



Research Article

Traffic injury severity prediction along with identification of contributory factors using learning vector quantization: a case study of the city of London

Meisam Siamidoudaran^{1,2} · Ersun Iscioglu³ · Mehdi Siamidodaran⁴

© Springer Nature Switzerland AG 2019

Abstract

This study presents learning vector quantization neural network modelling to predict injury severity of driver as well as riders, which applies to the backbone of traffic networks for London's central business district. The potential associations between injury severity classes and crash related factors that contribute to their generation are discovered. Accordingly, the model is addressed as an identification technique for contributory factors and range of interventions for road safety. Unsurprisingly, approaching a T/staggered junction is detected as an accident hotspot. Injuries caused by going ahead on a bend and turning manoeuvres are ranked as the next most important contributory factors. Likewise, the affect of most junction actions were almost triple compared to the other indexes. All other sensitive predictors approximately were held near as equal; injuries involving a stationary or parked vehicle, factors related to junction control, crossing facilities, alcohol involvement, rush hours, and vehicle type. Following this implication, with the purpose of maximising the likelihood of injury accuracy, the model is predicted through the most sensitive predictors.

Keywords Road safety · Crash prevention · Traffic accident prevention · Injury severity prediction · Contributory factors

1 Introduction

While motor vehicles deliver many profits, they also seriously harm people. In view of that, road traffic injuries are a health crisis that affects roughly 30–50,000,000 individuals which are seriously injured and about 1.2 million which are killed every year. Moreover, this complex health system is a leading cause of death among young people, aged 15–29 years [1]. Consequently, safety must be a priority for roads but unfortunately, traffic accidents are still frequent and fatalities and injuries are a common global concern along with developed countries such as UK.

UK now has one of the greatest road safety records in the world. However, there are still a number of road users

who die and a lot of people who are injured daily on Britain's roads, so there is much more to be done in relation to accident prevention and reduction performances [2]. As a result, due to the tragic consequences of road accidents' injuries and fatalities, the government's vision is a multi-million pound investment in terms of road safety. Accordingly, to quickly eliminate any concerns and to make sure that the country continues to remain as a strong global leader on road safety [3]. Part of this investment is the road danger reduction and active travel plan in the City of London. This long term strategy plans to set out the significant goals and objectives to sustain a safe environment for all road users. It targets to work towards eliminating the

✉ Meisam Siamidoudaran, meisam.siamidoudaran@emu.edu.tr; Ersun Iscioglu, ersun.iscioglu@emu.edu.tr; Mehdi Siamidodaran, siamimehdi@iauardabil.ac.ir | ¹Department of Civil Engineering, Eastern Mediterranean University, Famagusta, North Cyprus via Mersin 10, Turkey. ²Faculty of Economics, Management and Information Technology, Voronezh State Technical University, Ulitsa 20-Letiya Oktyabrya, 84, Voronezh, Russia 394006. ³Department of Computer Education and Instructional Technology, Eastern Mediterranean University, Famagusta, North Cyprus via Mersin 10, Turkey. ⁴Department of Industrial Engineering, Islamic Azad University, Basij Square, Ardabil, Iran.



SN Applied Sciences (2019) 1:1268 | <https://doi.org/10.1007/s42452-019-1314-6>

Received: 4 July 2019 / Accepted: 18 September 2019 / Published online: 24 September 2019

annual number of individuals killed and seriously injured in traffic collisions to zero before 2041. [4]

Some road users believe that the crashes are just one of those things that occur and they chalk them up to misfortune, or destiny. But road collisions don't have to truly happen and there are solutions to eliminate or mitigate the road safety problems, before they occur. With the goal of offering the road users the lowest probability of becoming a road casualty, it is vital to find factors before they contribute to the injury. These risk factors are related to a variety of reasons such as; roadway characteristics, vehicle features, environmental factors, human behaviour etc. In light of this, injury prediction model is an appropriate mechanism to figure out the complex relationship between severity of injury and accident related variables, thus it is able to deliver sophisticated analysis including the identification of the contributory factors. The factors alone can be a major issue or when combined with each other can contribute to personal injury severities but generally most of the accidents commonly occur because of multiple sources. Nevertheless, extra factors that are not available for road safety investigators are still remaining while they can contribute to the injuries [5]. Hence, this paper attempts to predict the injury severity by using big numbers of subdivision data [6]. In addition, due to the powerful relationship between injury risk and crash related factors, utilizing extra input factors leads to achieving superior prediction accuracy [7].

Being one of the major steps of accident management, injury severity prediction can forecast classes of the severity that may be estimated to happen in result of an accident. The injury severities are typically considered by several separate levels such as; fatality, serious injury, slight injury, and property damage only. Accordingly, the prediction model provides crucial information for emergency responders to evaluate the severity level of accidents, estimate the potential impacts, and implement efficient accident management procedures [8, 9]. The reliability and results of traffic injury prediction models that include the interaction of input and output variables and reviewing parameters involved in traffic accidents have an important meaning for the improvement of road safety management and can definitely help lower the number of traffic accidents [10]. Numerous applications have been developed to evaluate safety level of various types of road entities and to examine effects of safety countermeasures [11]. The outcomes of the injury prediction model can play a significant role in preventing or reducing casualties as well as solving many road safety problems.

In accordance with this, artificial neural network (ANN) models accommodate multiple input parameters to forecast several output classes. Accordingly, they display a superior performance in relation to prediction accuracy

[7–12]. This research makes an attempt to apply learning vector quantization neural network (LVQNN) model for prediction tasks which is not commonly used with ANN in previous injury related studies. In this context, the previous studies verified that LVQNN is feasible for predicting and determining the traffic parameters [13, 14]. Consequently, this model can be used to discover the relationship between accidents' factors in very complex circumstances, to predict the injury severity classes and to gain new insights for the field of road safety.

2 Literature review

Over the past decades, a large number of the injury prediction models have been proposed. Statistical models are the most traditional one and the commonly used models in these systems are perhaps the ordered probit (OP) model, multinomial logit (MNL) model, and the binary logit (BL) model. Statistical models are able to explicitly illustrate the effects of observed explanatory variables on the severity of crash occurrence and account for some characteristics in crash data. However, complex situation of traffic collisions in mass of noisy nonlinear dataset makes it very difficult to recognize the contributory factors while applying the statistical models efficiently. Furthermore, poor performance in employing several separable factors accompanied by the factors with a large number of categorical data is one more weakness for these methods [12, 15–18].

ANN prediction models have also been developed by many researches to overcome the disadvantage of statistical models. As a result, ANN methods are verified to be more valuable models in prediction of injury severity outcomes and attain a superior prediction accuracy in comparison to the statistical methods [5, 7, 17–21]. For example, Xie et al. [5] carried out a comparison between ANN and statistical models and the results of the predictions showed that back-propagation neural network (BPNN) and Bayesian neural network (BNN) models reached greater prediction accuracy compared to traditional negative binomial (NB) regression methods.

Among the ANN models, the multilayer perceptron neural network (MLPNN) is perhaps the most commonly used modelling technique in prediction of accident severities. For example, Abdelwahab and Abdel-Aty [22] used this model and fuzzy adaptive resonance theory (ART) with ordered logit models for accident prediction tasks. Their results showed that MLPNN predicted the injury severity outcomes better than other models. In another study, two researchers again compared the previous study outcomes and verified that the MLPNN had a much greater performance than fuzzy ART [23]. Delen et al. [24]

tried to discover the sensitive predictors using a series of binary MLPNN models. But, exploiting more injury classes along with the results of the forecast did not discover any improved outcomes than other previous studies [22, 23]. Another accident prediction related research was carried out using MLPNN, function fitting, and generalized regression neural networks. Accordingly, the performance of the ANN prediction models was measured using Mean Square Error (MSE) and multiple correlation coefficient (R). The comparison between the models verified that MLPNN model achieved superior performance on predicting than other prediction models, as a result of lower MSE and higher R [25]. Moreover, in MLPNN related predictions, Aghayan et al. compared the capability of fuzzy subtractive clustering, fuzzy C-means clustering and MLPNN models to predict injury severity classes along with response time. MLPNN was a good fit for traffic collision data due to achieving maximum R-value. Likewise, MLPNN had a greater accuracy in predicting the collisions between other techniques [26].

Support vector machine (SVM) model is also another frequently used form of the prediction which was reviewed in this literature review [18–20, 27]. For instance, Li et al. [19] SVM method in a crash injury related study and the findings of the prediction were compared with an OP model. The outcome of the comparison showed that, the ANN model achieved a better prediction accuracy. Another comparison has been done [18] using SVM model in the injury severity classes and compared the results of the prediction with NB regression. Accordingly, the comparison outcome demonstrated that SVM model produced a greater performance in terms of prediction accuracy.

In the last few years, Yu and Abdel-Aty [27] used SVM model, random parameter logit model, and fixed parameter logit model for predictions of crash injury severities. The comparison results showed that both the SVM and the random parameter methods demonstrated better prediction accuracy. In most recent comparisons related to the injury severities, Chen et al. [28] applied SVM model to predict injury severity levels. The researcher found out that the SVM models proved more prediction accuracy resulting from the comparison. In addition, the result displayed that polynomial kernel had a higher performance compared to the Gaussian RBF kernel.

Iranitalaba and Khattakb [29] used some statistical models and compared the predictions' outcomes with machine learning models. The findings showed that the nearest neighbour classification had the higher overall forecast performance. SVM and Random forests (RFs) methods had the next two acceptable performances and the poor performance referred to MNL model. Along this line, another ANN related study was done by Aghayan et al. [30] using SVM with different kernel functions for

injury severity prediction. Comparison of the overall prediction accuracy between the models showed that the SVM model was superior to the other models including MLP, genetic algorithm, combined genetic algorithm and pattern search. Moreover, they displayed that the constructed performance resulting from MLP was slightly superior to the SVM.

In numerous researches, non-parametric and artificial intelligence models have also been applied to overcome the weakness of statistical approaches. For instance, Chang and Wang, developed classification and regression trees (CART) to examine the association between different injury severity outcomes and contributory factors [31].

Yasin Çodur and Tortum [32] developed artificial intelligence methods including ANNs and genetic algorithms to examine the association between accident injury severity and several crash related factors. In this prediction, the sigmoid activation function was used with Levenberg–Marquardt algorithm. As a final point, the performance of the prediction models is measured by root mean square error (RMSE), MSE, and R [32].

Multiple logistic regression, Bayesian logistic and classification tree models were applied to analyse numerous contributing factors in fatal collisions. The outcomes gained from the models showed that controlling driver errors reduced the likelihood of motorist's fatality in traffic accidents [33].

A latest prediction was carried out by Wang and Kim [34] developing MNL and RF models. Using the potential of prediction models, they were able to show that only a few factors had a significant affect on the outcome of the accident severities. The comparison between models verified that, RF achieved a greater prediction accuracy than MNL. However, the outcomes of sensitivity analysis displayed that RF is less sensitive than other models.

In summary, the reviewed literature showed that ANN models had a greater performance in terms of prediction accuracy. Therefore, a different type of ANN, apart from the commonly reviewed ANNs is considered in this paper on the modelling of injury severity prediction. Within this framework, the learning vector quantization neural network (LVQNN) model is used by applying personal accident injury data for the city of London.

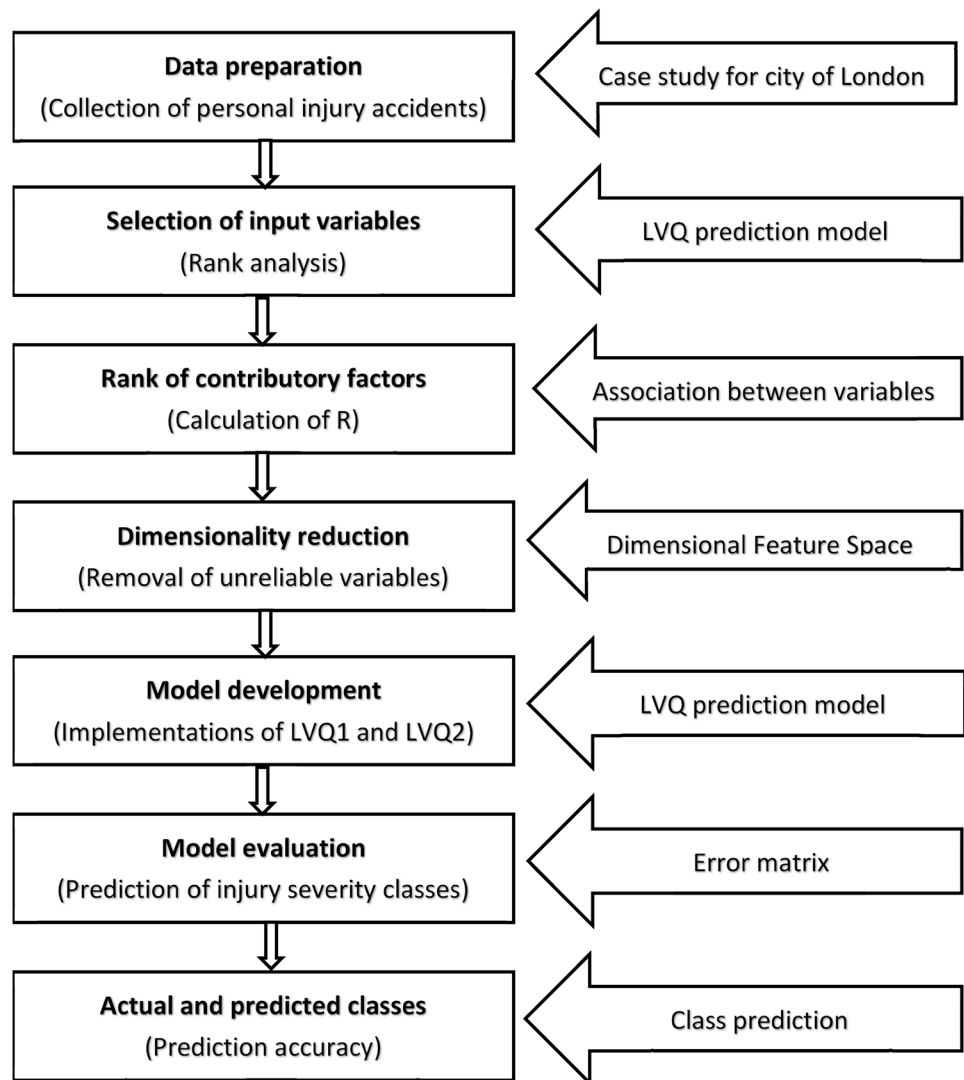
LVQNN has displayed a good pattern recognition performance in many more complex prediction tasks. Accordingly, the model is able to select from wide range of algorithms designed for improved classification effectiveness [35, 36]. LVQNN model is a successful pattern for classifying data with categorical values [37] and has reached the greatest overall accuracy in comparison to other AANs [38, 39]. In this paper, LVQNN model predicts the injury severity of driver/rider into either of the following four categories: fatality, serious injury, slight

injury, and only damage to property. Furthermore, the reviewed literature indicated that specific consideration has been emphasized on the prediction of the injury severities, but, injury related outcome of the predictions weren't the major focus in order to determine the contributory factors. Therefore, in response to this limitation, along with the outcomes of the prediction and the potential of injury severity analysis, we attempted to predict along with apprehensions of the associations between the injury severity classes and the influencing factors that contribute to their generation. Thus, the greatest sensitive predictors are ranked and measured as the contributory factors. Accordingly, the range of interventions for road safety and traffic collision clusters are recognised. Additionally, the second phase of the prediction has been made using the three sensitive predictors with the aim of maximising the model performance.

3 Materials and models

The data used in this study delivers detailed road safety statistics about the statuses of personal injury collisions on public roads of Great Britain. The LVQNN model treats an accurate modelling technique for prediction of injury severity outputs in this study. LVQNN is a powerful tool to solve various prediction problems as classification tasks [35, 36]. This model is adopted to set up a prediction model in many previous studies. As a result, numerous comparisons between LVQNN and other traditional models have proved that LVQNN approach has a greater prediction accuracy. In alignment with this, LVQNN method is used in this research to show the possibility of an effective application prospect in the field of traffic injury severity. Accordingly, schematic outline of this assignment is shown in Fig. 1.

Fig. 1 The flowchart of LVQNN prediction



3.1 Data description

There is constant association among the injury severity and the contributory factor. Likewise, using smaller number of input data refers to obtaining poor model performance in terms of prediction accuracy [7]. Therefore, we attempted to apply a large number of categorical data into the model with the aim of minimising prediction error.

STATS19 road safety data used in this prediction involved traffic accidents that were reported to the police within 30 days of the incident. The data provides details of the 3500 personal injury circumstances of diver/rider in the city of London that happened during 2014–2018. The consequential casualty of driver/rider, crash circumstances and types of vehicles involved are covered, which detail all the explanatory factors shown in Table 1. Furthermore, we refer to DFT [6] for full descriptive statistics and more detail information of the factors used in this research.

3.2 Models

3.2.1 Application of the LVQNN

LVQNN is one of the most powerful methods for classification tasks [40] and has achieved best overall accuracy in comparison to other AANs [38, 39]. Previous related studies show that this model is also a suitable tool for road traffic data analysis [13, 14] as well as it successfully being used for classifying data with categorical values [37]. Thus due to using a large number of subdivisions for variables in this study, LVQNN is considered as a modelling technique for the prediction of injury severity along with identification of significant predictors in the traffic collisions.

This algorithm was devised by Kohonen [35]. This model of neural network is a precursor to self-organizing maps (SOM) that can be used when there is labelled input data. As the value of the date used in this study is label, this learning technique is more appropriate for predicting the injury severity. The model utilizes the level data to relocate the Voronoi vectors slightly, so as to improve the quality of the classifier decision areas. It is a two phase procedure which consist of a SOM trailed by LVQNN as show in Fig. 2.

The model is an improved method of prediction and specifically suitable for clustering problems. The first stage is a selection of features that the unsupervised recognition of a reasonably minor set of specifications in which the important statistics content of the input data is focused. The second stage is the classification where the feature scopes are referred to individual levels. By using an encoder pattern for a big number of input vectors $x \in IR^n$, and transforming the input into an i -value which determine less significant factors and achieve a superior estimation to the unique input space.

Given the input vector x and suppose $x \in IR^n$, the model transforms the label input parameters into an i -value with an encoder form which $i \in \{1, 2, 3, \dots, k\}$. Perhaps the most efficient means to consider the LVQNN is concerning about common encoders and decoders. Figure 3 simply shows that the architecture involves two components as an encoder and a decoder.

Normally, x is elected at random in relation to some likelihood function $p(x)$. At that point the optimum encoding–decoding pattern is established by modifying the functions x and m_c to mitigate the expected distortion explained by Eq. 1.

$$E = \varepsilon \left\{ \|x - m_c\|^2 \right\} = \int \|x - m_c\|^2 p(x) d(x) \tag{1}$$

In the above equation, ε is the expected value (EV) and m_c is defined as centre of the winner. Once a decoder procedure is applied to i , the vector $m \in IR^n$, is gained and m remains an approximation of x , in the error of the vector quantization approximation equation.

The EV and the winning neuron are attained from the following equation in which C is the winner and obtained from Eq. 2.

$$C = \arg \min_i \|x - m_i\|^2 \tag{2}$$

To identify the limit of each level, it is essential to display the midline of the line segment designed for m_1, m_2 . In fact, the midline specifies a route that the space of all points on that route, is equal from the centres of m_1 and m_2 ($d_1 = d_2$). In terms of three-dimensional space, the midline performs as a midplane, and generally it is presented as a hyperplane. The algorithm initiates through a trained SOM with input vector and uses weight/Voronoi diagram if the requirement for a range of more centres is identified.

The classification labels of the inputs are used to discover the greatest classification label for each Voronoi neuron. As the Voronoi neuron boundaries do not match the classification boundaries, the model is attempts to fix this issue through shifting the boundaries.

If $x(t)$ does not exist on the boundary ($d_1 \neq d_2$), the associated centre encourages the classified integer level $[m_i(t)]$ to becomes nearer to $x(t)$ and informs as shown in the following equations.

$$m_i(t + 1) = m_i(t) + \Delta m_i(t), \quad i = 1, 2, 3, \dots, k \tag{3}$$

$$\Delta m_i(t) = \delta_{ci} \cdot \alpha(t) \cdot [x(t) - m_i(t)] \tag{4}$$

$$\delta_{ci} = \begin{cases} 1 & c = i \\ 0 & c \neq i \end{cases}, \quad 0 < \alpha(t) < 1$$

Table 1 Descriptive statistics of input parameters

Variable type	Label	Total (%)
Junction detail	T or staggered junction	55.91
	Crossroads	19.09
	Not at junction or within 20 m	16.33
	More than 4 arms (not roundabout)	04.11
	Roundabout	02.54
	Other junction	02.02
Junction location	Approaching junction or waiting in a queue	50.42
	Mid Junction	30.17
	Not at or within 20 m of junction	15.89
	Leaving/entering main road	03.18
	Leaving/entering roundabout	00.19
	Entering from slip road	00.15
Vehicle manoeuvre	Going ahead bend	48.11
	Turning/waiting to turn right	20.02
	Parked	09.21
	Overtaking	09.15
	Waiting to go/held up	03.23
	Moving off	03.01
	Slowing or stopping	02.57
	Changing lane	02.46
	Reversing	02.24
	Junction control	Auto traffic signal
Give way or uncontrolled		37.69
Not at junction or within 20 m		14.19
Crossing facilities ^a	Pedestrian phase at traffic signal junction	47.15
	No crossing facilities within 50 m	45.89
	Pedestrian light controlled crossing	04.01
	Zebra	02.03
	Central refuge	00.51
Alcohol involvement	Footbridge or subway	00.41
	Pedestrian	68.63
	Pedal rider	12.15
	Rider	10.03
Time	Driver	09.19
	16:00–19:59	28.12
	08:00–11:59	27.02
	12:00–15:59	18.86
	04:00–07:59	10.12
	20:00–11:59	08.86
Vehicle type	00:00–03:59	07.02
	Car	35.77
	Pedal cycle	27.45
	Public service vehicles	13.34
	Light and heavy goods	12.89
	Motorcycle	10.19
	Other vehicle	00.36
Light condition	Daylight	72.98
	Darkness—lights lit	26.32
	Darkness—lights unlit	00.39
	Darkness—lighting unknown	00.31
	Darkness—no lighting	00.00

Table 1 (continued)

Variable type	Label	Total (%)
1st point of impact	Front	38.91
	Nearside	24.01
	Offside	20.34
	Back	10.88
	Did not impact	05.86
2nd road class ^b	A	25.56
	B	02.56
	C	54.67
	Unclassified	17.21
Road surface condition	Dry	86.43
	Wet or damp	12.56
	Flood over 3 cm. deep	00.51
	Frost or ice	00.39
	Snow	00.11
	Mud	00.00
	Injury severity classes	Y4 Property damage only
	Y3 Slight injury	40.03
	Y2 Serious injury	06.02
	Y1 Fatal injury	00.16

^aThis variable is only applied when the crash is at junction, if not the value is blank

^bThis variable is only used for controlling pedestrians and pedal rider crossing [6]

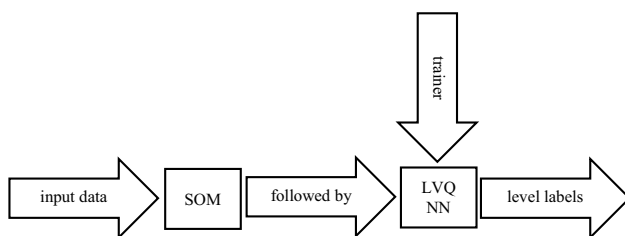


Fig. 2 Two phase process consist of a SOM followed by LVQNN

where $\alpha(t)$ is a learning rate that falls by the number of epochs or iteration of training process, and in each progress, this coefficient is dropped among 0 and 1.

LVQNN1 is an improved form of LVQNN and is updated similarly in the adjacent centre. However, if input $x(t)$ and associated Voronoi or weight such as, winning output node is correctly classified and has the similar label of class, encouragement the of $m_i(t)$ to $x(t)$ moves closer together as in the SOM network. If they have the different level labels, at that point, it is penalized, and $m_i(t)$ moves apart from $x(t)$. Voronoi vectors or weights corresponding to other input

areas are left unmoved with $\Delta m_i(t) = 0$. Consequently, the following equations are attained.

$$\Delta m_i(t) = \delta_{ci} \cdot f_i(t) \cdot \alpha(t) \cdot [x(t) - m_i(t)], \quad i = 1, 2, 3, \dots, k \tag{5}$$

$$m_i(t + 1) = m_i(t) + \delta_{ci} \cdot f_i(t) \cdot \alpha(t) \cdot [x(t) - m_i(t)] \tag{6}$$

$$f_i(t) = \begin{cases} +1 & \text{if } m_i(t), x(t) \text{ have the same class label} \\ -1 & \text{if } m_i(t), x(t) \text{ have the different class label} \end{cases}$$

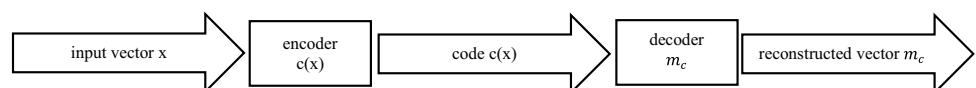
As a result, the following equations are considered.

$$m_i(t + 1) = m_i(t) + s_i(t) \cdot \alpha(t) \cdot [x(t) - m_i(t)] \tag{7}$$

$$\text{which } s_i(t) = \begin{cases} +1 & \text{correct Classification} \\ -1 & \text{incorrect Classification} \\ 0 & \text{if not} \end{cases}$$

In the case of optimised LVQNN1, $\alpha(t)$ in place of being similar for all centres, it performs as an individual learning rate for each centre. Thereby, the superior classification is achieved through the SOM alone. This will be reached if the ranking of the input data does not oppose in relation to the

Fig. 3 Encoder–decoder architecture in LVQNN



timeframe in a manner that the effect of the initial data has no significant difference with the last input, and also, all the data have the equal class labels. Therefore, in such circumstances, the following is acquired:

$$\text{Weight of } x(t) \rightarrow \alpha_i(t) \tag{8}$$

$$\text{Weight of } x(t - 1) \rightarrow [1 - s_i(t) \cdot \alpha_i(t)] \cdot \alpha_i(t - 1) \tag{9}$$

As a result of equating the above relationships, the following is obtained:

$$\alpha_i(t) = [1 - s_i(t) \cdot \alpha_i(t)] \cdot \alpha_i(t - 1) \tag{10}$$

$$\alpha_i(t) = \frac{\alpha_i(t - 1)}{[1 + s_i(t) \cdot \alpha_i(t - 1)]}, \quad 0 < \alpha_i(t) < 1 \tag{11}$$

$$\begin{cases} 0 < \alpha_i(t) < 1 \\ \alpha_i(0) = 0.3 \sim 0.5 \end{cases}$$

LVQNN2 is the second developed type of the LVQNN that is preferred in this study and it moves closer in influence to Bayesian decision theory. Likewise, LVQNN2 method which is opposite of LVQNN1 is updated at the parallel nearer to the centre. The method uses the correct and incorrect classification update equations. Therefore, in this case, the winners are two members as $m_i(t), m_j(t)$. The input vector x gives the correct classification through the associated Voronoi vector ($m_i(t)$) and the other nearest centre is incorrectly classified ($m_j(t)$). Additionally, the input vector x is well near to the decision boundary and $x(t)$ is in a specified range (W). Consequently, the following equations are obtained as below.

$$d_i = ||x(t) - m_i(t)|| \tag{12}$$

$$d_j = ||x(t) - m_j(t)|| \tag{13}$$

If d is the space between x and m , the followings are considered.

$$\min\left(\frac{d_i}{d_j}, \frac{d_j}{d_i}\right) > s \tag{14}$$

$$s = \frac{1 - w}{1 + w} \tag{15}$$

where w is the boundary width and normally is $0.2 \ll w \leq 0.3$ and that results to

$$\frac{7}{13} \leq s \leq \frac{2}{13}$$

$$0.5 < \min\left(\frac{d_i}{d_j}, \frac{d_j}{d_i}\right) \leq 1$$

where $d_i < d_j$ then we will have the following equations:

$$\frac{d_i}{d_j} = \frac{\frac{d_i+d_i}{2} - \frac{d_j-d_i}{2}}{\frac{d_i+d_i}{2} + \frac{d_j-d_i}{2}} = \frac{1 - \frac{d_j-d_i}{d_i+d_i}}{1 + \frac{d_j-d_i}{d_i+d_i}} \tag{16}$$

$$w = \frac{d_j - d_i}{d_j + d_i} \tag{17}$$

$$\frac{d_i}{d_j} = \frac{1 - w}{1 + w} \tag{18}$$

where m_i is considered as correct classification and m_j as an incorrect classification, we will have the equations as below.

$$m_i(t + 1) = m_i(t) + \alpha(t) \cdot [x(t) - m_i(t)] \tag{19}$$

$$m_j(t + 1) = m_j(t) + \alpha(t) \cdot [x(t) - m_j(t)] \tag{20}$$

LVQNN acts as a differential mode and moves one centre nearer together while moving another node apart. Alternatively, the preliminary selection of nodes for the LVQNN2 is more complicated, and to work out this weakness, initially the runs were made using the LVQNN1 and then retrieved by the LVQNN2.

LVQNN3 is another variation on this theme which is finally used for building more superior classification systems. Following this, where m_i, m_j and x have the same class label, the equation is obtained as below [36].

$$m_k(t + 1) = m_k(t) + \epsilon \alpha(t) \cdot [x(t) - m_k(t)], \quad k = i, j \tag{21}$$

ϵ is dependent on w and $0.1 \leq \epsilon \leq 0.5$

3.3 Sensitive predictors resulting by LVQNN

Typically, the dataset is divided into three parts. Accordingly, the LVQNN model is fit on training, validation and testing subsets. The 70% of the entire dataset is randomly divided for training, 15% of the data for validation and the remaining is separated for testing data. The aim of this process was to discover the model achieving the best performance on the data. Likewise, the model evaluated the error function by using testing data which is independent of that used for the training stage. The model is trained by minimization of an applicable error function specified analogous training data set. Model performance is then compared by assessing the error function via an independent validation set, and the model obtaining the slightest error according to the validation set is chosen. As a result,

the performance of the model is confirmed by evaluating its performance on the test set.

In this prediction, the Correlation coefficient (R) measured the strength and direction of the association between actual and predicted classes. R is a numerical measure of some type of association which is a powerful measurement for relationship between crash related factors [25, 26, 32].

Consequently, all the R values are used for assessment of comparison between training, validation, and a testing dataset. Table 2 shows the interpretation for the results of each correlation. As a result, sensitivity of the model is examined against the absence of each sub-variable on the output, and seventeen labels are discovered as the most sensitive predictors.

With respect to this, the highest R value refers to a strong relationship between the crash related factors and the injury severity outcomes as well as the superiority of the model [25, 26, 32].

As seen in Table 2, the most sensitive predictors have been identified and ranked in resulting of the LVQNN prediction model. Accordingly, the most contributory factors have been listed according to their R values. Using professional judgment, the threshold of R is considered as percentage of 25 which is 75% of the range. Hence, sub-variables over the threshold of 25% had a greater affect in

performance of the prediction in terms of accuracy. Likewise, the factors from X1 to X17 proved a stronger relationship with the injury severity outcomes. Accordingly, they have been considered as contributory factors. In different circumstances, the label after X17 refers to lower R value of 25% which are poor predictors. It should be noted that the unreliable factors after X21 haven't been mentioned in the table, due to their insignificant association with the injury severity classes. However, their variables have been listed in Table 1.

The significant findings demonstrate that the most important factors contributed to likelihood of injuries while vehicles approaching (X2) to T or staggered junctions (X1). In line with this, Curiel et al.'s recent research [41] found that approximately half of the car accidents occurred at 5% of the city's junctions. Following this, going ahead on a bend (X3) and turning (X4) manoeuvres in connection with junction actions immediately contributed to the injuries. These predicaments are injury severity results which are consistent with the previous researches displaying that the rise of the injury severity is associated with vehicle actions [7, 31].

The next main contributory factor attributes to stationary or parked vehicle (X5) which is typical in the central of London, particularly, queues of public service vehicles (PSV)s (X14) and traffic behind a road block on streets. This

Table 2 Sensitivity of LVQNN against the absence of each variables using R

Variable	Rank	Label	Correlation coefficient (R) %			
			Validation	All data	Test data	Train data
		Sensitive predictors				
Junction detail	X1	T or staggered junction	85.7	86.1	84.2	88.3
Junction location	X2	Approaching junction	84.9	85.3	84.2	86.5
Vehicle manoeuvre	X3	Going ahead bend	79.3	79.8	82.1	82.2
	X4	Turning	76.3	78.1	75.2	79.4
	X5	Stationary or parked vehicle	49.6	52.2	49.3	53.2
Junction control	X6	Auto traffic signal	46.3	48.4	46.1	49.2
	X7	Give way or uncontrolled	43.9	46.2	43.2	46.8
Crossing facilities	X8	Pedestrian or cyclist phase at traffic signal junction	39.7	40.2	38.2	41.1
	X9	No physical crossing facilities within 50 m	36.3	38.2	36.1	38.9
Alcohol involvement	X10	Pedestrian	34.7	35.6	34.2	36.8
Time	X11	08:00–11:59 and 16:00–19:59 Rush hours	31.6	32.3	31.5	33.7
Vehicle type	X12	Car or taxi	30.8	30.9	30.5	32.9
	X13	Bicycle	27.6	27.9	27.2	29.8
	X14	PSV	25.8	25.8	25.2	26.9
Light condition	X15	Daylight	26.1	26.1	25.9	26.6
1st point of impact	X16	Front	25.6	25.6	25.1	26.5
2nd road class	X17	C	25.2	25.2	25.3	26.1
Road surface	X18	Wet	17.6	16.3	16.9	19.1
Junction detail	X19	Crossroads	14.2	13.9	13.3	15.2
1st point of impact	X20	Nearside	11.8	10.6	9.9	13.9
Light condition	X21	Darkness—lights lit	4.6	2.8	2.3	6.5

finding fits the City of London’s report displaying that most collisions in the daylight having occurred around parked vehicles and a lot of pedestrians obscured while crossing from between stationary or queuing vehicles. Thus crossed road masked by the vehicles and oncoming drivers failed to see or anticipate a crossing vulnerable road user [4, 42]. In addition, cycling collisions when hit by an opening door of stationary vehicle was caused without looking at passing riders.

The next variable refers to the same contributory factors which associated with junction actions [41]. Accordingly, junction control contributed to the injury when driver/riders did not stop at the traffic signal set at red (X6). They also disobeyed give way signs or road markings (X7). Moreover, an uncontrolled intersection (X7) was one of the leading causes of accidents related to junction control.

The next variable again refers to the injuries at junctions. X8 donates to wrong use of the crossing facilities by pedestrian or cyclists at junctions controlled by traffic signals which has an indicator light for the vulnerable road users [4].

Furthermore, the most common injuries suffered by the vulnerable road users occurred where the crossing facilities were not available within 50 m (X9). In this regards, a previous study discovered that the vulnerable road users were considered to affect the high likelihood of being involved in crashes [31]. Furthermore, a Dutch accident study showed that more than half of the killed or seriously injury accidents which vulnerable road users were involved in occurred while crossing the road [43].

X10 attributes to pedestrian impaired by alcohol. In that context, alcohol involvement for pedestrians was reported by the city’s council as the most frequent contributing factors [4].

This study found that collisions at dawn and dusk were less than in the daytime. Time band between 08:00–11:59 and 16:00–19:59 during the weekday’s morning and evening was more likely to be assigned as a contributory factor.

Consequently, despite the typical morning and evening rush hours’ stats which injuries are exposed to, the severity of injuries did not decrease between 12.00 and 16.00 as most road users adjusted their travel throughout the daytime in tourist areas of London accordingly [4, 42].

According to the key findings of prediction, the impact of vehicle type is very sensitive and plays a major role in outcome of injury severities. This result fits previous researches in which vehicle type is recognised as a main role in occurrence of injury severity [22, 23, 31]. Further to this, the most significant finding is related to cycling, which after cars, became the most common type of accident on the City streets [4]. Moreover, road user injuries in crashes involving the vehicle group indicated that alarming rises in PSV accidents in the capital’s roads over the

recent years are revealed today [44]. Transport for London (TfL) has decided to perform a bus safety standard in order to decrease the frequency of accidents as well as to mitigate the injury severities related to PSVs [45]. In respect to the inconstant factors which were located on the bottom of the table, it can be noted that, unusually, results of injuries on a wet road surface were lower R value.

3.4 Result of the injury severity classes applying sensitive predictors

As a result of the prediction, with the intention of generalising the dimensional feature space, the unreliable factors with minor R values were eliminated and dropped to 17 factors. Following this, the reduced data is aimed at implementation of the final prediction classes. All the sensitive predictors are normalised between 0 and 1, and the run was completed using random division of 70% and 30% in the training and testing datasets.

The confusion matrix is used to summarise and evaluate the performance of the prediction mission. Consequence of that, accuracy (ACC), error parameters, and sensitivity (SEN) measures are used with the aim of calculating the number of correct and incorrect predictions of each level. Sample of the confusion matrix is shown in Fig. 4 and the related equations are defined as below.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \tag{23}$$

$$SEN = \frac{TP}{TP + FN} \tag{24}$$

As seen, TP and TN are true when observations for their values are positive and negative, respectively. FN appears to be a false negative, and it happens when

	Predicted class+	Predicted class-	Total
Actual class+	TP	FN	N
Actual class-	FP	TN	P
Total	N	P	

Fig. 4 A sample of confusion matrix

the class of observation is negative while the classifier label signifies as positive. Moreover, FP shows a false positive, and it takes place when the class of observation is positive, even though, the classifier label appears as negative.

Accordingly, the predicted outcomes of injury severity using training and testing data are applied and the obtained results are broken down into each class as seen in Fig. 5. Accordingly, the blue marks donate the actual class of data and the pink marks present the predicted classes by each network. The interpretation of the results indicates that, if the pink marks integrate with the blue marks, the network succeeded to predict injury severity highly accurate, however if there is no integration, this indicates that the network predicted with less accuracy.

The accuracy measure in the training and testing stages for the injury severity outputs are specified in the Fig. 4. Nonetheless, for Y1, due to lack of data for fatal injury, the model had a very poor performance and was able to evaluate only three correct predictions in the training phase. Therefore, in the test stage, the amount of sensitivity for death was equal to one.

As for the prediction of seriously injured (Y2), again due to the lack of data, the incorrect classification still remains.

On account of the sufficient number of the data associated to slight injury, the attained results of Y3 was very satisfactory compared to the Y2 and Y1. Thus the model was able to extremely increase the accuracy rate of prediction in this class. Likewise, the performance of the model had highly improved and the classification was practically desirable and the amount of sensitivity for the training and test phase is obtained approximately 80%.

As a final point, the LVQNN model was capable to maximise the accuracy rate of the injury severity prediction used for damage only levels (Y4). Accordingly,

the sensitivity values for the training and testing are achieved about 81% and 87%, respectively.

4 Conclusion

This research primarily focused on further developments in predicting of injury severity of driver or rider by applying learning vector quantization model. Based on the data which related only to personal injuries involving traffic accident in central London, the accident prevention method estimates maximum likelihood of the injury severity classes into; fatal injury, serious injury, slight injury, and damage only. Additionally, the outcome of the prediction led to better understanding of the relationship between the injury severity classes and the crash related factors. Following this, a number of sensitive predictors are recognised which we believe have contributed to the severity of injuries sustained by drivers or riders.

Along the lines of the key findings, the impact of the junction actions and vehicle manoeuvre are discovered as overhead factors. Thus, they had above double affect compared to the other key influences, and played a large role in the likelihood of injury severity outcomes. Style of T or staggered is the inferior performing junction and the stats of vehicles approaching the junction was certainly an accident hotspot for all the road users. After this, more significant finding was liked to going ahead on a bend and turning manoeuvres resulting in junction actions. Another factor related to manoeuvre refers to stationary or parked vehicle. The next main contributory factors again were results of junction actions. Drivers or riders contributed to the injury when they disobeyed at the automatic traffic signals and when they didn't give way at sign or road markings, or while they met an uncontrolled junction. Next factor refers to wrong use of the crossing facilities by the vulnerable road users at a

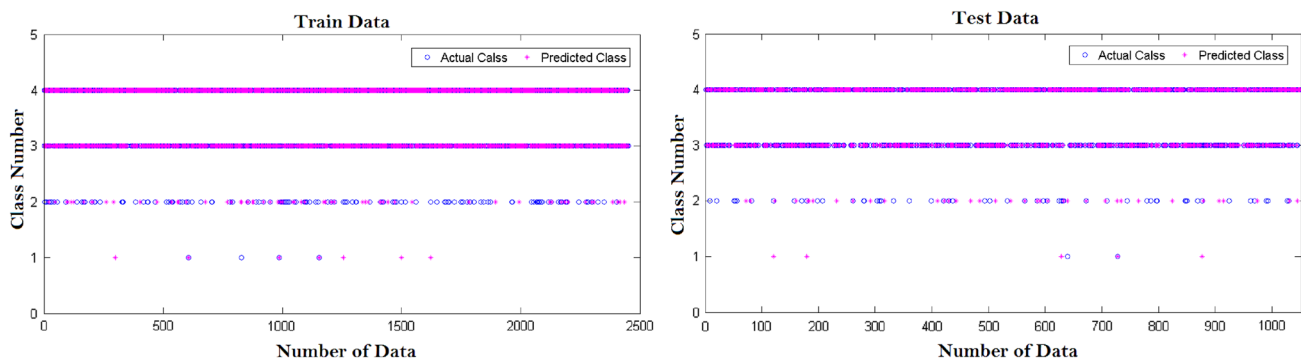


Fig. 5 Graph represents the correct and incorrect forecasts of actual and predicted classes

junction controlled by traffic signals. Moreover, the most common injuries suffered by unprotected road traffic participants which took place where a number of crossing facilities were not sufficient. Pedestrian impaired by alcohol, was a sensitive factor influencing the risk of the severity of injuries that result from accidents. The most dangerous times for all road users were between 08:00–11:59 and 16:00–19:59 and during night times seemed more likely to reduce the injury crashes. In the crash data, it is evident that type of vehicle contributed to injuries. In this respect cars and bikes presented the maximum concentration of the injuries.

This study ends by maximising the model's performance in terms of accuracy designed for the injury severity prediction. Consequently, the LVQNN model was conducted by applying the most sensitive predictors which considered as contributory factors. As a result, due to the lack of data for Y1 and Y2, killed and seriously injured classes led to poor classifications. On the other hand, as the data for slight injury and damage only classes in training stage was sufficient enough, the best levels were connected to Y3 and Y4. Also, for Y2, quantitative effects of each input sub variables on the injury severity could not predict properly and it tended to work with Y1. So, in terms of future work, with the purpose of achieving improved outcomes in classes used for killed and seriously injured, the Y1 and Y2 would be merged together as killed or seriously injured. Furthermore, as in this study, the detection of sensitive locations and groups were most in requirement of a road safety intervention, it would be valuable to focus on valuable road users—motor vehicle collisions approaching T/staggered intersections.

Acknowledgements The authors wish to thank DFT for providing the road safety data used in this article.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- World Health Organization (2018) Global status report on road safety 2018. World Health Organization
- RoSPA (2018) Road safety advice, information and resource. Royal Society for the Prevention of Accidents. <https://www.rospa.com/road-safety/>. Accessed 22 May 2019
- DfT (2018) Road safety statement progress report (Moving Britain Ahead). Department for Transport. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719259/road-safety-statement-progress-report.pdf. Accessed 1 July 2019
- City of London Corporation (2018) Road danger reduction and active travel plan 2018–2023. Department of the Built Environment City Transportation. <https://www.cityoflondon.gov.uk/services/transport-and-streets/road-safety/Documents/road-danger-reduction-plan-2018.pdf>. Accessed 28 June 2019
- Xie Y, Lord D, Zhang Y (2007) Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis. *Accid Anal Prev* 39(5):922–933
- DFT (2011) Instructions for the completion of road accident reports from non-CRASH sources. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/230596/stats20-2011.pdf. Accessed 20 March 2019
- Li Y, Ma D, Zhu M, Zeng Z, Wang Y (2018) Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. *Accid Anal Prev* 111:354–363
- Zhang J, Li Z, Pu Z, Xu C (2018) Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEE Access* 6:60079–60087
- Zong F, Xu H, Zhang H (2013) Prediction for traffic accident severity: comparing the Bayesian network and regression models. *Math Probl Eng* 2013:47519. <https://doi.org/10.1155/2013/475194>
- Zheng L, Meng X (2011) An approach to predict road accident frequencies: application of fuzzy neural network. In: 3rd international conference on road safety and simulation
- Dong N, Huang H, Zheng L (2015) Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. *Accid Anal Prev* 82:192–198
- Tabachnick BG, Fidell LS, Ullman JB (2012) Using multivariate statistics, 6th edn. Pearson, Boston
- Shen X, Chen (2009) Study on prediction of traffic congestion based on LVQ neural network. In: 2009 International conference on measuring technology and mechatronics automation, vol 3. IEEE, pp 318–321
- Priyono A, Ridwan M, Alias AJ, Rahmat RA, Hassan A, Ali MA (2005) Application of LVQ neural network in real-time adaptive traffic signal control. *J Teknol* 42(1):29–44
- Karlaftis MG, Vlahogianni EI (2011) Statistical methods versus neural networks in transportation research: differences, similarities and some insights. *Transp Res Part C Emerg Technol* 19(3):387–399
- Principe JC, Euliano NR, Lefebvre WC (2000) Neural and adaptive systems: fundamentals through simulations. Wiley, New York
- Abdelwahab HT, Abdel-Aty MA (2002) Artificial neural networks and logit models for traffic safety analysis of toll plazas. *Transp Res Rec* 1784(1):115–125
- Li X, Lord D, Zhang Y, Xie Y (2008) Predicting motor vehicle crashes using support vector machine models. *Accid Anal Prev* 40(4):1611–1618
- Li Z, Liu P, Wang W, Xu C (2012) Using support vector machine models for crash injury severity analysis. *Accid Anal Prev* 45:478–486
- Zhang J, Li Z, Pu Z, Xu C (2018) Comparing prediction performance for crash injury severity among various machine learning and statistical methods. *IEEE Access* 6:60079–60087
- Zeng Q, Huang H (2014) A stable and optimized neural network model for crash injury severity prediction. *Accid Anal Prev* 73:351–358
- Abdelwahab HT, Abdel-Aty MA (2001) Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. *Transp Res Rec* 1746(1):6–13
- Abdel-Aty MA, Abdelwahab HT (2004) Predicting injury severity levels in traffic crashes: a modeling comparison. *J Transp Eng* 130(2):204–210

24. Delen D, Sharda R, Bessonov M (2006) Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accid Anal Prev* 38(3):434–444
25. Aghayan I, Noii N, Kunt MM (2012) Extended traffic crash modelling through precision and response time using fuzzy clustering algorithms compared with multi-layer perceptron. *Promet-Traffic Transp* 24(6):455–467
26. Can Yilmaz A, Aci C, Aydin K (2016) Traffic accident reconstruction and an approach for prediction of fault rates using artificial neural networks: a case study in Turkey. *Traffic Inj Prev* 17(6):585–589
27. Yu R, Abdel-Aty M (2014) Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Saf Sci* 63:50–56
28. Chen C, Zhang G, Qian Z, Tarefder RA, Tian Z (2016) Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accid Anal Prev* 90:128–139
29. Iranitalab A, Khattak A (2017) Comparison of four statistical and machine learning methods for crash severity prediction. *Accid Anal Prev* 108:27–36
30. Aghayan I, Hosseinlou MH, Kunt MM (2015) Application of support vector machine for crash injury severity prediction: a model comparison approach. *J Civ Eng Urban* 5(5):193–199
31. Chang LY, Wang HW (2006) Analysis of traffic injury severity: an application of non-parametric classification tree techniques. *Accid Anal Prev* 38(5):1019–1027
32. Yasin Çodur M, Tortum A (2015) An artificial neural network model for highway accident prediction: a case study of Erzurum, Turkey. *PROMET-Traffic Transp* 27(3):217–225
33. Tazik M, Aghayan I, Sadeghi M (2017) Application of multiple logistic regression, bayesian logistic and classification tree to identify the significant factors influencing crash severity. *J Eng Sci Technol* 12(11):3044–3056
34. Wang X, Kim SH (2019) Prediction and factor identification for crash severity: comparison of discrete choice and tree-based models. *Transp Res Rec*. <https://doi.org/10.1177/0361198119844456>
35. Kohonen T (1997) *Self-organizing maps*. Springer, Berlin
36. Nova D, Estévez PA (2014) A review of learning vector quantization classifiers. *Neural Comput Appl* 25(3–4):511–524
37. Chen N, Marques NC (2009) Extending learning vector quantization for classifying data with categorical values. In: International conference on agents and artificial intelligence. Springer, Berlin, pp 124–136. https://doi.org/10.1007/978-3-642-11819-7_10
38. Thanasarn T, Warisarn C (2013) Comparative analysis between BP and LVQ neural networks for the classification of fly height failure patterns in HDD manufacturing process. In: Proceeding conference paper ICEAST
39. Al-Daoud E (2009) A comparison between three neural network models for classification problems. *J Artif Intell* 2(2):56–64
40. Villmann T, Bohnsack A, Kaden M (2017) Can learning vector quantization be an alternative to SVM and deep learning?—recent trends and advanced variants of learning vector quantization for classification learning. *J Artif Intell Soft Comput Res* 7(1):65–81
41. Curiel RP, Ramírez HG, Bishop SR (2018) A novel rare event approach to measure the randomness and concentration of road accidents. *PLoS ONE* 13(8):e0201890
42. City of London Corporation (2017) Road danger reduction—City Etiquette Guide: analysis of KSI collisions. <https://www.cityoflondon.gov.uk/services/transport-and-streets/road-safety/Documents/city-etiquette-guide-report.pdf>. Accessed 28 June 2019
43. SWOV (2010) SWOV Fact sheet: crossing facilities for cyclists and pedestrians. Leidschendam: Institute for Road Safety Research. https://www.swov.nl/sites/default/files/publicaties/gearchiveerde-factsheet/uk/fs_crossing_facilities_archived.pdf. Accessed 01 May 2019
44. Murray D (2017) Revealed: bus crashes on capital's roads rise by 5,000 in only three year. *Evening Standard*. <https://www.standard.co.uk/news/london/revealed-bus-crashes-on-capital-s-roads-rise-by-5000-in-only-three-years-a3648196.html>. Accessed 11 April 2019
45. TRL (2018) TRL published project report PPR819: Analysis of bus collisions and identification of countermeasures. Transport Research Laboratory. https://trl.co.uk/sites/default/files/Bus%20Safety%20Standard%20Phase%201%20Report_TRL.pdf. Accessed 07 June 2019

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.