Research Article

Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models

Vahid Najafi Moghaddam Gilani ,[1] **Seyed Mohsen Hosseinian** ,[1] **Meisam Ghasedi** ,[2] **and Mohammad Nikookar** [2]

[1]*School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran*
[2]*Department of Civil Engineering, University of Guilan, Rasht, Iran*

Correspondence should be addressed to Meisam Ghasedi; meisam_ghasedi@msc.guilan.ac.ir

Modeling the severity of accidents based on the most effective variables accounts for developing a high-precision model presenting the possibility of occurrence of each category of future accidents, and it could be utilized to prioritize the corrective measures for authorities. The purpose of this study is to identify the variables affecting the severity of the injury, fatal, and property damage only (PDO) accidents in Rasht city by collecting information on urban accidents from March 2019 to March 2020. In this regard, the multiple logistic regression and the pattern recognition type of artificial neural network (ANN) as a machine learning solution are used to recognize the most influential variables on the severity of accidents and the superior approach for accident prediction. Results show that the multiple logistic regression in the forward stepwise method has R^2 of 0.854 and an accuracy prediction power of 89.17%. It turns out that the accidents occurred between 18 and 24 and KIA Pride vehicle has the highest effect on increasing the severity of accidents, respectively. The most important result of the logit model accentuates the role of environmental variables, including poor lighting conditions alongside unfavorable weather and the dominant role of unsafe and poor quality of vehicles on increasing the severity of accidents. In addition, the machine learning model performs significantly better and has higher prediction accuracy (98.9%) than the logit model. In addition, the ANN model's greater power to predict and estimate future accidents is confirmed through performance and sensitivity analysis.

1. Introduction

Transportation, like any other industry and phenomenon, along with its advantages, has its disadvantages and limitations for road users [1, 2]. Traffic and its related predicaments have been on the rise all over the world and have had a detrimental effect on the lives and property of the people of the community. Urban pollution, rising fuel and energy consumption, wasting millions of hours a day in traffic congestion, wasting community service facilities and national assets, and, ultimately, the occurrence of accidents resulting in injury, death, and property damage are the result of poor traffic facilities and conditions [3–5].

Road traffic accidents now represent the eighth leading cause of death globally. In addition, road traffic fatalities have increased to 1.35 million a year and caused up to 50 million injuries in 2016; that is, nearly 3700 people die on the world's roads every day. In Iran, the reported number of road traffic deaths is 15932 in 2016 [6]. The statistics also reveals the high number of injury and property damage only (PDO) accidents in urban and suburban roads in the country. Therefore, the effect of factors affecting the severity of accidents should be investigated meticulously to provide practical solutions for improving safety and reducing the high number of accidents.

Various studies have been carried out on the subject area of traffic safety in recent years using multiple logistic regression. Sherafati et al. explored road traffic fatalities after receiving emergency services using multiple logistic regression in Langarud. Results showed that males,

motorbikes, and pedestrians had a positive and significant relationship with fatal accidents [7]. Intini et al. conducted a study to investigate the relationships between road familiarity/unfamiliarity and the occurrence of accidents using multiple logistic regression. Factor analysis is a very vital step in many applications. The factors of minor intersections/driveways, autumn/winter, and speed limits less than 80 km/h were more related to familiar driver crashes, but the factors of head-on and rear-end accidents, summer, and heavy vehicles were more related to unfamiliar drivers involved in accidents [8]. Also, Casado-Sanz et al. applied a logistic regression model to investigate the impact of age on accident severity in the rural crosstown road and concluded that female drivers and motorbikes had a negative impact on the likelihood of accidents [9].

Generally, metaheuristic algorithms and machine learning techniques have been widely used in different engineering studies, especially in transportation problems, which they are in desperate need of complex and accurate solutions to provide more accurate prediction models than statistical methods due to their capability of handlig more complex functions and classification problems [10–25]. Pattern recognition tools and their accurate analysis using optimized prediction tasks are a trendy topic in the two recent years [26–30]. Supplementary to this, various prediction methods have been used in different engineering problems by the emergence of various datasets [31–35]. Some studies have applied traditional methods such as regression models alongside machine learning approaches to validate the ANN-based solutions as an effective alternative method to predict the severity of accidents with higher accuracy [36–44]. Because of the close connection of accidents to human life, machine learning approaches would be applied in important functions, like predicting the type and severity of accidents, due to their higher precision [45, 46]. In order to understand the effects of the determining factors in an activity and predict incidents in the future, nonlinear links between variables with different forms of ANN may be modeled [47–49].

Chimba and Sando investigated the severity of traffic accidents using ANN and compared this method's accuracy with the ordered probit model. Results indicated that the prediction accuracy of ANN was higher [50]. Khair et al. tried to present an accident prediction model using a new framework of an artificial neural network (ANN). They affirmed that the predicted collisions were similar to the number of real incidents and thus found the proposed model [51]. Omrani presented multiple logistic regression and ANN models to predict individuals' travel mode in Luxembourg. Results showed that ANN models had better performance [52]. In order to provide a decent prediction for traffic collisions in metropolitan areas of Nuevo León, Contreras et al. utilized an innovative ANN model. The programming feature of Scilab software has been used in this analysis to verify the highest sensitivity on the expected neural network [53]. Amin used the backpropagation ANN approach to investigate gender characteristics of older driver accidents and model the variables of their accidents and finally illustrated that journey purpose was the highest

contributor factor of accident risk for older drivers and lighting condition was the second most important factor [54]. Ghasedi et al. used the logit model, factor analysis, and machine learning approaches to recognize the most effective variables on suburban accidents and generate the most accurate model for predicting future accidents in the busiest suburban highway of Guilan Province located in the north of Iran. Their studies showed the outstanding role of environmental factors, such as rainy weather and inadequate lighting condition, on increasing the severity of pedestrian accidents [55].

2. Study Area and Methodology

Rasht is one of the most congested cities in Iran and also the most populous city in the north of Iran, with high traffic volume in most of the days during the year. According to the official census in 2016, its density was announced to be 414 people per square kilometer [56]. Besides, more than 60% of the accidents between 2019 and 2020 in the city of Rasht occurred in the inner ring road of this city. The effective factors on increasing the probability of accidents vary from city to city, and due to the congested traffic flows even in ring roads and high traffic interferences with other flows in several parts of this city, a separate study is needed to be conducted in Rasht metropolis. Due to the above-mentioned information and very dense urban texture and numerous traffic interferences between nonmotorized users and passing vehicles, the need to identify the most influential variables on the severity of accidents and provide the most accurate prediction model for future accidents are of great importance. As it is mentioned earlier, different approaches have been used worldwide for analyzing the number and severity of accidents considering the specific condition of crashes, including crashes with unknown severity, high-severity crashes, or the severity of crashes that may be expected to occur sometime in the future. However, in this study, due to a balanced number of datasets in terms of three categories of fatal, injury, and PDO accidents, we decided to use logit and pattern recognition type of ANN as a machine learning approach to analyze and provide a prediction model. In order to conduct research on road safety improvements, accident information that includes factors such as the time of the accident, human characteristics, environmental characteristics, and accident types should be collected meticulously. The results of consecutive visits to the Rasht Traffic Police Statistic Center led to the collection of information on the 12-month accidents from March 2019 to March 2020. In total, the statistical population includes 965 accidents that occurred in Rasht, 738 of which are related to urban accidents and the rest are related to suburban accidents. The dependent variable in this study is the severity of accidents, which are classified into three categories: fatal, injury, and PDO accidents. Since the amount of fatal accidents is small in comparison to total accidents, and the independent variables significance and goodness of fit of a model cannot be satisfied by considering the three types of dependent variables, fatal accidents have been merged with injury accidents, and the

dependent variable has been split into two categories: PDO and injury/fatal accidents [55].

Table 1 classifies the independent variables influencing the occurrence of accidents in Rasht city and the appropriate coding for each of them for modeling purposes. Moreover, the remainder of this article is structured as indicated in Figure 1.

3. Statistical and Machine Learning Forecasting Approaches

3.1. Multiple Logistic Regression. To connect a set of variables X and a dependent variable such as Y, one has a multivariate problem. In analyzing such a problem, different types of mathematical models have been used to consider the relationship's complexity. The logistics regression method is a mathematical method used to describe the relationship between several variables x and a binary dependent variable. The function used in this method is an S-shaped function called a logistics function. However, the application of logistics analysis is not just limited to the above-mentioned issue. Expanding the logistics function can also be used to solve multifaceted problems. Therefore, the logistic regression method can be used to define the variable Y in a multifaceted way. In the simplest case, $P(Y = i)$ can be considered a linear function of x_i ($P_i = x_i\beta$), where β is the vector of regression coefficients. However, one of the considerations is that the probability of P_i mentioned in the left part of the equation must be between zero and one, but the linear multiplication of $x_i\beta$ in the right part includes all real numbers. A simple way to solve this problem is to use the probability transfer function to remove the distance constraints and model the transfer function as a linear function of the parameters. This conversion takes place in two stages [57]. First, the probability of P_i becomes a chance of success from

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

In the second step, the logarithm of equation (1) is performed so that logit or logarithm of the chance of success is obtained from

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

The reverse transfer function, called antilogistics, is used to calculate probability in terms of logistics from

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

Logistics is a transfer function that extends the probabilities of the range (0, 1) to all real numbers. Negative logistics indicate probabilities of less than 50% and positive logistics indicate probabilities of more than 50%. Therefore, the logistics model is a general linear model that has a logistics transfer function. In other words, the probability logistics of P_i instead of the probability follow the linear model [58].

3.2. Neural Network. Neural networks have significant abilities to detect complicated data relations and could be used to extract patterns and classify methods that humans and other computer methods find extremely difficult to comprehend [45]. ANNs have been interpreted as a nonlinear system based on human's brain activity [59, 60]. The researchers could take advantage of these networks' precious ability in recognizing the unknown relationship of natural and complex systems. Due to the close relationship between accidents and the well-being of society and their direct and indirect influences on human lives, ANNs have been considered a powerful and highly precise approach to deal with accidents dilemma [61]. Therefore, in this study, to find the most accurate prediction model, urban accidents of Rasht city have been modeled by pattern recognition type of ANN, which can result in various hands-on outputs to reduce the severity and number of accident.

4. Result Analysis and Discussion

4.1. Modeling Using Multiple Logistic Regression. In this section, the multiple logistic regression is used to determine the effect of each independent variable on the dependent variables (severity of accidents). To model the severity of accidents in Rasht, initially, 63 independent variables and 3 dependent variables are defined. Due to the small number of fatal accidents (7 cases), the dependent variables are summarized into two categories, including fatal/injury and PDO accidents. To build the logit model, the entering, the forward stepwise, and the backward stepwise methods can be used. Now, it should be examined which of the three methods has the most suitable output, or in other words, which of the methods can provide a better model for accidents in Rasht. To determine this, the criteria of higher prediction accuracy and the goodness of fit index of the model are considered to identify the best model. The coefficient of determination parameter (R^2) indicates the goodness of fit of the model. To be more specific, the closer R^2 to 1, the greater the model's goodness of fit, and the higher the correct percentage for the model, the higher the strength of the model to predict accidents. Table 2 presents a summary of the three methods, including two criteria of the prediction accuracy and the goodness of fit (R^2) of the model. The forward stepwise method with the correct percentage of 89.17% and the R^2 value of 0.854 is selected as the best method to build the logit model of the severity of accidents in Rasht.

Table 3 indicates the chi-square, degree of freedom (df), and significance (sig) of the forward method in the modeling process. The chi-square statistic index is used to determine the effect of independent variables on the dependent variable and, in general, the fit of the model and it is comparable to the F-statistic in normal regression analysis. The chi-square of the model in Step 23 is equal to 61.038 with a significant value of less than 5%. Since the significance of the forward model used to predict accidents is less than 5 percent, the capability of the model to predict accidents is confirmed. Thus, independent variables affect the dependent variable and indicate a good fit.

TABLE 1: Description of variables used in the study.

Variable	Variable levels
Accident severity	1. PDO 2. Injury/fatal
Accident time	1. 00 : 00 to 06 : 00 2. 06 : 00 to 12 : 00 3. 12 : 00 to 18 : 00 4. 18 : 00 to 24 : 00
Accident day	1. Saturday to Tuesday (major workdays in Iran) 2. Wednesday to Friday and holidays
Road category	1. Arterial road 2. Access road
Road surface condition	1. Dry 2. Humid 3. Wet
Geometry of accident location	1. Alignment 2. Roundabout 3. U-turn/J-turn 4. Uphill 5. Downhill 6. Alley 7. Intersection
Daylight condition	1. Day 2. Night 3. Sunset and sunrise
Accident type	1. Light vehicle-light vehicle 2. Light vehicle-motorcycle 3. Light vehicle-heavy vehicle/pickup truck-truck/bus
Type of collision	1. Head-on collision 2. Rear-end collision 3. Side-impact collision
Driver gender	1. Male 2. Female
Driver age	1. Less than 18 2. 18 to 30 3. 30 to 45 4. 45 to 60 5. 60 and over
Weather condition	1. Clear 2. Cloudy 3. Rainy
Reason of accident	1. Lack of attention 2. Backover movement 3. Inability to control the vehicle 4. Exceeding lawful speed 5. Crossing a forbidden place 6. Failure to observe longitudinal spacing 7. Failure to observe lateral spacing 8. Unsafe lane changes 9. Hasty-caused accident 10. Unsafe lane changes 11. Failure to yield the right-of-way 12. Improper turns 13. Wrong way movements 14. Yaw motion of the vehicle to the left 15. Technical defect in the vehicle

Table 1: Continued.

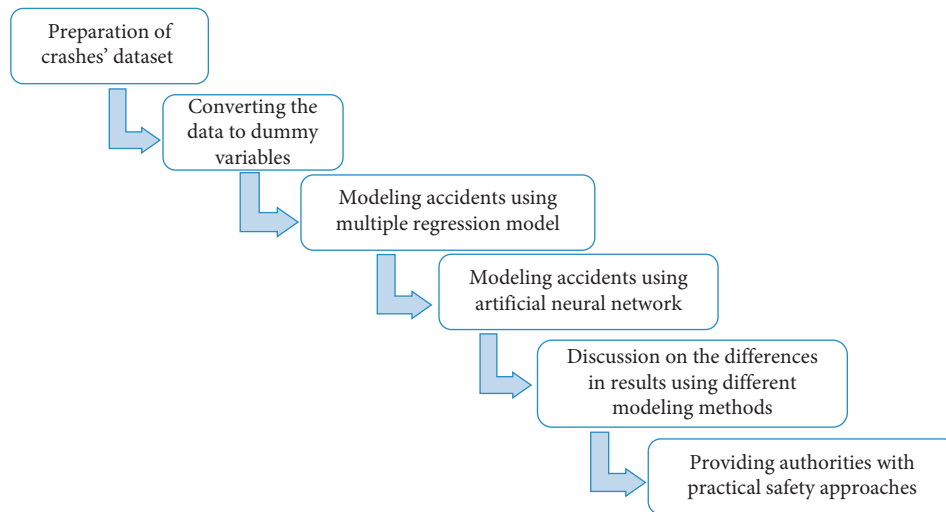| Variable	Variable levels
Vehicle	1. KIA Pride 2. Paykan 3. Renault 4. Peugeot 5. Taxi 6. Pickup truck 7. Minibus 8. Bus 9. Truck 10. Trailer

Figure 1: The overview of the proposed modeling process.

Table 2: Summary of the entering, the forward stepwise, and the backward stepwise methods.

Logistic regression type	Prediction accuracy	Goodness of fit (R^2)
Entering	82.52	0.686
Forward stepwise	89.17	0.854
Backward stepwise	86.93	0.738

Table 3: The result of the forward stepwise method.

		Chi-square	df	Sig.
Step 23	Step	−1.992	1	0.158
	Block	61.038	35	0.000
	Model	61.038	35	0.000

By selecting the forward stepwise method and entering all 63 selected variables in the modeling process, this method's final model is obtained in 23 steps. In this model, 11 variables are identified as the most effective variables in the severity of accidents leading to injury/fatal and PDO accidents in Rasht. Table 4 shows the variables entered in the model and their statistical indexes. According to Wald and sig, the Wald test examines the significance of the regression equation's variables and is comparable to the t-statistic in normal regression.

According to Table 4, the most effective variables increasing the severity of vehicle accidents are accident time, 18 : 00–24 : 00 and 12 : 00–18 : 00, KIA Pride vehicle, and rainy weather, respectively. Conversely, the most affecting variables with negative coefficients decreasing the severity of accidents include day time, accident time (6 : 00–12 : 00), and sunny weather. The most important result of the logit model underlines the role of environmental factors, including poor lighting conditions alongside unfavorable weather and the dominant role of unsafe and poor quality of vehicles on increasing the severity of accidents.

4.2. Modeling Using Artificial Neural Network. An ANN prediction model can be created using a variety of neural networks. Given the qualitative data used in this analysis, the prediction model is created using a neural network with pattern recognition capabilities. Using either supervised or unsupervised grouping, pattern recognition divides input data into objects or classes based on main characteristics [55].

The machine learning method uses the same input attributes and output labels as the variables described in

TABLE 4: Variables of the severity of accident logit model.

Variables	β	Standard deviation	Wald statistic	Significance	Odds ratio
6:00 to 12:00	−3.345	1.326	6.364	0.012	0.035
12:00 to 18:00	2.939	1.286	5.201	0.023	18.897
18:00 to 24:00	7.373	2.488	8.783	0.003	1592.404
Autumn	−1.918	0.725	7.010	0.008	0.147
Sunny weather	−2.446	0.806	9.205	0.002	0.087
Rainy weather	2.751	1.157	5.656	0.017	15.658
Lack of attention	1.119	0.644	3.020	0.082	3.062
Female driver	−2.063	0.783	6.946	0.008	0.127
Head-on collision	2.204	0.771	8.176	0.004	9.061
Day time	−4.343	1.422	9.323	0.002	0.013
KIA Pride vehicle	3.795	1.287	8.696	0.003	44.478
Constant	6.734	2.914	5.338	0.021	840.503

TABLE 5: Details of data entry.

Number of inputs	Number of hidden layers	Number of outputs	Number of training-validation-testing samples			
			—	Samples	MSE	%E
63	11	2	Training	516	$1.07429e^{-1}$	$12.81204e^{-0}$
			Validation	111	$1.36493e^{-1}$	$24.35571e^{-0}$
			Testing	111	$1.74612e^{-1}$	$24.72239e^{-0}$

Table 1. It is worth noting that the dependent variable (output class) is the different levels of accident severity, as mentioned in the "study route and methodology" section. For vehicle accidents, it has been split into two categories: fatal/injury and PDO. The ANN used in this analysis is the utilization of an existing algorithm in the software. The neural network's input data is divided into three categories:

(1) Training: these are presented to the network during training for the learning process, and the network is adjusted according to its error.

(2) Validation: these are used to measure network generalization and to halt training when generalization stops improving

(3) Testing: these have no effect on training and so provide an independent measure of network performance during and after training. In other words, it is the main criterion to realize how much the neural network's findings are similar to the actual result.

Table 5 shows the details of the accident data entry into the software and the Mean Squared Error and Percent Error. Since the number of accidents is adequate for the network training process, 70% of the data are used for network training and 15% for validation, and the remaining 15% are considered as a test of the developed network.

4.2.1. The Results of the Confusion Matrix. As can be seen in Figure 2, according to the all confusion matrix, which represents the result of the three processes of training, validation, and testing of the network, out of 632 property damage accidents, all 632 cases, and out of 106 injury/fatal accidents, 98 cases are predicted correctly by the model. The

prediction accuracy of property damage accidents in the model is 100% and the prediction accuracy of injury/fatal accidents is 92.5%. Therefore, the accuracy of the model in the classification and separation of injury/fatal accidents and property damage accidents from each other is at a high level. Also, the accuracy of the whole model in determining the severity of accidents is 98.9%. With confidence of 98.9%, the model can predict the severity of accidents in terms of effective parameters.

4.2.2. The Results of the Performance of Neural Network. Figure 3 shows that the training process stopped after 92 repetitions. The point marked on the diagram shows that the answers will no longer improve from this point on, and this point with the mean squared error (MSE) 0.0231 indicates the best point for the end of the calculations and creation of the ANN for the information given.

4.2.3. Sensitivity and Specificity Analysis of the Neural Network for the Given Accident Data. Figure 3 shows the sensitivity analysis of the true positive rate of the generated ANN model compared to the false positive rate using a receiver operating characteristic (ROC) graph for accidents in Rasht. The ROC graph is a technique for visualizing, organizing, and selecting classifiers based on their performance [62]. Its popularity comes from several well-studied characteristics, such as intuitive visual interpretation of the curve and easy comparisons of multiple models [63]. When it is needed to check or visualize the multiclass classification problem's performance, the ROC curve is used. In this model, 70% of the data are considered for training, 15% for testing, and 15% for validation. As shown in Figure 4, Class 1 indicates the accuracy of the network prediction for existing

Training confusion matrix

	1	2	
1	447 86.6%	4 0.8%	99.1% 0.9%
2	0 0.0%	65 12.6%	100% 0.0%
	100% 0.0%	94.2% 5.8%	99.8% 0.8%

Output class / Target class

(a)

Validation confusion matrix

	1	2	
1	89 80.2%	3 2.7%	96.7% 3.3%
2	0 0.0%	19 17.1%	100% 0.0%
	100% 0.0%	86.4% 13.6%	97.3% 2.7%

Output class / Target class

(b)

Test confusion matrix

	1	2	
1	96 86.5%	1 0.9%	99.0% 1.0%
2	0 0.0%	14 12.6%	100% 0.0%
	100% 0.0%	93.3% 6.7%	99.1% 0.9%

Output class / Target class

(c)

All confusion matrix

	1	2	
1	632 85.6%	8 1.1%	98.8% 1.2%
2	0 0.0%	98 13.3%	100% 0.0%
	100% 0.0%	92.5% 7.5%	98.9% 1.1%

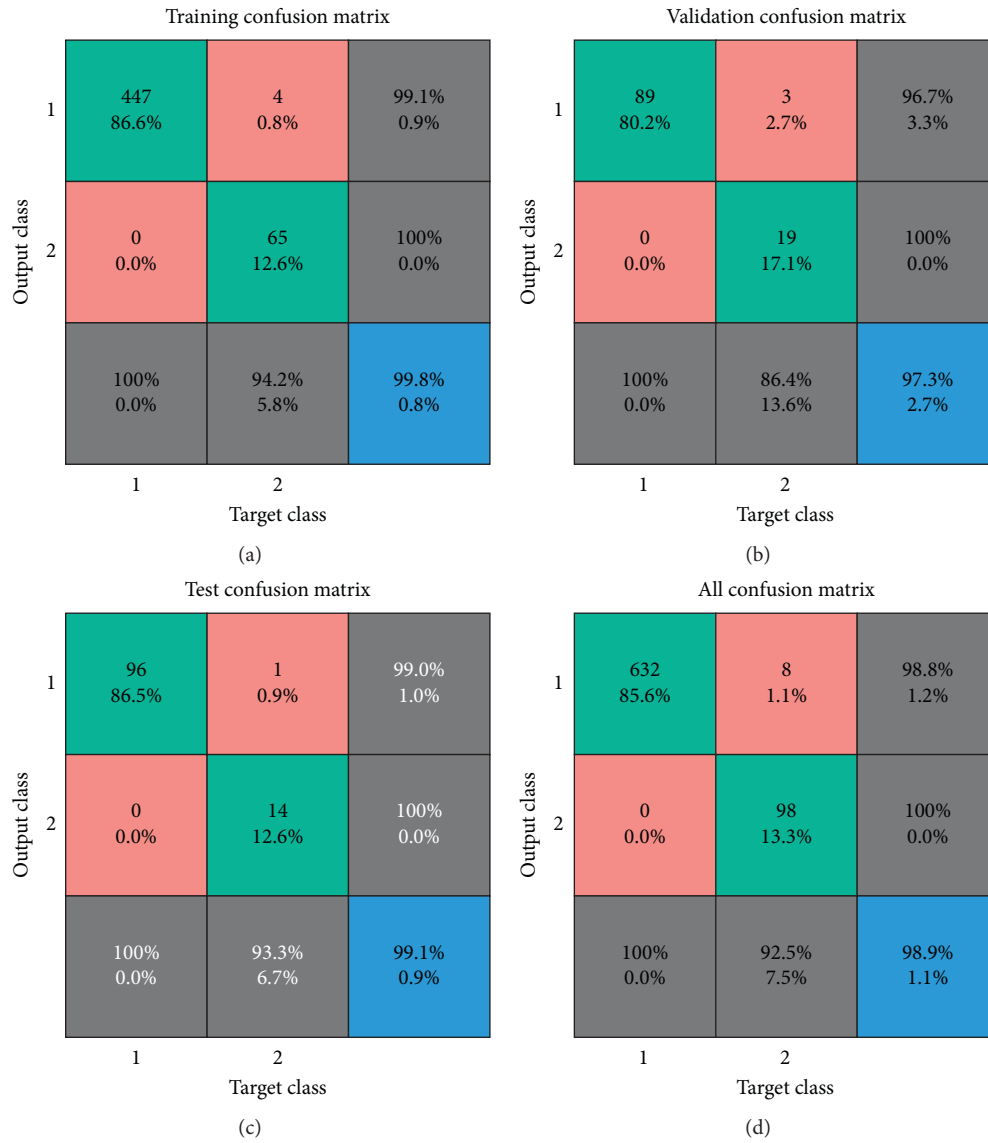Output class / Target class

(d)

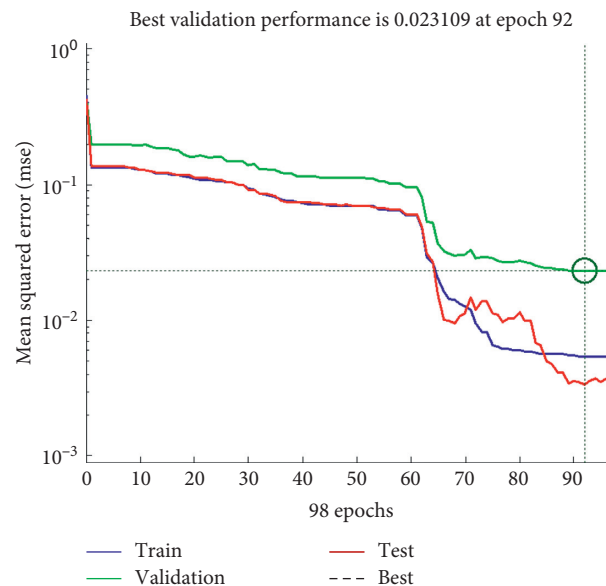FIGURE 2: The confusing matrix of accidents.

FIGURE 3: Performance of the neural network training process of accidents.
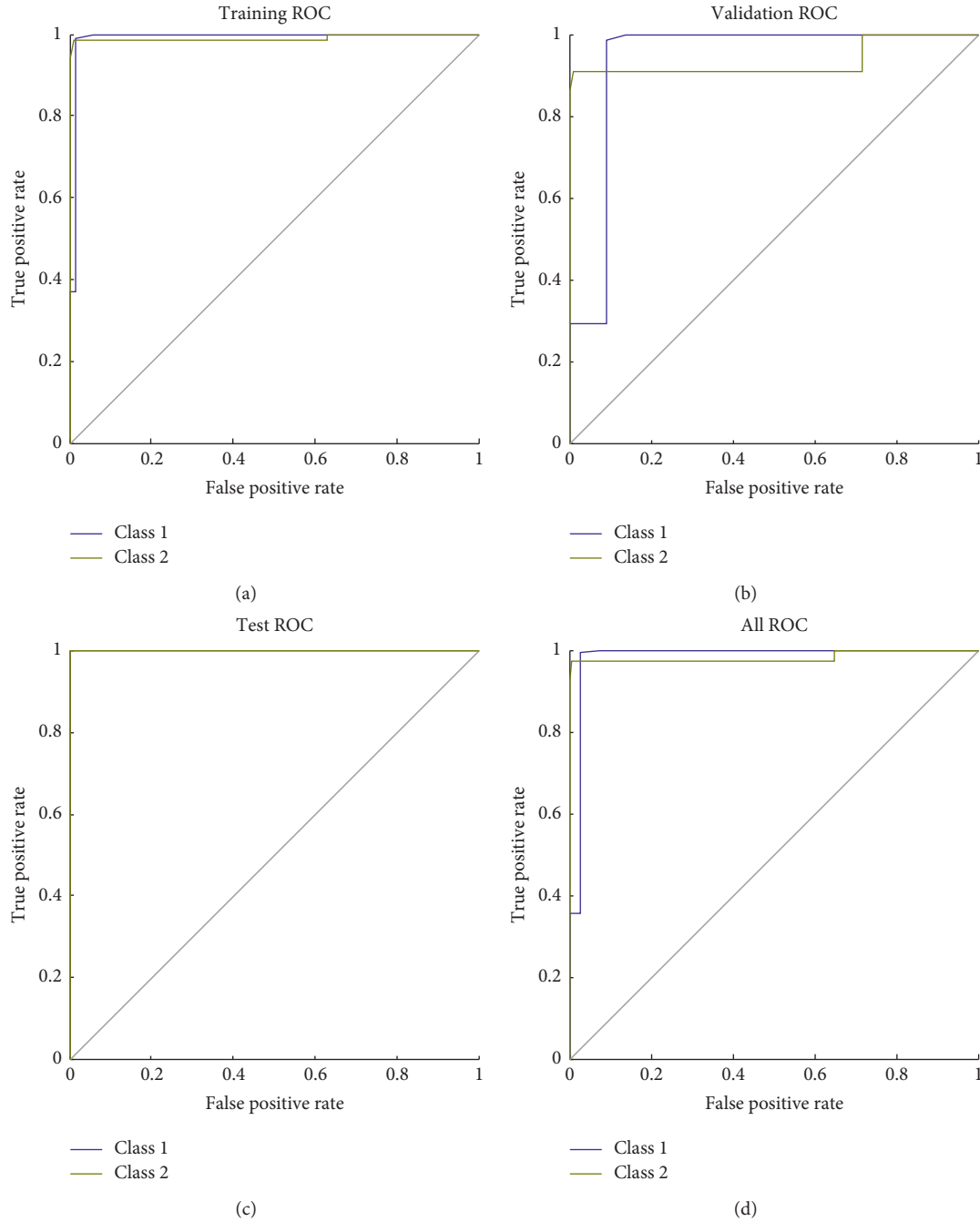
FIGURE 4: Sensitivity and specificity diagrams.

accidents, and Class 2 indicates the network's accuracy for future accidents. The more inclined the top and left curves, the more powerful the network in estimating and predicting more accurately [64].

5. Conclusion and Safety Approaches

In this study, aiming to investigate the main causes and severity of urban accidents in Rasht city, two models of accident analysis approaches are adopted and compared to provide practical solutions to increase overall safety and

reduce the number of accidents within the city. Therefore, using the multiple logistic regression and pattern recognition type of ANN, the variables affecting the severity of accidents and the most powerful approach to predict the accidents in Rasht have been presented; the most important results and prevention strategies of accidents are as follows:

(1) Comparing the correct percentage of prediction in the multiple logistic regression and the machine learning model, the results showed that the ANN model performed better and has a higher prediction power than the logit model. To be more specific, the

prediction accuracy of the utilized ANN is 98.9%, while the logit model's prediction accuracy is 89.17%. In other words, the prediction error rate of the ANN model is 1.1%, while the logit model prediction error rate is 10.83%, which justifies the utilization of the ANN model. On the other hand, the network sensitivity analyzer diagram of ANN approach proves the high power of this model in the prediction of urban accidents.

(2) The multiple logistic regression results show that the forward stepwise method is the best method for making the logit model of accident severity in Rasht city, considering the two criteria of the goodness of fit (R^2) of 0.854 and the prediction accuracy of 89.17%.

(3) According to the logit model results, the variables of accident time (12–18) and (18–24), rainy weather, the accident reason (lack of attention), head-on collision, and KIA Pride vehicle increased the severity of accidents with a positive coefficient. In other words, the significant role of KIA Pride vehicle in the occurrence of accidents, especially at (18–24) accident time (night peak hours' traffic), is noticeable. Therefore, it is expected that the officials should conduct major improvements and corrections in lighting facilities and urban routes infrastructures and improve the quality of mass vehicle production in collaboration with car manufacturing companies.

(4) The logit model also shows that the interactive effect of the darkness of the air and head-on collision increased the likelihood of accidents, which may be due to poor visibility of drivers and cognitive distraction. A greater presence of police and speed control strategies, especially at night (between the hours of 18 and 24) and rainy weather days, is one of the best ways to reduce traffic accidents. Last but not least, increasing penalties for mobile phone users to decrease accidents due to lack of attention and providing warning signs or other pavement-based warning techniques, including pavement markers and rumble strips, may help reduce accident risk.

Since statistical analysis and programming models are usually not able to consider all the required details of the problem, for future studies, it is suggested that to utilize Geographic Information System (GIS) roadway profile data along with more deep learning and optimization techniques to have an in-depth analysis and find the most desirable solutions [65–68]. Moreover, due to the significant importance of pedestrian accidents and permanent interference of nonmotorized users and vehicles flow within a city, it is recommended to use more analytical methods and pattern recognition type of machine learning approach to provide better decision making approaches and present the most accurate prediction model for pedestrian accidents separately occurring in urban environments [55, 69]. Last but not least, to have a more in-depth analysis about pedestrian accidents and find the most dangerous conflicts, it is highly

suggested to use AI-based object detection and image processing approaches.

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.

References

[1] Y. Liu, C. Yang, and Q. Sun, "Thresholds based image extraction schemes in big data environment in intelligent traffic management," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2020.

[2] Q. Zhu, "Research on road traffic situation awareness system based on image big data," *IEEE Intelligent Systems*, vol. 35, no. 1, pp. 18–26, 2019.

[3] A. Abdi, O. Nassimi, R. Salehfard, and V. N. Moghaddam, "Analysing the influence of encroachment angle and median parameters on safety of rural highways using vehicle dynamics performance," *IOP Conference Series: Materials Science and Engineering*, vol. 471, p. 62043, 2019.

[4] I. Bargegol, V. N. M. Gilani, M. Ghasedi, and M. Ghorbanzadeh, "Delay modeling of un-signalized roundabouts using neural network and regression," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, no. 1, pp. 28–34, 2016.

[5] S. M. Hosseinian and V. Najafi Moghaddam Gilani, "Analysis of factors affecting urban road accidents in Rasht metropolis," *ENG Transactions*, vol. 1, no. 1, pp. 1–4, 2020.

[6] World Health Organization, *Global Status Report on Road Safety 2018*, World Health Organization, Geneva, Switzerland, 2018.

[7] F. Sherafati, E. Homaie-Rad, A. Afkar, R. Gholampoor-Sigaroodi, and S. Sirusbakht, "Risk factors of road traffic accidents associated mortality in northern Iran; a single center experience utilizing oaxaca blinder decomposition," *Bulletin of Emergency & Trauma*, vol. 5, no. 2, pp. 116–121, 2017.

[8] P. Intini, N. Berloco, P. Colonna, V. Ranieri, and E. Ryeng, "Exploring the relationships between drivers' familiarity and two-lane rural road accidents. a multi-level study," *Accident Analysis & Prevention*, vol. 111, pp. 280–296, 2018.

[9] 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, p. 100377, 2019.

[10] M. Yazdi, N. A. Golilarz, A. Nedjati, and K. A. Adesina, "An improved lasso regression model for evaluating the efficiency of intervention actions in a system reliability analysis," *Neural Computing and Applications*, 2021.

[11] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing*, vol. 88, p. 105946, 2020.

[12] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced moth-flame optimizer with mutation strategy for global optimization," *Information Sciences*, vol. 492, pp. 181–203, 2019.

[13] M. Yazdi, "A perceptual computing–based method to prioritize intervention actions in the probabilistic risk

assessment techniques," *Quality and Reliability Engineering International*, vol. 36, no. 1, pp. 187–213, 2019.

[14] C. Li, L. Hou, B. Y. Sharma et al., "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, 2018.

[15] M. Mirmozaffaria, "Filtering in image processing," *ENG Transactions*, vol. 1, pp. 1–5, 2020.

[16] J. Xia, H. Chen, Q. Li et al., "Ultrasound-based differentiation of malignant and benign thyroid Nodules: an extreme learning machine approach," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 37–49, 2017.

[17] N. A. Golilarz, H. Gao, S. Pirasteh, M. Yazdi, J. Zhou, and Y. Fu, "Satellite multispectral and hyperspectral image denoising with enhanced adaptive generalized Gaussian distribution threshold in the wavelet domain," *Remote Sensing*, vol. 13, no. 1, p. 101, 2021.

[18] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, 2016.

[19] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine*, vol. 59, pp. 116–124, 2015.

[20] N. A. Golilarz, M. Mirmozaffari, T. A. Gashteroodkhani et al., "Optimized wavelet-based satellite image de-noising with multi-population differential evolution-assisted harris hawks optimization algorithm," *IEEE Access*, vol. 8, pp. 133076–133085, 2020.

[21] A. Addeh, A. Khormali, and N. A. Golilarz, "Control chart pattern recognition using RBF neural network with new training algorithm and practical features," *ISA Transactions*, vol. 79, pp. 202–216, 2018.

[22] J. Hu, H. Chen, A. A. Heidari et al., "Orthogonal learning covariance matrix for defects of grey wolf optimizer: insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, p. 106684, 2021.

[23] A. Anaei and A. A. Kalteh, "A new method for dental caries diagnosis using convolutional neural networks and bees algorithm," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 5, no. 2, pp. 52–57, 2019.

[24] I. Bargegol, M. Ghorbanzadeh, M. Ghasedi, and M. Rastbod, "Evaluation of effective factors on travel time in optimization of bus stops placement using genetic algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 245, 2017.

[25] M. Ghasedi, M. Ghorbanzadeh, and I. Bargegol, "Robust optimization of bus stop placement based on dynamic demand using meta heuristic approaches: a case study in a developing country," *Transport and Telecommunication Journal*, vol. 22, no. 1, pp. 39–52, 2021.

[26] W. Zhu, C. Ma, X. Zhao et al., "Evaluation of sino foreign cooperative education project using orthogonal sine cosine optimized kernel extreme learning machine," *IEEE Access*, vol. 8, pp. 61107–61123, 2020.

[27] G. Liu, W. Jia, M. Wang et al., "Predicting cervical hyperextension injury: a covariance guided sine cosine support vector machine," *IEEE Access*, vol. 8, pp. 46895–46908, 2020.

[28] Y. Wei, H. Lv, M. Chen et al., "Predicting entrepreneurial intention of students: an extreme learning machine with Gaussian barebone harris hawks optimizer," *IEEE Access*, vol. 8, pp. 76841–76855, 2020.

[29] H. Tang, Y. Xu, A. Lin et al., "Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted K-nearest neighbor classifiers," *IEEE Access*, vol. 8, pp. 35546–35562, 2020.

[30] A. Lin, Q. Wu, A. A. Heidari et al., "Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy K-nearest neighbor classifier," *IEEE Access*, vol. 7, pp. 67235–67248, 2019.

[31] B. Wang, L. Zhang, H. Ma, H. Wang, and S. Wan, "Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction," *Complexity*, vol. 2019, Article ID 7414318, 13 pages, 2019.

[32] K. Qu, L. Wei, and Q. Zou, "A review of DNA-binding proteins prediction methods," *Current Bioinformatics*, vol. 14, no. 3, pp. 246–254, 2019.

[33] Z. Zhang, C. Luo, and Z. Zhao, "Application of probabilistic method in maximum tsunami height prediction considering stochastic seabed topography," *Natural Hazards*, vol. 104, no. 3, pp. 2511–2530, 2020.

[34] S. Wang, K. Zhang, L. P. H. van Beek, X. Tian, and T. A. Bogaard, "Physically-based landslide prediction over a large region: scaling low-resolution hydrological model results for high-resolution slope stability assessment," *Environmental Modelling & Software*, vol. 124, p. 104607, 2020.

[35] S. Zheng, Z. Lyu, and L. K. Foong, "Early prediction of cooling load in energy-efficient buildings through novel optimizer of shuffled complex evolution," *Engineering with Computers*, 2020.

[36] 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.

[37] A. Abdulhafedh, "Crash frequency analysis," *Journal of Transportation Technologies*, vol. 6, no. 4, pp. 169–180, 2016.

[38] S. Mokhtarimousavi, J. C. Anderson, A. Azizinamini, and M. Hadi, "Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and artificial neural networks," *International Journal of Transportation Science and Technology*, vol. 9, no. 2, pp. 100–115, 2020.

[39] J. Golrokh and A. Hasan, "A comparison of machine learning clustering algorithms based on the DEA optimization approach for pharmaceutical companies in developing countries," *ENG Transactions*, vol. 1, pp. 1–8, 2020.

[40] A. Mahani, P. Bazoobandi, S. M. Hosseinian, and H. Ziari, "Experimental investigation and multi-objective optimization of fracture properties of asphalt mixtures containing nanocalcium carbonate," *Construction and Building Materials*, vol. 285, p. 122876, 2021.

[41] W. Zhu, J. Miao, J. Hu, and L. Qing, "Vehicle detection in driving simulation using extreme learning machine," *Neurocomputing*, vol. 128, pp. 160–165, 2014.

[42] 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.

[43] M. Yazdi, "Ignorance-aware safety and reliability analysis: a heuristic approach," *Quality and Reliability Engineering*, vol. 36, no. 2, pp. 652–674, 2020.

[44] Y. Zhang, R. Liu, A. A. Heidari et al., "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 430, pp. 185–212, 2020.

[45] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[46] H. Behbahani, H. Ziari, A. Amini, V. N. M. Gilani, and R. Salehfard, "Investigation of un-signalized roundabouts delay with adaptive network-based fuzzy inference system and fuzzy logic," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, pp. 140–149, 2016.

[47] J. de Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 402–411, 2011.

[48] A. Pashaei, M. Ghatee, and H. Sajedi, "Convolution neural network joint with mixture of extreme learning machines for feature extraction and classification of accident images," *Journal of Real-Time Image Processing*, vol. 17, no. 4, pp. 1051–1066, 2019.

[49] B. Yu, Y. Chen, S. Bao, and D. Xu, "Quantifying drivers' visual perception to analyze accident-prone locations on two-lane mountain highways," *Accident Analysis & Prevention*, vol. 119, pp. 122–130, 2018.

[50] D. Chimba and T. Sando, "The prediction of highway traffic accident injury severity with neuromorphic techniques," *Advances in Transportation Studies*, vol. 19, pp. 17–26, 2009.

[51] S. J. Khair, A. F. Muaath, and F. G. Hala, "Prediction of road traffic accidents in Jordan using artificial neural network (ANN)," *Journal of Traffic and Logistics Engineering*, vol. 2, pp. 92–94, 2014.

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

[53] I. E. Contreras, L. Torres-Treviño, and F. Torres, "Prediction of car accidents using a maximum sensitivity neural network," in *Smart Technology Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, G. F. Torres, J. Lozoya-Santos, E. Gonzalez Mendivil, L. Neira-Tovar, P. Ramírez Flores, and J. Martin-Gutierrez, Eds., Springer, Cham, Switzerland, 2018.

[54] S. Amin, "Backpropagation-artificial neural network (BP-ANN): understanding gender characteristics of older driver accidents in West Midlands of United Kingdom," *Safety Science*, vol. 122, p. 104539, 2020.

[55] 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, p. 13, 2021.

[56] Iran Statistical Center, *Official Results of Census of Populations and Houses of Iran*, Iran Statistical Center, Tehran, Iran, 2016.

[57] C. Kosun and S. Ozdemir, "The modified Brandeis dice for the traffic states dependency: the I-dice problem," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 6, no. 2, pp. 76–83, 2020.

[58] K. Rosaiah, N. S. Kolli, and N. S. Sanjeeva Rao, "An application of multiple logistic regression for identifying lipid profile changes towards assessing maternal and fetal outcomes," *International Journal of Community Medicine and Public Health*, vol. 7, no. 8, p. 3014, 2020.

[59] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "Bayesian hierarchical model-based information fusion for degradation analysis considering non-competing relationship," *IEEE Access*, vol. 7, pp. 175222–175227, 2019.

[60] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "A Bayesian approach for degradation analysis with individual differences," *IEEE Access*, vol. 7, pp. 175033–175040, 2019.

[61] X. Zhang, M. Fan, D. Wang, P. Zhou, and D. Tao, "Top-k feature selection framework using robust 0-1 integer programming," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2020.

[62] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[63] T. Saito and M. Rehmsmeier, "Precrec: fast and accurate precision-recall and ROC curve calculations in *R*," *Bioinformatics*, vol. 33, no. 1, pp. 145–147, 2017.

[64] A. Addeh and B. M. Maghsoudi, "Control chart patterns detection using COA based trained MLP neural network and shape features," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, no. 1, pp. 5–8, 2016.

[65] F. Jiang, K. K. R. Yuen, and E. W. M. Lee, "Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology," *Journal of Safety Research*, vol. 75, pp. 292–309, 2020.

[66] R. U. Khan, X. Zhang, R. Kumar, A. Sharif, N. A. Golilarz, and M. Alazab, "An adaptive multi-layer botnet detection technique using machine learning classifiers," *Applied Sciences*, vol. 9, no. 11, p. 2375, 2019.

[67] A. Addeh and M. Iri, "Brain tumor type classification using deep features of MRI images and optimized RBFNN," *ENG Transactions*, vol. 2, no. 1, pp. 1–7, 2021.

[68] N. Amiri Golilarz and H. Demirel, "Thresholding neural network (TNN) based noise reduction with a new improved thresholding function," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 3, no. 2, pp. 81–84, 2017.

[69] S. M. Hosseinian, V. N. M. Gilani, H. T. Amoli, M. Nikookar, and A. Oroue, "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, 16 pages, 2021.