

# DEVELOPMENT OF MODELS TO STUDY TRAFFIC ACCIDENTS ON THE FINAL SECTIONS OF ACCESS ROADS TO THE CITIES: A CASE STUDY OF THREE MAJOR IRANIAN CITIES

Mehdi FALLAH TAFTI<sup>1</sup>, Reza ROSHANI<sup>2</sup>

<sup>1,2</sup> Department of Civil Engineering, Yazd University, Yazd, Iran

## Abstract:

The final sections of main access roads to the cities require especial attention as the frequency of accidents in these road sections are considerably higher than other parts of interurban roads. These road sections operate as an interface between the rural roads and urban streets. The previous researches available on this subject are limited and they have also mainly focused on a narrow range of factors contributing to the accidents in these areas. The main contribution of this research is to consider a relatively comprehensive range of potential factors, and to examine their impacts through the development and comparison of both conventional probabilistic models and Artificial Neural Network (ANN) models. For this purpose, information related to the main access roads of three major Iranian cities were collected. This information consisted of accident frequency data together with the field observations of traffic characteristics, roadway conditions and roadside features of these roads. Various ANN and probabilistic models were developed. The frequency of accidents, i.e. fatal, injured, or damaged accidents, was considered as the output of the developed models. The results indicated that a hybrid of ANN models, each comprised of 10 input variables representing traffic, roadway and roadside conditions, outperformed several probabilistic models, i.e. Poisson, Negative binomial, Zero-truncated Poisson, and Zero-truncated Negative Binomial models, also developed under similar conditions in this study. Moreover, effective roadway width, roadway lighting condition, the standard deviation of vehicles speed, percentage of drivers violating the speed limit, average annual daily traffic, percentage of heavy goods vehicles, the density of roadside commercial and industrial landuses, the density of median U-turns, the density of local access roads, and the effective width of the left-side shoulder were identified as the most effective factors contributing to the accidents in these areas. The developed ANN model can be used as a tool to predict accident rates in these road sections, and to estimate a potential reduction in the accident rates, following any improvements in the major factors contributing to the traffic accidents in these areas.

**Keywords:** city access roads, city entrance roads, traffic safety, accident prediction models, neural networks

## To cite this article:

Fallah Tafti, M., Roshani, R., 2021. Development of models to study traffic accidents on the final sections of access roads to the cities: a case study of three major Iranian cities. Archives of Transport, 59(3), 129-148. DOI: <https://doi.org/10.5604/01.3001.0015.2646>



## Contact:

1) Fallah.Tafti@Yazd.ac.ir [<https://orcid.org/0000-0001-6844-3316>] - corresponding author,  
2) R.Roshani@stu.yazd.ac.ir [<https://orcid.org/0000-0002-1088-1556>]

## 1. Introduction

Road safety has been one of the main concerns of road users and highway authorities over the years. The Global Status Report on Road Safety published by the World Health Organization in 2018 (WHO, 2018), indicated that the annual number of road accident fatalities was around 1.35 million people in 2016 and it is the leading cause of fatality among the children and young adults aged 5 to 29 years. It is anticipated that by 2030, the road accidents will be the fifth most important cause of human fatality worldwide (WHO, 2018). Traffic accidents would also impose a great burden on the families, economy, health sectors, insurance companies, etc. in each country.

In Iran, road transportation mode nationally has a share of around 94% among all transportations modes and therefore, the provision of safety for this mode is an essential issue. (Khalili and Pakgozar, 2013). The spatial pattern of accident distribution in this country, indicates that around 40% and 70% of traffic accidents in rural roads have occurred within the 5km and 30km distances from the city entries, respectively. Considering the fact that these main access roads to the cities only constitute around 17% of the rural roads in this country, it can be concluded that these road sections are more prone to the traffic accidents than other rural road sections (Afandizadeh and Golshan Khavas, 2006). An access road to a city acts as an interface and transitional section between the rural roads and urban streets. Therefore, its functionality, physical characteristics and surroundings are a mixture of these two different environments. This situation would create confusion and inappropriate driving behavior along these road sections which in turn will result in reduced safety and increased accidents (Davoodi and Ahmadi, 2015). Thus, the safety of such road sections should be treated with a higher priority than other rural road sections.

An initial but important issue in the study of traffic safety along the main access roads to the cities is to decide on the overall study length. Through a literature search, the authors could not find any concrete guidance or criteria for this. However, a road length of 30km is widely adopted by the Iranian officials in the Ministry of Roads and Urban Development and also the Police as the influence area. Thus, this length was adopted in this research as well.

Due to the transitional status and mixed function of main access roads to the cities, a wider range of factors comprising both rural and urban related factors, which may contribute to the occurrence of accidents along these road sections, should be considered. Pedestrians and two-wheeled motorized and non-motorized vehicles, such as motorcycles and bicycles, are highly involved in the urban mobility and due to their vulnerability features, are highly subjected to the traffic accidents. On this basis, a number of studies have been conducted in the past to estimate the frequency and severity of accident accidents in urban areas (Abdel-Aty and Pande, 2005; Gomes et al., 2012; Pulugurtha and Sambhara, 2011; Khishdari and Fallah Tafti, 2017; Alarifi et al., 2018; Yuan, et al., 2019; Thakur and Biswas, 2019; Carey, et al., 2020). In some other studies, the high interaction between urban traffic and inconsistent landuses located along the streets are enumerated as one the major causes of traffic accidents (Edquist et al., 2012; Ukkusuri et al., 2012; Narayanamoorthy et al., 2013; Das and Burger, 2015). However, factors contributing to the accidents on rural roads could be different from urban streets. This may be attributed to their high-speed driving environment and much lower presence of pedestrians and other non-motorized traffic. Furthermore, the roadway and traffic characteristics of rural roads and their surrounding environment are different from the urban streets. This is another reason for the apparent differences between the nature of accidents in urban and rural roads (Holdridge et al., 2005; Mustakim et al., 2008; Huang et al., 2014; Zou et al., 2014; Elvik, 2017; Gaca and Pogodzińska, 2017; Wang, et al., 2019; Saeed, et al., 2019; Yang, et al., 2019; Ambros et al., 2020; Abounoas et al., 2020). Moreover, the geometric features of rural roads has a considerable effect on their safety conditions. In this regard, previous researchers have shown that the geometric factors such as roadway width, median width, shoulder width and horizontal and vertical curves have a significant role on the occurrence of accidents in these road sections (Caliendo et al., 2007; Ben-Bassat and Shinar, 2011; Lee et al., 2015; Mecheri et al., 2017; Hou, et al., 2018; Ambros et al., 2019; Wu, et al., 2019; Pu, et al., 2020; Elamrani Abou El Assad, et al., 2020; Moomen, et al., 2020; Ma, et al., 2020).

As mentioned earlier, due to the transitional role of the major access roads to the cities, a combination of factors contributing to the accidents in the urban

streets and rural roads should be considered for these roads. A few researches available on this subject have also confirmed this assumption (Fitzpatrick et al., 2001; Afandizadeh and Golshan Khavas, 2006; Potts et al., 2007; Dadashova et al., 2016; Vayalankuzhi and Amirhalingam, 2016). For instance, Afandizadeh and Golshan Khavas (2006), in their research on the safety of access roads to the cities, identified six parameters including roadway width, Average Annual Daily Traffic (AADT), number of local access roads along these road sections, population of the destination city, road longitudinal gradient, and the road section length as the major factors contributing to the frequency of accidents in these road sections (Afandizadeh and Golshan Khavas, 2006). Also, Khalili and Pakgozar (2013), identified the road lighting, density of retailing and other landuse activities along these road sections and the frequency of local access roads, as the major factors contributing to the accidents in these areas. A number of other researches have reported that unexpected factors such as the existence of various retailing, industrial and agricultural activities along these road sections, roadside parked vehicles, and turning movements of vehicles at the local intersections would complicate the situation in these areas, causing further deteriorations to the through traffic flow and degrading the safety of road users in these areas. (Wang, Song, et al., 2014; Saha et al., 2015; Elyasi et al., 2016; Wang, Yuan, et al., 2018).

In terms of study approach, previous researches have widely relied on the development of accident prediction models as a tool to investigate the underlying factors contributing to the occurrence of traffic accidents. For this purpose, various statistical and probabilistic models and more recently, artificial intelligence based models are used. Among these models, the probabilistic models are widely used by researchers for both urban and rural road areas (Caliendo et al., 2007; Potts et al., 2007; Gomes, Geedipally, and Lord, 2012; Wang, Song, et al., 2014; Khishdari and Fallah Tafti, 2017; Saeed et al., 2019; Moomen et al., 2020). More recently, artificial intelligent based models, e.g. Artificial Neural Networks (ANN) models are also used as a potential approach to predict the number of traffic related accidents by some reseachers (Dougherty, 1995; Abdelwahab and Abdel-Aty, 2001; Abdel-Aty and Pande, 2005; Chang, 2005; Xie et al., 2007; Akgungor et al., 2008; Bayata et al., 2011; Khishdari and Fallah Tafti,

2017; Rûžička et al., 2019; Yuan et al., 2019; Tercan, 2020).

Considering higher rates of traffic accidents on these road sections and the fact that previous researches available on this subject are insufficient, mainly focusing on a narrow range of factors contributing to the accidents in these areas, further researches in which a wider range of sites and factors are considered, would be beneficial. Hence, a major contribution of this research is to consider a relatively comprehensive range of factors that may have a significant impact on the frequency of traffic accidents (crashes) along these road sections. The factors considered in this research are: lane width, number of carriageway lanes, shoulder width, road lighting condition, road pavement condition, traffic composition, median width, actual driving speeds and deviation from speed limits, along with some other factors that would indicate surrounding land use and road traffic characteristics. Another major contribution of this research is to examine a range of probabilistic accident prediction models for these areas and to identify the outperformed model and compare its performance with an appropriate ANN model, also developed for these road sections. From the outperformed model, it would be possible to identify the dominant factors that are contributing to the accident occurrence, to use the developed model to predict the accident rates, and to estimate a potential reduction in the accident rates, following any improvements in the major factors contributing to the traffic accidents in these road areas.

In addition to the parameters mentioned above, the driver characteristics and driving behavior are also important features that should be taken into account. However, in the this research, these factors are not considered as the main purpose of this research was to investigate only the impacts of traffic, geometric and surrounding environment of the main access roads to the cities on the frequency of accidents in these road sections.

## 2. Material and Methods

### 2.1. Data collection process

The data collection process used for the development of accident prediction models and to identify the models with overall best performance is presented in this section.

In this research, the data related to the main access roads of three cities namely Isfahan, Kerman, and

Yazd were collected. All three cities are the capital of their provinces in Iran with around 2.4, 0.6 and 0.6 million population, respectively. Fig. 1 shows the bird's-eye views of these cities and their main access roads. The leading parts of rural roads, providing main access to each city, were considered. A road length of 30km from the city entrance was considered for each access road unless there was a small town along this distance, which in this case, a shorter length was adopted for that access road. In total, 565km stretches of road carriageways related to the access roads of these three cities were examined and the required information was gathered through the site visits and further data acquired from their corresponding road authorities. The data related to 325km of these roads were used for the model development and the remaining data was used for the validation of the developed models. In order to consider the variations in the factors contributing to the traffic accidents in these road areas, each 30km access road was subdivided into 6 road sections, each with 5km length. Therefore, in total, 113 carriageway sections were investigated in this research. The details of access roads that were examined in this study are presented in Table 1.

As indicated in previous section, one of the main objectives of this research was to identify the major factors contributing to the traffic accidents on the access roads to the cities. To achieve this objective, a set of 16 potential factors or variables were identified. These factors were initially extracted from the factors identified in the previous researches (e.g. Abdel-Aty and Pande, 2005; Holdridge et al., 2005; Mustakim et al., 2008; Pulugurtha and Sambhara, 2011; Edquist et al., 2012; Gomes et al., 2012; Ukkusuri et al., 2012; Narayanamoorthy et al., 2013; Huang et al., 2014; Zou et al., 2014; Das and Burger, 2015; Davoodi and Ahmadi, 2015; Elvik, 2017; Khishdari and Fallah Tafti, 2017) and then complemented by additional likely factors that were identified through the site visits accompanied by the experienced staff in the Iranian Local Road Authorities. The overall variables considered in this study and their assigned codes for data analysis are shown in Table 2. The accidents data related to each carriageway section were collected from the local road authorities and the police for a period of 3 years (2015 to 2017). The definition of accident in this study is any incident in which when a vehicle collides with another vehicle, pedestrian, animal, road debris, or

other stationary obstructions, such as a tree, pole or building. Traffic accident may result in injury, disability, death, and property damage. A Summary of descriptive statistics related to the data gathered for this study is presented in Table 3. It should be noted that, because of the difference in the nature of these data items, they have been normalized, using a proven method presented in Equation 1 (Sukirty et al., 2018; Jiawei and Micheline, 2006):

$$X = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where:

$X$  – the normalized variable, the real value of  $i^{\text{th}}$  variable;

$x_i$  – the real observed minimum value of  $i^{\text{th}}$  variable;

$x_{min}$  – the real observed minimum value of  $i^{\text{th}}$  variable;

$x_{max}$  – the real observed maximum value of the  $i^{\text{th}}$  variable.

## 2.2. Model development

In this section, the methodology used for the development of accident prediction models and to identify the models with overall best performance is presented. Then using these models, underlying traffic, roadway and roadside features contributing to the traffic accidents along the main access roads to the cities are identified.

The normalized data, obtained from the procedure explained in the previous section, was used for the development and comparison of probabilistic and ANN models which will be described in the following sections. The probabilistic and ANN models were developed using STATA Ver.14. and MATLAB 2018.b software, respectively.

In order to compare the performance of the developed statistical models, the Goodness Of Fit (GOF) measures including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), McFadden pseudo R, McFadden adjusted pseudo R and Root Mean Square Error (RMSE) were used. These indicators have widely been used in the past, and it has been proven that they are powerful measures to compare the fitness level of the probabilistic models such as Poisson and Negative Binomial family of models (Khishdari and Fallah Tafti, 2017). This measure is calculated using Equation 2:

$$RMSE = \sqrt{\frac{1}{t} (\sum_{i=1}^t (O_i - E_i)^2)} \quad (2)$$

Where:

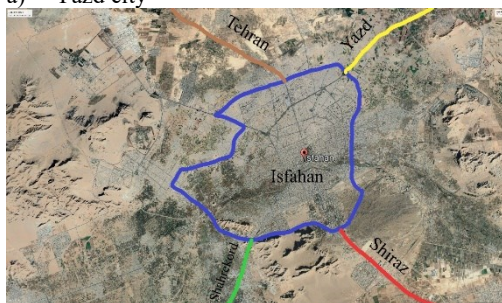
$O_i$  – the annual average number of observed accidents in the  $i^{th}$  road carriageway section;

$E_i$  – the average annual number of expected accidents in that section which has been estimated by the developed model;

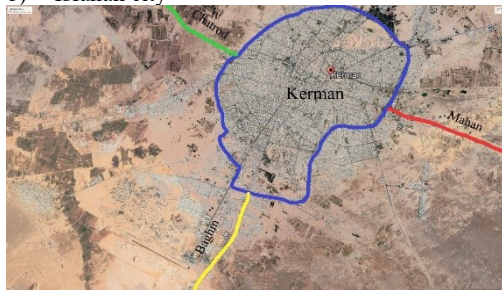
' $t$ ' – the number road carriageway sections considered for each access road to the city.



a) Yazd city



b) Isfahan city



c) Kerman city

Fig. 1. A bird's eye view of the Yazd, Isfahan and Kerman cities

Table 1. The main access roads to the cities considered in this study

City	Access Road	Number of Carriageway Sections	Total Carriageways Length (km)
Isfahan	Isfahan - Tehran	12*	60**
	Isfahan - Shiraz	12	60
	Isfahan – Shahre Kord	12	60
	Isfahan - Yazd	12	60
Yazd	Yazd – Meybod	12	60
	Yazd – Mehriz	12	60
	Yazd - Taft	9	45
Kerman	Kerman – Baghin	12	60
	Kerman – Chatrod	10	50
	Kerman - Mahan	10	50

Moreover, overfitting control and sensitivity analysis were used to identify the input variables with significant impact on the model output and also for the ranking of these variables. It would then also be possible to develop further models in which only variables with significant impact are used and thereby the number of input variable required by the model can be reduced. A model with the reduced number of input variables would be preferable as it would require less data items, provided that it provides a level of accuracy comparable to the original model. In this research, due to the difference in the nature of probabilistic and ANN models, their sensitivity analysis was performed differently. The procedure used for the sensitivity analysis of probabilistic models and ANN models are shown in Figs. 2 and 3., respectively. The sensitivity analysis of the probabilistic and ANN models was carried out using the step-backward analysis and the randomization test, respectively. Finally, based on the sensitivity magnitude of the model output to the input variables (e.g.  $p$  values obtained for each input variable), the non-significant variables were eliminated, while still maintaining the same overall performance of the model. Through this process, the model with the minimum number of input variables and almost the same level of accuracy was selected as the final model.

### 2.3. Results of probabilistic models Development

Two probabilistic models that are widely used for traffic accident frequency prediction namely, the

Poisson and Negative Binominal and their deviations were developed in this research. The Poisson probability distribution model can be deviated from the Binomial distribution. When the number of

binomial tests reaches to infinity, its corresponding Binomial distribution tend to approach a Poisson distribution (Hilbe, 2011).

Table 2. The characteristics of variables used in this study

Variable	Type	Assigned Code	Unit
Average Annual Number of Accidents over 3 years (2015- 2017)	Dependent	TA	Number
Effective Roadway Width	Independent	RWidth	m
Effective Width of Right-Side Shoulder	Independent	Right Sh	m
Effective Width of Left-side Shoulder	Independent	Left Sh	m
Median Width	Independent	Med Width	m
Percentage of the Road Section Provided with Lighting	Independent	Lighting St	%
Pavement Condition	Independent	Con Pavement	Qualitative (1,2,3,4)
Density of Commercial and Industrial landuses Along the Road Section	Independent	Commercial	m/m
Density of Agricultural landuses Along the Road Section	Independent	Agricultural	m/m
Density of Access Roads	Independent	Den Access	Numbers per unit kilometer
Density of Roadside Parking Spaces	Independent	Numb Parking	Numbers per unit kilometer
Density of Median U-turns	Independent	Uturns	Numbers per unit kilometer
Average Distance of each section from the City Entrance	Independent	Distance	kilometer
Average Annual Daily Traffic (AADT)	Independent	AADT	Number
Percentage of Heavy Goods Vehicles	Independent	Heavy Traffic	%
Percentage of Drivers Violating Speed Limit	Independent	Unallowable Speed	%
Standard Deviation of Vehicle Speeds	Independent	SpeedDeviation	Km/hr

Table 3. Summary of descriptive statistics related to the data gathered for this study

Variable	Mean	Std. dev.	Min	Max	Cor. Coeff.	P - Value
Average Annual Number of Accidents over 3 years (2015- 2017)	5,969	3,588	0	16	NA	NA
Effective Roadway Width	8,74	1,402	7	11,5	0,359	0.027 <0.05
Effective Width of Right-Side Shoulder	1,65	0,203	1,1	2,2	-0,054	0.049=0.05
Effective Width of Left-side Shoulder	0,455	0,12	0,2	0,8	-0,215	0.039 <0.05
Median Width	7,15	3,173	0,3	10	-0,331	0.092 >0.05
Percentage of the Road Section Provided with Lighting	0,658	0,425	0	1	0,127	0.020 <0.05
Pavement Condition	2,877	0,331	2	3	-0,032	0.352>0.05
Density of Commercial and Industrial landuses Along the Road Section	0,017	0,027	0	0,091	0,085	0.028 <0.05
Density of Agricultural landuses Along the Road Section	0,019	0,03	0	0,096	-0,172	0.082 >0.05
Density of Access Roads	1,354	0,896	0	3,4	0,131	0.033<0.05
Density of Roadside Parking Spaces	0,271	0,216	0	0,8	-0,2	0.214>0.05
Density of Median U-turns s	0,431	0,277	0	1	-0,237	0.042 <0.05
AADT	9395	3602	4904	14944	0,308	0.036 <0.05
Percentage of Heavy Goods Vehicles	17,9	7,1	4	28	0,058	0.282>0.05
Standard Deviation of Vehicle Speeds	19,9	9,4	8,3	35,3	-0,051	0.014 <0.05
Percentage of Drivers Violating Speed Limit	9,5	4,7	0,3	17,7	0,019	0.042 <0.05
Average Distance of each section from the City Entrance	14,1	8,3	2,5	27,5	-0,127	0.171>0.05

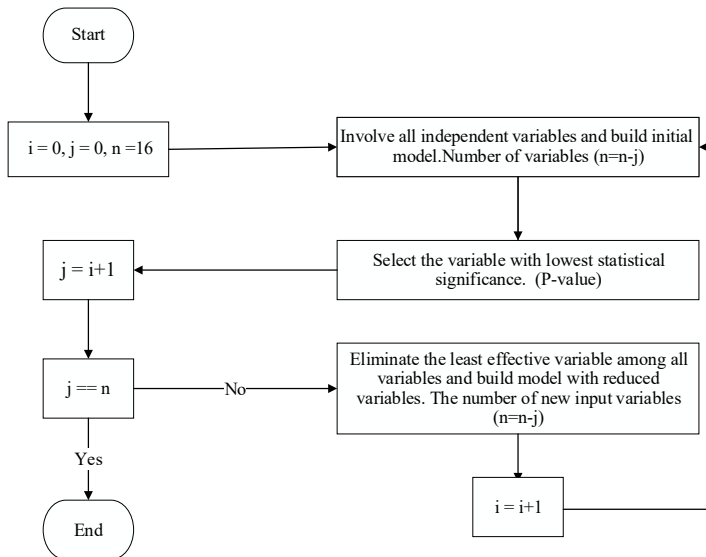


Fig. 2. Flowchart of iterative back-forward process used for input variable selection of the examined Probabilistic Models (where  $n$  is the number of input variables  $i$  and  $j$  are parameters used to indicate iteration process of the methodology)

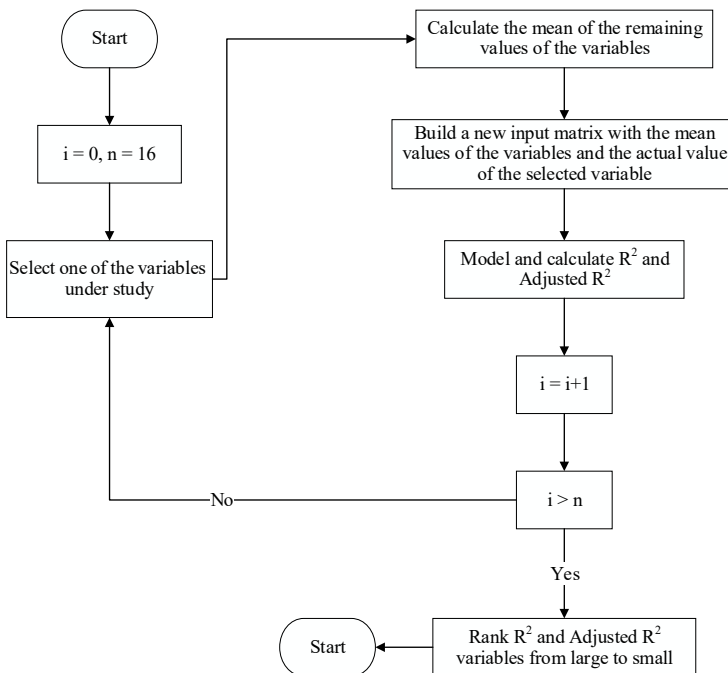


Fig. 3. Flowchart of the Sensitivity Analysis process used for the Neural Network Models (where  $n$  is the number of input variables  $i$  and  $j$  are parameters used to indicate iteration process of the methodology)

The overall form of Poisson probability distribution functions is as Equation 3 (Olkin et al., 1980):

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \tag{3}$$

Where:

$y_i$  – the number of accidents occurred in the  $i^{th}$  carriageway section of the studied road;

$\mu_i$  – the average number of accidents occurred in that section, which can be calculated using Equation 4:

$$\mu_i = \exp(\beta X_j) \quad j = 1, 2, 3, \dots, n \tag{4}$$

Where:

$\beta$  – a  $(1 \times n)$  vector of regression coefficients;

$X_j$  – the  $j^{th}$  study variable.

For the determination of the regression coefficients of Poisson model, the Logarithm Likelihood should be used. One issue associated with the Poisson model is that it assumes that the variance and average values of the dependent variable (model output) are the same (Hilbe, 2011). This assumption has rarely been observed for accident data in practice. Depending on the data types used in the modeling process, the variance of dependent variable can be lower or higher than the average (Cameron and Trivedi, 1998).

To overcome this issue, the Negative Binomial model, first introduced by Cameron and Trivedi, is usually used (Cameron and Trivedi, 1998). In this model, the over-dispersion parameter ( $\alpha$ ) was used to include the inequality of average and variance of data. The  $\alpha$  parameter can be calculated using Equation 5.

$$Var(y_i) = E(y_i)[1 + \alpha E(y_i)] \tag{5}$$

Where:

$y_i$  – the average frequency of the expected number of occurred accidents in the  $i^{th}$  carriageway section of the road under study;

$Var(y_i)$  – the variance of accidents in the  $i^{th}$  section of the road under study.

Then, the Negative Binomial distribution can be obtained using Equation 6:

$$f(y_i; \mu_i, \alpha) = \frac{\Gamma(y_i+1/\alpha)}{\Gamma(y_i+1) \Gamma(1/\alpha)} \left( \frac{1}{1+\alpha\mu_i} \right)^{\frac{1}{\alpha}} \left( 1 - \frac{1}{1+\alpha\mu_i} \right)^{y_i} \tag{6}$$

Where all parameters of this equation are previously defined.

In practice, another type of the over-or-under-dispersion can occur due to the outliers, parts of data that are extremely higher or lower than the overall data, leading to over-dispersion and under-dispersion, respectively (Saffari et al. 2011). To resolve this issue, the Zero-truncated Poisson and the Zero-truncated Negative Binomial models were introduced.

Following the development of the Poisson, Negative binomial, Zero-truncated Poisson, and Zero-truncated Negative Binomial models, the sensitivity analysis was performed in accordance with the process presented in Fig. 4. Finally, Fig. 5. was plotted to determine the appropriate number of input variables and to overcome overfitting issue. In order to address the overfitting issue, it would be more appropriate to consider adjusted  $R^2$  values rather than  $R^2$  values, as the earlier would be able to identify the number of input variables with sensible impact on the overall performance of the model. Furthermore, the number of input variables should be limited to the maximum number of input variables beyond which the Adjusted  $R^2$  values would be reduced. On this basis, the number of appropriate variables to prevent overfitting issue can be observed from second order polynomial curves fitted to the results in Fig. 5. According to this Figure, it can be observed that the best performance of Poisson and Negative Binomial regression models was achieved by 9 input variables and the best performance of Zero-truncated Poisson and Zero-truncated Negative Binomial models was achieved by 11 input variables.

For each probabilistic model type, the appropriate input variables were selected according to the ranking list obtained from the sensitivity analysis to prevent the overfitting problem. New probabilistic models were developed based on these reduced number of input variables and compared using appropriate measures presented in Table 4. According to the values of performance measures presented in Table 4, it can be concluded that the Zero-truncated Poisson model comprising of 11 input variables has demonstrated an overall better performance than other probabilistic models. The input variables of this model in descending order of their significance are: standard deviation of vehicle speeds, road lighting status, the density of commercial and industrial



landuses, the density of local access roads, percentage of heavy goods vehicles, effective roadway width, AADT, the effective width of left-side

shoulder, the density of agricultural landuses, the effective width of right-side shoulder, and density of median U-turns.

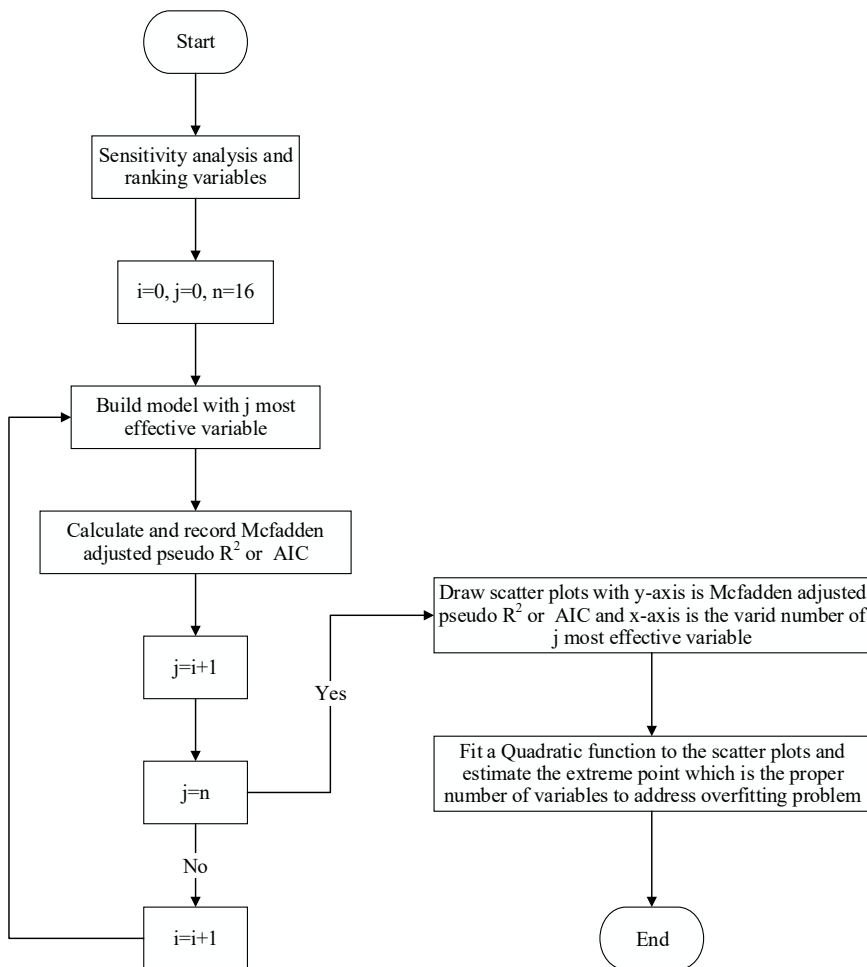


Fig. 4. Methodology used for the illumination of over-fitting issue of models (where n is the number of input variables *i* and *j* are parameters used to indicate iteration process of the methodology)

Table 4. Comparison of the performance of Poisson and negative binomial models

Performance measure	Poisson	Zero-truncated poisson	Negative Binomial	Zero-truncated Negative Binomial
McFadden pseudo R <sup>2</sup>	0,091	0,342	0,119	0,253
McFadden adjusted pseudo R <sup>2</sup>	0,04	0,231	0,061	0,163
AIC	208,09	202,19	207,43	2.4.28
BIC	216,89	217,64	216,97	217,37
RMSE	0,125	0,112	0,139	0,116

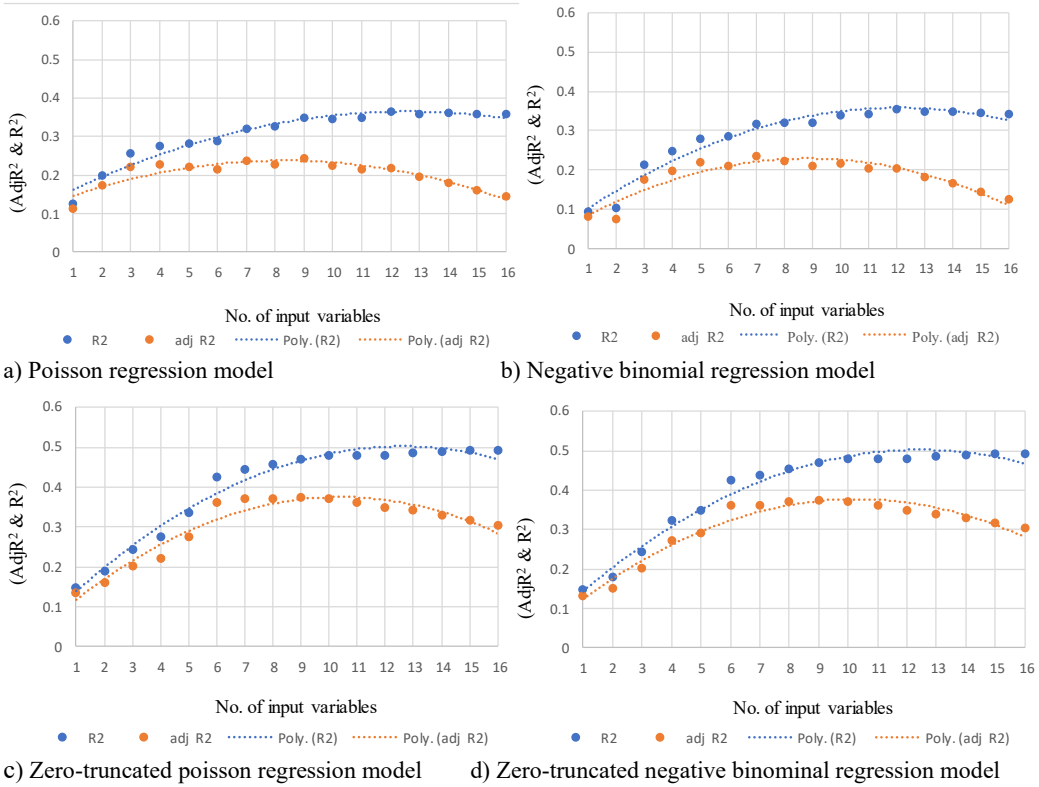


Fig. 5. Results of investigating the effect of the number of input variables to avoid over-fitting issue in developed models

**2.4. Results of ANN models development**

Generally, an ANN model consists of three layers: the input layer, the hidden layer, and the output layer. Here, the input layer is related to the independent variables which are effective on the frequency of accidents, the output layer is related to the average annual number of accidents in the  $i^{th}$  carriageway section of the road under study. The hidden layer provides an interface between the input and output layers. The feedforward backpropagation neural network is one of the most common types of ANN models, which is widely used for the prediction type models and was therefore used in this research. In this research, a hybrid structure was used for the ANN model. The hybrid structure consists of 10 ANN model each with the same input variables and the same number of neurons in the hidden layer but each model was trained with a different initial random number or seed number. The hybrid model

calculates the average output of these 10 ANN models as the final model result. For the development of this model, the "trainlm" function in MATLAB 2018.b was used to train the ANN models. In this process, Levenberg-Marquardt optimization technique was used to train the ANN models. The structure of the developed hybrid ANN model is shown in Fig. 6.

During the ANN modeling process, the collected data is usually divided into three separate parts for training, testing and validation of the model. Usually, a predefined percentage of the available data are used for each of the above-mentioned parts and each part of the data is randomly selected from the available data. To minimize the impact of data ratio selection on the model performance, two commonly used combinations of data proportions were tested in this research as follows:

- 70% (train) + 15% (test) + 15% (validation)

– 60% (train) + 20% (test) + 20% (validation)

To develop the ANN model, 16 independent variables were considered as the initial input variables of the model. As for the number of neurons in the hidden layer, it has been suggested to choose the number of neurons in the hidden layer in the range of number of output variables and the sum of input and output variables (Heaton, 2010). Thus, for each of the above-mentioned data proportions, the number of neurons in the hidden layer was increased in each step by 2, in the range of 1 to 17 neurons. In each step, the hybrid structure for the neural network shown in Fig. 6. was used to develop 10 similar ANN models each trained with a different initial seed number. The best hybrid model was then identified by comparing the performance of these models. Table 5 shows the codes assigned to each one of the developed hybrid models. As mentioned earlier, the difference between in these models are associated with the number of neurons used in the hidden layer as well as the data proportions used to train, test and validate the model.

The RSME was used as a measure of performance to compare the performance of the developed hybrid ANN models with each other and to identify outperformed model. This measure has successfully been used in previous researches as well and is calculated using Equation 2. (Chang, 2005; Xie et al., 2007; Li et al., 2008; Khishdari and Fallah Tafti, 2017; Liu et al., 2020).

Fig. 7. shows the RMSE values of the developed hybrid ANN models. According to the results presented in this Figure, it can be concluded that the ANN model comprising 16 input variables and 11 neurons in the hidden layer (NN6), developed with the train-test-validation percentages of 70-15-15, has demonstrated the better performance among the developed models.

### 2.5. Sensitivity analysis of the best Hybrid ANN model

In this section, the results of the sensitivity analysis, performed on the best developed ANN model (NN6) using the randomization test procedure shown in Fig. (3) is presented. The results presented in Fig. 8. indicate the degree of effectiveness of each input variable to the model output. . In this Figure, the variables shown on the horizontal axis are previously defined in Table 2.

Table 5. The coding used for the hybrid ANN models developed in this research

ANN model code	Train-Test- Validation % Ratio	No. of neurons in the hidden layer
NN1	70-15-15	1
NN2	70-15-15	3
NN3	70-15-15	5
NN4	70-15-15	7
NN5	70-15-15	9
NN6	70-15-15	11
NN7	70-15-15	13
NN8	70-15-15	15
NN9	70-15-15	17
NN10	60-20-20	1
NN11	60-20-20	3
NN12	60-20-20	5
NN13	60-20-20	7
NN14	60-20-20	9
NN15	60-20-20	11
NN16	60-20-20	13
NN17	60-20-20	15
NN18	60-20-20	17

The results of sensitivity analysis were then used to evaluate the performance of the new ANN models developed with reduced number of input variables. In each step, the least effective variable was eliminated and a range of new ANN models, comprising different neurons in the hidden layer, were developed. The optimum ANN model for that step was identified as the model with the maximum R<sup>2</sup> value. The various conditions in which these ANN models were developed and evaluated are listed in Table 6. Moreover, the characteristics of the best ANN models developed with different number of input variables are presented in Table 7. The performance of outperformed ANN models in each step were then compared to identify the ANN model with the best overall performance. The ANN model with the best overall performance was identified as the model with the least RMSE and no overfitting issue. The variation of the RMSE against the reduced number of input variables is shown in Fig. 9. The overfitting issue was investigated using the plots of R<sup>2</sup> and adjusted R<sup>2</sup> values against the number of input variables in accordance with Fig. 10. According to these Figures, the hybrid ANN model comprising 10 input variables, 11 neurons in the hidden layer, developed with the data ratio of 70-15-15, provided the least RMSE value, and with no overfitting issue. Thus, this model was identified as the best overall Hybrid

ANN model. The input variables in this model in descending order of their significance are: effective roadway width, road lighting status, standard deviation of vehicle speeds, percentage of drivers violated speed limit, AADT, percentage of heavy goods vehicles, density of roadside commercial and industrial landuses, density of median U-turns, density of local access roads, and effective width of the left-side shoulder.

A comparison of the two outperformed model types namely, the Zero-truncated Poisson model and the hybrid ANN model, indicate that these models possess 9 input variables in common. These common input variables are: standard deviation of vehicle speeds, road lighting status, density of commercial and industrial landuses, density of local access roads,

percentage of heavy goods vehicles, effective roadway width, AADT, effective width of left-side shoulder and density of median U-turns.

### 3. Summary of results, discussion and conclusions

In this research, a series of two model types were developed to anticipate the frequency of traffic accidents along the road carriageway sections located at the leading parts of the main access roads to three major cities of Iran. These two model types were: the probabilistic models and the ANN models.

A Zero-truncated Poisson model, comprising 11 input variables, was identified as the best probabilistic model. All of these variables are measured along each road carriageway section.

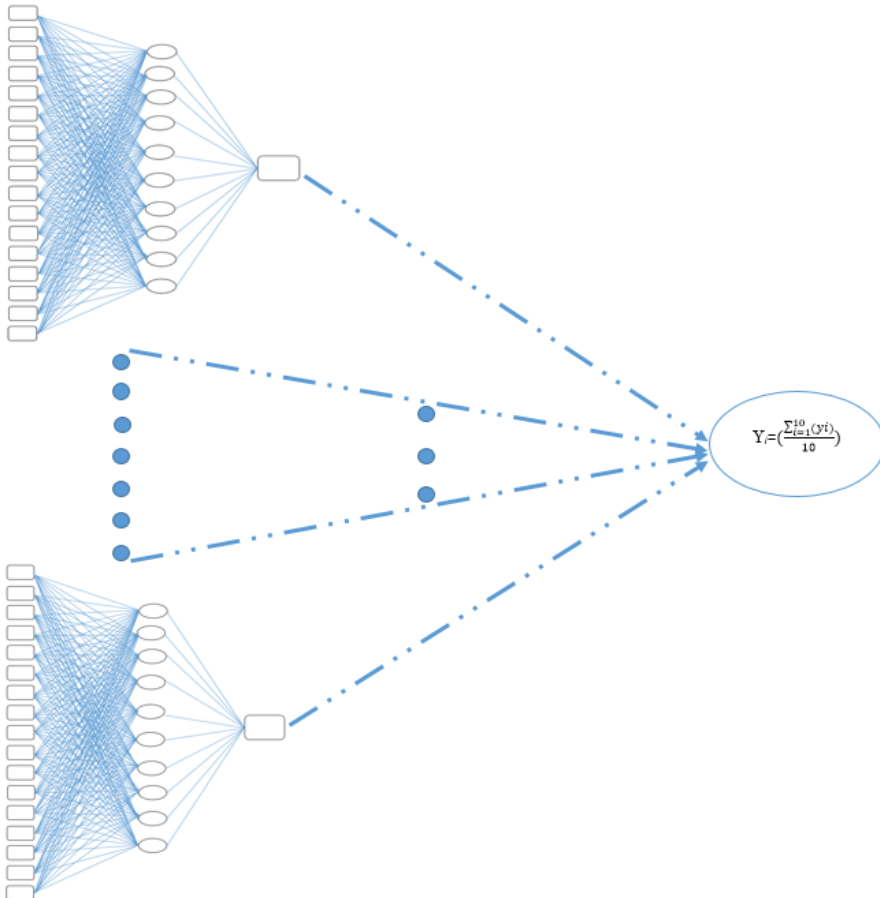
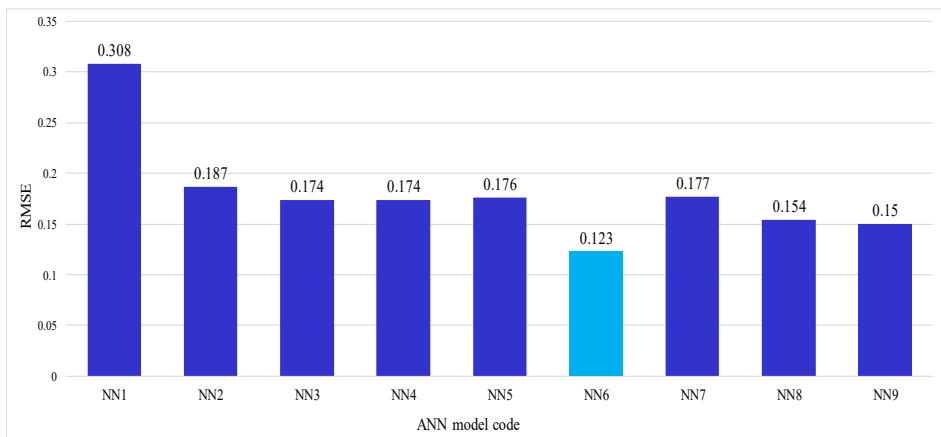
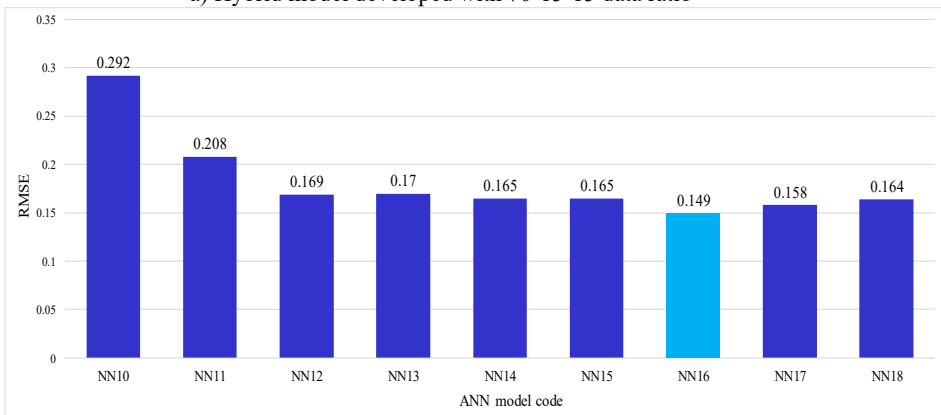


Fig. 6. The Structure of the adopted hybrid ANN model in this research



a) Hybrid model developed with 70-15-15 data ratio



b) Hybrid model developed with 60-20-20 data ratio

Fig. 7. The RMSE values of the developed hybrid ANN models

As for the ANN models, a Hybrid ANN model, consisting of 10 similar ANN models and each comprising of 10 input variables, 11 neurons in the hidden layer, developed with the train-test-validation data proportions of 70-15-15 percentages, demonstrated the best performance in comparison with other developed ANN models. All of these variables are measured along each road carriageway section.

The best ANN and probabilistic models were then validated with the data related to the main access roads of the Isfahan city. This fresh data was not used throughout the model development process. This data was used to validate the generalization ability of the best Zero-truncated Poisson and Hybrid ANN models. For this purpose, the RMSE and  $R^2$  values of these two models were calculated under

the modeling and validation data sets. The results are presented in Table 8. Furthermore, a comparison of the normalized values of accident frequencies in different sections obtained from the final Zero-truncated Poisson and Hybrid ANN models with their corresponding observed values are presented in Fig. 11, indicating an overall better performance of the Hybrid ANN model.

As for the ANN models, a Hybrid ANN model, consisting of 10 similar ANN models and each comprising of 10 input variables, 11 neurons in the hidden layer, developed with the train-test-validation data proportions of 70-15-15 percentages, demonstrated the best performance in comparison with other developed ANN models. All of these variables are measured along each road carriageway section.

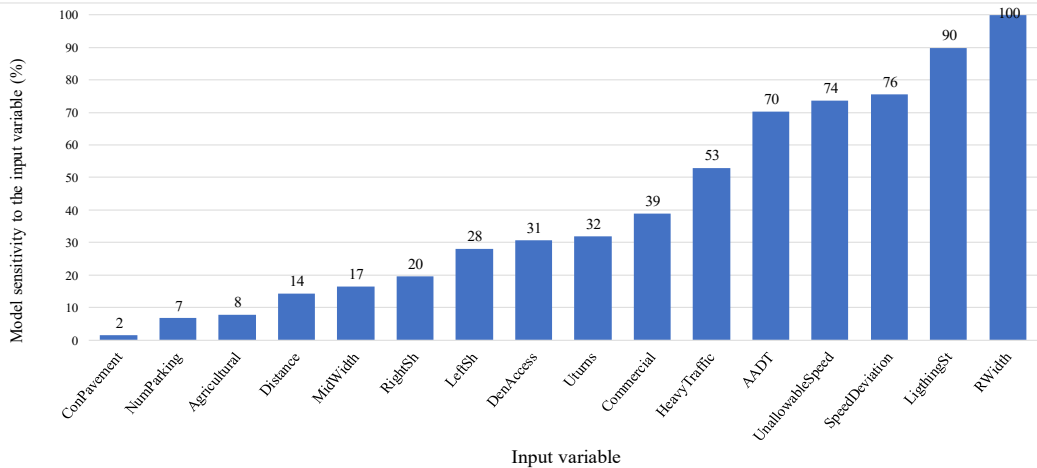


Fig. 8. The Sensitivity of each input variable to the NN6 model output

Table 6. Various ANN models examined to determine the best model

No. of input variables	Different neurons considered for the hidden layer	Number of train-test-validation sets	Total No. developed ANN models
16	1-3-5-7-9-11-13-15-17	2	18
15	1-3-5-7-9-11-13-15	2	16
14	1-3-5-7-9-11-13-15	2	16
13	1-3-5-7-9-11-13	2	14
12	1-3-5-7-9-11-13	2	14
11	1-3-5-7-9-11	2	12
10	1-3-5-7-9-11	2	12
9	1-3-5-7-9	2	10
8	1-3-5-7-9	2	10
7	1-3-5-7	2	8
6	1-3-5-7	2	8
5	1-3-5	2	6
4	1-3-5	2	6
3	1-3	2	4
2	1-3	2	4
1	1	2	2
Total No. of developed ANN models		160	

The best ANN and probabilistic models were then validated with the data related to the main access roads of the Isfahan city. This fresh data was not used throughout the model development process. This data was used to validate the generalization ability of the best Zero-truncated Poisson and Hybrid ANN models. For this purpose, the RMSE and  $R^2$  values of these two models were calculated under

the modeling and validation data sets. The results are presented in Table 8. Furthermore, a comparison of the normalized values of accident frequencies in different sections obtained from the final Zero-truncated Poisson and Hybrid ANN models with their corresponding observed values are presented in Fig. 11, indicating an overall better performance of the Hybrid ANN model.

Table 7. The characteristics of the best ANN models developed with different no. of input variables

Number of input variables	Train-test-validation %	No. of neurons in the hidden layer	R <sup>2</sup> value
16	70-15-15	17	0.528
15	70-15-15	13	0.528
14	70-15-15	15	0.530
13	60-20-20	7	0.529
12	70-15-15	11	0.533
11	70-15-15	7	0.520
10	70-15-15	11	0.516
9	70-15-15	5	0.519
8	70-15-15	9	0.494
7	70-15-15	5	0.490
6	60-20-20	3	0.457
5	70-15-15	3	0.453
4	70-15-15	3	0.444
3	70-15-15	3	0.427
2	70-15-15	3	0.369
1	70-15-15	1	0.296

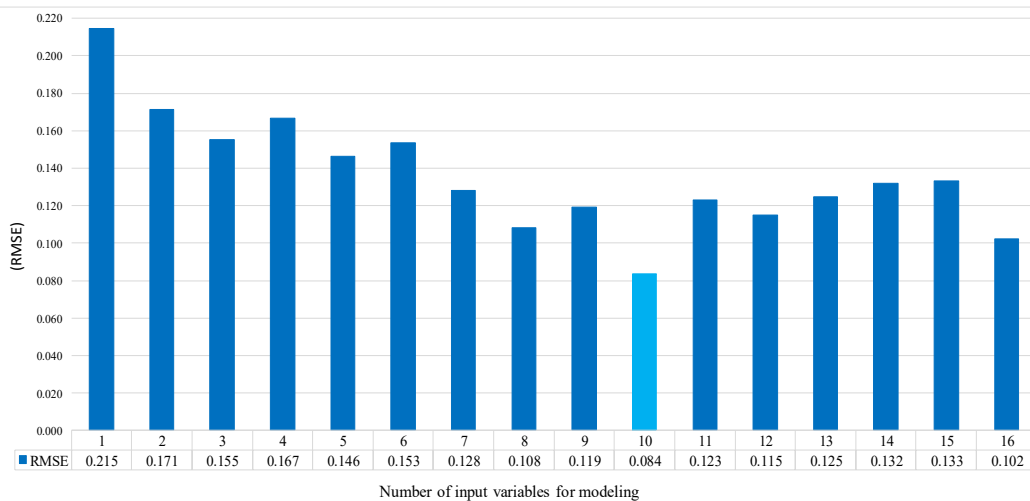


Fig. 9. The RMSE values of the best Hybrid ANN models associated with different number of input variables

The results indicate that the Zero-truncated Poisson model has resulted in almost the same RMSE values under both modelling and validation data sets. The RMSE value for the ANN model, using the validation data, is even less than the RMSE value obtained using modeling data set. Furthermore, as the RMSE values are calculated based on the normalized data, its complement value in percentage, somehow represents the accuracy of the model. Thus, it may be concluded that the results indicate an accuracy 88 to 89% for the Zero-truncated Poisson model and an

accuracy of the 92 to 94% for the best overall ANN model. Moreover, the R<sup>2</sup> value for the ANN model under the validation data set is even higher than the R<sup>2</sup> value obtained using the modeling data set and it is higher than its corresponding value for the Zero-truncated Poisson model. Hence, it could be concluded that the performance of both models are satisfactory in terms of accuracy and transferability. However the performance of the best overall ANN model is somewhat better than the Zero-truncated Poisson model.

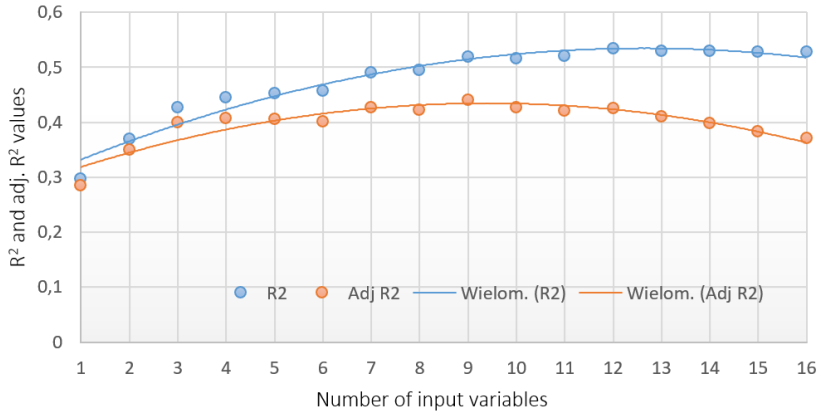


Fig. 10. The variation of R<sup>2</sup> and adjusted R<sup>2</sup> values for the best ANN models developed with different number of input variables

Table 8. Comparison of the RMSE and R<sup>2</sup> values obtained for the best models under modeling and validation data sets

Model Type	RMSE		R <sup>2</sup>	
	Modeling Data	Validation Data	Modeling Data	Validation Data
Zero-truncated Poisson Regression	0,114	0,116	0,951	0,888
Hybrid ANN model	0,084	0,063	0,919	0,985

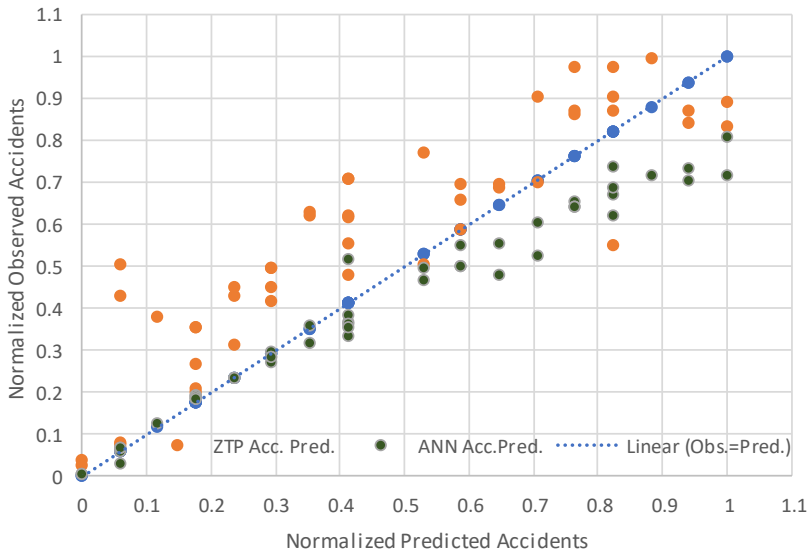


Fig. 11. Comparison of the performance of final Zero-truncated Poisson and Hybrid ANN models

It should be noted that in this research, the frequency of accidents was considered as a dependent variable or the output of the developed models, meaning that no distinguish was made between different types of

accidents, i.e. fatal, injured, or damaged accidents. Thus, future researches in which the severity of accidents is also taken into account would be very useful.



## Acknowledgements

The authors would like to express their gratitude towards the staff at Isfahan, Kerman and Yazd Highways Authorities and also the local Police Organizations for their advice and their assistance in gathering and collecting the information required for this study.

## References

1. ABDEL-ATY, M., PANDE, A., 2005. *Identifying crash propensity using specific traffic speed conditions*. Journal of Safety Research, 36(1): 97 – 108.
2. ABDELWAHAB, H.T., ABDEL-ATY, M., 2001. *Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections*. Transportation Research Record: Journal of the Transportation Research Board, 1746, 6–13. DOI: <http://dx.doi.org/10.3141/1746-02>.
3. ABOUNOAS, Z., RAPHAEL, W., BADR, Y., FADDOUL, R., GUILLAUME, A., 2020. *Crash data reporting systems in fourteen Arab countries: challenges and improvement*. Archives of Transport, 56(4), 73-88. DOI: <https://doi.org/10.5604/01.3001.0014.5628>.
4. AFANDIZADEH, S., GOLSHAN KHAVAS, R., 2006. *Design of a Safety Entrance Model for City Entrance Roads*. Journal of Transportation Research. 3(4), 233-241. (in Persian).
5. AKAIKE, H., 1974. *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19(6), 716–723, DOI: 10.1109/TAC.1974.1100705.
6. AKGUNGOR, A.P., DOGAN, E., 2008. *Estimating road accidents of Turkey based on regression analysis and artificial neural network approach*. Advances in Transportation Studies, 16, 11–22. DOI: <https://trid.trb.org/view.aspx?idD873702>.
7. ALARIFI, A.S., ABDEL-ATY, M., LEE, J., 2018. *A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors*. Accident Analysis and Prevention, 119:263–273.
8. AMBROS, J., ALTMANN, J., JUREWICZ, C., CHEVALIER, C., 2019. *Proactive assessment of road curve safety using floating car data: An exploratory study*. Archives of Transport, 50(2), 7-15. DOI: <https://doi.org/10.5604/01.3001.0013.5570>.
9. AMBROS, J., ELGNER, J., TUREK, R., VALENTOVÁ, V., 2020. *Where and when do drivers speed? A feasibility study of using probe vehicle data for speeding analysis*. Archives of Transport, 53(1), 103-113. DOI: <https://doi.org/10.5604/01.3001.0014.1747>.
10. BAYATA, H.F., HATTATOGLU, F., KARSLI, N., 2011. *Modeling of monthly traffic accidents with the artificial neural network method*. International Journal of the Physical Sciences, 6(2), 244–254.
11. BEN-BASSAT, T., SHINAR, D., 2011. *Effect of shoulder width, guardrail and road-way geometry on driver perception and behavior*. Accident Analysis and Prevention, 43(6): 2142–2152.
12. CALIENDO, C., GUIDA, M., PARISI, A., 2007. *A crash-prediction model for multilane roads*. Accident Analysis and Prevention, 39(4):657 – 670.
13. CAMERON, A.C., TRIVEDI, P., 1998. *Regression analysis of count data*. Cambridge: Cambridge University Press.
14. CAREY, E. M., ANDERSON, D. E., MANSOUR, R., SLOAN, J., CURRY, E. A., 2020. *Missed opportunities to advance knowledge on traffic safety: Accessibility of driver licensing and crash data for scientific research*. Accident Analysis and Prevention, 139: 105500.
15. CHANG, L. Y., 2005. *Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network*. Safety Science, 43(8), 541–557. DOI: <http://dx.doi.org/10.1016/j.ssci.2005.04.004>.
16. DADASHOVA, B., RAMÍREZ, B.A., MC WILLIAMS, J.M., IZQUIERDO, F.A., 2016. *The identification of patterns of interurban road accident frequency and severity using road geometry and traffic indicators*. Transportation Research Procedia, 14:4122 – 4129.
17. DAS, D., BURGER, E., 2015. *Evaluation of the role of the urban planning related factors on the occurrence of traffic accident in urban areas in South Africa - A case study in Bloemfontein city*. Proceedings of the 34th Southern African Transport Conference (SATC 2015), 553–567.

18. DAVOODI, S.R., AHMADI, P., 2015. *Prioritizing suggested strategies to reduce road accidents at the cities entrance using analytic hierarchy process*. International Journal of Asian Social Science, 5: 706–714.
19. DOUGHERTY, M., 1995. *A review of neural networks applied to transport*. Transportation Research Part C: Emerging Technologies, 3(4), 247–260. DOI: [http://dx.doi.org/10.1016/0968-090X\(95\)00009-8](http://dx.doi.org/10.1016/0968-090X(95)00009-8).
20. EDQUIST, J. RUDIN-BROWN, C., LENNÉ, M., 2012. *The effects of on-street parking and road environment visual complexity on travel speed and reaction time*. Accident Analysis and Prevention, 45: 759 – 765.
21. ELAMRANI ABOU ELASSAD, Z., MOUSANNIF, H., AL MOATASSIME, H., 2020. *A proactive decision support system for predicting traffic crash events: A critical analysis of imbalanced class distribution*. Knowledge-Based Systems, DOI: <https://doi.org/10.1016/j.knosys.2020.106314>.
22. ELVIK, R., 2017. *A synthesis of studies of access point density as a risk factor for road accidents*. Accident Analysis and Prevention, 107:1 – 10.
23. ELYASI, M., SAFFARZADEH, M., BOROJERDIAN, A. M., SEMNARSHAD, M., MAZAEHRI, M., 2016. *Prioritization of suburban accident factors based on analytical network process*. International Journal of Transportation Engineering, 5:197–209.
24. FITZPATRICK, K., CARLSON, P., BREWER, M., WOOLDRIDGE, M., 2001. *Design factors that affect driver speed on suburban streets*. Transportation Research Record: Journal of the Transportation Research Board 1751, p. 18–25.
25. GACA, S., POGODZIŃSKA, S., 2017. *Speed management as a measure to improve road safety on Polish regional roads*. Archives of Transport, 43(3).
26. GOMES, S.V., GEEDIPALLY, S.R., LORD, D., 2012. *Estimating the safety performance of urban intersections in Lisbon, Portugal*. Safety Science, 50(9):1732 – 1739.
27. HEATON, J.T., 2010). *Introduction to neural networks with Java* (2nd Ed.). Washington: Heaton Research.
28. HILBE, J., 2011. *Negative Binomial Regression*. Cambridge University Press, New York. United States of America, 2nd ed.
29. HOLDRIDGE, J., SHANKAR, V., ULFARSSON, G., 2005. *The crash severity impacts of fixed roadside objects*. Journal of Safety Research, 36(2):139 – 147.
30. HOU, Q., MENG, X., LENG, J., YU, L., 2018. *Application of a random effects negative binomial model to examine crash frequency for freeways in China*. Physica A 509: 937–944.
31. HUANG, B., ZHANG, Y., LU, L., LU, J.J., 2014. *A new access density definition and its correlation with crash rates by microscopic traffic simulation method*. Accident Analysis and Prevention, 64:111 – 122.
32. JIAWEI, H., MICHELINE, K., 2006. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publisher.
33. KHALILI, M., PAKGOHAR, A., 2013. *Logistic regression approach in road defects impact on accident severity*. Journal of emerging technologies in web intelligence, 5(2).
34. KHISHDARI, A., FALLAH TAFTI, M., 2017. *Development of crash frequency models for safety promotion of urban collector streets*, Journal of Injury Control and Safety Promotion, 24(4), 519-533, DOI: 10.1080/17457300.2016.1278237.
35. LAFERRIÈRE, R., 1999. *The air network itinerary and trip aggregation (ANITA) model*. In M. Gaudry R. Mayes (Eds.), Taking stock of air liberalization, 223–228. Boston, MA: Springer US.
36. LEE, C., ABDEL-ATY, M., PARK, J., WANG, J.H., 2015. *Development of crash modification factors for changing lane width on roadway segments using generalized nonlinear models*. Accident Analysis and Prevention, 76: 83 – 91.
37. LI, X., LORD, D., ZHANG, Y., XIE, Y., 2008. *Predicting motor vehicle crashes using Support Vector Machine models*. Accident Analysis and Prevention, 40(4), 1611–1618. DOI: <http://dx.doi.org/10.1016/j.aap.2008.04.010>.
38. LIU, Y., ZOU, B., NI, A., GAO, L., ZHANG, C., 2020. *Calibrating microscopic traffic simulators using machine learning and particle swarm optimization*. Transportation Letters, DOI: 10.1080/19427867.2020.1728037.

39. MA, Q., YANG, H., WANG, Z., XIE, K., YANG, D., 2020. *Modeling crash risk of horizontal curves using large-scale auto-extracted roadway geometry data*. *Accident Analysis and Prevention*, 144: 105669.
40. MECHERI, S., ROSEY, F., LOBJOIS, R., 2017. *The effects of lane width, shoulder width, and road cross-sectional reallocation on drivers' behavioral adaptations*. *Accident Analysis and Prevention*, 104: 65 – 73.
41. MOOMEN, M., REZAPOUR, M., RAJA, M., KSAIBATI, K., 2020. *Predicting downgrade crash frequency with the random-parameters negative binomial model: Insights into the impacts of geometric variables on downgrade crashes in Wyoming*. *IATSS Research* 44: 94-102.
42. MUSTAKIM, F., YUSOF, I., RAHMAN, I., SAMAD, A.A.A., SALLEH, N.E.B.M., 2008. *Blackspot study and accident prediction model using multiple liner regression*. *First International Conference on Construction in Developing Countries (ICCICD-I)*, (1): 121-136.
43. NARAYANAMOORTHY, S., PALETI, R., BHAT, C., 2013. *On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level*. *Transportation Research Part B: Methodological*, 55: 245 – 264.
44. OLKIN, I., GLESER, L.J., DERMAN, C., 1980. *Probability models and applications*. New York, NY: MacMillan.
45. POTTS, I., HARWOOD, D., RICHARD, K., 2007. *Relationship of lane width to safety on urban and suburban arterials*. *Transportation Research Record: Journal of the Transportation Research Board*, 2023: 63-82.
46. PU, Z., LI, Z., KE, R., HUA, X., WANG, Y., 2020. *Evaluating the Nonlinear Correlation between Vertical Curve Features and Crash Frequency on Highways Using Random Forests*. *Journal of Transportation Engineering, Part A: Systems*, 146(10): 04020115.
47. PULUGURTHA, S. SAMBHARA, V. (2011). *Pedestrian crash estimation models for signalized intersections*. *Accident Analysis and Prevention*, 43(1):439 – 446.
48. RŮŽIČKA, J., BĚLINOVÁ, Z., KOREC, V., CIKHARDTOVÁ, K., TICHÝ, T., 2019. *Prediction of Traffic Accidents by Using Neural Network*. *Archives of Transport System Telematics*, 12(4).
49. SAEED, U.T., HALL, T., BAROUD, H., VOLOVSKI, J.M., 2019. *Analyzing road crash frequencies with uncorrelated and correlated random-parameters count models: An empirical assessment of multilane highways*. *Analytic Methods in Accident Research* 23: 100101.
50. SAFFARI, M., SEYED ABRISHAMI, E., ADNAN, R., GREENE, W., 2011. *Handling of over-dispersion of count data via truncation using Poisson regression model*. *Journal of Computer Science Computational Mathematics*, 1(1), 1-4.
51. SAHA, D., ALLURI, P., GAN, A., 2015. *Prioritizing Highway Safety Manual's crash prediction variables using boosted regression trees*. *Accident Analysis and Prevention*, 79: 133 – 144.
52. SPIEGELHALTER, D.J., BEST, N.G., CARLIN, B.P., VAN DER LINDE, A., 2002. *Bayesian measures of model complexity and fit*. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639, DOI: 10.1111/1467-9868.00353.
53. SUKIRTY J., SANYAM, S., RAJESH, W., 2018. *Dynamic selection of Normalization Techniques using Data Complexity Measures*. *Expert Systems with Applications*.
54. TERCAN, E. , BEŞDOK, E., TAPKIN , S., 2020. *Heuristic Modelling of traffic accident characteristics*”, *Transportation Letters*, DOI: 10.1080/19427867.2020.1734273.
55. THAKUR, S., BISWAS, S., 2019. *Assessment of pedestrian-vehicle interaction on urban roads: a critical review*. *Archives of Transport*, 51(3), 49-63. DOI: <https://doi.org/10.5604/01.3001.0013.6162>.
56. UKKUSURI, S., MIRANDA-MORENO, L., RAMADURAI, G., ISA-TAVAREZ, J., 2012. *The role of built environment on pedestrian crash frequency*. *Safety Science*, 50(4): 1141 – 1151.
57. VAYALAMKUZHI, P., AMIRTHALINGAM, V., 2016. *Influence of geometric design characteristics on safety under heterogeneous traffic flow*. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(6): 559 – 570.

58. WANG, J., KONG, Y., FU, T., 2019. *Expressway crash risk prediction using back propagation neural network: A brief investigation on safety resilience*. Accident Analysis and Prevention 124: 180–192.
59. WANG, X., SONG, Y., YU, R., SCHULTZ, G., 2014. *Safety modeling of sub urban arterials in shanghai, China*. Accident Analysis and Prevention, 70: 215 – 224.
60. WANG, X., YUAN, J., SCHULTZ, G., FANG, S., 2018. *Investigating the safety impact of roadway network features of suburban arterials in Shanghai*. Accident Analysis and Prevention, 113: 137 – 148.
61. WHO, 2018. WORLD HEALTH ORGANIZATION, 2018. *Global status report on road safety*. Geneva, Licence: CC BYNC-SA 3.0 IGO.
62. WU, P., MENG, X., SONG, L., 2019. *A novel ensemble learning method for crash prediction using road geometric alignments and traffic data*. Journal Of Transportation Safety Security. DOI: 10.1080/19439962.2019.1579288.
63. XIE, Y., LORD, D., ZHANG, Y., 2007. *Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis*. Accident Analysis Prevention,. 39(5), 922–933. DOI: <http://dx.doi.org/10.1016/j.aap.2006.12.014>.
64. YANG, B., LIU, P., CHAN, C., XU, C., GUO, Y., 2019. *Identifying the crash characteristics on freeway segments based on different ramp influence areas*. Traffic Injury Prevention, DOI: 10.1080/15389588.2019.1588965.
65. YUAN, J., ABDEL-ATY, M., GONG, Y., CAI, Q., 2019. *Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network*. Journal of the Transportation Research Board. 1-13. DOI: 10.1177/0361198119840611.
66. ZOU, Y., TARKO, A., CHEN, E., ROMERO, M., 2014. *Effectiveness of cable barriers, guardrails, and concrete barrier walls in reducing the risk of injury*. Accident Analysis and Prevention, 72: 55 – 65.