IJTE, Vol. 10, No. 4, Spring 2023, pp. 1163-1173
Research Paper

Study of Urban Taxi-related Accident Analysis Using the Multiple Logistic Regression and Artificial Neural Network Models

Amir Izadi[1,*], Faramarz Jamshidpour[2], Iraj Bargegol[3]

Received: 2021/10/22 *Accepted: 2022/01/22*

Abstract

In this research, factors affecting the severity of property damage only (PDO) and injury/fatal accidents were examined using taxi-related accident data from March 2015 to March 2021 in urban sites of Rasht city. The multiple logistic regression and artificial neural network (ANN) were applied to recognize the most influential variables on the severity of accidents. Results indicated that the multiple logistic regression in the backward stepwise method had a prediction accuracy of 88.54% and R^2 value of 0.871. Moreover, the regression analysis revealed that the wet surface condition, night without sufficient light, rainy weather, Kia Pride taxi and lack of attentions increased the severity of accidents, respectively. The most important result of the logit model was the significant role of environmental factors, including slippery road surface, unfavorable weather as well as poor lighting condition, and also indicated the dominant role of poor quality of vehicles along with human factors in increasing the severity of accidents. Comparing the correct percentage of prediction in the multiple logistic regression and ANN model, the results showed that ANN model performed better so that the prediction accuracy of ANN was 95.8%.

Keywords: Safety; Urban accidents; Multiple logistic regression; Artificial neural network

[*] Corresponding author. E-mail: amirizadi@shomal.ac.ir

[1] Assistant Professor, Department of Civil Engineering, Shomal University, Amol, Iran

[2] Ph.D. Candidate, Department of Civil Engineering, Shomal University, Amol, Iran

[3] Assistant Professor, Department of Civil Engineering, University of Guilan, Rasht, Iran

1. Introduction

Traffic flow management is certainly one of the main problems in modern society in which the governments should make considerable investments in order to have a well-organized transportation system and enhance its performance [Hosseinian et al., 2021b]. The development of transportation facilities in Iran has remarkably increased the damages caused by accidents [Nadimia et al., 2017; Nejad et al., 2019], which can be due to environmental, road or pavement conditions [Hosseinian et al., 2020]. The fatalities caused by traffic accidents in Iran reported in 2016 were 15,932, of which 78% were male and 22% were female [Hosseinian and Gilani, 2020; Nikookar et al., 2021]. So the severity of accidents and the effect of factors affecting that should be examined to present safety solutions for enhancing safety and decreasing the occurrence of accidents.

Many previous researches have been conducted on traffic safety and accidents in recent years using multiple logistic regression. A study was carried out by Sherafati et al., in which the road traffic fatalities after receiving emergency services were explored by the use of multiple logistic regression. They concluded that males, motorbikes and pedestrians had a positive and significant relationship with fatalities caused by road traffic accidents [Sherafati et al., 2017]. In another research, Intini et al. utilized multiple logistic regression to examine the relationship between road familiarity/unfamiliarity and the occurrence of accidents. Results indicated that the variables of autumn/winter, speed limit less than 80 km/h and minor intersections/driveways had greater odds of accidents for familiar drivers. However, the variables of head-on and rear-end accidents, summer, heavy vehicles and young drivers had higher odds of having unfamiliar drivers involved in accidents [Intini et al., 2018]. By investigating the effect of age on accident severity using a logistic regression model, Casado-Sanz et al. revealed that female drivers and motorbikes had a negative impact on the likelihood of accidents [Casado-Sanz, Guirao and Gálvez-Pérez, 2019]. In another study conducted by Hosseinian et al. in recognizing the most effective variables on the severity of accidents by the use of the multiple logistic regression and a type of artificial neural network (ANN), results indicated that clothing colors and weather conditions had the greatest impact on the occurrence of accidents [Hosseinian et al., 2021a].

Various studies have been performed on the severity of accidents using ANN models. Contreras et al. applied an ANN model in order to provide a suitable prediction method for traffic collisions in urban areas. The programming feature of Scilab software was used in this analysis to verify the greatest sensitivity on the expected neural network [Contreras, Torres-Treviño, Torres, 2018]. In a study carried out by Amin, the back-propagation ANN approach was presented to examine the gender characteristics of older driver accidents and model the variables affecting their accidents. They demonstrated that journey purpose was the highest effective factor for older drivers and lighting condition was the second most important factor on the accident severity [Amin, 2020]. Ghasedi et al. used the multiple logistic regression and the pattern-recognition type of ANN models to determine the most effective variables on the occurrence of accidents in order to present the most accurate model for predicting future accidents. The results of the study indicated the important role of environmental factors, such as rainy weather and inadequate lighting condition on increasing the severity of accidents [Ghasedi, Sarfjoo, Bargegol, 2021].

In this study, the effect of various factors affecting the severity of injury/fatal and PDO accidents of taxi vehicles in Rasht city was investigated by collecting accident data from March 2015 to March 2021 by the use of the multiple logistic regression and ANN model. Finally, the most effective factor on accident occurrence was presented.

2. Methodology

2.1. Data Collection

In this research, 272 injury/fatal and PDO taxi accidents were used on urban roads of Rasht city, Iran, from March 2015 to March 2021. Table 1 presents the variables influencing the occurrence of accidents in Rasht city, as well as the appropriate codes for each variable.

Table 1. The variables applied in this research

Variables	Codes
Accident severity	1- Injury/fatal 2- PDO
Accident season	1- Spring 2- Summer 3- Autumn 4- Winter
Accident time	1- 07:00 to 13:00 2- 13:00 to 17:00 3- 17:00 to 21:00 4- 21:00 to 01:00 5- 01:00 to 07:00
Accident day	1- Weekdays 2- Weekends and holidays
Road classification	1- Arterial road 2- Access road
Road surface condition	1- Dry 2- Wet 3- Frost
Geometry of accident location	1- Alignment 2- Curve 3- Roundabout 4- U-turn/J-turn 5- Intersection- 4 Legs 6- Intersection- 3 Legs
Daylight condition	1- Day 2- Sunset and sunrise 3- Night with sufficient light 4- Night without sufficient light
Vehicle type	1- Peykan taxi 2- Kia Pride taxi 3- Samand taxi 4- Peugeot taxi 5- Van taxi

Variables	Codes
	6- Other
Driver age	1- Less than 18 2- 18 to 30 3- 30 to 45 4- 45 to 60 5- 60 and over
Driver gender	1- Male 2- Female
Driver job	1- Driver 2- Retiree 3- Second job
Collision type	1- Rear-end collisions 2- Head-on collisions 3- Side collisions
Weather condition	1- Clear 2- Cloudy 3- Rainy
Accident reason	1- Lack of attentions 2- Back over movements 3- Inability to control the taxi 4- Exceeding lawful speed 5- Crossing a forbidden place 6- Failure to observe longitudinal space 7- Failure to observe lateral space 8- Unsafe lane changes 9- Hasty-caused accident 10- Failure to comply with the right of way 11- Improper turns 12- Wrong way movements

2.2. Multiple Logistic Regression

The logistic regression technique is a mathematical approach that has been applied to determine the relationships between several variables of x and a binary target variable. An S-shaped function is used in this model, known as logistic function [Connelly, 2020; Shah, 2020]. Let consider P(Y=i) as a linear function of x_i. The probability of P_i becomes a chance of success by the use of Equation 1. The logarithm of Equation 1 (logarithm of the chance of success) is then obtained as the logit function through Equation 2. Anti-logistic as the reverse transfer function is then applied to measure the probability in terms of logistic using Equation 3 [Gilani et al. 2021a; Austin and Merlo, 2017].

$$Odds = \frac{P_i}{1-P_i} \tag{1}$$

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

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

2.3. Neural Network

ANNs are nonlinear systems based on the activity of human's brain, which have various advantages in engineering areas. These methods have the precious capacity to recognize the unknown relationships in the natural and complex systems

[Gilani et al., 2021b; Gilani et al., 2020a; Gilani et al., 2020b; Mahani et al., 2021]. ANNs have been considered as a powerful and high-accuracy approach that can deal with accident-related problems in the transportation fields because of their robust relationships between the accidents and their direct or indirect effects on the human lives. So in this research, urban accidents of Rasht city have been modeled in order to find the most accurate prediction model by the multi-layer perceptron type of ANN to reduce the severity and occurrence of accidents.

3. Results

3.1. Multiple Logistic Regression

In order to model the accident severity in Rasht, initially, 62 independent variables and 2 dependent variables were described. The entering, forward stepwise and backward stepwise methods were used to examine which method has the most appropriate results. The criteria of prediction accuracy and R^2 were considered. The higher the R^2 and the prediction accuracy, the greater the strength of the model to predict accidents. Table 2 represents a summary of the three methods.

Table 2. Summary of regression modeling results

Modeling type	Prediction accuracy	R^2
Entering	76.93%	0.681
Backward stepwise	88.54%	0.871
Forward stepwise	86.27%	0.839

Results indicated that the backward stepwise method with the correct percentage of 88.54% and R^2 of 0.871 was selected as the best method to create the logit model.

Results of Table 3 indicate that the significance level of the model was less than 5%, demonstrating the ability of the model to predict accidents. Moreover, it was confirmed that independent variables had an impact on the dependent variable and indicated a good fit in the backward stepwise method.

Table 3. Backward stepwise results

		Chi-square	df	Sig.
Step 17	Step	-1.992	1	0.164
	Block	61.038	32	0.000
	Model	61.038	32	0.000

Table 4 shows the output variables of the model and their statistical indexes.

Table 4. Regression results in the seventeen step

Variables	β	S.E.	Wald	Sig.	Odds ratio
Winter	2.194	0.762	8.290	0.003	8.971
01:00 to 07:00	-3.629	1.057	11.789	0.016	0.027
17:00 to 21:00	2.827	1.254	5.082	0.029	16.836
Weekends and holidays	1.043	0.361	8.347	0.002	2.838
Wet surface condition	4.172	1.138	13.441	0.017	64.845
Roundabout	-2.416	0.875	7.624	0.035	0.089
Night without sufficient light	3.752	1.349	7.736	0.003	42.606
Kia Pride taxi	2.973	1.154	6.637	0.028	19.551
Age 45 to 60	-2.036	0.894	5.187	0.001	0.131
Second job	0.941	0.306	9.457	0.000	2.563
Rainy	3.617	1.182	9.364	0.021	37.226
Lack of attentions	2.861	0.774	13.663	0.018	17.479

Variables	β	S.E.	Wald	Sig.	Odds ratio
Unsafe lane changes	2.392	0.588	16.549	0.004	10.935
Constant	6.829	2.362	8.359	0.031	924.266

According to Table 4, the most effective variables increasing the severity of accidents were wet surface condition, night without sufficient light, rainy weather, Kia Pride taxi, lack of attentions, accident time 17:00 to 21:00, unsafe lane changes, winter season, weekends and holidays and second job, respectively. However, the most influential variables with negative coefficients that decrease the severity of accidents were accident time 01:00 to 07:00, roundabout and age 45 to 60, respectively. As can be seen, the most important results of the multiple logistic regression model revealed the significant role of environmental factors, including slippery road surface, unfavorable weather as well as poor lighting condition, and also indicated the dominant role of poor quality of vehicles along with human factors in increasing the severity of accidents.

3.2. Neural Network

The accident data were modeled by the multi-layer perceptron in this research and the results of the model are presented in this section. The number of neurons in the input, hidden and output layers was 38, 12 and 6, respectively. The prediction accuracy of the data is indicated in Table 5.

Table 5. Artificial neural network results

Sample	Observed	Predicted		
		PDO	Injury/fatal	Accuracy
Train	PDO	63	2	96.9%
	Injury/fatal	7	118	94.4%
	Total			95.8%
Test	PDO	25	3	89.2%
	Injury/fatal	5	49	90.7%
	Total			90.1%

According to Table 5, 63 and 118 accidents were correctly predicted as PDO and injury/fatal accidents, respectively. In general, the accuracy of the neural network model was 95.8%. Furthermore, the cross entropy error in the train and test samples were 38.132 and 32.457, respectively.

In Figure 1, the illustrative representation of sensitivity against specificity for the accident categories, including PDO and injury/fatal accidents, is presented.

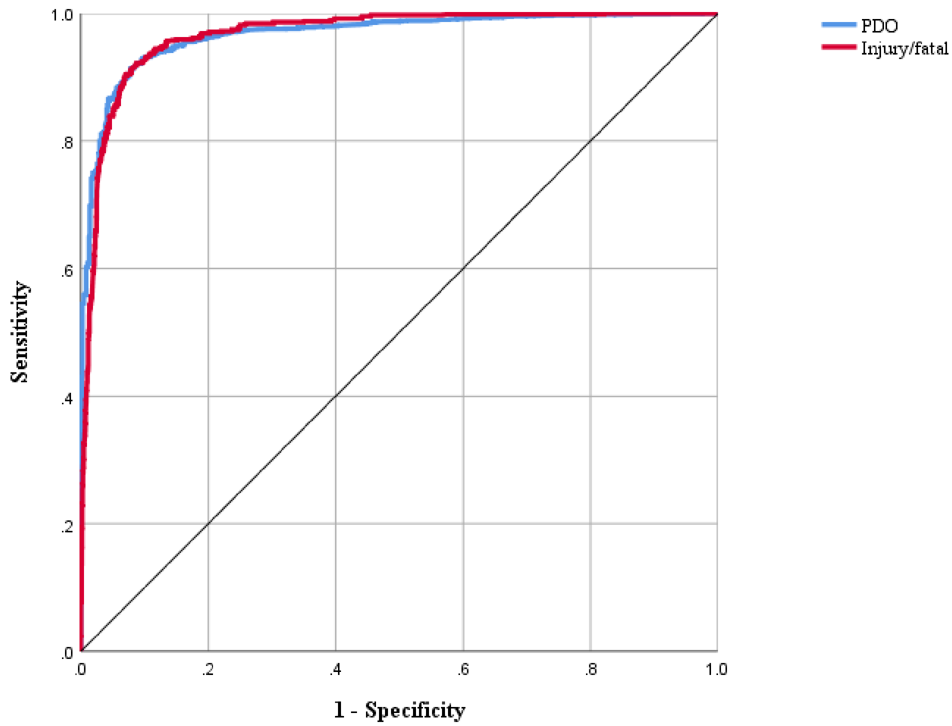

Figure 1. **Sensitivity analysis in the neural network model**

Area under the curve (AUC) is another parameter that is applied to examine the efficiency of the prediction model [LeDell, Petersen and van der Laan, 2015]. The higher the area under the curve, the greater the power of prediction model. Table 6 illustrates the probability of an accident resulting in injury/fatal or PDO accidents that was calculated 0.967, more than 0.5, so the response of the neural network was positive [Fawcett, 2006].

Table 6. AUC in neural network model

		AUC
Accident severity	PDO	0.967
	Injury/fatal	0.967

The influence of each independent variable on the dependent variable is presented in Table 7. Moreover, the importance of variables affecting the severity of accidents is also illustrated in Figure 2.

Table 7. Importance of independent variables in the neural network model

Variable	Importance	Normalized importance
Road surface condition	0.148	100.00%
Daylight condition	0.117	79.20%
Weather condition	0.09	61.30%
Accident time	0.087	58.60%
Accident reason	0.085	57.80%
Vehicle type	0.085	57.50%
Geometry of accident location	0.083	56.40%
Accident season	0.072	48.60%
Driver age	0.065	43.90%
Accident day	0.055	36.90%
Driver job	0.043	29.00%
Collision type	0.035	23.90%
Driver gender	0.026	17.40%
Road classification	0.01	6.80%

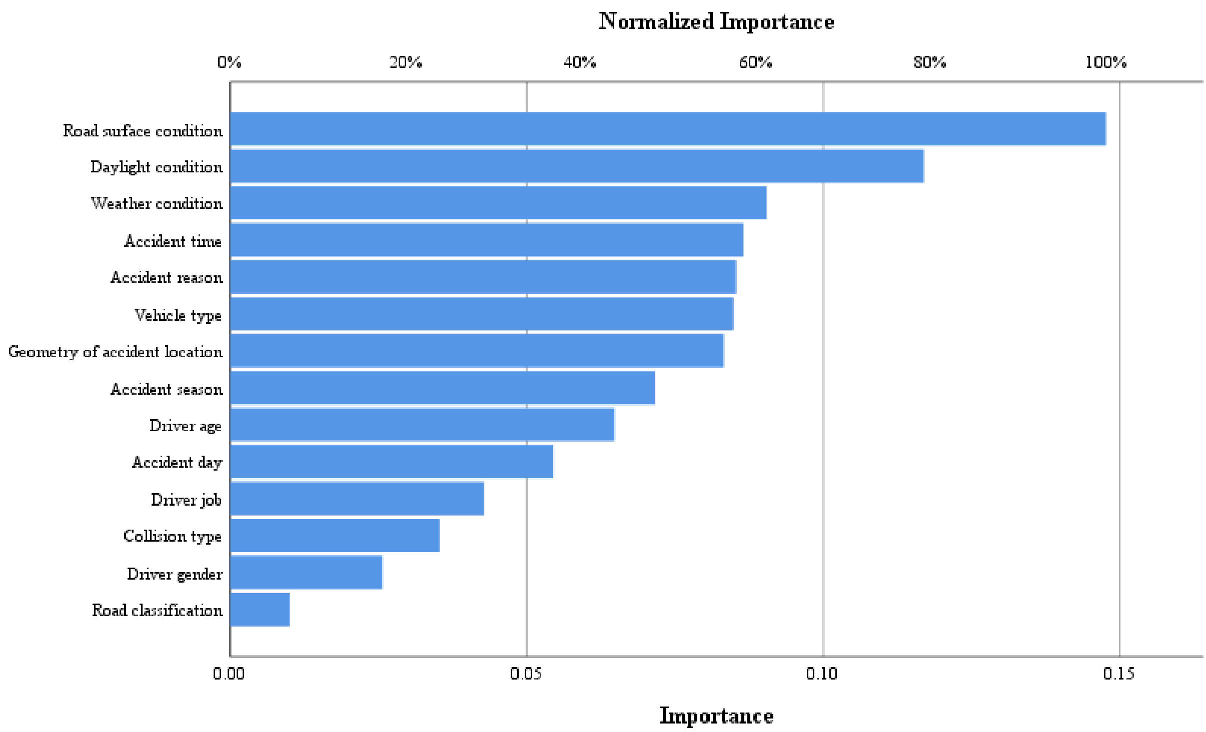

Figure 2. Independent variable importance diagram in the neural network model

As can be seen from Table 7 and Figure 2, the variables of road surface condition (14.8%), daylight condition (11.7%) and weather condition (9%) had the highest effect on the severity of accidents, respectively.

4. Conclusion

In this study, the effect of various factors affecting the severity of accidents of taxi-related vehicles in Rasht city was examined using multiple logistic regression and artificial neural network.

The results of multiple logistic regression showed that the backward stepwise method was the best modeling method to construct the logit model, regarding the two criteria of R2 of 0.871 and the prediction accuracy of 88.54%.

Based on the logit model results, wet surface condition, night without sufficient light, rainy weather, Kia Pride taxi, lack of attentions, accident time 17:00 to 21:00, unsafe lane changes, winter season, weekends and holidays and second job increased the severity of accidents with a positive coefficient. Therefore, the contribution of environmental factors, including slippery road surface, unfavorable weather as well as poor lighting condition, will increase the likelihood of the occurrence of accidents intensely. Moreover, the poor quality of vehicles along with human factors indicated a dominant role in increasing the severity of accidents.

The results of neural network model indicated that the prediction accuracy was 95.8%. Moreover, the sensitivity analysis of the model showed a great power of the model to predict accidents.

In the neural network model, road surface condition, daylight condition and weather condition had the greatest impact on the severity of accidents, respectively.

For future research works, other machine learning techniques [Taleghani and Taleghani, 2013; Behbahani et al., 2016; Gilani, Hosseinian and Nikookar, 2021c; Hamedi et al., 2021; Ziari et al., 2021] and deep learning methods [Ahmed and Hanif, 2020; Addeh and Iri, 2021] can be provided to be incorporated into the proposed approaches to obtain more accurate results.

5. References

-Addeh, A. and Iri, M. (2021) "Brain tumor type classification using deep features of MRI images and optimized RBFNN", Engineering Transactions, Vol. 2, pp. 1-7.

-Ahmed, A., Jalal, A. and Kim, K. (2020) "A novel statistical method for scene classification based on multi-object categorization and logistic regression", Sensors, Vol. 20, No. 4, pp. 3871.

-Ahmed, S.S. and Hanif, U. (2020) "News Recommendation Algorithm Based on Deep Learning", Computational Research Progress in Applied Science & Engineering (CRPASE), Vol. 6, pp. 132-140.

-Amin, S. (2020) "Backpropagation–Artificial Neural Network (BP-ANN): Understanding gender characteristics of older driver accidents in West Midlands of United Kingdom", Safety science, Vol. 122, pp. 104539.

-Austin, P. C. and Merlo, J. (2017) "Intermediate and advanced topics in multilevel logistic regression analysis", Statistics in Medicine, Vol. 36, pp. 3257-3277.

-Behbahani, H., Ziari, H., Amini, A., Gilani, V. N. M. and Salehfard, R. (2016) "Investigation of un-signalized roundabouts delay with adaptive-network-based fuzzy inference system and fuzzy logic", Computational Research Progress in Applied Science & Engineering, Vol. 2, pp. 1-7.

-Casado-Sanz, N., Guirao, B. and Gálvez-Pérez, D. (2019) "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, pp. 100377.

-Connelly, L. (2020) "Logistic regression", Medsurg Nursing, Vol. 29, pp. 353-354.

-Contreras, E., Torres-Treviño, L. and Torres, F. (2018) “Prediction of car accidents using a maximum sensitivity neural network”, Smart Technology, Springer, pp. 86-95.

-Ghasedi, M., Sarfjoo, M. and Bargegol, I. (2021) “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, pp. 1-16.

-Gilani, V. N. M., Hosseinian, S. M., Behbahani, H. and Hamedi, G. H. (2020a) “Prediction and pareto-based multi-objective optimization of moisture and fatigue damages of asphalt mixtures modified with nano hydrated lime”, Construction and Building Materials, Vol. 261, pp. 120509.

-Gilani, V. N. M., Hosseinian, S. M., Behbahani, H., Ghasedi, M. and Nikookar, M. (2021a) “Data-driven urban traffic accident analysis and prediction using logit and machine learning-based pattern recognition models”, Mathematical problems in engineering, Vol. 2021.

-Gilani, V. N. M., Hosseinian, S. M., Hamedi, G. H. and Safari, D. (2021b) “Presentation of predictive models for two-objective optimization of moisture and fatigue damages caused by deicers in asphalt mixtures”, Journal of Testing and Evaluation, Vol. 49, No. 6.

-Gilani, V. N. M., Hosseinian, S. M. and Nikookar, M. (2021c) “Presentation of a new deicer with the least moisture and fatigue failures in asphalt mixtures”, Arabian Journal for Science and Engineering, pp. 1-15.

-Gilani, V. N. M., Hosseinian, S. M., Safari, D. and Movahhed, M. B. (2020b) “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, pp. 1-14.

-Hamedi, G. H., Esmaeeli, M. R., Najafi Moghaddam Gilani, V. and Hosseinian, S. M. (2021) “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.

-Hosseinian, S. M. and Gilani, V. N. M. (2020) “Analysis of factors affecting urban road accidents in rasht metropolis”, ENG Transactions, Vol. 1, pp. 1-4.

-Hosseinian, S. M., Najafi Moghaddam Gilani, V., Mehraban Joobani, P. and Arabani, M. (2020) “Investigation of moisture sensitivity and conductivity properties of inductive asphalt mixtures containing steel wool fiber”, Advances in Civil Engineering, Vol. 2020.

-Hosseinian, S. M., Najafi Moghaddam Gilani, V., Mirbaha, B. and Abdi Kordani, A. (2021a) “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.

-Hosseinian, S. M., Najafi Moghaddam Gilani, V., Tahmasbi Amoli, H., Nikookar, M. and Orouei, A. (2021b) “Presentation of analytical methods for better decision making about the most important factor influencing rural accidents”, Mathematical Problems in Engineering, Vol. 2021.

-Intini, P., Berloco, N., Colonna, P., Ranieri, V. and Ryeng, E. (2018) "Exploring the relationships between drivers' familiarity and two-lane rural road accidents. A multi-level study", Accident Analysis & Prevention, Vol. 111, pp. 280-296.

-LeDell, E., Petersen, M. and van der Laan, M. (2015) "Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates", Electronic journal of statistics, Vol. 9, pp. 1583.

-Mahani, A. G., Bazoobandi, P., Hosseinian, S. M. and Ziari, H. (2021) "Experimental investigation and multi-objective optimization of fracture properties of asphalt mixtures containing nano-calcium carbonate", Construction and Building Materials, Vol. 285, pp. 122876.

-Mishra, K. N. and Pandey, S. C. (2021) "Fraud Prediction in Smart Societies Using Logistic Regression and k-fold Machine Learning Techniques", Wireless Personal Communications, pp. 1-27.

-Nadimia, N., Alavia, S. S. N., Mirhosseinia, A. H. Z. and Amirib, A. M. (2017) "Proposing a Model for Ranking Hotspots in Rural Roads Using a Multi-criteria Decision-making Method", Computational Research Progress in Applied Science & Engineering (CRPASE), Vol. 3.

-Nikookar, M., Bagheri Movahhed, M., Ayoubinejad, J., Gilani, V. N. M. and Hosseinian, S. M. (2021) "Improving the moisture sensitivity of asphalt mixtures by simultaneous modification of asphalt binder and aggregates with carbon nanofiber and carbon nanotube", Advances in Civil Engineering, Vol. 2021.

-Nejad, M. A. R. S., Agha, F. N. J., Chenari, H. M., and Badkoobeh, F. (2019) "Analysis of Parking Situation in Rasht Using Multi-criteria Evaluation Method", Computational Research Progress in Applied Science & Engineering (CRPASE), Vol. 5.

-Omrani, H. (2015) "Predicting travel mode of individuals by machine learning", Transportation Research Procedia, Vol. 10, pp. 840-849.

-Shah, S. K. (2020) "Binary Logistic Model to Identify the Factors Associated with Households with Bank Accounts in Nepal", Quest Journal of Management and Social Sciences, Vol. 2, pp. 323-336.

-Sherafati, F., Rad, E. H., Afkar, A., Gholampoor-Sigaroodi, R. and Sirusbakht, S. (2017) "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, pp. 116.

-Taleghani, M. and Taleghani, A. (2013) "Identification and Ranking of Factors affecting the implementation of Knowledge Management based on TOPSIS Technique", ENG Transactions, Vol. 1.

-Ziari, H., Divandari, H., Seyed Ali Akbar, S. M. and Hosseinian, S. M. (2021) "Investigation of the effect of crumb rubber powder and warm additives on moisture resistance of SMA mixtures", Advances in Civil Engineering, Vol. 2021.