. 128, Article ID 103176, 2021.
[71] F. Faghihinejad and S. Monajem, "Pathology of the Disabled People Access to Public Transport and Prioritizing Practical Solution," *Tobacco Regulatory Science (TRS)*, vol. 7, pp. 7634–7643, 2021.
[72] H. Cheng, L. Sun, Y. Wang, and X. Chen, "Effects of actual loading waveforms on the fatigue behaviours of asphalt mixtures," *International Journal of Fatigue*, vol. 151, Article ID 106386, 2021.
[73] H. Cheng, L. Liu, and L. Sun, "Bridging the gap between laboratory and field moduli of asphalt layer for pavement design and assessment: a comprehensive loading frequency-based approach," *Frontiers of Structural and Civil Engineering*, vol. 16, no. 3, pp. 267–280, 2022.
[74] G. H. Hamedi, M. R. Esmaeeli, V. Najafi Moghaddam Gilani, and S. M. Hosseinian, "The effect of aggregate-forming minerals on thermodynamic parameters using surface free energy concept and its relationship with the moisture susceptibility of asphalt mixtures," *Advances in Civil Engineering*, vol. 2021, Article ID 8818681, 2115 pages, 2021.
[75] S. M. Hosseinian, V. Najafi Moghaddam Gilani, P. Mehraban Joobani, and M. Arabani, "Investigation of moisture sensitivity and conductivity properties of inductive asphalt mixtures containing steel wool fiber," *Advances in Civil Engineering*, vol. 2020, pp. 1–9, Article ID 8890814, 2020.
[76] V. Najafi Moghaddam Gilani, S. M. Hosseinian, D. Safari, and M. Bagheri Movahhed, "Investigation of the impact of deicer materials on thermodynamic parameters and its relationship with moisture susceptibility in modified asphalt mixtures by carbon nanotube," *Arabian Journal for Science and Engineering*, vol. 46, no. 5, pp. 4489–4502, 2021.
[77] H. Ziari, H. Divandari, S. M. Seyed Ali Akbar, and S. M. Hosseinian, "Investigation of the effect of crumb rubber powder and warm additives on moisture resistance of SMA mixtures," *Advances in Civil Engineering*, vol. 2021, Article ID 6653594, 2112 pages, 2021.
[78] T. Gao, C. Li, Y. Zhang et al., "Dispersing mechanism and tribological performance of vegetable oil-based CNT nanofluids with different surfactants," *Tribology International*, vol. 131, pp. 51–63, 2019.
[79] Y. Zhang, C. Li, D. Jia et al., "Experimental study on the effect of nanoparticle concentration on the lubricating property of nanofluids for MQL grinding of Ni-based alloy," *Journal of Materials Processing Technology*, vol. 232, pp. 100–115, 2016.
[80] B. Li, C. Li, Y. Zhang et al., "Heat transfer performance of MQL grinding with different nanofluids for Ni-based alloys using vegetable oil," *Journal of Cleaner Production*, vol. 154, pp. 1–11, 2017.
[81] S. Guo, C. Li, Y. Zhang et al., "Experimental evaluation of the lubrication performance of mixtures of castor oil with other vegetable oils in MQL grinding of nickel-based alloy," *Journal of Cleaner Production*, vol. 140, pp. 1060–1076, 2017.
[82] Y. Wang, C. Li, Y. Zhang et al., "Experimental evaluation of the lubrication properties of the wheel/workpiece interface in MQL grinding with different nanofluids," *Tribology International*, vol. 99, pp. 198–210, 2016.
[83] T. Gao, C. Li, D. Jia et al., "Surface morphology assessment of CFRP transverse grinding using CNT nanofluid minimum quantity lubrication," *Journal of Cleaner Production*, vol. 277, Article ID 123328, 2020.
[84] T. Gao, C. Li, M. Yang et al., "Mechanics analysis and predictive force models for the single-diamond grain grinding of carbon fiber reinforced polymers using CNT nano-lubricant," *Journal of Materials Processing Technology*, vol. 290, Article ID 116976, 2021.
[85] D. Jia, C. Li, D. Zhang, Y. Zhang, and X. Zhang, "Experimental verification of nanoparticle jet minimum quantity lubrication effectiveness in grinding," *Journal of Nanoparticle Research*, vol. 16, no. 12, pp. 2758–2815, 2014.
[86] X. Cui, C. Li, W. Ding et al., "Minimum quantity lubrication machining of aeronautical materials using carbon group nanolubricant: from mechanisms to application," *Chinese Journal of Aeronautics*, vol. 1, 2021.
[87] X. Wu, C. Li, Z. Zhou et al., "Circulating purification of cutting fluid: an overview," *International Journal of Advanced Manufacturing Technology*, vol. 117, no. 9-10, pp. 2565–2600, 2021.
[88] T. Gao, Y. Zhang, C. Li et al., "Grindability of carbon fiber reinforced polymer using CNT biological lubricant," *Scientific Reports*, vol. 11, no. 1, pp. 22535–22614, 2021.
[89] T. Gao, C. Li, Y. Wang et al., "Carbon fiber reinforced polymer in drilling: from damage mechanisms to suppression," *Composite Structures*, vol. 286, Article ID 115232, 2022.
[90] H. Li, Y. Zhang, C. Li et al., "Extreme pressure and antiwear additives for lubricant: academic insights and perspectives," *International Journal of Advanced Manufacturing Technology*, vol. 120, no. 1-2, pp. 1–27, 2022.
[91] D. Jia, Y. Zhang, C. Li et al., "Lubrication-enhanced mechanisms of titanium alloy grinding using lecithin biolubricant," *Tribology International*, vol. 169, Article ID 107461, 2022.
[92] T. Gao, Y. Zhang, C. Li et al., "Fiber-reinforced composites in milling and grinding: machining bottlenecks and advanced

strategies," *Frontiers of Mechanical Engineering*, vol. 17, no. 2, pp. 24–35, 2022.

[93] D. Zhang, C. Li, Y. Zhang, D. Jia, and X. Zhang, "Experimental research on the energy ratio coefficient and specific grinding energy in nanoparticle jet MQL grinding," *International Journal of Advanced Manufacturing Technology*, vol. 78, no. 5-8, pp. 1275–1288, 2015.

[94] C. H. Li, J. Y. Li, S. Wang, and Q. Zhang, "Modeling and numerical simulation of the grinding temperature field with nanoparticle jet of MQL," *Advances in Mechanical Engineering*, vol. 5, Article ID 986984, 2013.

[95] Z. Said, M. Ghodbane, B. Boumeddane et al., "Energy, exergy, economic and environmental (4E) analysis of a parabolic trough solar collector using MXene based silicone oil nanofluids," *Solar Energy Materials and Solar Cells*, vol. 239, Article ID 111633, 2022.

[96] Q. Yin, C. Li, L. Dong et al., "Effects of physicochemical properties of different base oils on friction coefficient and surface roughness in MQL milling AISI 1045," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 8, no. 6, pp. 1629–1647, 2021.

[97] Y. Zhang, C. Li, D. Jia, D. Zhang, and X. Zhang, "Experimental evaluation of the lubrication performance of MoS2/CNT nanofluid for minimal quantity lubrication in Ni-based alloy grinding," *International Journal of Machine Tools and Manufacture*, vol. 99, pp. 19–33, 2015.

[98] Y. Zhang, C. Li, D. Jia, D. Zhang, and X. Zhang, "Experimental evaluation of MoS2 nanoparticles in jet MQL grinding with different types of vegetable oil as base oil," *Journal of Cleaner Production*, vol. 87, pp. 930–940, 2015.

[99] Y. Wang, C. Li, Y. Zhang et al., "Experimental evaluation of the lubrication properties of the wheel/workpiece interface in minimum quantity lubrication (MQL) grinding using different types of vegetable oils," *Journal of Cleaner Production*, vol. 127, pp. 487–499, 2016.

[100] Y. Zhang, C. Li, H. Ji et al., "Analysis of grinding mechanics and improved predictive force model based on material-removal and plastic-stacking mechanisms," *International Journal of Machine Tools and Manufacture*, vol. 122, pp. 81–97, 2017.

[101] J. Chen, Q. Wang, and J. Huang, "Motorcycle ban and traffic safety: evidence from a quasi-experiment at Zhejiang, China," *Journal of Advanced Transportation*, vol. 2021, Article ID 7552180, 2113 pages, 2021.

[102] Z. Li, L. Chen, L. Nie, and S. X. Yang, "A novel learning model of driver fatigue features representation for steering wheel angle," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 269–281, 2022.

[103] F. Liu, G. Zhang, and J. Lu, "Multisource heterogeneous unsupervised domain adaptation via fuzzy relation neural networks," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 11, pp. 3308–3322, 2021.

[104] W. Zhou, Q. Guo, J. Lei, L. Yu, and J. N. Hwang, "IRFR-Net: interactive recursive feature-reshaping network for detecting salient objects in RGB-D images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 2021, pp. 1–13, Article ID 3105484, 2021.

[105] Y. Li, P. Che, C. Liu, D. Wu, and Y. Du, "Cross-scene pavement distress detection by a novel transfer learning framework," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 11, pp. 1398–1415, 2021.

[106] Y. Xiao, X. Zuo, J. Huang, A. Konak, and Y. Xu, "The continuous pollution routing problem," *Applied Mathematics and Computation*, vol. 387, Article ID 125072, 2020.

[107] E. E. Miller, S. Pourfalatoun, A. Nylen, and T. Weldon, *Evaluating the Human-Automated Maintenance Vehicle Interaction for Improved Safety and Facilitating Long-Term Trust*, 2021.

[108] G. Liu, "Data collection in mi-assisted wireless powered underground sensor networks: directions, recent advances, and challenges," *IEEE Communications Magazine*, vol. 59, no. 4, pp. 132–138, 2021.

[109] Y. Du, B. Qin, C. Zhao, Y. Zhu, J. Cao, and Y. Ji, "A novel spatio-temporal synchronization method of roadside asynchronous MMW radar-camera for sensor fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 112 pages, 2021.

[110] X. Wu, W. Zheng, X. Chen, Y. Zhao, T. Yu, and D. Mu, "Improving high-impact bug report prediction with combination of interactive machine learning and active learning," *Information and Software Technology*, vol. 133, Article ID 106530, 2021.

[111] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M. W. Wu, and T. Luo, "Local and global feature learning for blind quality evaluation of screen content and natural scene images," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2086–2095, 2018.

[112] R. Sun, J. Wang, Q. Cheng, Y. Mao, and W. Y. Ochieng, "A new IMU-aided multiple GNSS fault detection and exclusion algorithm for integrated navigation in urban environments," *GPS Solutions*, vol. 25, no. 4, pp. 147–217, 2021.

[113] R. Sun, Z. Zhang, Q. Cheng, and W. Y. Ochieng, "Pseudorange error prediction for adaptive tightly coupled GNSS/IMU navigation in urban areas," *GPS Solutions*, vol. 26, no. 1, pp. 28–13, 2022.

[114] W. Zheng, Y. Xun, X. Wu, Z. Deng, X. Chen, and Y. Sui, "A comparative study of class rebalancing methods for security bug report classification," *IEEE Transactions on Reliability*, vol. 70, no. 4, pp. 1658–1670, 2021.

[115] D. Li, S. S. Ge, and T. H. Lee, "Simultaneous arrival to origin convergence: sliding-mode control through the norm-normalized sign function," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 1966–1972, 2022.

[116] X. Gong, L. Wang, Y. Mou et al., "Improved Four-channel PBTDPA control strategy using force feedback bilateral teleoperation system," *International Journal of Control, Automation and Systems*, vol. 20, no. 3, pp. 1002–1017, 2022.

[117] S. Liu, X. Zhang, W. Zheng, and B. Yang, "Adaptive control for time-delay teleoperation systems with uncertain dynamics," *Journal of Physics: Conference Series*, vol. 887, 2017.

[118] M. Yang, C. Li, L. Luo, R. Li, and Y. Long, "Predictive model of convective heat transfer coefficient in bone micro-grinding using nanofluid aerosol cooling," *International Communications in Heat and Mass Transfer*, vol. 125, Article ID 105317, 2021.

[119] Z. Duan, C. Li, W. Ding et al., "Milling force model for aviation aluminum alloy: academic insight and perspective analysis," *Chinese Journal of Mechanical Engineering*, vol. 34, no. 1, pp. 18–35, 2021.

[120] H. Wang, Q. Gao, H. Li, H. Wang, L. Yan, and G. Liu, "A structural evolution-based anomaly detection method for generalized evolving social networks," *The Computer Journal*, vol. 65, no. 5, pp. 1189–1199, 2022.

[121] X. Wu, W. Zheng, X. Xia, and D. Lo, "Data quality matters: a case study on data label correctness for security bug report prediction," *IEEE Transactions on Software Engineering*, vol. 48, no. 7, pp. 2541–2556, 2022.

[122] Y. Wang, C. Li, Y. Zhang et al., “Processing characteristics of vegetable oil-based nanofluid MQL for grinding different workpiece materials,” *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 5, no. 2, pp. 327–339, 2018.

[123] Y. Zhang, H. N. Li, C. Li et al., “Nano-enhanced biolubricant in sustainable manufacturing: from processability to mechanisms,” *Friction*, vol. 10, no. 6, pp. 803–841, 2022.

[124] Z. Said, L. S. Sundar, H. Rezk, A. M. Nassef, S. Chakraborty, and C. Li, “Thermophysical properties using ND/water nanofluids: an experimental study, ANFIS-based model and optimization,” *Journal of Molecular Liquids*, vol. 330, Article ID 115659, 2021.

[125] Y. Guo, T. Sayed, and L. Zheng, “A hierarchical bayesian peak over threshold approach for conflict-based before-after safety evaluation of leading pedestrian intervals,” *Accident Analysis & Prevention*, vol. 147, Article ID 105772, 2020.

[126] Y. Guo, T. Sayed, and M. Essa, “Real-time conflict-based Bayesian Tobit models for safety evaluation of signalized intersections,” *Accident Analysis & Prevention*, vol. 144, Article ID 105660, 2020.

[127] Y. Guo, Z. Li, P. Liu, and Y. Wu, “Modeling correlation and heterogeneity in crash rates by collision types using full Bayesian random parameters multivariate Tobit model,” *Accident Analysis & Prevention*, vol. 128, pp. 164–174, 2019.

[128] Y. Guo, A. Osama, and T. Sayed, “A cross-comparison of different techniques for modeling macro-level cyclist crashes,” *Accident Analysis & Prevention*, vol. 113, pp. 38–46, 2018.