JOURNAL OF TRAFFIC AND TRANSPORTATION ENGINEERING (ENGLISH EDITION) 2020; 7 (6): 775–790

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/jtte

Review Article

Machine learning applied to road safety modeling: A systematic literature review

Philippe Barbosa Silva [a,b,*], Michelle Andrade [b], Sara Ferreira [c]

[a] Department of Civil Engineering, Goiano Federal Institute-Rio Verde Campus, Rio Verde 75901-970, Brazil
[b] Department of Civil and Environmental Engineering, University of Brasilia, Brasilla 70910-900, Brazil
[c] Research Centre for Territory, Transports and Environment, University of Porto, Porto 4200-465, Portugal

HIGHLIGHTS

- The review study explored three different approaches to predict crashes.
- The use of machine learning techniques in crash prediction models are promising.
- Neural networks is the most used machine learning technique for crash prediction.
- The road-environmental factors are the most used in the three modeling approaches.

ARTICLE INFO

Article history:
Received 18 September 2019
Received in revised form
6 July 2020
Accepted 12 July 2020
Available online 31 October 2020

Keywords:
Transportation engineering
Road safety modeling
Crash prediction
Crash injury severity
Machine learning
Systematic literature review

ABSTRACT

Road safety modeling is a valuable strategy for promoting safe mobility, enabling the development of crash prediction models (CPM) and the investigation of factors contributing to crash occurrence. This modeling has traditionally used statistical techniques despite acknowledging the limitations of this kind of approach (specific assumptions and prior definition of the link functions), which provides an opportunity to explore alternatives such as the use of machine learning (ML) techniques. This study reviews papers that used ML techniques for the development of CPM. A systematic literature review protocol was conducted, that resulted in the analysis of papers and their systematization. Three types of models were identified: crash frequency, crash classification by severity, and crash frequency and severity. The first is a regression problem, the second, a classificatory one and the third can be approached either as a combination of the preceding two or as a regression model for the expected number of crashes by severity levels. The main groups of techniques used for these purposes are nearest neighbor classification, decision trees, evolutionary algorithms, support-vector machine, and artificial neural networks. The last one is used in many kinds of approaches given the ability to deal with both regression and classification problems, and also multivariate response models. This paper also presents the main performance metrics used to evaluate the models and compares the results, showing the clear superiority of the ML-based models over the statistical ones. In addition, it identifies the main explanatory variables used in the models, which shows the predominance of road-environmental aspects as the most important factors contributing to crash occurrence. The review fulfilled its objective, identifying the various approaches and

* *Corresponding author*. Department of Civil Engineering, Goiano Federal Institute-Rio Verde Campus, Rio Verde 75901-970, Brazil. Tel.: +55 64 3620 5600.
E-mail addresses: philippe.silva@ifgoiano.edu.br (P.B. Silva), michelleandrade@unb.br (M. Andrade), sara@fe.up.pt (S. Ferreira).
Peer review under responsibility of Periodical Offices of Chang'an University.
https://doi.org/10.1016/j.jtte.2020.07.004

the main research characteristics, limitations, and opportunities, and also highlighting the potential of the usage of ML in crash analyses.

1. Introduction

The growth of countries and populations has given rise to various externalities, such as the increase of road crash. There are millions of deaths from traffic accidents every year, besides severe economic, social, and environmental consequences. Many efforts have been made to reduce the frequency and severity of traffic accidents. The most efficient way to tackle the problem is by means of an extensive program of road safety management (Nodari and Lindau, 2007), in which road safety modeling is essential. The modeling process attempts to adjust a model to the crash data, the geometric and operational characteristics of the road, and the environmental conditions, incorporating the most important factors (Chang, 2005; Hauer, 2004).

Different modeling techniques have been developed to improve the representation of reality in the models, which allows the employment of techniques that are more appropriate to the problem's data (Costa et al., 2016). Traditionally, statistical modeling techniques have been used to predict crashes and classify their severity (Kidando et al., 2019; Lord and Mannering, 2010; Savolainen et al., 2011). However, the limitations of this approach have been widely explored, offering an opportunity to use new approaches, such as machine learning (ML) techniques.

To the best of the authors' knowledge, papers addressing the state-of-the-art of road safety modeling using ML techniques are unknown, despite the current importance of this topic. In general, ML techniques are superficially mentioned in road safety papers. The objective of this paper is to present a review of the most recent papers reporting the use of ML techniques to analyze crash data, predict crash frequency, and classify severity. This paper will first make some observations concerning crash modeling both for frequency and severity and then discuss the main features of each of the methodological approaches presented.

2. Review methodology

This study made use of the systematic literature review (SLR) methodology to search for, identify, and select appropriate papers concerning the use of ML techniques to analyze crash occurrence. The aim of SLR is to identify good quality references of real interest to a study. This paper follows the procedure proposed by Kitchenham and Charters (2007), which is conducted in three stages, namely planning, conducting, and reporting, as shown in Fig. 1.

The search terms were divided into two groups: terms associated to crash prediction models and terms related to machine learning techniques. The strings were defined to identify any term associated to crash prediction models (e.g., “crash prediction”, “injury severity”, “road traffic crash”, “crash injury”, combined with the function OR) with the function and to a term related to machine learning (e.g., “machine learning”, “artificial intelligence”, “expert system”). The platforms selected for conducting the search were CAPES Periodicals Portal (the Brazilian platform with the largest

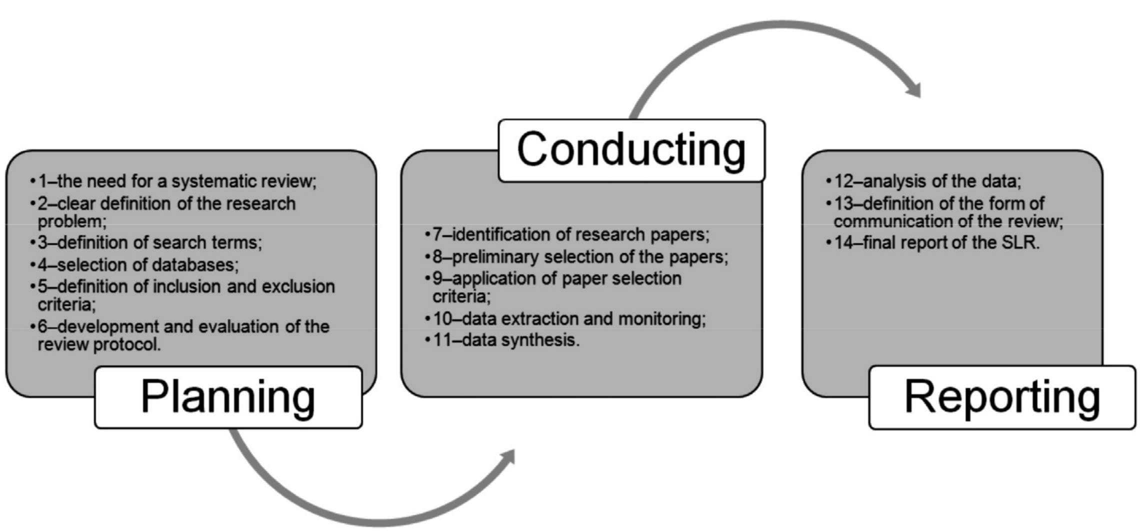


Fig. 1 – General framework of the review procedure (Kitchenham and Charters, 2007).

selection of international journals) and Google Scholar. The CAPES Periodicals Portal was selected because it covers many widely known databases, namely: Web of Science, Scopus, Journal Citation Reports (JCR), Engineering Village, MAS, ASTM International, SciFinder, ProQuest, Britannica Academic Edition, Thomson Reuters, Eighteenth Century Collections Online and Begell House. Google Scholar is a platform with a wider reach, capturing results not contemplated by the other databases.

The set of results was further refined by applying inclusion criteria (works obtained by snowball sampling) and exclusion criteria (inaccessible works, works that were not papers, repetitions, works outside the scope of the transportation area, and works in languages other than English), resulting in 122 papers. Those with a JCR impact factor or a Scientific Journal Rankings (SJR) index higher than 0.5 were then selected. Finally, a preliminary analysis of the remaining papers was conducted, and 26 papers were compatible with the purpose of this SLR. Fig. 2 displays the publication timeline of the systematized papers with an identification of their models. Reporting is the last stage of SLR, in which the papers are systematized and analyzed. This paper is the result of that reporting process.

3. Road safety modeling

Wang et al. (2011) stated that crash prediction models have been widely used to estimate crash frequency for a given location during a specified period. Savolainen et al. (2011) highlighted the importance of prediction models for injury severity as they contribute to the proposal of countermeasures to reduce crash severity. Kim and Washington (2006) and Hauer (2004) described how road safety modeling can provide two kinds of results: estimates of crash frequency (or severity) based on the infrastructure characteristics and estimates of how the characteristics of the infrastructure can influence the expected frequency (or severity) of crashes.

Different approaches can be used to predict either crash severity or crash frequency. The response variable in severity analyses is crash classification, which can be a binary problem (injury or non-injury; injury or property damage; severe injury or non-severe injury; possible/non-incapacitating injury or incapacitating/fatal injury) or a multiclassification problem (no injury, injury or fatal injury; no injury, possible injury, evident injury or incapacitating/fatal injury; no injury, possible injury, evident injury or incapacitating injury or fatal injury). Researchers have also investigated the relationship between crash severity and risk factors (human factors, road-environmental, and/or vehicle-related factors) including analyses of specific types of crash (e.g., vehicle rollover) or the vehicle involved (e.g., crash between two light vehicles). Researchers have used crash frequency prediction models as an attempt to detect a relationship between the number of crashes and the risk factors, mostly the road-environmental. The response variable of these models is the number of crashes per segment or the number of crashes per segment per year.

Traditionally, statistical techniques have been used to model road safety. Many models have been used (Lord and Mannering, 2010): Poisson regression, binomial regression, negative binomial regression, Poisson-lognormal regression, gama regression, zero-inflated regression, generalized estimation equations, negative multinomial model, random effects model, and random parameters model. For crash severity, the following models have been proposed (Savolainen et al., 2011): binary logit, binary probit, Bayesian ordered probit, Bayesian hierarchical binomial logit, generalized ordered logit, log-linear model, multinomial logit, multivariate probit, ordered logit, and ordered probit.

However, the limitations of statistical modeling are widely acknowledged since each model has its own assumptions and predefined relationships between dependent and independent variables (Zeng et al., 2016a), despite the progress achieved with these techniques. Mussone et al. (1999), Li et al. (2012), and Chang (2005) also pointed out that statistical modeling requires assumptions related to data distribution. Such premises may be untrue and, being violated, might lead to mistaken estimates and incorrect inferences. The use of artificial neural networks (ANN) does not require that kind of predefined relationship between the variables (Abdelwahab and Abdel-Aty, 2001; Chang, 2005; Li et al., 2012; Mussone et al., 1999). Instead of defining an analytical functional form, which may be laborious, a model is reconstructed after learning from real crash data, obtaining the weights of the model's variables. In that context, researchers have been making considerable efforts to explore the applicability of machine learning techniques to road safety modeling, which is the object of analysis of this paper.

3.1. *Methodological approaches*

Machine learning is a sub-division of artificial intelligence and is widely used as a powerful tool for solving problems in various domains. ML algorithms involve knowledge of various

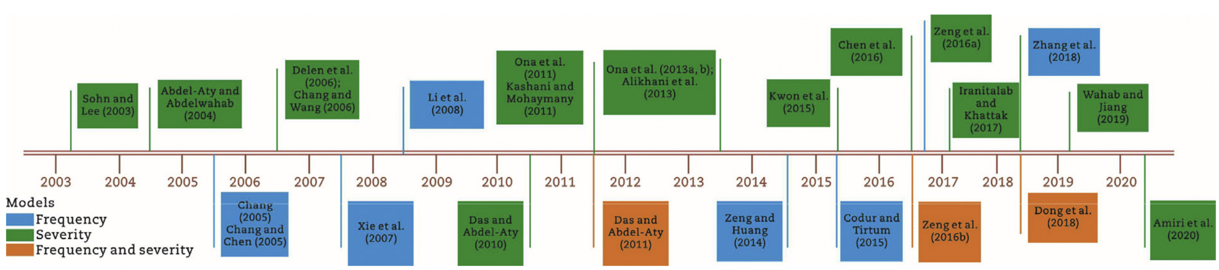


Fig. 2 – Publication timeline of the systematized papers.

areas such as probability and statistics, computational complexity, information theory, psychology, neurobiology, and control theory (Basgalupp et al., 2009).

Learning can be divided into supervised, unsupervised, and semi-supervised learning. In supervised learning, the response values of the examples in the training set are known; in other words, the goal is to learn a mapping of x to y given a training set with the pairs (x_i, y_i). In this case, the model's response to the current pattern of inputs is evaluated, which allows changes that bring the model response closer to the expected (known) response. In unsupervised learning, the data are non-labeled (unknown classes) and the objective is to find a structure (relationships or patterns) in the data (x_1, … , x_n) of the n examples. Semi-supervised learning is an intermediary situation between the other two; in addition to supplying non-labeled data (without a known response), some of the supervision information is supplied, but not necessarily to all examples (Chapelle et al., 2006; Decker and Focardi, 1995).

Given a set of input data described by pairs (x_i, y_i), where x is the vector of the variables that represent the predictive attributes and y is the label of the class that the example belongs to, the task is to learn a target function f that maps each set of attribute x in one of the predefined y classes (Tan et al., 2005). The class (label) is a special attribute that describes a characteristic of the phenomenon of interest. If the class is continuous, it is a regression problem, and, if it is discrete, it is a classification problem (Chapelle et al., 2006).

The major groups of techniques identified for road safety modeling were nearest neighbor classification, decision trees, evolutionary algorithms, support-vector machines, and artificial neural networks.

Nearest neighbor classification (KNN) is a simple and pioneering technique in ML. In a prediction task, the KNN classifies an observation based on the closest k observations. The nearest neighbor decision rule is used to assign a new sample point with the classification associated to the nearest of a set of previously classified points. Therefore, the class of the observation of interest should include the majority of the k closest observations (Devroye et al., 1994).

Decision tree (DT) technique is very useful for classification tasks. In the construction of a tree, a training set made up of inputs and outputs (i.e., classes) is formed. The tree structure consists of a root node that begins the tree, decision nodes that divide an attribute and form ramifications, and leaves that contain the classification information. Each node represents the test of an attribute and the criterion for ramification is the attribute's utility for classification. Thus, the selected attribute, one of the tree nodes, generates the greatest information gain (entropy); i.e., it provides the best quality for classification. The tree path (from the root node to each leaf node) corresponds to an association rule (Quinlan, 1986; Trabelsi et al., 2019). In decision trees, the induction algorithms seek the attributes that better generate the examples, generating sub-trees.

Evolutionary algorithms (EA) are stochastic search methods based on natural selection mechanisms in which the fittest individuals survive (Holland, 1975). Each individual corresponds to a candidate solution for a problem and is evaluated by a fitness function, that measures the quality of the solution. For each iteration (generation), the best individuals are more likely to be chosen for reproduction. The selected individuals are subject to crossover (parts of the genetic material of two individuals are exchanged) and mutation (part of the genetic material of an individual is replaced by other random genetic material), generating new individuals (offspring) that will replace the parents and form a new generation of the population. That process is iteratively repeated until a stop criterion is satisfied (Floreano and Mattiussi, 2008; Yu and Gen, 2010). The two main types of evolutionary algorithms are genetic algorithms and genetic programming, and they are especially useful for optimizing problems, usually associated with other techniques.

Support-vector machine (SVM) technique is based on statistical learning theory (Scholkopf and Smola, 2002). It constructs a hyper plane as a decision surface to maximize the margin of separation between examples. The model uses the hyper plane to discriminate the set of test samples in two groups, namely, positive samples and negative samples. Although it was originally conceived as a classification technique, it has been extended to solve regression problems and problems with non-linearly separable data (Burges, 1998; Smola and Scholkopf, 2004; Trafalis and Gilbert, 2006; Üstün et al., 2005).

An artificial neural network (ANN) is a highly complex, non-linear, parallel processor with a natural propensity for storing experimental knowledge and making it available afterward (Haykin, 2009). A multi-layer perceptron ANN is typically made up of three kinds of layers: an input layer, an output layer, and one or more hidden layers. The input layer receives the values of the explanatory variables, i.e., the input data. The hidden layer, made up of m neurons, adds up the weights of the input values of the various explanatory variables, and calculates the complex association patterns. A single hidden layer is usually enough for crash analysis applications, but the definition of the number of neurons in it is generally the object of experimentation (Chang, 2005; Villiers and Barnard, 1993). For the output layer, the values of the various hidden neurons are summed and the network's output values are presented. Feedforward is the most common type of network architecture, in which the propagation of signals is always from the previous layers to the posterior ones. In terms of training, the back propagation algorithm is the most used to minimize errors by adjusting the weights of the network (Haykin, 2009). The gradient descent method is generally used. In this case, the cost function is in the direction in which the function's variation rate is minimal and it guarantees that the network surface trends in the direction that leads to the greatest error reduction. Lastly, the main activation function used is related to the representational capacity of the neural network and it introduces a non-linear component. Sigmoid-type functions are generally employed Fig. 3 displays examples of DT, EA, SVM, and ANN.

The development of ML culminated in a new approach described as deep learning (DL). It explores many layers of non-linear information, supervised or unsupervised, to analyze or classify patterns (Deng and Yu, 2014). DL is a ML subarea that builds models capable of extracting

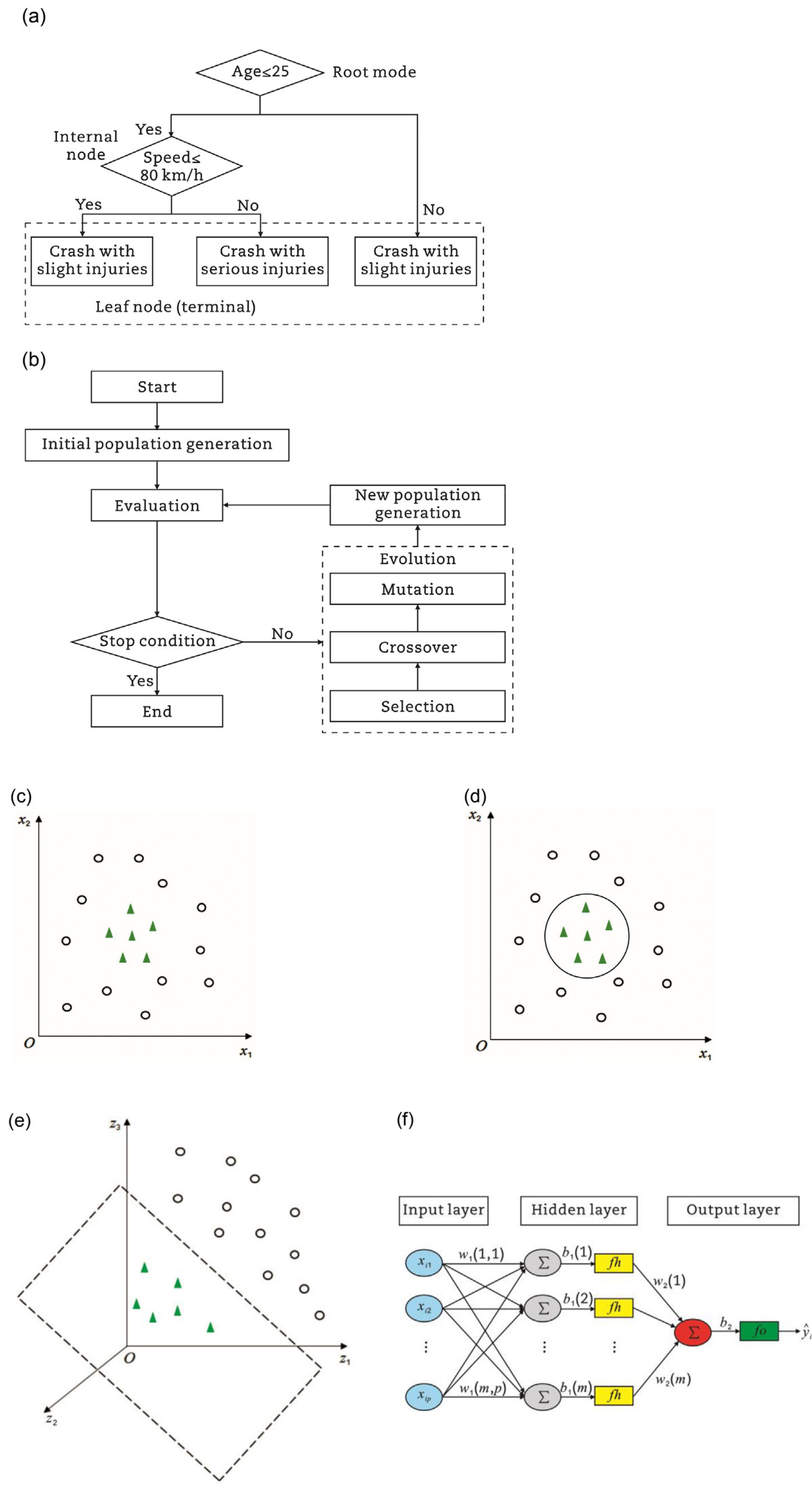


Fig. 3 – Techniques identified for road safety modeling. (a) Example of a DT (Abellán et al., 2013). (b) Typical flow diagram of EA (Das and Abdel-Aty, 2010). (c) Example of how SVM works for non-linearly separable data (non-linear data set) (Lee and Park, 2011). (d) Example of how SVM works for non-linearly separable data (non-linear borders in the inputs space) (Lee and Park, 2011). (e) Example of how SVM works for non-linearly separable data (linear borders in the characteristic space) (Lee and Park, 2011). (f) Typical structure of three-layered feedforward ANN (Xie et al., 2007).

characteristics from the lowest (deepest) level to the highest (most superficial) level (LeCun et al., 2015; Schmidhuber, 2015). This ability overcomes the limitation of raw data processing, a common problem with many ML techniques. The most common examples of this approach are deep neural networks (DNNs), recurrent neural networks (RNNs) and convolutional neural networks (CNNs), generally applied in speech recognition, visual object recognition, and object detection.

As ML techniques have different working principles, road safety modeling may also take advantage of different approaches. The use of DT, for example, is indicated and useful for classifying the crash severity, especially when the objective is to interpret the interaction of the factors. SVM are efficient for use in classification problems, with better results than ANN, but they provide less interpretability than DT. SVM can also be used for regression problems. Genetic programming is more successful in combination with other machine learning techniques to optimize the results. Finally, ANN is a robust technique, with high computational cost, but suitable for complex problems, such as simultaneous modeling with multiple outputs. Table 1 displays the papers identified by the systematic literature review by type of technique and general characteristics. The main aspects of each work will be presented in the following sections.

3.1.1. *Crash frequency modeling*

ML crash prediction models have generally been compared to negative binomial regression models (NB). In most cases, the performance of ML models was greater (Chang, 2005; Chang and Chen, 2005; Li et al., 2008; Xie et al., 2007; Zeng et al., 2016a).

Among the ML techniques, ANN have been mostly used in crash prediction models with the traditional back propagation training algorithm (Chang, 2005; Xie et al., 2007; Zeng et al., 2016a). Çodur and Tortum (2015) compared ANN using two different training algorithms, back propagation and its variant, the Levenberg-Macquardt algorithm. The usage of the latter brought convergence improvement for the ANN, which confirmed its properness to model crash prediction.

Other ML techniques have been explored, such as CART (Chang and Chen, 2005) and SVM (Li et al., 2008). Their performance was also greater than traditional statistical models. In addition, Xie et al. (2007) investigated the usage of BNN for a crash prediction model and compared it to ANN. The BNN's performance was better than an ANN with back propagation.

3.1.2. *Modeling crashes by severity*

Crashes can be modeled by severity levels using bivariate or multivariate structures depending on the number of proposed levels. Bivariate models only have two levels, such as crashes with injured (any severity level) and property damage only (no injuries), as in Sohn and Lee (2003), Alikhani et al. (2013), and Kwon et al. (2015). Multivariate models have more than two levels. Abdel-Aty and Abdelwahab (2004) proposed 4 severity levels: no injury, possible injury, evident injury, and severe/fatal.

The reviewed studies had different goals with the usage of ML to model crash severity. Most works explored different techniques or attempted to improve a single technique by changing its structure, the training algorithms, the activation functions, or using auxiliary mechanisms (e.g., data clustering). Clustering brought improvements in numerous applications, such as genetic programming (Das and Abdel-Aty, 2011), ANN (Alikhani et al., 2013; Sohn and Lee, 2003), DT (Sohn and Lee, 2003), and latent class analysis (LCA) and Bayesian networks (RB) (Oña et al., 2013b).

Works that evaluated the application of ML techniques for crash prediction by severity compared the results with statistical models. Many models have been used for this purpose. The logit model was used to evaluate the results of the bivariate model of Kwon et al. (2015); the ordered probit model and multinomial logit were used in comparison to the multivariate models, such as the models developed by Abdel-Aty and Abdelwahab (2004), Zeng and Huang (2014), Iranitalab and Khattak (2017), Zhang et al. (2018), and Wahab and Jiang (2019). The results indicated the greater performance of the ML models.

Numerous ML techniques, associated or not, have been used to explore not only how ML models can be improved, but also the influence of explanatory variables in crash occurrence by severity. ANN, DT, and CART were the most explored techniques, but the list of models also include linear genetic algorithm (LGA), adaptive neuro-fuzzy inference system (ANFIS), Bayesian network (BN), latent class analysis (LCA), naive bayes classifiers (NBC), NHPF, support vector machine (SVM), random forest (RF), *k*-nearest neighbor (KNN), and hybrid intelligent genetic algorithm (HIGA). Among the main conclusions of these studies, some observations can be mentioned as follows.

- Sensitivity analysis can be used with ML models to identify the most relevant explanatory variables for crash classification. This technique was employed with ANN with bivariate outputs (Delen et al., 2006), ANN with multivariate outputs (Zeng and Huang, 2014) and with CART and SVM (Chen et al., 2016).
- Multi-class classification problem is most efficiently addressed using binary response variables than using one multi-categorical response (Chen et al., 2016; Delen et al., 2006).
- The performance of decision trees was greater than ANN in Sohn and Lee (2003) and NBC in Kwon et al. (2015). It presented similar results to CART in Oña et al. (2013a). These three studies developed bivariate models. The decision trees also presented an interesting application for visualizing the importance factors of each explanatory variable, such as reported by Chang and Wang (2006).
- However, the performance of RF was greater than DT in the works of Wahab and Jiang (2019) and Zhang et al. (2018). In these studies, other techniques were employed, KNN and SVM, but RF presented a better approach in all cases. The tested models were all multivariate.

3.1.3. *Modeling crashes by both frequency and severity*

To obtain a broader view of road safety on highways, Das and Abdel-Aty (2011) combined analyses of crash frequency and

Table 1 – General description of the systematized papers.

Reference	Dependent variable		Technique used	Number of crashes (period)	Data set (training/validation) (%/%)	Study area	
Sohn and Lee (2003)	Classification	Injury; property damage only	ANN; DT	11,564 (1 year)	60/40	Urban road	
Abdel-Aty and Abdelwahab (2004)	Classification	No injury; possible injury; evident injury; incapacitating/fatal injury	ANN	7891 (2 years)	51.9/48.1	Urban road and highway	
Chang (2005)	Frequency	Number of crashes per segment per year	ANN	1338 (2 years)	75/25	Multilane highway	996 segments (from 0.1 km to 4.2 km)
Chang and Chen (2005)	Frequency	Number of crashes per segment per year	CART	1075 (2 years)	75/25	Multilane highway	742 segments of 1 km
Delen et al. (2006)	Classification	No injury; possible injury; non-incapacitating injury; incapacitating injury; fatal injury	ANN	30,358 (6 years)	–	Urban road and highway	
Chang and Wang (2006)	Classification	No injury; injury; fatal injury	CART	12,604 (1 year)	–	Urban road and highway	
Xie et al. (2007)	Frequency	Number of crashes per segment	ANN; BNN	122 (5 years)	60/40; 70/30; 80/20	Two-way two-lane highway	88 segments (from 1.11 km to 8.59 km)
Li et al. (2008)	Frequency	Number of crashes per segment	SVM	122 (5 years)	60/40; 70/30; 80/20	Two-way two-lane highway	88 segments (from 1.11 km to 8.59 km)
Das and Abdel-Aty (2010)	Classification	No injury; possible/non-incapacitating injury; incapacitating/fatal injury	GP	104,952 (3 years)	70/30	Expressway	
Oña et al. (2011)	Classification	Slight injury; severe/fatal injury	BN	1536 (3 years)	66.67/33,33	Two-way two-lane highway	
Kashani and Mohaymany (2011)	Classification	Slight injury; severe injury; fatal injury	CART	21,025 (3 years)	70/30	Two-way two-lane highway	
Das and Abdel-Aty (2011)	Classification	Non-severe injury; severe/fatal injury	GP	57,155 (3 years)	70/30	Urban road	Segments of 850 m
	Frequency	Number of crashes per segment					
Oña et al. (2013a)	Classification	Non-severe injury; severe/fatal injury	CART; DT	1801 (7 years)	70/30	Two-way two-lane highway	
Oña et al. (2013b)	Classification	Non-severe injury; severe/fatal injury	BN	3229 (4 years)	–	Two-way two-lane highway	
Alikhani et al. (2013)	Classification	Injury; property damage only	ANN; ANFIS	7035 (1 year)	80/20	Highway	
Zeng and Huang (2014)	Classification	No injury/property damage only; possible injury; non-incapacitating injury; incapacitating/fatal injury	ANN	53,732 (1 year)	80/20	Highway	
Kwon et al. (2015)	Classification	Property damage only; fatal injury	DT, NBC	1,350,958 (7 years)	70/30	Highway	
Çodur and Tortum (2015)	Frequency	Number of crashes per segment	ANN	7285 (8 years)	70/30	Multilane highway	16 segments (from 4 km to 18.4 km)

(*continued on next page*)

Table 1 – (continued)

Reference	Dependent variable		Technique used	Number of crashes (period)	Data set (training/validation) (%/%)	Study area	
Chen et al. (2016)	Classification	No injury; non-incapacitating injury; incapacitating/fatal injury	CART; SVM	3106 (2 years)	60/40; 70/30; 80/20	Urban road and highway	
Zeng et al. (2016a)	Frequency	Number of crashes per segment per year	ANN	1612 (5 years)	–	Highway	211 segments (from 0.15 km to 9.07 km)
Zeng et al. (2016b)	Frequency by severity	Number of crashes with slight injuries per segment per year; number of crashes with severe or fatal injuries per segment per year	ANN	1612 (5 years)	–	Highway	211 segments (from 0.15 km to 9.07 km)
Iranitalab and Khattak (2017)	Classification	Property damage only; possible injury; severe injury; disabling/fatal injury	KNN; SVM; RF	68,448 (4 years)	70/30	Local, interstate and highway	
Zhang et al. (2018)	Classification	No injury; possible injury; non-capacitating injury; incapacitating injury; fatal injury	KNN; DT; RF; SVM	5538 (3 years)	75/25	Freeway	
Dong et al. (2018)	Frequency by severity	Number of major injury crashes per segment per year; number of minor injury crashes per segment per year; number of no injury crashes per segment per year	DL; SVM	5365 (5 years)	80/20	Highway	635 segments (from 0.032 km to 19,81 km)
Wahab and Jiang (2019)	Classification	Damage injury; injured; hospitalized; fatal injury	DT; RF; KNN	8516 (5 years)	10-fold cross validation	Urban road	
Amiri et al. (2020)	Classification	Property damage only; complaint of pain; visible injury; severe injury; fatal injury	ANN; HIGA	4070 (1 year)	70/30	Highway	

Note: DT means decision trees; CART means classification and regression trees; GP means genetic programming; BN means Bayesian networks; BNN means Bayesian neural networks; ANN means artificial neural networks; SVM means support vectors machine; DL means deep learning; KNN means k-nearest neighbor; HIGA means hybrid intelligent genetic algorithm; NBC means naïve Bayes classifiers; ANFIS means adaptive neuro-fuzzy inference system.

severity. Genetic Programming was used to investigate those two aspects, frequency and severity, but a joint prediction model was not developed. For frequency prediction, they selected the models with the least errors and, for severity classification, the models with the highest accuracies. Their results showed the overlap of a set of significant factors (median, skid resistance, and road width) for both the frequency and the classificatory models. Their hypothesis for this overlap was the existence of a complex relationship between apparently different problems. The authors underscored that modeling with GP provides independence for the development of models free from any data distribution restrictions.

Zeng et al. (2016b) used ANN to explore the nonlinear relationship between crash frequency by severity and risk factors. They proposed a network structure optimization and a rule extraction method to eliminate the possibility of over-fitting and to deal with the network's "black box" characteristic. Their results indicated that, if trained and optimized, neural networks have a better fit and also better prediction power than the Poisson-lognormal multivariate model. The authors stated that the extracted rules implied a nonlinear relationship between each explanatory variable and crash frequency by severity levels in different conditions. Accordingly, they believe that the use of optimization algorithms and rule extraction can provide the modified neural networks considerable improvements for modeling crash frequency by severity levels.

As a last example of crash frequency modeling by severity, Dong et al. (2018) proposed the usage of deep learning in two steps. First, with an unsupervised step to establish a relationship between explanatory variables. Second, with a supervised step to predict the number of crashes by severity level. The authors also incorporated the unobserved heterogeneity issues with a layer consisting of a multivariate negative binomial (MVNB) model. The results, which were compared to a SVM model, suggest that deep learning is a better approach to predict crashes since it is capable of simultaneously modeling crash prediction by severity levels.

3.2. *Explanatory variables*

The choice of the explanatory variables is an essential step of the modeling process and it depends on the purpose of the model. The inclusion of a variable assumes that it has a degree of association with the dependent variable of interest. Therefore, the selection of predictors depends on previous judgment and knowledge about the data, prior modeling experience, and availability of data (Hauer, 2015).

Crashes are complex events that involve the interactions of various contributing factors. Many studies have investigated aspects that are expected to have influence on crashes, such as the roadways' geometrical and operational characteristics, the environment, the condition of the vehicles, lightning and human factors (Abdel-Aty and Radwan, 2000; Carson and Mannering, 2001; Elvik et al., 2009; Miaou and Lum, 1993; Rolison et al., 2018; Wang et al., 2013). To continue research in this area, it is important to know which variables the various studies have used in their models. To evaluate the models, this review grouped the variables in four major classes: human factors, road-environmental factors, vehicle-related factors, and crash characterization.

Fig. 4 displays the variables' distribution for crash classification by severity studies. It shows that all studies incorporated road-environmental factors into their modeling even though, in some cases, such as Sohn and Lee (2003), only one variable was considered. In addition to this study, only Delen et al. (2006) and Kwon et al. (2015) did not have most of the variables in their studies related to environmental conditions. The latter used factors vehicle-related factors which were entirely absent from the models of Alikhani et al. (2013), Das and Abdel-Aty (2010), Iranitalab and Khattak (2017), Kashani and Mohaymany (2011), Oña et al. (2011), Oña et al. (2013b), Zhang et al. (2018). In other models, they were inexpressive (i.e., small number of variables compared to the other groups). In general terms, human factors and crash characterization were used to the same extent in developing the models reported in the literature.

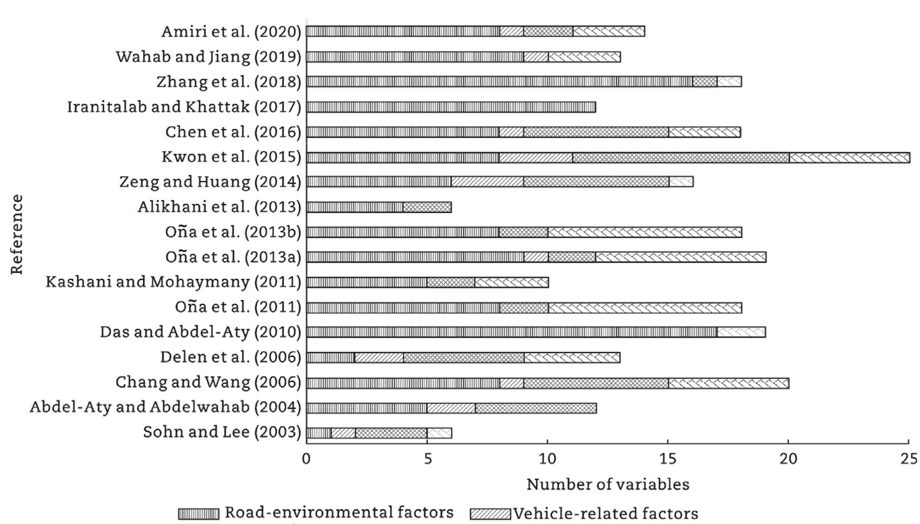


Fig. 4 – Group of variables used in crash modeling by severity.

Fig. 5 displays the distribution of the variables for each crash prediction model. There is an evident dominance of road-environmental factors in all models, and most of them actually only present this group of variables. All the analyzed studies made use of road-environmental factors to develop their models. In crash prediction models, the data corresponds to highway segments and not for individual crash occurrences (and their respective characteristics). This justifies the relative lack of expression of vehicle-related factors, human factors, and crash characterization factors for crash prediction models. Individual crash characteristics and human- and vehicle-related are hard to include in road segments attributes. In most cases, these factors make no sense or have no significance in the model.

Joint crash frequency and severity models also segmented the highway, such as Das and Abdel-Aty (2011), Zeng et al. (2016b), and Dong et al. (2018). Because of that, these studies also basically made use of road-environmental factors. Table 2 sets out the most utilized explanatory variables for each kind of model.

In addition to a general overview of the explanatory variables used in previous studies, it is essential to recognize the most important variables to give an indication that may assist modelers when choosing the variables to use in future studies. Table 3 displays the most strongly related variables to crash frequency or severity for each of the works reviewed. For the models designed to classify crash severity, the variables considered to be most important are posted speed limit, traffic volume, land use, traffic flow separation devices (median or median barrier), pavement surface, horizontal signaling, roadway width, and the number of lanes. Also, traffic volume, segment length, horizontal alignment, and posted speed limit are the most important factors for models designed to predict crash frequency per segment.

3.3. *Performance metrics*

Evaluating the performance of the models is as important as developing them. Certain metrics of performance (e.g., accuracy (A), sensitivity (R), specificity (S), and F-measure) have been used to evaluate the classifiers in the reviewed models.

Those metrics are based on examples that have been correctly and incorrectly classified, which are stored in a confusion matrix. Four possible situations can occur in this confusion matrix: (i) true positive (TP) – the example is correctly predicted as belonging to a positive class; (ii) false positive (FP) – the example is predicted as belonging to the positive class but actually belongs to the negative one; (iii) true negative (TN) – the example is correctly predicted as belonging to the negative class; and (iv) false negative (FN) – the example is predicted as belonging to the negative class but actually belongs to the positive one. These values make possible to measure a model's performance. The accuracy (hit rate), recall (sensitivity), specificity, precision (P), and F-measure are obtained using Eqs. (1)–(5).

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$S = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{3}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

$$F\text{-measure} = \frac{2PR}{P + R} \tag{5}$$

Accuracy refers to the extent to which the test is capable of determining the true value, or, in other words, to which extent it is capable of estimating the probability of the classifier being correct in its predictions. Recall (sensitivity) measures the capacity of making a positive prediction of a class in which the prediction turns out to be correct, in other words, it is a measure of how many positive examples were correct out of the total number of examples. The specificity metrics refer to the capacity of predicting a negative class in which the prediction is correct, in other words, the number of negative examples that were predicted out of the total number of examples. Precision calculates the probability of a positive prediction being correct in relation to all the samples. Finally,

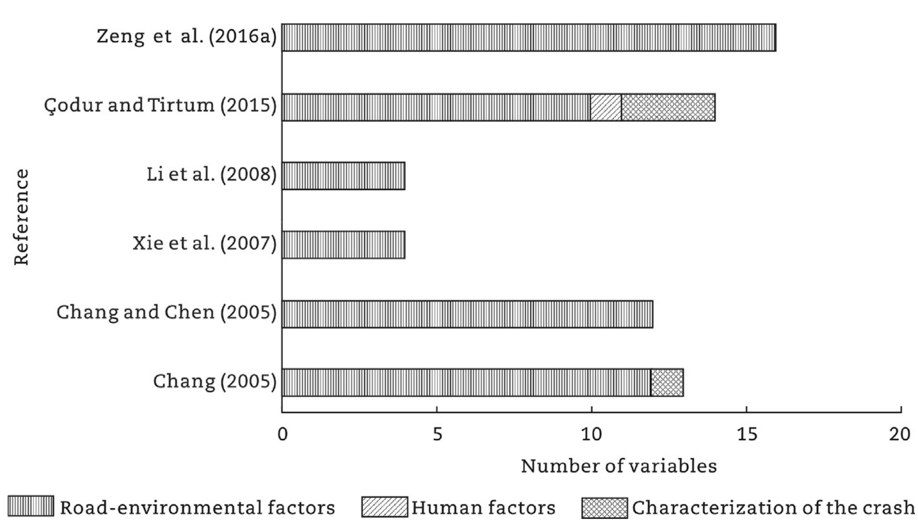


Fig. 5 – Group of variables used for modeling crash frequency.

Table 2 – Main explanatory variables used in the models.

Model	Road-environmental variable	Human variable	Vehicular variable	Crash characterization variable
Severity model	Weather conditions; lightning; roadway segment; type of shoulder	Sex; age; seat belt use	Vehicle type; vehicle age	Crash type; time; day of the week
Frequency model	Traffic volume; segment length; horizontal alignment; shoulder width; roadway segment	Sex		Year; season; vehicles involved in crash
Frequency and severity models	Traffic volume; shoulder width; speed limit; road width; pavement conditions; segment length			Time; day of the week

the F-measure is a balanced combination of the precision and sensitivity metrics (Sun et al., 2003).

The receiver operating characteristics (ROC) curve can also be used to evaluate the performance of models. It expresses the relationship between the sensitivity and specificity metrics by providing an aggregate performance metric in all possible classification boundaries. ROC analysis is extensively used in machine learning and data mining techniques. Geometrically, it is a probability curve displaying pairs of values of the FP and the TP rates of a model considered to be a good fit. The area under the ROC curve (AUC) represents the separability degree, i.e., it indicates the model's ability to distinguish between classes. The AUC ranges from 0 to 1, in which the null value represents a model with 100% misclassifications and AUC = 1 corresponds to a model with totally correct classifications (Fawcett, 2006).

Table 3 – Main contributing factors for crash occurrence or severity.

Reference	Main contributing factor
Crash severity prediction model	
Abdel-Aty and Abdelwahab (2004)	Sex; posted speed limit; seat belt use; type of vehicle; point of impact; land use
Delen et al. (2006)	Seat belt use; rollover; sex; age; use of alcohol or drugs; type of vehicle
Chang and Wang (2006)	Type of collision; contributing circumstances; vehicle/driver action
Das and Abdel-Aty (2010)	Presence of parking area; posted speed limit; percentage of heavy vehicles; median barrier; traffic volume
Oña et al. (2011)	Type of crash; age; lightning; number of injuries
Kashani and Mohaymany (2011)	Seat belt use; cause of crash; pavement surface; weather conditions
Oña et al. (2013a)	Lightning; type of crash; sex; weather conditions; cause of crash; time
Oña et al. (2013b)	Vehicles involved; number of injuries; weather conditions; horizontal signaling; roadway width
Zeng and Huang (2014)	Sex; age; safety devices (e.g., seat belt); age of vehicle; percentage of heavy vehicles; point of impact
Kwon et al. (2015)	Type of collision; type of traffic violation; movement prior to collision; type of highway
Chen et al. (2016)	Seat belt use; pavement surface; weather conditions; maximum damage to vehicle; alcohol or drugs use; age; number of lanes; demographic characteristics of driver
Zhang et al. (2018)	Lighting; collision of type "sideswipe"; road surface type; weather
Wahab and Jiang (2019)	Location type; time of the crash; collision partner; collision type; road separation; road surface type; day of the week; road shoulder
Amiri et al. (2020)	Light condition; existence of the right and left shoulders; cause of collision; average annual daily traffic (AADT); number of involved vehicles; age; road surface condition; gender
Crash frequency prediction model	
Chang (2005)	Segment in military area; existence of intersections; percentage of heavy vehicles; number of lanes; traffic volume
Chang and Chen (2005)	Traffic volume; precipitation; percentage of heavy vehicles; horizontal curvature
Xie et al. (2007)	Segment length; volume of traffic; lane width
Li et al. (2008)	Traffic volume; shoulder width
Çodur and Tortum (2015)	Vertical curvature; traffic volume; horizontal curvature; segment length
Zeng et al. (2016a)	Traffic volume; posted speed limit; annual precipitation; segment length; median barrier; bus stop
Crash frequency and severity prediction model	
Das and Abdel-Aty (2011)	Traffic volume; posted speed limit; roadway width; skid resistance
Zeng et al. (2016b)	Traffic volume; segment length; posted speed limit; bus stop; annual precipitation

Table 4 – Performance metrics for crash severity prediction models.

Reference	Performance metric	
Sohn and Lee (2003)	Maximum accuracy	i) DT: 72.30%; ii) ANN: 70.86%; iii) DT (cluster): 76.10%; iv) ANN (cluster): 73.94%
Abdel-Aty and Abdelwahab (2004)	Maximum accuracy	i) ANN MLP: 73.5%; ii) ordered probitmodel: 61.7%
Chang and Wang (2006)	Maximum accuracy	CART: 96.4%
Delen et al. (2006)	Maximum accuracy	i) ANN with five class: 53.78%; ii) ANN with two class: 99.52%
Das and Abdel-Aty (2010)	Maximum accuracy	GP1–angle/turning movements: 83.58%; GP2–rear-end: 84.53%; GP3–head-on: 91.48%
Oña et al. (2011)	Maximum accuracy	BN: 59%
Kashani and Mohaymany (2011)	Average accuracy	i) CART with three class: 34.06%; ii) CART with two class: 60.94%
Oña et al. (2013a)	Maximum accuracy	i) CART: 56%; ii) C4.5: 54%; iii) ID3: 53%
Oña et al. (2013b)	Maximum accuracy	i) Full dataset: 60%; ii) cluster 1: 64%; iii) cluster 2: 58%; iv) cluster 3: 59%
Alikhani et al. (2013)	Maximum accuracy	i) ANFIS: 76.11%; ii) ANN: 77.26%; iii) ANFIS (*k*-means): 77.76%; iv) ANN (*k*-means): 79.53%; v) ANFIS (self-organizing maps): 77.47%; vi) RNA (self-organizing maps): 81.81%
Zeng and Huang (2014)	Maximum accuracy	i) ANN: 54.84%; ii) optimized ANN: 54.91%; iii) ordered logit model: 51.78%
Kwon et al. (2015)	ROC curve	DT is better than other two models (logistic regression and naive Bayes classifier)
Chen et al. (2016)	Maximum accuracy	(1) For SVM with Gaussian function, i) three class: 45.76%; ii) two class: 53.92%; (2) for SVM with polynomial function, i) three class: 50.91%; ii) two class: 62.63%
Iranitalab and Khattak (2017)	Average accuracy	i) KNN: 44% ii) SVM: 26% iii) RF: 26%; iv) multinomial logit model: 25%
Zhang et al. (2018)	Average accuracy	i) KNN: 77.6%; ii) DT: 79.8%; iii) RF: 78.6%; iv) SVM: 58.1%; v) ordered probit model: 44.7%; vi) multinomial logit model: 51.4%
Wahab and Jiang (2019)	Average accuracy	i) DT: 73.64%; ii) RF: 73.91%; iii) KNN: 73.71%; iv) multinomial logit model: 52.04%
Amiri et al. (2020)	Average accuracy	i) ANN: 22.18%; ii) HIGA: 21.66%

Almost all severity models used accuracy as a performance metric. An exception was the work of Kwon et al. (2015). This metric will be used as the parameter for comparing the different models by observing the maximum or average accuracy of the validation. Table 4 demonstrates the results.

For regression models, the performance metrics generally compute the difference between predicted (expected) and observed (real) values. The main metrics used in regression models are the mean absolute deviation (MAD), the mean squared error (MSE), and the root mean square error (RMSE). The MAD evaluates the prediction error after calculating the mean absolute error. It does not consider the error direction and the deviations are equally weighted. The MSE is similar to the MAD, but it is more sensitive to greater errors because the deviations are squared. Finally, the RMSE can be understood as the residuals' standard deviation. It indicates how dispersed are the data compared to the model, or how close the data points are to a fitted line. Their equations are as follows.

$$\mathrm{MAD} = \frac{\sum_{i=1}^{n} \left| y_i - \widehat{y}_i \right|}{n} \tag{6}$$

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y}_i \right)^2}{n} \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{8}$$

where $\widehat{y}_i$ and y_i are the predicted and observed values, respectively, and n is the size of the set of training or test. Table 5 summarizes the results of the studies that developed crash prediction models.

Finally, for models that joint predicted crash frequency and severity, Table 6 demonstrates the results.

Table 5 – Performance metrics for crash prediction models.

Reference	Performance metric	
Chang (2005)	Maximum accuracy	i) ANN: 61.4%; ii) regression model: 60.8%
Chang and Chen (2005)	Average accuracy	i) CART: 52.6%; ii) negative Binomial model: 52.3%
Xie et al. (2007)	Minimum error value	i) For ANN, a) MAD: 0.84; b) MSE: 0.83; ii) for BNN, a) MAD: 0.73; b) MSE: 0.77; iii) for negative binomial model, a) MAD: 0.80; b) MSE: 0.87
Li et al. (2008)	Minimum error value	i) For SVM, a) MAD: 0.73; b) MSE: 0.75; ii) for negative binomial model, a) MAD: 0.80; b) MSE: 0.87
Çodur and Tortum (2015)	Minimum error value	For ANN, MSE: 4.11; RMSE: 2.02; R^2: 0.98
Zeng et al. (2016a)	Average MAD value	i) ANN: 3.57; ii) optimized ANN: 3.43; iii) statistical model: 3.7

Table 6 – Performance measurements for crash frequency and severity prediction models.

Reference		Performance metric
Das and Abdel-Aty (2011)	Severity: maximum accuracy	i) Cluster 1: 91.85%; ii) cluster 3: 91.48%; iii) cluster 4: 89.43%
	Frequency: minimum MSE	10.264
Zeng et al. (2016b)	Minimum error value	For slight injury, i) ANN-MAD: 2.76; ii) optimized ANN-MAD: 2.76; iii) Poisson-lognormal model-MAD: 2.82
		For serious or fatal injury, i) ANN-MAD: 1.07; ii) optimized ANN-MAD: 1.03; iii) Poisson-lognormal model-MAD: 1.0
Dong et al. (2018)	Average MAD value	i) SVM: 0.26; ii) DL: 0.06

4. Conclusions

This paper's objective was to conduct a systematic review of the main papers addressing road safety modeling using ML techniques and obtain a broader and more detailed perspective of several aspects. The main conclusions are listed in the following subsections.

4.1. *Methodological approaches*

The major ML techniques used in crash modeling can be grouped as follows: nearest neighbor classification, decision trees, genetic programming, support-vector machines, and artificial neural networks. Several studies found that the ML techniques improved the models' performances in comparison to statistical models. As these studies used different forms of measuring the error, a general comparison of each technique's performance was not possible. However, considering the aspect of applicability and the results obtained, it seems that ANN are identified as the most appropriate technique for modeling crash frequency. ANN are also useful for crash analysis by severity, and encouraging results were also obtained with the use of CART analysis. Many algorithm proposals (e.g., training and network structure optimization, sensitivity analysis, and rule set extraction) confer greater potential to the use of ANN. They seek to eliminate model over-fitting and also reveal patterns that exist among the explanatory variables and the outputs (in an attempt to open the “black box”), which are the most criticized aspects of the ANN.

Cluster analysis associated to a ML technique for modeling has shown promising results (Alikhani et al., 2013; Das and Abdel-Aty, 2011; Iranitalab and Khattak, 2017; Oña et al., 2013b; Sohn and Lee, 2003). In addition, Abdel-Aty and Abdelwahab (2004) and Alikhani et al. (2013) demonstrated that the use of fuzzy logic does not seem to bring improvements to the models.

Deep learning has also been employed for road safety modeling in the past years. Dong et al. (2018) proposed a hybrid model with unsupervised and supervised steps, in addition to adding a layer to deal with the network's unobserved heterogeneity issues. As a result, three crash prediction response variables were modeled simultaneously. The performance metrics using deep learning were lower than a quarter of the metrics using SVM. Therefore, this application shows the potential of using deep learning in road safety.

4.2. *Explanatory variables*

The main explanatory variables used in crash modeling by severity consist of road-environmental factors, human factors, crash characteristics, and vehicle-related factors, in descending order of importance. However, Das and Abdel-Aty (2010), Das and Abdel-Aty (2011), and Iranitalab and Khattak (2017) only considered road-environmental factors. Furthermore, those authors underscored that there is no need to divide the roadway into segments for modeling for classification purposes, which means that these models can be based on a greater amount of data (i.e., each crash occurrence is a single observation for classification, whereas crash occurrences must be grouped for modeling frequency) and also lead to an improved generalization capacity.

There is a clear predominance of road-environmental factors as input variables in crash frequency modeling studies. The four most important variables (traffic volume, segment length, horizontal alignment, and posted speed limit) belong to that group. Also, many studies divided the highway into segments, either homogeneously (with fixed geometrical and operational characteristics) or of fixed-length. In addition, the exposure variables (i.e., annual average daily traffic and segment length, in case of homogeneous segments) are covariates in all the frequency models, as expected. Other studies that developed crash prediction models by severity, such as Zeng et al. (2016b) and Dong et al. (2018), also segmented the highway and included roadway-environmental factors as their explanatory variables.

The models' investigation allowed the establishment of degrees of importance of each variable because of their frequent use in crash frequency prediction models. However, the choice of a group of variables may have been based on prior data analyses, on convenience, or on data limitations. However, this explanatory variables' analysis serves as a guide for new works related to road safety, which could even investigate the potential of variables that have been ignored or hardly explored at all.

Three crash frequency models investigated were based on data of multilane highways in rural environments: Chang (2005), Chang and Chen (2005), and Çodur and Tortum (2015). Other models included multi-lane highways, but associated with other kinds of roadway, such as urban arterial roads. In addition, some authors failed to specify the type of roadway for which they undertook the study. The joint frequency-severity approach in Zeng et al. (2016b) and Dong et al. (2018) was developed with various types of roadways. Thus, the proposal of the present paper is useful as it helps to fill a gap in crash frequency models by severity in multilane highways using ML techniques.

4.3. *Final remarks*

The promising results from ML techniques in crash prediction models, even comparable to traditional statistical modeling, led to an increase in the exploration – and publication – of ML applied to road safety analysis studies. This paper shows the progress achieved and the opportunities for further investigation. It is, therefore, expected that this paper encourages and provides a general overview for researchers interested in studying this area.

Conflict of interest

The authors do not have any conflict of interest with other entities or researchers.

Acknowledgments

The first author wishes to thank the Instituto Federal Goiano (IFGoiano) (Goiano Federal Institute) for the financial support it provided. The study also received support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brazil (CAPES) – Financing Code 001 (Coordination of Improvement of Higher Education Personnel) and the Fundação para a Ciência and Tecnologia-Portugal-(FCT) (Science and Technology Foundation) under the project "Mobilidade Urbana SustentáveleSegura" (Safe and Sustainable Urban Mobility) of which this research is a part.

REFERENCES

Abdel-Aty, M.A., Abdelwahab, H.T., 2004. Predicting injury severity levels in traffic crashes: a modelling comparison. Journal of Transportation Engineering 130 (2), 204–210.
Abdel-Aty, M.A., Radwan, A.E., 2000. Modelling traffic crash occurrence and involvement. Accident Analysis & Prevention 32 (5), 633–642.
Abdelwahab, H.T., Abdel-Aty, M.A., 2001. Development of artificial neural network models to predict driver injury severity in traffic crashes at signalized intersections. Transportation Research Record 1746, 6–13.
Abellán, J., López, G., Oña, J., 2013. Analysis of traffic crash severity using decision rules via decision trees. Expert Systems with Applications 40 (15), 6047–6054.
Alikhani, M., Nedaie, A., Ahmadvand, A., 2013. Presentation of clustering-classification heuristic method for improvement accuracy in classification of severity of road crashes in Iran. Safety Science 60, 142–150.
Amiri, A.M., Sadri, A., Nadimi, N., et al., 2020. A comparison between artificial neural network and hybrid intelligent genetic algorithm in predicting the severity of fixed object crashes among elderly drivers. Accident Analysis & Prevention 138, 1–10.
Basgalupp, M.P., Carvalho, A.C.P.L.F., Barros, R.C., et al., 2009. Lexicographic multi-objective evolutionary induction of decision trees. International Journal of Bio-Inspired Computation 1 (1), 105–117.
Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2, 121–167.
Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-crash frequencies and severities. Accident Analysis & Prevention 33 (1), 99–109.
Chang, L., 2005. Analysis of freeway crash frequencies: negative binomial regression versus artificial neural network. Safety Science 43, 541–557.
Chang, L., Chen, W., 2005. Data mining of tree-based models to analyze freeway crash frequency. Journal of Safety Research 36 (4), 365–375.
Chang, L., Wang, H., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis & Prevention 38, 1019–1027.
Chapelle, O., Schölkopf, B., Zien, A., 2006. Semi-supervised Learning. MIT Press, Cambridge.
Chen, C., Zhang, G., Qian, Z., et al., 2016. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accident Analysis & Prevention 90, 128–139.
Çodur, M.Y., Tortum, A., 2015. An artificial neural network model for highway crash prediction: a case study of Erzurum, Turkey. PROMET-Traffic & Transportation 27 (3), 217–225.
Costa, J.O., Jacques, M.A.P., Soares, F.E.C., et al., 2016. Integration of geometric consistency contributory factors in three-leg junctions collision prediction models of Portuguese two-lane national highways. Accident Analysis & Prevention 86, 59–67.
Das, A., Abdel-Aty, M., 2010. A genetic programming approach to explore the crash severity on multi-lane roads. Accident Analysis & Prevention 42 (2), 548–557.
Das, A., Abdel-Aty, M., 2011. A combined frequency-severity approach for the analysis of rear-end crashes on urban arterials. Safety Science 49 (8–9), 1156–1163.
Decker, K.M., Focardi, S., 1995. Technology Overview: a Report on Data Mining. Technical Report CSCS TR-95-02. Swiss Scientific Computing Center, Bern.
Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic crashes using a series of artificial neural networks. Accident Analysis & Prevention 38 (3), 434–444.
Deng, L., Yu, D., 2014. Deep learning: methods and applications. Foundations and Trends® in Signal Processing 7 (3–4), 197–387.
Dong, C., Shao, C., Li, J., et al., 2018. An improved deep learning model for traffic crash prediction. Journal of Advanced Transportation 2018, 1–13.
Devroye, L., Gyorfi, L., Krzyzak, A., et al., 1994. On the strong universal consistency of nearest neighbor regression function estimates. Annals of Statistics 22, 1371–1385.
Elvik, R., Vaa, T., Høye, A., et al., 2009. The Handbook of Road Safety Measures. Emerald Group Publishing Limited, Bingley.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27 (8), 861–874.
Floreano, D., Mattiussi, C., 2008. Bio-inspired Artificial Intelligence: Theories, Methods, and Technologies. MIT Press, Cambridge.
Hauer, E., 2004. Statistical road safety modelling. Transportation Research Record 1897, 81–87.
Hauer, E., 2015. The Art of Regression Modelling in Road Safety. Springer, Cham.
Haykin, S., 2009. Neural Networks and Learning Machines, third ed. Prentice Hall, New York.
Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accident Analysis & Prevention 108, 27–36.

Kashani, A.T., Mohaymany, A.S., 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Safety Science 49, 1314–1320.

Kidando, E., Moses, R., Ozguzen, E.E., et al., 2019. Incorporating travel time reliability in predicting the likelihood of severe crashes on arterial highways using non-parametric random-effect regression. Journal of Traffic and Transportation Engineering (English Edition) 6 (5), 470–481.

Kim, D., Washington, S., 2006. The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. Accident Analysis & Prevention 38, 1094–1100.

Kitchenham, B., Charters, S., 2007. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01. School of Computer Science and Mathematics, Keele University, Keele.

Kwon, O.H., Rhee, W., Yoon, Y., 2015. Application of classification algorithms for analysis of road safety risk factor dependencies. Accident Analysis & Prevention 75, 1–15.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lee, C., Park, S., 2011. Damage classification of pipelines underwater flow operation using multi-modeactuated sensing technology. Smart Materials and Structures 20 (11), 115002.

Li, H., Graham, D.J., Majumdar, A., 2012. The effects of congestion charging on road traffic casualties: a causal analysis using difference-in-difference estimation. Accident Analysis & Prevention 49, 366–377.

Li, X., Lord, D., Zhang, Y., et al., 2008. Predicting motor vehicle crashes using support vector machine models. Accident Analysis & Prevention 40 (4), 1611–1618.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44 (5), 291–305.

Miaou, S., Lum, H., 1993. Modelling vehicle crashes and highway geometric design relationships. Accident Analysis & Prevention 25 (6), 689–709.

Mussone, L., Ferrari, A., Oneta, M., 1999. An analysis of urban collisions using an artificial intelligence model. Accident Analysis & Prevention 31 (6), 705–718.

Nodari, C.T., Lindau, L.A., 2007. Proactive method for safety evaluation of two-lane rural highway segments. Advances in Transportation Studies 11, 51–61.

Oña, J., López, G., Abellán, J., 2013a. Extracting decision rules from police crash reports through decision trees. Accident Analysis & Prevention 50, 1151–1160.

Oña, J., López, G., Mujalli, R.O., et al., 2013b. Analysis of traffic acidentes on rural highways using latent class clustering and Bayesian networks. Accident Analysis & Prevention 51, 1–10.

Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic crash injury severity on Spanish rural highways using Bayesian networks. Accident Analysis & Prevention 43 (1), 402–411.

Quinlan, J.R., 1986. Induction of decision trees. Machine Learning 1, 81–106.

Rolison, J.J., Regev, S., Moutari, S., et al., 2018. What are the factors that contribute to road crashes? An assessment of lawenforcement views, ordinary drivers' opinions, and road crash records. Accident Analysis & Prevention 115, 11–24.

Savolainen, P., Mannering, F., Lord, D., et al., 2011. The statistical analysis of crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis & Prevention 43 (5), 1666–1676.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Networks 61, 85–117.

Scholkopf, B., Smola, A.J., 2002. Learning with Kernels. MIT Press, Cambridge.

Smola, A.J., Scholkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing 14, 199–222.

Sohn, S., Lee, S., 2003. Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic crash in Korea. Safety Science 41 (1), 1–14.

Sun, A., Lim, E.-P., Ng, W.-K., 2003. Performance measurement framework for hierarchical text classification. Journal of the American Society for Information Science and Technology 54 (11), 1014–1028.

Tan, P.N., Steinbach, M., Kumar, V., 2005. Introduction to Data Mining, first ed. Addison-Wesley Longman Publishing Co., Boston.

Trabelsi, A., Elouedi, Z., Lefevre, E., 2019. Decision tree classifiers for evidential attribute values and class labels. Fuzzy Sets and Systems 366, 46–62.

Trafalis, B.T., Gilbert, R.C., 2006. Robust classification and regression using support vector machines. European Journal of Operational Research 173 (3), 893–909.

Üstün, B., Melssena, W.J., Oudenhuijzenb, M., et al., 2005. Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. Analytica Chimica Acta 544 (1–2), 292–305.

Villiers, J., Barnard, E., 1993. Back propagation neural nets with one and two hidden layers. IEEE Transactions on Neural Networks 4 (1), 136–141.

Wahab, L., Jiang, H., 2019. A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. PLoS One 14 (4), 1–17.

Wang, C., Quddus, M.A., Ison, S.G., 2011. Predicting crash frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accident Analysis & Prevention 43 (6), 1979–1990.

Wang, C., Quddus, M.A., Ison, S.G., 2013. The effect of traffic and road characteristics on road safety: a review and future research direction. Safety Science 57, 264–275.

Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural networks: an empirical analysis. Accident Analysis & Prevention 39 (5), 922–933.

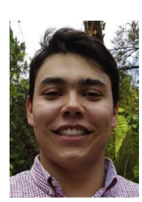
Yu, X., Gen, M., 2010. Introduction to Evolutionary Algorithms. Springer, London.

Zeng, Q., Huang, H., 2014. A stable and optimized neural network model for crash injury severity prediction. Accident Analysis & Prevention 73, 351–358.

Zeng, Q., Huang, H., Pei, X., et al., 2016a. Rule extraction from an optimized neural network for traffic crash frequency modelling. Accident Analysis & Prevention 97, 87–95.

Zeng, Q., Huang, H., Pei, X., et al., 2016b. Modelling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. Analytic Methods in Crash Research 10, 12–25.

Zhang, J., Li, Z., Pu, Z., et al., 2018. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEE Access 6, 60079–60087.



Philippe Barbosa Silva holds a PhD in transportation from the University of Brasilia and is an assistant professor in the Department of Civil Engineering and head of Pavement and Transportation Laboratory at the Goiano Federal Institute-Rio Verde Campus. Silva's research interest is road safety, especially highway road safety modelling.

Michelle Andrade has been an adjunct professor at the postgraduate program in transportation of the Faculty of Technology, University of Brasilia (UnB), since 2013. She is a civil engineer and teaches highway design and road safety at UnB. She heads the research group on road safety, where academic and technical professionals discuss road safety challenges to the society, methods and results of the academic and technical studies.

Sara Ferreira holds a PhD in civil engineering from the Faculty of Engineering of University of Porto and is an assistant professor in University of Porto. The doctoral research focused on urban road network safety model at the transportation planning process. Her main fields of expertise, with several published papers in international peer-reviewed journals, relate to road safety. She coordinated an EU project and she is now coordinating two national projects and is involved in an H2020 project, all focused on traffic safety. She is thematic editor for the Revista Transportes from Brasil and member of the international association ICTCT.