Accident Analysis & Prevention

Special Issue: "Crash Precursors and Surrogate Measures of Safety: Measurement Technology, Theory and Applications"

A conflict-based approach for real-time road safety analysis: comparative evaluation with crash-based models

Federico Orsini [a, *], Gregorio Gecchele [b], Riccardo Rossi [a], Massimiliano Gastaldi [a, c]

[a] Department of Civil Environmental and Architectural Engineering, University of Padua, Via Marzolo 9, 35131 Padua, Italy.

[b] Atraki s.r.l., via Diaz 4, 37015 S. Ambrogio di Valpolicella (Verona), Italy.

[c] Department of General Psychology, University of Padua, Via Venezia 8, 35131 Padua Italy.

* corresponding author. E-mail: federico.orsini@dicea.unipd.it

To cite this article:

Orsini, F., Gecchele, G., Rossi, R., & Gastaldi, M. (2021). A conflict-based approach for real-time road safety analysis: Comparative evaluation with crash-based models. *Accident Analysis & Prevention*, *161*, 106382.

Link to the article:

https://doi.org/10.1016/j.aap.2021.106382

Abstract

An innovative approach for real-time road safety analysis is presented in this work. Unlike traditional real-time crash prediction models (RTCPMs), in which crash data are used in the training phase, a real-time conflict prediction model (RTConfPM) is proposed. This model can be trained using surrogate measures of safety, and can therefore be applied even in situations in which highly spatial/temporal-accurate crash data are unavailable or unreliable. The application of an RTConfPM consists of using a set of input variables recorded during a given time interval, to predict whether there will be an increased risk of unsafe situations in the following interval.

This paper presents an RTConfPM to predict rear-end crashes, using time-to-collision values recorded with radar sensors on multiple motorway cross-sections to define unsafe situations, and traffic conditions recorded on the same sections as input to the model. The RTConfPM is compared to a traditional RTCPM, trained with a dataset of crashes located on the same motorway, and using the same traffic data as input. In both approaches, variable selection is performed with Pearson's correlation test and random forest; synthetic minority oversampling technique (SMOTE) is used to balance the classes in the training dataset, support vector machine (SVM) is used as classifier, and Monte Carlo cross-validation is adopted for robustness. The two approaches are evaluated considering accuracy, recall, specificity/false alarm rate, and area under the curve (AUC).

As shown by the results of this paper, the conflict-based approach appears promising, and is able to predict the occurrence of unsafe situations within 5 minutes with more than 93% accuracy, recall and specificity, significantly outperforming the RTCPM.

Keywords: real-time conflict prediction model; real-time crash prediction; traffic conflicts; road safety; machine learning; rear-end crashes.

1. Introduction

In recent years, several studies have dealt with real-time crash prediction models (RTCPMs), the concept of which relates to the hypothesis that "the probability of a crash occurring on a specific road section within a very short time window can be predicted using the instantaneous traffic dynamics" (Hossain et al., 2019). In these models, crash data are fundamental, as they are used to calibrate or train the model. However, several issues may affect crash data: lack of availability, lack of spatial and/or temporal precision, under-reporting, misclassification. Moreover, crashes are relatively rare events, so data must be collected for several years and/or in several different locations to obtain enough data to apply certain statistical models (Tarko, 2018a). The reliability of an RTCPM is strongly dependent on the reliability of the crash data; in practice, it is not possible to develop and apply an RTCPM if crash data are unavailable or unreliable.

To avoid the use of crash data, in road safety analysis there is an alternative approach, which consists in analyzing conflicts. The use of the term "traffic conflict" to describe a near-crash situation dates back to the 1960s and 70s (Hayward, 1972; Perkins and Harris, 1968). A formal definition was proposed by the Association for International Cooperation on Traffic Conflict Techniques: "A traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged" (Güttinger, 1984). As underlined by Davis et al. (2011), this definition implies that an event can be considered a conflict only if it satisfies a counterfactual analysis:[1] if the users had not modified their movement, then a collision probably would have occurred. Tarko (2019) further elaborated this concept, explicitly affirming that conflicts are "precursors of crashes and not alternative outcomes." This is visually represented in Figure 1: a set of varied causes may induce a traffic conflict between road users; if there is no intervention, this results in a

[1] Counterfactual analysis can be defined as "a comparison between what actually happened and what would have happened in the absence of the intervention" (White, 2006).

crash, whereas an effective intervention can prevent it. The presence of some kind of system "failure" (e.g., high traffic volume, driver distraction, etc.) is a necessary condition for a conflict to happen, which in turn is necessary for a collision to happen.

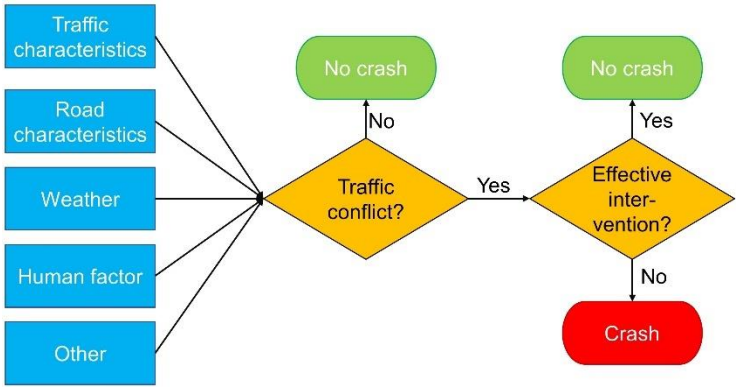

Figure 1. Causality model of traffic conflicts and crashes (adapted from (Tarko, 2019)*).*

Although conflict analysis is well established in road safety, it has not yet found its way into real-time road safety prediction models, except for very few works, which are discussed in Section 2.2. To fill this gap, in this paper, we propose a methodology to predict conflict-prone situations based on real-time traffic data.

In practical terms, a real-time conflict prediction model (RTConfPM) is structured similarly to an RTCPM. The flowchart in Figure 2 shows its main features:

- The model is developed offline using historical traffic, weather, and road data as input and historical crash/conflict data for calibration.
- The model is applied in real time using traffic, weather, and road data collected at the present interval to predict whether an unsafe situation will exist in the next future.
- If an unsafe situation is predicted, then an intervention strategy is applied to prevent it. Several solutions can be applied, such as variable message signs or ramp metering (Pande and Abdel-Aty, 2005).

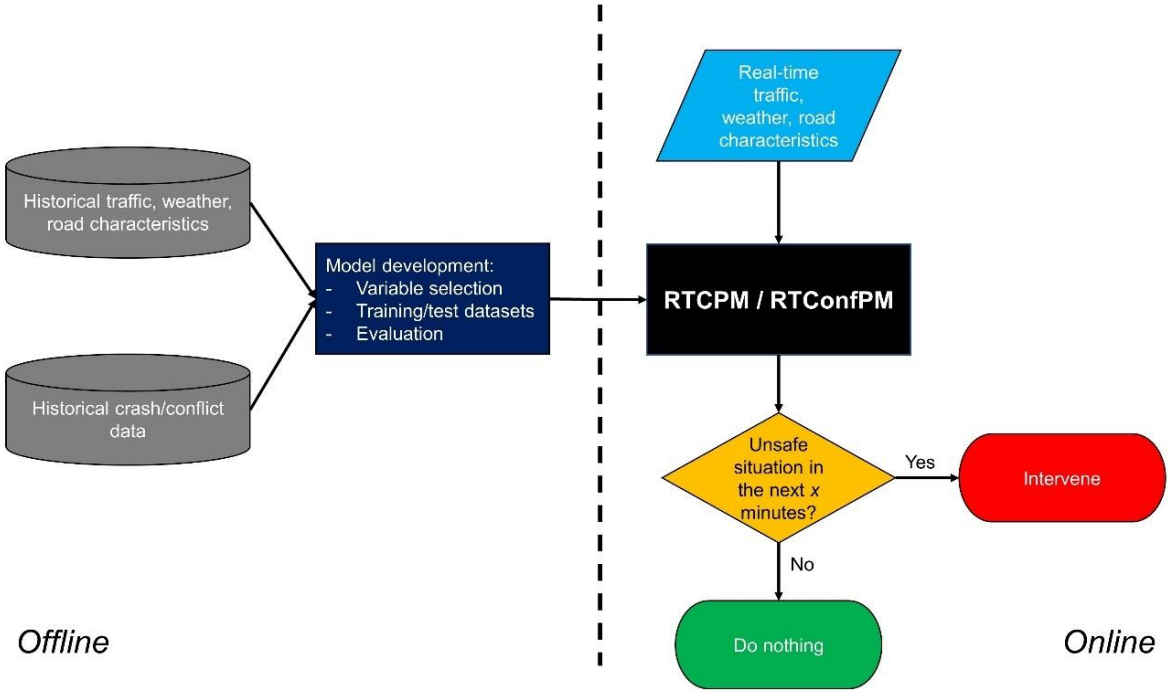

Figure 2. Flowchart representing the offline development of an RTCPM/RTConfPM (left) and the real-time application of an RTCPM/RTConfPM.

A major advantage of an RTConfPM is that it can be developed even in absence of highly spatial/temporal-accurate crash data,[2] which can be a huge benefit in practical applications, since inaccuracies in crash location and time are among the most serious quality issues of crash data (Imprialou and Quddus, 2019). In this paper, we further suggest that this approach can identify real-time risky conditions more reliably than traditional RTCPMs. This is discussed in quantitative terms in the next sections of the paper, but is also intuitively illustrated in Figure 3. Let us consider a first situation, in which a set of traffic characteristics induced the occurrence of several traffic conflicts within a certain time window, but no actual collision ensued. Let us now consider a second situation, in which the same set of traffic characteristics produced the same amount of traffic conflicts, and, due to a road user's ineffective intervention, a crash ensued. With the RTConfPM that set of traffic characteristics induced an unsafe situation in both cases; with the

[2] However, crash data are still needed, at least in some aggregate form, to initially validate the definition of conflicts and unsafe situations used in the RTConfPM. This is further discussed in Appendix A.

RTCPM, the same set of traffic conditions was considered, in one case, a crash precursor, and in the other case not, although the only difference between the two situations was the effectiveness of a road user's intervention, which is independent from the input data of the model (i.e., traffic characteristics). This inconsistency can have negative consequences on the effectiveness of model training.

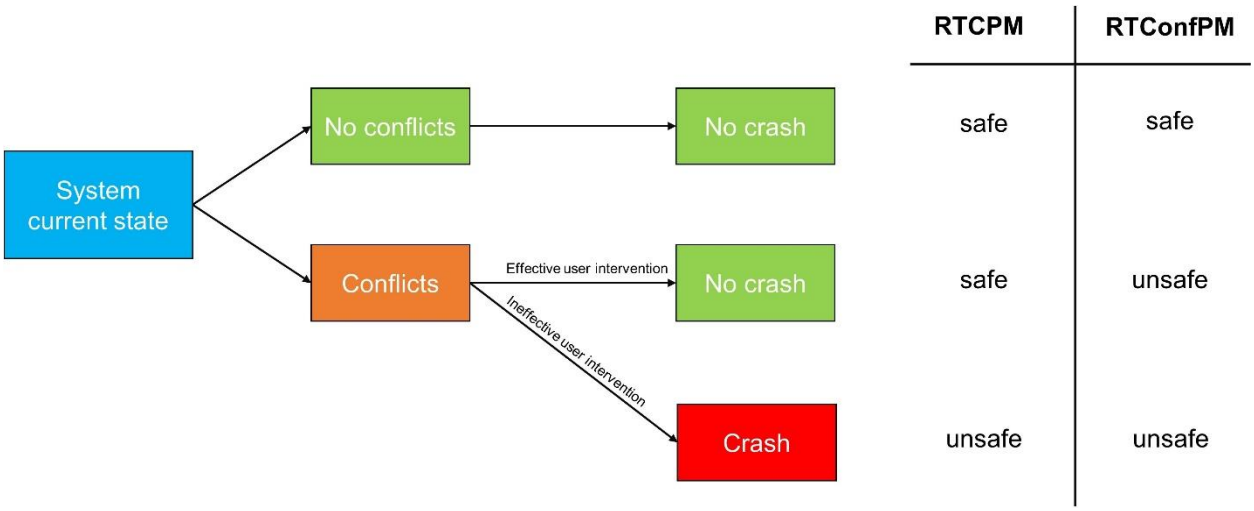

Figure 3. Safe and unsafe situations for RTCPMs and RTConfPMs.

Another advantage of dealing with near-crashes is that they are much more frequent than crash events (see the classic "safety pyramid" introduced by Hydén, 1987), resulting in shorter observation times to collect enough data to train an RTConfPM. Moreover, recent developments in sensing and ITS technologies significantly improved the availability and accuracy of traffic measures (Sadeq and Sayed, 2016; Tarko, 2012), especially on highways (Eltrass and Khalil, 2018; St-Aubin et al., 2015).

Starting from the considerations discussed above, here we aim to develop an RTConfPM to predict rear-end crashes on motorways.

The first main novelty introduced by this work is that the results of the proposed conflict-based approach are compared to those of a traditional crash-based model calibrated on the same dataset. To the best of

the authors' knowledge, this represents the first direct comparison between the two approaches in real-time road safety applications.

The second main novelty is that cross-sectional vehicle-by-vehicle data were used both to retrieve input traffic variables and to detect traffic conflicts. This was done by means of microwave radar sensors, which can produce reliable information with relatively low data collection and elaboration costs, and has great potential in view of future large-scale practical implementations.

Finally, as discussed in Section 7, this work can be viewed as a starting point for several research developments.

2. Literature review

The rationale behind an RTConfPM is not dissimilar to that of an RTCPM: given the system current state (composed by a set of road, traffic, weather characteristics, etc.), an RTConfPM aims to predict the occurrence of an unsafe situation in the immediate future. The only difference, as discussed in Section 1, consists of how unsafe situations are defined in the two approaches. Therefore, methodologically, an RTConfPMs adopts the same tools as RTCPMs do to identify relationships between the system current state and unsafe situations and, therefore, the choice of which tools to use cannot be made without an appropriate review of the methods employed in RTCPMs.

2.1 Real-time crash prediction models

The first works to deal with real-time assessment of road safety date back to the late 90s/early 00s (Abdel-Aty et al., 2004; Hughes and Council, 1999; Lee et al., 2002; Madanat and Liu, 1995). Since then, several RTCPMs have been developed. Recently, a systematic review by Hossain et al. (2019) identified and analyzed 78 papers published between 2003 and 2018, dealing with real-time prediction of crashes.

The main objective of RTCPMs is, of course, to predict crashes; some studies also considered crash severity (Park et al., 2018; Xu et al., 2013a) and/or crash type (Wu et al., 2018a; Yu and Abdel-Aty, 2013a). Other papers dealt with crash mechanisms (Yu et al., 2013) or intervention strategies (Abdel-Aty and Wang, 2017; Pande and Abdel-Aty, 2005).

The source of traffic data is, more frequently, loop detectors (Dimitriou et al., 2018; Roy et al., 2018); some studies employed microwave vehicle detection systems (Wang et al., 2017; Yasmin et al., 2018), automated vehicle identification (Ahmed et al., 2012; Basso et al., 2018), Bluetooth detectors (Yuan et al., 2018; Yuan and Abdel-Aty, 2018), remote traffic microwave stations (Wu et al., 2018a; Yu et al., 2014), probe vehicles (Park and Haghani, 2016), or combinations of these sources.

The main input in RTCPMs consists of traffic flow variables, typically a subset of average, standard deviation, and other statistics or transformations, of speed, flow and occupancy; in addition to these, it has been shown that considering traffic flow composition can improve model performance (Basso et al., 2020). Data are aggregated at different temporal scales, depending mainly on the available raw data, usually in the range of 1 to 5 minutes, with the vast majority of studies using 5-minute aggregation (Hossain et al., 2019). Moreover, several other non-traffic variables have been used as input in RTCPMs, including: geometrical features of the road, such as the presence of curves (Ahmed et al., 2012) and pavement condition (Wang et al., 2017); weather conditions, such as the presence of fog (Wu et al.,

2018b) and lighting conditions (Yasmin et al., 2018); and other variables, such as the presence of neighborhoods with younger populations (Paikari et al., 2014).

This large amount of variables often produces a large variable space, compared to the usually relatively small sample size of crashes, leading to the need to apply variable selection methods, to discard the less meaningful ones. The recent trend is to apply artificial intelligence and data mining techniques, such as classification trees (Pande and Abdel-Aty, 2006), random forest (Abdel-Aty and Haleem, 2011), or random multinomial logit (Hossain and Muromachi, 2012).

Modeling methods used to develop RTCPMs are generally divided into statistical and machine learning approaches. Among the former, multinomial logistic regressions (Abdel-Aty et al., 2004) and Bayesian logistic regression (Ahmed et al., 2012) have been widely applied. Among the latter, several classification methods have been used: support vector machine (SVM - Qu et al., 2013; Yu and Abdel-Aty, 2013b), classification trees/random forest (Hossain and Muromachi, 2011), k-Nearest Neighbor (kNN - Lv et al., 2009a), Bayesian networks (Hossain and Muromachi, 2012), and various forms of neural networks (Pande and Abdel-Aty, 2006) and deep neural networks (P. Li et al., 2020; Yuan et al., 2019). Recently, Theofilatos et al. (2019) applied and compared several machine learning and deep learning techniques on the same dataset: kNN, Naïve Bayes, classification tree, random forest, SVM, shallow neural network, and deep neural network, observing that the deep learning approach yielded the best results, but other far less complex methods such as Naïve Bayes performed only marginally worse.

Regardless of the model employed, an RTCPM is, in general, a binary classification problem, in which the two classes are crash and no-crash. As crashes are rare events, the dataset used to train the model is unbalanced, with far more time intervals in which no crashes were observed, than those in which a crash happened. In an unbalanced classification problem, models tend to classify all of the sample into the major class, due to the inherent motivation of their objective functions, which minimize the sum of errors by

assigning the same weight to both the major and minor classes, whereby samples in the minor class make little contribution (Sun et al., 2007). There are two possible approaches to deal with this issue: algorithm-level solutions and data-level solutions (Ke et al., 2018). The former approach consists in applying cost-sensitive learning techniques, in which different weights are assigned to the classes in the objective function, to penalize the misclassification of minority class samples. In the latter approach, the training dataset is resampled to obtain a more balanced problem, with several sampling methods existing in the literature (He and Garcia, 2009). In RTCPM literature, data-level solutions are usually employed, typically with a matched case-control method (Abdel-Aty et al., 2004), although recently, other more sophisticated approaches, such as synthetic minority oversampling technique (SMOTE), have started to be used (Basso et al., 2018; P. Li et al., 2020).

2.2 Real-time conflict prediction models

There is a limited, but rapidly growing, literature on real-time conflict prediction models. To the best of the authors' knowledge, the only attempts to predict conflict-prone conditions on a road segment in real-time were made by (Caleffi et al., 2017; Formosa et al., 2020; Katrakazas et al., 2018, 2017).

Similarly to this work, Katrakazas et al. (2017, 2018) motivated the need for an RTConfPM on the difficulty of retrieving the high spatial-temporal precision crash data that an RTCPM needs. Contrarily to the present work, Katrakazas et al. (2017, 2018) did not have access to real-world vehicle-by-vehicle traffic data, but only to 15-minute aggregated data; to overcome this, they used such data as input to a microsimulation model, to reproduce vehicle trajectories and retrieve simulated conflicts. Conflicts were identified based on some surrogate measures of safety thresholds/rules of thumb from the literature, but no ad hoc analysis of crash-conflict relationships was performed. They applied three machine learning classifiers:

SVM, random forest, and kNN; however, differently from the authors of this work, they did not compare the results of their RTConfPM to those of an RTCPM trained on the same dataset.

Caleffi et al. (2017) applied linear discriminant analysis to predict conflicts on a Brazilian freeway section. They were able to detect conflicts in a disaggregate manner, retrieving them from video recordings. Their approach to detecting conflicts, however, was significantly different from that of the present work: conflicts were defined in qualitative terms and detected only by observing evasive actions, instead of being quantitatively determined by surrogate measures of safety (see Section 4.1.3).

Recently, Formosa et al. (2020), proposed an RTConfPM in which input traffic data were collected with both loop detectors and in-vehicle sensors, and conflicts were identified with an instrumented vehicle. As in Katrakazas et al. (2018), conflicts were defined according to general rules and thresholds from the literature. The predictive model adopted was a deep neural network, and the results obtained were promising. The data acquisition process in their study is quite different from that of the present study, as it involves instrumented vehicles; its complexity can be regarded as providing an opportunity to handle complex relationship between heterogeneous variables, as well as acting as a potential limitation for a large-scale real-world implementation, at least in the short-medium term. No direct comparison with a crash-based approach was made in their work.

In a different environment (i.e., intersections), the recent works of Essa and Sayed (2018, 2019) and Zheng and Sayed (2020) developed a conflict-based model for real-time safety evaluation of signalized intersections, using generalized linear models, full-Bayesian and extreme value theory approaches. Traffic input variables included queue length, shock wave speed and area, and the platoon ratio; time-to-collision (TTC), modified TTC, and the deceleration rate to avoid the crash were used as traffic conflict indicators. Data were gathered at several intersections for some hours using video cameras. Their works lacked a direct comparison with crash-based approaches.

Another work worth mentioning is by Osman et al. (2019), who predicted near-crashes based on the vehicle kinematics variables recorded immediately beforehand, using several machine learning classifiers. A similar approach was previously proposed by Wang and Chan (2017). The point of view adopted by these studies is quite different from that of the present work, as it was exclusively vehicle-based, rather than infrastructure-based: their aim was not to predict unsafe situations on a certain road segment, but rather to develop a tool to support crash avoidance systems and/or ADAS.

3. Methodology

In this work, we developed and compared an RTConfPM and an RTCPM. Methodologically, they both follow the same typical scheme, as shown in Figure 2.

3.1 Variable selection model: Random forest

Random forest is an ensemble machine learning classification model composed of multiple tree classifiers. Its main advantage over a simpler decision tree is that it can provide estimates of the classification error that are more stable and unbiased. Moreover, as the number of trees within the random forest grows, the results become more robust against over-fitting.

Each tree of the random forest is grown following this procedure (Breiman, 2001): (i) given N records in the training set, a bootstrap sample n of N is drawn to grow the tree; (ii) given M variables, a constant number of $m<M$ variables are selected at each node of the tree, and the best split among m is used; (iii) the tree is grown fully, without pruning.

The $N - n$ left-out records are called out-of-bag (OOB) data, and they are, in fact, an internal test dataset. OOB data are used to obtain an unbiased error estimate, and to assess the variables importance. An alternative measure of importance is the total decrease in Gini node impurities from splitting on the variable p, averaged over all trees. The two importance metrics are generally consistent with each other; however, the former has the drawback of tending to overestimate the variable importance of highly correlated variables (Strobl et al., 2008).

Random forests are widely used in RTCPMs for variable selection (Abdel-Aty et al., 2008; Abdel-Aty and Haleem, 2011; Ahmed et al., 2012; Basso et al., 2018; Shi and Abdel-Aty, 2015; Xu et al., 2013b), and in some cases, also as a classifier itself (Theofilatos et al., 2019).

3.2 Training dataset resampling: SMOTE

As mentioned in Section 2.1, the unbalanced classification issue can be resolved by resampling the training dataset, in order to make it more balanced. The easiest way to achieve this is by randomly under-sampling the majority class or by randomly over-sampling the minority class; however, such methods may lead, respectively, to under-fitting and over-fitting (He and Garcia, 2009).

A more advanced solution to the problem is to apply the synthetic minority oversampling technique (SMOTE), proposed by Chawla et al. (2002). Instead of using duplicate samples, SMOTE generates synthetic samples of the minority class. The algorithm follows a relatively simple and computationally inexpensive procedure:

- For each minority class sample, k nearest neighbor minority class samples are selected.
- For each neighbor, the difference between the original sample feature vector and the neighbor feature vector is computed.

- The difference is multiplied by a random number between 0 and 1 and added to the original sample feature vector, generating a new synthetic sample of the minority class.

For example, in a bidimensional feature space, the synthetic sample would be a random point on the line segment between the original sample and the neighbor sample.

SMOTE has been applied, with good results, in several recent RTCPMs (Basso et al., 2018; Katrakazas et al., 2019; Ke et al., 2018; P. Li et al., 2020; Parsa et al., 2019; Yuan et al., 2019).

3.3 Classifier: support vector machine

Vapnik (1995) introduced SVM as a supervised machine learning method for classification and regression analysis. In a classification problem, the basic idea behind SVM is to find the hyperplane that separates all data points, maximizing the distance between the classes (i.e., the "margin").

This results in a simple optimization problem: choosing the coefficient of the linear boundary to maximize the distance between the boundary and the nearest observations, subject to the constraint that all observations must be on the correct side of the boundary. In the end, the optimal solution is determined only by the observations nearest to the boundary (i.e., the "support vectors"). If, as it often happens, data from the two classes are not linearly separable, then SVM can use a soft margin; that is, a hyperplane that separates many (yet not all) data points. For more details, the reader is referred to Cortes and Vapnik (1995) and Vapnik (1995).

SVM is one of the most commonly used classifiers in RTCPMs. It was introduced in the works of Lv et al., (2009b), Qu et al. (2013), and Yu and Abdel-Aty (2013b), and recently applied in Theofilatos et al. (2019), J. Wang et al. (2019), and L. Wang et al. (2019), sometimes in combination with SMOTE (Basso et al., 2018; Parsa et al., 2019).

3.4 Performance indicators

Several performance metrics can be used to evaluate an RTCPM or an RTConfPM. Many of the indicators can be derived from the confusion matrix, which, in the case of binary classification, is a 2x2 array composed of the following components (see Figure 4): true positive, true negative, false positive, and false negative.

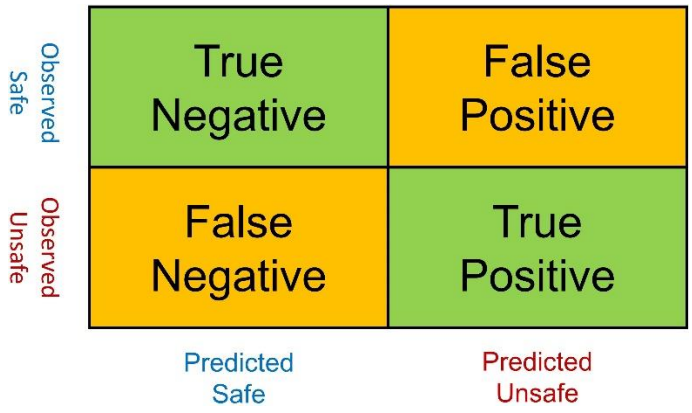

Figure 4. Confusion matrix for the binary unsafe/safe classification problem.

The indicators considered in this work, and computed as in Table 1, are as follows (Stehman, 1997; Tharwat, 2018):

- Accuracy: the number of safe and unsafe situations correctly predicted by the model, divided by the total number of instances. This measure should be treated very cautiously when classes are unbalanced.
- Recall (also known as sensitivity or true positive rate): the number of correctly predicted unsafe situations divided by the total number of observed unsafe situations. A low-recall model indicates the model missed some unsafe situations.
- Specificity (also known as true negative rate): the number of correctly predicted safe situations divided by the total number of observed safe situations. A low-specificity model indicates the

model considered many of the safe situations unsafe. The complementary measure (1 − specificity) is the false alarm rate (or false positive rate).

Another useful tool to evaluate a classification model is the receiver operating characteristic (ROC) curve. The ROC curve is created by plotting the true positive rate (i.e., recall) against the false positive rate (i.e., false alarm rate) at various threshold settings. From this curve it is possible to derive a widely used performance indicator: the area under the ROC curve (AUC). AUC is the probability that a classifier will rank a randomly chosen positive instance higher than it would a randomly chosen negative one (Fawcett, 2006). AUC is considered a measure of aggregated classification performance (Flach et al., 2011).

To summarize, accuracy and AUC can be used for an overall evaluation of the model performance, whereas recall and specificity can assess specific aspects of the model. In road safety applications, recall is particularly important because it quantifies the influence of the number of false negatives: a model that marks situations “safe” when they are actually unsafe must be avoided because it precludes the possibility of intervening and avoid a crash. Specificity quantifies the impact of false positives; although high-specificity models are preferable, a model that triggers too many false alarms is a problem mainly from an operational point of view instead of a safety point of view.

This set of performance indicators is commonly used to evaluate RTCPMs and RTConfPMs (Caleffi et al., 2017; P. Li et al., 2020; Wang et al., 2020; Yuan et al., 2019).

Table 1. Equations of performance indicators used to evaluate the models developed in this work. TP=true positive; TN=true negative; FP=false positive; FN=false negative.

Indicator	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Recall	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
False alarm rate	$\frac{FP}{TN+FP}=1-\frac{TN}{TN+FP}$

4. Real-time conflict prediction

An RTConfPM was developed using data collected on a three-lane Italian motorway for 1 year, from January 1st to December 31st, 2013. The motorway is about 150km long, and there are 20 cross-sections in each direction (40 in total) that are equipped with microwave Doppler radars. The average spacing of the detectors is 7.58 km, with minimum spacing of 0.76 km and maximum of 16.07 km.

4.1 Data Preparation

In an RTConfPM, input data (see Section 4.1.1) collected during a given interval are used to predict whether the risk of an unsafe situation occurring will increase (defined according to Sections 4.1.2 and 4.1.3) in the following interval. Conflict data are derived from surrogate measures of safety for model calibration and validation, whereas they are not used during the live real-world application of the model (see Figure 2).

4.1.1 Traffic data

Over each lane of each motorway cross-section, a dedicated radar collects the following vehicle-by-vehicle information: speed, time gap, time stamp and vehicle class. These data, originally available in disaggregated form, were then aggregated in 5-minute intervals, a length chosen to align this paper with the vast majority of works in the RTCPM/RTConfPM literature (see Section 2).

Data aggregation provided the following variables for each of the three lanes:

- Traffic volume (*Vol*): the total number of vehicles.
- Heavy-duty traffic volume (*HVol*): the percentage of vehicles classified as heavy-duty trucks.
- Speed (*Sp*): harmonic mean of vehicle speeds.
- Speed variance (*SdSp*): standard deviation of vehicle speeds.

Therefore, for each cross-section, 12 variables are available in 5-minute aggregation. In the remainder of the paper, suffixes R, C, and L will indicate the rightmost lane, the center lane, and leftmost lane, respectively. Descriptive statistics are presented in Table 2.

Traffic volume, average speed, and standard deviation of speed are variables commonly used in real-time road safety analysis, and they are particularly relevant in the case of rear-end crashes (Dimitriou et al., 2018). In this case study, speeds are recorded on a cross-section, but the aim of the model is eventually to detect conflict-prone and crash-prone situations near the cross-section (i.e., on a road segment). For this reason, to avoid overestimating the influence of faster vehicles on the mean speed of such segment, we considered space-mean-speed instead of time-mean-speed (Knoop et al., 2009). The former, under the assumption of stationary conditions, is given by the harmonic mean of speed (Daganzo, 1997).

Information on vehicle type is not often used in real-time road safety analysis. However, Basso et al. (2020) underlined the importance of considering flow composition in an RTCPM. Another peculiarity of the

variables adopted in this work is that they were considered separately for each of the three motorway lanes: As can be observed in Table 2, each of the three lanes has specific characteristics and aggregating variables over the whole cross-section would cause the loss of such information.

Table 2. Descriptive statistics of candidate input traffic variables

Traffic variable	Description	Unit	Mean (SD)	95% quantile range
Vol_R	Traffic volume on right lane	veh/5min	30.0 (27.9)	0; 72
Vol_C	Traffic volume on central lane	veh/5min	42.3 (38.8)	0; 111
Vol_L	Traffic volume on left lane	veh/5min	27.4 (62.6)	0; 111
HVol_R	% of heavy-duty vehicles on right lane	%	51.2 (27.4)	0.00; 92.3
HVol_C	% of heavy-duty vehicles on central lane	%	7.19 (10.2)	0.00; 36.7
HVol_L	% of heavy-duty vehicles on left lane	%	0.47 (2.83)	0.00; 3.41
Sp_R	Harmonic mean of speed on the right lane	km/h	90.2 (9.58)	72.2; 110.0
Sp_C	Harmonic mean of speed on the central lane	km/h	112.3 (11.1)	90.6; 131.4
Sp_L	Harmonic mean of speed on the left lane	km/h	128.4 (14.4)	104.0; 154.5
SdSp_R	SD of speed on the right lane	km/h	10.5 (4.27)	3.84; 19.6
SdSp_C	SD of speed on the central lane	km/h	13.7 (4.01)	8.04; 22.6
SdSp_L	SD of speed on the right lane	km/h	11.5 (5.78)	0.00; 23.4

4.1.2 Surrogate measures of safety data

Considering the above-mentioned traffic data (see Section 4.1.1) in disaggregate, vehicle-by-vehicle form, it is possible to calculate the TTC between each couple of vehicles consecutively detected on the same lane, with the well-known definition:

$$TTC = \frac{R}{RR} = \frac{v_L * t_{GAP}}{v_F - v_L} \quad (1)$$

where R (range) is the separation between the leading and the following vehicle, computed assuming constant speed of the vehicles, and RR (range rate) is the speed difference between the following and the leading vehicle. In (1), v_L and v_F are the speeds of the leading and the following vehicles, respectively; t_{GAP} is the time gap between them, recorded at the cross-section.

TTC is a surrogate measure of safety commonly used to analyze rear-end crashes. It represents "the time required for two vehicles to collide if they continue at their present speed on the same path" (Hayward, 1972). Two vehicles are on a collision course if the speed of the following vehicle is higher than the leading vehicle (RR must be higher than zero); TTC equal to zero represents a collision. For more details on how to compute it and on alternative definitions, the reader is referred to Nadimi et al. (2020).

4.1.3 Conflict data

In each 5-minute interval, a TTC value for each couple of consecutive vehicles on each lane of the cross-section is available. With this kind of information, determining whether one or more conflicts have happened in that 5-minute interval is possible. However, it is neither a straightforward nor a trivial operation.

A TTC threshold *u* should be set to separate non-risky interaction events and proper conflict events; however, such a threshold cannot be defined univocally, because certain low TTC values may represent intentionally aggressive behaviors, which may still produce close but controlled interactions between vehicles (Tarko, 2019). Setting a high threshold allows consideration of all the "true conflicts" (in accordance with the definition provided in Section 1) as conflicts, but also many controlled interactions; setting a low threshold has the consequence of discarding some real conflicts, but also excludes all controlled interactions. Tarko (2018b) suggests choosing "a threshold separation sufficiently short to claim only traffic conflicts," because including controlled interactions may bias the results.

A further complication arose in this work: we needed to define if, considering the *n* conflicts observed in the 5-minute interval, that interval should be considered safe or unsafe. In this case, operational needs come into play: the lower value of *n* is selected, the lower crash risk is accepted, but the higher numbers of alarms are triggered. In addition, the choice of *n* may significantly influence model performance.

In this work, we applied a statistical approach (i.e., extreme value theory [EVT]), to investigate the probabilistic relationship between conflicts and crashes in the motorway. This analysis, whose details are reported in Appendix A, allowed us to determine a TTC threshold (u = 0.78 s) to identify rear-end conflicts. Subsequently, a sensitivity analysis on model performance, and some considerations on the effect of n on the size of the "unsafe" class of observations, lead to the selection of n = 3 (see Appendix A). In other words, in the RTConfPM, a 5-minute interval was considered unsafe if at least three TTC values under 0.78 seconds were recorded during it.

As a result, 2,241 unsafe situations were detected and matched with the traffic variables recorded on the same cross-section in the previous 5-minute interval.

4.1.4 Training and test datasets

From the procedure of Section 4.1.3, a huge dataset with almost 3.5 million records was built, one for each of the 5-minute intervals and each cross-section of the year. The total number could potentially have been 4,204,800 records (288 5-minute intervals in 1 day × 365 days × 40 cross-sections), but some of the radars were not able to record all of the data during the year due to technical or operational problems. As mentioned before, only 2,241 of these records were matched to an unsafe situation, producing an extremely unbalanced classification problem (one unsafe record every 1,554 safe records). With reference to Figure 5, this is the full dataset.

This dataset was then split into a full training dataset (80% of the full dataset) and in a full test dataset (20%), keeping the unsafe-to-safe ratio of 1:1,554 in both datasets. The latter was then used exclusively for model evaluation.

On the other hand, the full training dataset was used for variable selection (Section 4.2) and then divided into unsafe situation and safe situation datasets, respectively containing only unsafe records and only safe records.

SMOTE was then applied on the unsafe situations dataset to increase the number of unsafe records, whereas a limited number of records was randomly sampled from the safe situations dataset to have a 1:1 ratio between unsafe and safe records. The two sets of records were merged in the SMOTEd dataset. The combination of SMOTE and under-sampling is commonly adopted (Chawla et al., 2002). The choice of random under-sampling was suggested by recent literature praising its performance over that of the matched case-control method (Peng et al., 2020).

This dataset was again split into a SMOTEd Training dataset (80% of SMOTEd dataset) and a SMOTEd test dataset (20%), with the same 1:1 unsafe-to-safe ratio in both datasets. The former was used to train the SVM (see Section 4.3.2), the latter for model evaluation.

It is worth underlining that although the original full dataset contained more than 3.5 million records, the SMOTEd Training dataset is much smaller (about 14,000 records), therefore the SVM is able to handle it without the need to adopt advanced algorithmic strategies required for large-scale learning (Menon, 2009).

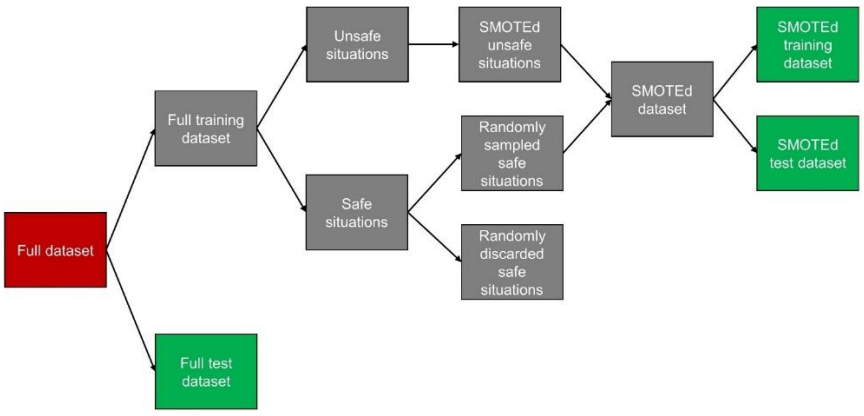

Figure 5. Overview on training, test and full-test dataset formation.

4.2 Variable selection

Before applying a classifier, it is good practice to select a subset of variables, excluding those less important and those correlated with other variables. In the case of unbalanced datasets, the variable selection procedure should be performed before applying the resampling procedure, therefore on the full training dataset (see Figure 5). This is especially true when applying SMOTE, which introduces correlation between samples, whereas most variable selection methods assume sample independence (Blagus and Lusa, 2013).

Pearson's correlation was computed for each pair of the 12 recorded variables to test for linear correlation. Results are presented in Figure 6. In RTCPMs/RTConfPMs there is not a clear rule on when to consider two variables as "too much correlated": for example, L. Wang et al. (2019) excluded variables with correlation coefficients higher than 0.3, whereas Li et al. (2020) used a 0.5 threshold, Yuan et al. (2019) a 0.65 threshold, and Basso et al. (2018) a 0.95 threshold. Therefore, we decided not to remove any of the variables a priori, and to perform further investigations with random forest.

	Vol_R	Vol_C	Vol_L	HVol_R	HVol_C	HVol_L	Sp_R	Sp_C	Sp_L	SdSp_R	SdSp_C	SdSp_L
Vol_R	1.00	0.59	0.28	0.19	0.26	0.06	-0.17	-0.24	-0.16	-0.28	-0.01	0.12
Vol_C	0.59	1.00	0.40	-0.08	-0.06	0.05	-0.01	-0.13	-0.14	-0.13	-0.26	0.14
Vol_L	0.28	0.40	1.00	0.05	0.04	0.04	-0.09	-0.14	-0.11	-0.12	-0.08	0.03
HVol_R	0.19	-0.08	0.05	1.00	0.57	0.06	-0.39	-0.23	-0.06	-0.56	0.36	-0.01
HVol_C	0.26	-0.06	0.04	0.57	1.00	0.17	-0.36	-0.48	-0.22	-0.44	0.37	0.04
HVol_L	0.06	0.05	0.04	0.06	0.17	1.00	-0.14	-0.19	-0.21	0.05	0.15	0.09
Sp_R	-0.17	-0.01	-0.09	-0.39	-0.36	-0.14	1.00	0.68	0.53	0.34	-0.16	0.02
Sp_C	-0.24	-0.13	-0.14	-0.23	-0.48	-0.19	0.68	1.00	0.74	0.34	-0.08	0.01
Sp_L	-0.16	-0.14	-0.11	-0.06	-0.22	-0.21	0.53	0.74	1.00	0.20	0.03	0.00
SdSp_R	-0.28	-0.13	-0.12	-0.56	-0.44	0.05	0.34	0.34	0.20	1.00	0.02	0.11
SdSp_C	-0.01	-0.26	-0.08	0.36	0.37	0.15	-0.16	-0.08	0.03	0.02	1.00	0.16
SdSp_L	0.12	0.14	0.03	-0.01	0.04	0.09	0.02	0.01	0.00	0.11	0.16	1.00

Figure 6. Correlation matrix for the full training dataset.

The random forest model for variable selection was built using the RandomForest package in R (Liaw and Wiener, 2002). Three variables were randomly sampled at each split and 500 trees were grown. Figure 7 presents variable importance as the total decrease in Gini node impurities.

According to the model, the most important variables are those related to speed. In addition to the information on total traffic volumes, the presence of heavy-duty trucks on the rightmost and central lanes appears to be significant. The least important variable is the percentage of heavy-duty trucks in the leftmost lane; this is not surprising because this percentage is always close to zero, even during congestion periods, as trucks are forbidden to use that lane, according to the Italian Highway Code (Art. 176-9).

This "importance ranking" does not determine, on its own, the optimal set of variables to include in the model; instead, it can be used to define several candidate sets to test and evaluate using performance indicators (see Sections 4.3 and 4.4).

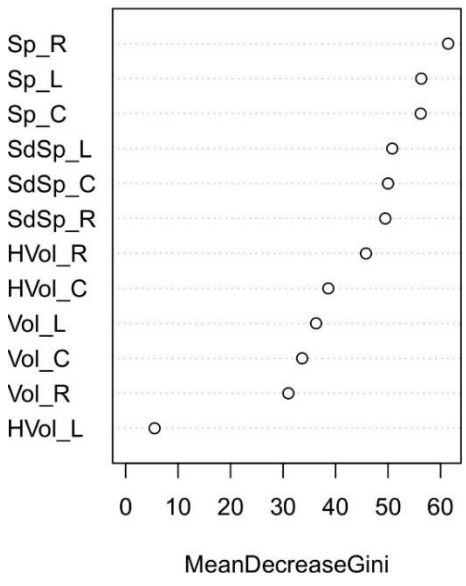

Figure 7. Change in Gini impurity index to determine variable importance in the RTConfPM.

4.3 Model development

4.3.1 Resampling

A combination of SMOTE over-sampling of unsafe records and random under-sampling of safe records was applied to achieve a balanced dataset. Two parameters should be chosen in this phase:

- The factor *k* of the SMOTE (i.e., how many new samples are generated for each original sample).
- The desired ratio between unsafe and safe records.

In the literature, a typical value of *k* is 5, which was originally proposed by Chawla et al. (2002). For this work, we carried out a sensitivity analysis, which showed that a further increase in the value of *k* did not produce a significant improvement in performance. Another sensitivity analysis was carried out to choose the unsafe/safe ratio. A 1:1 ratio was chosen because higher ratios of safe situations would have produced a significant decrease in the recall indicator. For the sake of brevity, details on this sensitivity analysis are not reported in this paper. SMOTE procedure was implemented in MATLAB by the authors.

4.3.2 SVM

Having obtained a SMOTEd dataset (see Figure 5), 80% of this dataset was used to train several SVM classifiers, using combinations of different sets of input variables and kernel functions. SVM models were implemented in MATLAB using 'fitcsvm' function (cost = 1; gamma = 1).

Different sets of input variables were selected:

- Set #1, with all 12 recorded variables (i.e., without variable selection).
- Set #2, with 11 variables, which excluded the least important according to random forest (i.e., *HVol_L*; see Section 4.2).

- Set #3, with the six most important variables according to random forest (i.e., average and standard deviation of speeds; see Section 4.2).

Moreover, SVMs with either a linear or radial basis (RBF) kernel function were trained, resulting in six possible combinations.

For each combination of input variable set and kernel function, Monte Carlo cross-validation (Simon, 2007; Xu and Liang, 2001) with 100 repetitions was employed. Therefore, for each combination, 100 different 80%-20% partitions of the SMOTEd dataset (see Figure 5) were performed, resulting in the training of 100 different SVMs.

4.4 Results

4.4.1 Predictive performance

For each of the 100 SVMs of each input variable – kernel function combination, accuracy, recall, specificity, and AUC were computed on SMOTEd training and test datasets and on the full test dataset; in the remainder of the paragraph, we will refer to these three datasets simply as training, test, and full-test datasets, respectively. Averages across the 100 repetitions are reported in Table 3.

Train and test indicators indicate how well the SVM models adapted to the SMOTEd dataset; then, full-test is used to evaluate the models' performance on a proper real-world dataset, with a realistic, strongly unbalanced ratio between crashes and non-crashes.

In particular, it is possible to observe that, for all the models, training and test indicators have similar values, meaning that none of the SVM models suffers from overfitting.

Table 3. Performance indicator values, averaged on the 100 Monte Carlo repetitions, for training, test, and full-test dataset of the RTConfPM

Training dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
Linear	Set #1	0.958	0.982	0.934	0.066	0.952
	Set #2	0.957	0.981	0.934	0.066	0.951
	Set #3	0.788	0.975	0.601	0.399	0.945
Radial basis	Set #1	0.981	0.992	0.970	0.030	0.995
	Set #2	0.981	0.992	0.971	0.029	0.995
	Set #3	0.825	0.998	0.651	0.349	0.995

Test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
Linear	Set #1	0.958	0.982	0.934	0.066	0.953
	Set #2	0.957	0.981	0.933	0.067	0.952
	Set #3	0.788	0.976	0.600	0.400	0.945
Radial basis	Set #1	0.972	0.986	0.959	0.041	0.985
	Set #2	0.972	0.986	0.959	0.041	0.985
	Set #3	0.809	0.979	0.639	0.361	0.981

Full-test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
Linear	Set #1	0.934	0.983	0.933	0.067	0.948
	Set #2	0.933	0.981	0.933	0.067	0.947
	Set #3	0.600	0.978	0.600	0.400	0.942
Radial basis	Set #1	0.958	0.945	0.958	0.042	0.972
	Set #2	0.959	0.947	0.959	0.041	0.973
	Set #3	0.639	0.904	0.639	0.361	0.966

Overall, big differences do not exist between the linear and RBF models, with the latter having marginally better indicators, except for recall in the full-test dataset.

Input variable Set #3 produces significantly worse results, compared to the others, meaning that some variables, whose effects are indeed significant, are not included in that set. In practical terms, this means that despite the speed variables being the most important (see Section 4.2), they cannot, on their own, produce a reliable model, which confirms that the percentage of heavy-duty trucks and total traffic volumes are indeed significant, as intuitively expected.

Set #1 and Set #2 return almost identical results, confirming that *HVol_L* is irrelevant, as expected. By focusing on the full-test dataset indicators, the two best models appear to be Linear-Set#1 and RBF-Set#2. The first has a high recall of 98.3%, the second a quite lower 94.7%; nevertheless, the latter is better in all other indicators, with a 95.9% specificity and accuracy and 0.973 AUC.

In addition to showing great potential from a safety point of view, their operational performance is remarkable. In RBF-Set#1, the false alarm rate is only 4.1%; therefore, a false alarm is triggered only once out of 24.4 five-minute intervals or, in other words, fewer than 12 times per day.

4.4.2 Contributing factors

SVM is often regarded as a "black-box" technique because the effects of the explanatory variables cannot be directly assessed. To overcome this issue and investigate the factors contributing to the risk of unsafe situations, we applied the sensitivity analysis developed by Fish and Blodgett (2003) and applied in several other road safety works (Dong et al., 2015; Gu et al., 2020; J. Li et al., 2020; Yu and Abdel-Aty, 2013b). According to this procedure, each variable is changed by a pre-specified unit while keeping the others unchanged; then SVMs are recalibrated with the modified datasets, and comparisons are made on the mean predicted probabilities of unsafe situations occurring. This allows retrieval of positive and negative dependence for each input variable and unsafe situation, as well as their functional relationship, but "direct comparison among [input] variables is precluded" (Fish and Blodgett, 2003).

In this work, we considered 10% of the mean value of each of the 12 input variables as changing unit; SVMs were trained with both linear and radial-basis kernel function. As expected, the kernel function type did not have an effect on the interpretation of contributing factors. Results of this analysis are reported in Table 4, highlighting which relationships exist between input variables and unsafe situations, with the same style adopted by Gu et al. (2020) and Yu and Abdel-Aty (2014).

A positive relationship exists between unsafe situations that may lead to rear-end collisions and traffic flow; this result is expected because higher flows imply more interactions between vehicles, and therefore an increased propensity for rear-end conflicts (Dimitriou et al., 2018). Average speed is negatively correlated with traffic density; therefore, it is not surprising that a higher chance of unsafe situations exists as average speed decreases (Wang et al., 2017; Yu and Abdel-Aty, 2013b). Standard deviation of speed has a positive relationship with unsafe situations, with higher differences in speeds among vehicles resulting in an increase in risk. This has been observed in several works (Theofilatos et al., 2019; Wang et al., 2017; Zheng et al., 2010), and it is particularly relevant in the case of rear-end conflicts (Dimitriou et al., 2018). Finally, an increase in the percentage of heavy-duty vehicles is linked to an increase in risk. The presence of such vehicles is correlated to an increase in congestion (especially in the rightmost lane), a decrease of average speed, and an increase of standard deviation of speed in the central lane (see Figure 6), which are all factors that contribute to an increase of rear-end conflicts.

Table 4. Sensitivity analysis results - positive relationship: if the input variable increases, the risk of unsafe situation increases; negative relationship: if the input variable decreases, the risk of unsafe situation increases.

Variable	**Relationship**
Vol_R	positive
Vol_C	positive
Vol_L	positive
HVol_R	positive
HVol_C	positive
HVol_L	positive
Sp_R	negative
Sp_C	negative
Sp_L	negative
SdSp_R	positive
SdSp_C	positive
SdSp_L	positive

5. Real-time crash prediction

A traditional RTCPM was developed, using the same traffic data of the RTCPM described in Section 4.1.1.

5.1 Data preparation

5.1.1 Crash data

A 6-year crash database was available for the same motorway. Data were collected by the administrative department of the motorway from 1st January 2011 to 31st December 2016. During this period, 5,362 crashes occurred; 1,575 of them were rear-end collisions.

In the RTCPM developed for this paper only rear-end collisions which occurred in 2013 within 2.5 kilometers of distance (either upstream or downstream) from any of the radar-equipped cross-sections were considered. One hundred and twenty-seven crashes satisfied these conditions. The decision to focus exclusively on rear-end crashes was motivated by the need to compare this RTCPM directly to an RTConfPM which can only predict rear-end collisions, as explained in Section 4.

The selected crashes were matched with the traffic variables recorded at the closest cross-section in the first complete 5-minute interval before the crash. The decision to refer crashes to the closest detector is common in RTCPM literature; however more sophisticated RTCPMs also considered traffic data from one or more upstream and downstream detectors (Hossain et al., 2019). In the present case study, this choice was forced by the irregular disposition and spacing of detectors on the motorway.

Not all the crashes could be matched with traffic conditions because some sensors were not active at the time of the crash: 39 collision events had to be discarded for this reason, resulting in a dataset of 88 usable crashes.

5.1.2 Training and test datasets

The same procedure depicted in Figure 5 was followed. The full test dataset of RTCPM and RTConfPM are different. In the first case, the full dataset's partition was performed so that a full training dataset and full test dataset had the same ratio of crash and non-crash records; in the second case, they had the same ratio of safe and unsafe records (where unsafe situations were defined in accordance with Section 4.1.3).

5.2 Variable selection

The same procedure described in Section 4.2 was used for the RTCPM.

Because different records are assigned to the two response classes (safe versus unsafe) in RTCPM and RTConfPM, random forest variable selection could, in principle, and from a statistical point of view only, produce different results. However, as explained in Section 1, if it is assumed that traffic conditions are precursors to conflicts, which in turn are precursors to crashes, then expecting variables that are "important" to predicting conflicts to be "important" to predicting crashes is also reasonable.

Figure 8 presents variable importance as the total decrease in Gini node impurities and, as the reader can observe when comparing this figure to Figure 7, the order of importance and the relative importance between variables is very similar for the two approaches. The random forest used for this analysis had the same parameters (500 trees and three variables randomly sampled at each split) as that of Section 4.2.

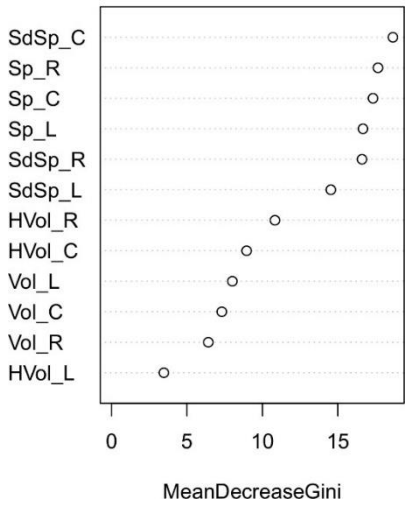

Figure 8. Change in Gini impurity index to determine variable importance in the RTCPM.

5.3 Model development

The number of crash records in the RTCPM is much lower than the unsafe records of RTConfPM, leading to an even more unbalanced dataset (crash/non-crash ratio is 1:39,567); therefore, SMOTE is also applied in this case. k = 5 and a 1:1 ratio were chosen for RTConfPM, following a similar sensitivity analysis.

The same six combinations of input variables and kernel functions were analyzed. For each of them, a Monte Carlo cross-validation procedure was applied, the SMOTEd dataset was partitioned 100 times, and 100 independent SVMs were trained for each combination.

5.4 Results

Accuracy, recall, specificity and AUC values were averaged over the 100 repetitions for training, test, and full-test datasets (Table 5).

RBF models have significantly higher performance in training than linear models do, but, contrarily to the RTConfPM case (see Section 4.4.1), they seem to suffer from overfitting because test and full-test performances are much lower.

The removal of an irrelevant (according to random forest variable selection and intuition) variable (i.e., *HVol_L*) provided an increase, albeit small, of all indicators of the linear models. Considering only the six most important variables according to random forest did not result in good performance: a steep increase in recall was indeed matched by a great decrease in all other indicators, producing unsatisfactory overall accuracy values.

Table 5. Performance indicator values, averaged on the 100 Monte Carlo repetitions, for training, test, and full-test dataset of the RTCPM.

Training dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
Linear	Set #1	0.652	0.490	0.814	0.186	0.702
	Set #2	0.662	0.493	0.831	0.169	0.700
	Set #3	0.500	0.809	0.190	0.810	0.604
Radial basis	Set #1	0.967	0.940	0.995	0.005	1.000
	Set #2	0.963	0.944	0.982	0.018	0.998
	Set #3	0.806	0.985	0.627	0.373	0.997

Test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
Linear	Set #1	0.628	0.465	0.791	0.209	0.656
	Set #2	0.636	0.469	0.806	0.194	0.651
	Set #3	0.494	0.800	0.183	0.817	0.580
Radial basis	Set #1	0.856	0.818	0.894	0.106	0.942
	Set #2	0.831	0.774	0.887	0.113	0.931
	Set #3	0.676	0.802	0.594	0.406	0.850

Full-test dataset

Kernel function	Input variables	Accuracy	Recall	Specificity	False alarm rate	AUC
Linear	Set #1	0.788	0.395	0.788	0.212	0.612
	Set #2	0.795	0.529	0.795	0.205	0.682
	Set #3	0.175	0.824	0.175	0.825	0.583
Radial basis	Set #1	0.899	0.219	0.899	0.101	0.615
	Set #2	0.875	0.177	0.875	0.125	0.569
	Set #3	0.485	0.353	0.485	0.515	0.537

Overall, among the six combinations, the best seems to be the linear model which used input variable Set #2. Considering the full-test dataset, it has lower specificity than the RBF-Set#2 model, but the latter

greatly struggles with recall. With a 0.68 AUC, 52.9% recall, and 79.5% specificity (corresponding to a 20.5% false alarm rate), this model is not the best in the literature, but it is in line with others (Theofilatos et al., 2019; Yuan et al., 2019). Moreover, it must be taken into account that it was calibrated starting from a full dataset that only included 88 crashes. For comparison, in Hossain et al.'s (2019) systematic review, 30 studies out of 78 had a crash sample higher than 500.

Moreover, as underlined by Basso et al. (2018), many RTCPMs in the literature are only evaluated on artificially balanced datasets, instead of a real-world dataset, and that could provide overoptimistic results for models that are actually not very good. In addition, the results presented in this paper are robust because they were obtained with Monte Carlo cross-validation, and therefore not influenced by particularly favorable dataset partitions, which again may produce overoptimistic results (Santos et al., 2018).

Regarding the analysis of the contributing factors on the risk of crash occurrence, the same procedure described in Section 4.4.2 was applied. As expected, the same positive and negative relationships were found (see Table 4).

6. Comparison between RTConfPM and RTCPM

The previous sections showed how the RTConfPM is able to predict unsafe situations with significantly more accuracy than the RTCPM can. In the two approaches, unsafe situations are defined in different terms; nevertheless, these are very important results because conflicts can be considered precursors to crashes, thus predicting them and intervening to prevent them from happening would in turn prevent crashes from happening.

The scenario analyzed in this paper allows us to perform a further direct comparison that is of great practical interest and answers the following question: how many real-world observed crashes would the RTConfPM predict as unsafe situations?

To answer this question, the same full-test dataset used for RTCPM validation was considered. Records in that dataset are divided into two unbalanced classes: crashes and non-crashes. Instead of making predictions on that dataset using the RTCPM as in Section 5.4, the RTConfPM is now applied to determine whether real crashes would be predicted as “unsafe situations” with the conflict-based approach.

Linear-Set#1 was applied (with 100 Monte Carlo repetitions) and produced the following averaged indicators:

- 93.3% accuracy
- 64.7% recall
- 93.3% specificity
- 0.775 AUC

These indicators should be compared with those of the best RTCPM model (Linear-Set#2; see Table 5), and they show that the RTConfPM is able to predict more crashes correctly than the RTCPM can (+11.8% in recall indicator), and do it with a lower false alarm rate (+13.8% in specificity). Overall, the model also has better accuracy (+13.8%) and AUC (+0.093).

This leads to several meaningful considerations. First, a model trained exclusively with conflicts was able to predict more crashes correctly than a model trained with crashes was, confirming what was suggested in theoretical terms in Section 1.

Second, this increase in recall is accompanied by an even bigger decrease in the false alarm rate, meaning the conflict-based approach can provide not only safety but also operational benefits.

Finally, there is still a fairly high percentage of crashes not predicted by the RTConfPM: not all crashes are caused by traffic conditions and therefore the model cannot predict all of them. This can be partially improved in the future by using additional input information (e.g., weather conditions), but a share of crashes will always be primarily caused by unprovoked human error, which would be very challenging to predict from an exclusively cross-sectional point of view.

7. Conclusion

This work introduced a model to predict conflict-prone situations in real-time. Although the concept is not entirely new (Caleffi et al., 2017; Formosa et al., 2020; Katrakazas et al., 2018), this paper represents the first attempt at developing an RTConfPM based on real-world observed surrogate measures of safety and comparing it to a traditional RTCPM.

Both models were developed using a similar and robust procedure. The results were promising, with the best RTConfPM achieving accuracy, recall, and specificity values above 93%, significantly outperforming the crash-based approach.

This work has some limitations that future studies will address.

The RTConfPM was compared with an RTCPM that had below-average performance due to a limited number of observed rear-end collisions and the detectors' layout. Comparing the two approaches in case studies in which the RTCPM provides better performance would be interesting.

To keep the present paper within a reasonable length, only one classifier, SVM, was applied. In the future, other machine learning approaches or different approaches (e.g., Bayesian) should be used to compare RTCPM and RTConfPM.

This work can be considered a starting point for several research developments that can overcome the limitations of this work and extract the full potential of RTConfPM:

- Applying other statistical and machine learning classifiers.
- Using alternative definitions of unsafe situations, possibly considering different levels of risk.
- Collecting traffic and surrogate safety data using different technologies (e.g., V2X).
- Applying the model to other types of infrastructures (e.g., intersections).
- Including in RTConfPM other types of input (e.g., weather conditions).
- Estimating how short the observation period for an RTConfPM can be while still providing good performance.
- Studying transferability and seasonality issues.
- Testing different time intervals for data aggregation and/or different prediction horizons.
- Analyzing real-time intervention strategies to prevent conflicts from happening.

Appendix A. Defining unsafe situations

As mentioned in Section 4.1.3, to define a situation as "unsafe," two parameters must be chosen:

1. The TTC threshold *u* under which an interaction between vehicles is considered a conflict
2. The minimum number *n* of conflicts within the 5-minute interval to consider it "unsafe."

A statistical approach is adopted to determine the threshold, exploiting the probabilistic connection of traffic conflicts with crashes. Such connection can be modelled with extreme value theory (EVT), which is a branch of statistics dealing with the stochastic behavior of extreme events (Coles, 2001). EVT has been applied in several road safety studies to estimate crash probabilities with surrogate measures of safety (Cavadas et al., 2020; Farah and Azevedo, 2017; Fu et al., 2020; Gastaldi et al., 2021; Orsini et al., 2020a, 2019; Songchitruksa and Tarko, 2006; Wang et al., 2018; Zheng et al., 2014, 2019, 2018). The core of the idea is that the observed surrogate measures of safety can be fitted with an extreme value distribution; from that, extrapolating a crash probability is possible by considering crashes as "extreme" surrogate events. Recently, EVT has also been applied in a conflict prediction model to determine the threshold identifying near-crash events between motor vehicles and pedestrians at signalized intersections (Zhang et al., 2020).

The extreme value distribution adopted in the present study is the generalized Pareto (GP), which models the distribution of exceedances *X* over a fixed threshold *u*, provided *u* is sufficiently high. In that case, $F_u(x) = \Pr\,(X - u \leq x | X > u)$ can be approximated by a GP distribution, whose cumulative distribution function is as follows:

$$GP(x) = 1 - \left(1 + \xi \frac{x - u}{\sigma}\right)^{-\frac{1}{\xi}} \qquad (A.1)$$

where $\sigma > 0$ is the scale parameter and $-\infty < \xi < +\infty$ the shape parameter. The reader is referred to the above-mentioned works dealing with road safety applications of extreme value theory for further methodological details.

The aim of the analysis in this Appendix is to select an appropriate threshold u for a GP distribution fitted on the TTC values recorded in this work's case study. Such threshold will be used in the remainder of the paper to identify traffic conflicts.

There are several techniques to select the best threshold for an EVT model (Scarrott and MacDonald, 2012). The most used approach is the graphical diagnostic method, in which the GP distribution is fitted multiple times on the data, progressively increasing the threshold; the best threshold is then chosen by visually inspecting mean excess and stability plots. The drawback of this method is it requires a certain expertise from the analyst, and it may not give a precise and unambiguous value. An alternative, more objective approach was proposed by Thompson et al. (2009): the automatic threshold selection method (ATSM). In this method, a GP is fitted for each potential threshold and the differences in the modified scale parameters for neighboring thresholds is calculated. Asymptotic normality, with mean equal to zero, of these scale differences is assumed, if the GP is the correct distribution. A forward selection procedure is then applied, until the assumption of normality and mean zero is not rejected. At this point, the threshold is said to be consistent with a GP distribution. More details are available in Thompson et al. (2009).

With respect to the present study, a preliminary operation must be performed on the TTC data. EVT is designed to estimate the probability of extremely high values of the variable of interest; however, in this case, extreme events are represented by extremely low TTC values. Therefore, data were transformed, considering negated values of TTC. This is a common practice in EVT applications to road safety.

The negated TTC values were used to fit several different GP distributions, progressively increasing the threshold value by 0.01 seconds. Note that we were allowed to fit a single distribution for all cross-sections because a transferability study on data from the same motorway showed that "it is possible to aggregate data from several road sections, provided that they share similar geometric, traffic and weather characteristics" (Orsini et al., 2020b). Otherwise, separate GP distributions should be fitted for each cross-section.

The stability plot is presented in Figure A.1a, which shows that the shape parameter remains constant until the threshold is around −0.80 s. ATSM returns a value of −0.78 seconds, which is consistent with the stability plot. Visual inspection of the goodness-of-fit of the GP distribution with threshold equal to −0.78 seconds is satisfactory (see Figure A.1b). The estimated value of the scale parameter is 0.352 (standard error 0.0011); the estimated value of the shape parameter is −0.445 (standard error 0.0013).

From this distribution, extracting the crash probability is possible under the assumption that –TTC ≥ 0 represents a crash, resulting in $P_{inc} = 1 - GP(0) = 7.22 * 10^{-5}$. This probability is associated with crashes in the close vicinity of the cross-sections (about 250 m before and after the section, according to Orsini et al., 2020a). Assuming a constant crash probability within 2500 m before and after the cross-section (i.e., considering a 10× multiplier), and multiplying this probability by the number of −TTC values above −0.78 (i.e., 111,859), we obtain an estimated number of 80.7 crashes within 2,500 m from the cross-sections during the observation period, with a 95% confidence interval [73.4, 90.9]. The observed number of crashes in the same area and period is 88, which falls within the confidence interval, therefore validating the goodness of the model estimated with a threshold value of 0.78. More details on these procedures are available in Orsini et al. (2020a).

Note that, in order to validate this procedure and choose the threshold, crash data were needed. So, even in this conflict-based RTConfPM, crash data are still necessary. It is worth pointing out, however, that

these data need not to have a high spatial/temporal precision, as they are only used in aggregated form. From a practical point of view this is crucial, because it means that, contrarily to a traditional RTCPM, the proposed conflict-based approach can be adopted even in absence of highly spatial/temporal-accurate data, which may not be available for many real-world highways. Furthermore, after appropriate transferability studies, EVT models could be transferred even to highway segments were crash data are completely unavailable.

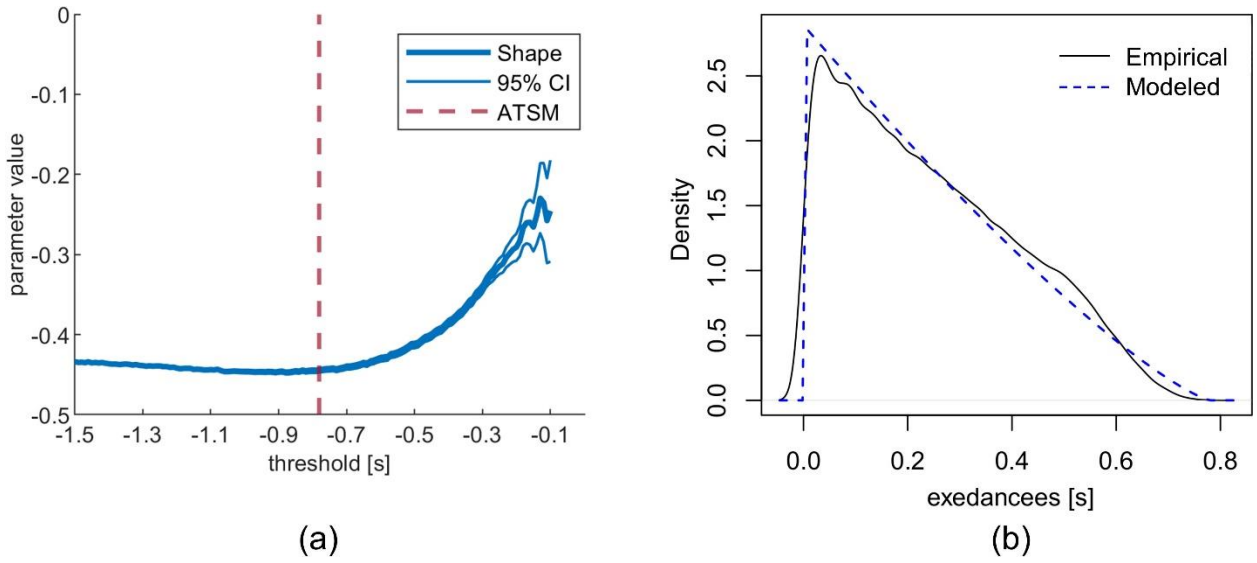

Figure A.1. (a) Stability plot for shape parameter; (b) Probability density plot for a model fitted with u = −0.78s.

P_{inc} can be interpreted as the probability of a conflict resulting into a crash. Therefore, assuming independence between conflicts, if we predict that n conflicts will happen within a given 5-minute interval, then the probability of observing a crash during such interval is $n \times P_{inc}$.

Contrary to the selection of parameter u, which should be supported by a study on the conflict–crash relationship, the choice of n primarily depends on operational aspects. The choice is a compromise between the level of risk one wishes to accept, the amount of alarms that one can receive compatibly with intervention strategies, and model predictive performance.

Increasing *n* (i.e., accepting a higher crash risk) quickly decreases the number of "unsafe" intervals, as can be observed in Table A.1. Having a high value of *n* may reduce the number of unsafe intervals to the point that it becomes comparable or even lower than the actual number of observed crashes, cancelling one of the advantages of the conflict-based approach over the crash-based approach (i.e., retrieving more data for training the model). However, the more intervals are considered as "unsafe," the more often alarms are triggered (regardless of the predictive power of the model), which may be undesirable under an operational point of view.

Table A.1. Effect of parameter n *on the number of unsafe intervals.*

n	**Number of unsafe intervals**
1	98,397
2	10,738
3	2,241
4	408
5	64
6	10
7	1
8+	0

In terms of model performance, choosing *n* = 1 may loosen the relationship between the input traffic variable and unsafe situations. Conflicts, as crashes, may be caused not just by traffic characteristics, which are input variables to the RTConfPM, but also, for example, by unprovoked human error (see Figure 1), which is not and cannot be an input variable in the RTConfPM presented in this work. If a 5-minute interval is considered unsafe because a single conflict occurred, then we might include in the unsafe set some situations which are not strictly unsafe from a traffic variables point of view, with negative consequences on model performance. However, choosing a high value of *n* reduces the numerosity of the unsafe class of intervals, which in turn reduces the size of the training dataset, potentially leading to a reduction of model performance. Figure A.2 presents a sensitivity analysis of the performance of the model in the full-test dataset, while increasing *n* from 1 to 5. Parameters used in this sensitivity analysis

are 0.78-second threshold, *k* = 5, 1:1 ratio of safe versus unsafe situations, linear kernel function, and Set #1 input variables. For each *n*, 100 SVMs were trained with Monte Carlo cross-validation technique, and average AUC, specificity and recall were computed. As *n* increases, the relationship between unsafe situations and traffic conditions becomes stronger and the predictive power of the model increases. However, this increase becomes marginal when *n* is higher than 2, and it starts to decrease when *n* is 5, because the size of the training dataset becomes small enough to affect the reliability of the SVM.

Therefore, according to this sensitivity analysis, the best recall performance is obtained with *n* = 3, and because about 25 times more unsafe intervals than crashes exist for the value of *n* (therefore significantly increasing the amount of data available for training the model), we chose to adopt this parameter value in the analyses presented in Section 4.

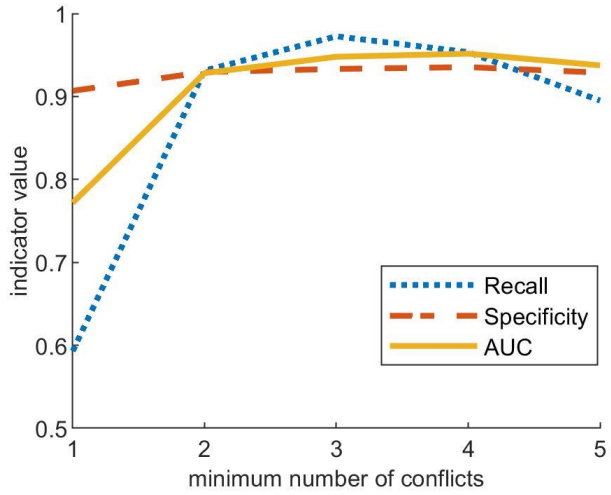

Figure A.2. Sensitivity analysis of the effect of parameter n *on performance indicators (full-test dataset).*

Funding

This work was supported by the University of Padova (project: “Advances in proactive and reactive approaches to road safety: development of innovative crash risk prediction and automatic incident detection systems". Grant number: BIRD177377/17).

Declaration of interests

None

References

Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accid. Anal. Prev. 43 1 , 461–470. doi:10.1016/j.aap.2010.10.002

Abdel-Aty, M., Pande, A., Das, A., Knibbe, W.J., 2008. Assessing safety on Dutch freeways with data from infrastructure-based intelligent transportation systems. Transp. Res. Rec. 2083 1 , 153–161. doi:10.3141/2083-18

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M.F., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. Transp. Res. Rec. 1897 1 , 88–95. doi:10.3141/1897-12

Abdel-Aty, M., Wang, L., 2017. Implementation of Variable Speed Limits to Improve Safety of Congested Expressway Weaving Segments in Microsimulation. Transp. Res. Procedia 27, 577–584. doi:10.1016/j.trpro.2017.12.061

Ahmed, M., Abdel-Aty, M., Yu, R., 2012. Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. Transp. Res. Rec. 2280 1 , 51–59. doi:10.3141/2280-06

Basso, F., Basso, L.J., Bravo, F., Pezoa, R., 2018. Real-time crash prediction in an urban expressway using disaggregated data. Transp. Res. Part C Emerg. Technol. 86, 202–219. doi:10.1016/j.trc.2017.11.014

Basso, F., Basso, L.J., Pezoa, R., 2020. The importance of flow composition in real-time crash prediction. Accid. Anal. Prev. 137, 105436. doi:https://doi.org/10.1016/j.aap.2020.105436

Blagus, R., Lusa, L., 2013. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 14 1 , 106. doi:10.1186/1471-2105-14-106

Breiman L., 2001. Random Forests. Mach. Learn. 45 1 , 5–32. doi:10.1023/A:1010933404324

Caleffi, F., Anzanello, M.J., Cybis, H.B.B., 2017. A multivariate-based conflict prediction model for a Brazilian freeway. Accid. Anal. Prev. 98, 295–302. doi:10.1016/j.aap.2016.10.025

Cavadas, J., Azevedo, C.L., Farah, H., Ferreira, A., 2020. Road safety of passing maneuvers: a bivariate extreme value theory approach under non-stationary conditions. Accid. Anal. Prev. 134, 105315. doi:10.1016/j.aap.2019.105315

Chawla, N. V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. doi:10.1613/jair.953

Coles, S.G., 2001. An introduction to statistical modeling of extreme values, Springer Series in Statistics. London. doi:10.1007/978-1-4471-3675-0

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. doi:10.1007/bf00994018

Daganzo, C.F., 1997. Fundamentals of Transportation and Traffic Operations, Fundamentals of Transportation and Traffic Operations. doi:10.1108/9780585475301

Davis, G.A., Hourdos, J., Xiong, H., Chatterjee, I., 2011. Outline for a causal model of traffic conflicts and crashes. Accid. Anal. Prev. 43 6 , 1907–1919. doi:10.1016/j.aap.2011.05.001

Dimitriou, L., Stylianou, K., Abdel-Aty, M.A., 2018. Assessing rear-end crash potential in urban locations based on vehicle-by-vehicle interactions, geometric characteristics and operational conditions. Accid. Anal. Prev. 118, 221–235. doi:10.1016/j.aap.2018.02.024

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. doi:10.1016/j.aap.2015.05.018

Eltrass, A., Khalil, M., 2018. Automotive radar system for multiple-vehicle detection and tracking in urban environments. IET Intell. Transp. Syst. 12 8 , 783–792. doi:10.1049/iet-its.2017.0370

Essa, M., Sayed, T., 2018. Traffic conflict models to evaluate the safety of signalized intersections at the cycle level. Transp. Res. Part C Emerg. Technol. 89, 289–302. doi:10.1016/j.trc.2018.02.014

Farah, H., Azevedo, C.L., 2017. Safety analysis of passing maneuvers using extreme value theory. IATSS Res. 41 1 , 12–21. doi:10.1016/j.iatssr.2016.07.001

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27 8 , 861–874. doi:10.1016/j.patrec.2005.10.010
Fish, K.E., Blodgett, J.G., 2003. A visual method for determining variable importance in an artificial neural network model: An empirical benchmark study. J. Targeting, Meas. Anal. Mark. 11 3 , 244–254. doi:10.1057/palgrave.jt.5740081
Flach, P., Hernández-Orallo, J., Ferri, C., 2011. A coherent interpretation of AUC as a measure of aggregated classification performance, in: Proceedings of the 28th International Conference on Machine Learning, ICML 2011.
Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., Yuan, J., 2020. Predicting real-time traffic conflicts using deep learning. Accid. Anal. Prev. 136, 105429. doi:10.1016/j.aap.2019.105429
Fu, C., Sayed, T., Zheng, L., 2020. Multivariate Bayesian hierarchical modeling of the non-stationary traffic conflict extremes for crash estimation. Anal. Methods Accid. Res. 28, 100135. doi:10.1016/j.amar.2020.100135
Gastaldi, M., Orsini, F., Gecchele, G., Rossi, R., 2021. Safety analysis of unsignalized intersections: a bivariate extreme value approach. Transp. Lett. 13 3 , 209–218. doi:10.1080/19427867.2020.1861503
Gu, X., Abdel-Aty, M., Lee, J., Xiang, Q., Ma, Y., 2020. Identification of contributing factors for interchange crashes based on a quasi-induced exposure method. J. Transp. Saf. Secur. doi:10.1080/19439962.2020.1812783
Güttinger, V.A., 1984. Conflict Observation in Theory and in Practice, in: International Calibration Study of Traffic Conflict Techniques. doi:10.1007/978-3-642-82109-7_3
Hayward, J.C., 1972. Near-miss determination through use of a scale of danger. Highw. Res. Rec. doi:TTSC 7115
He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 9 , 1263–1284. doi:10.1109/TKDE.2008.239
Hossain, M., Abdel-Aty, M., Quddus, M.A., Muromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. Accid. Anal. Prev. 124, 66–84. doi:10.1016/j.aap.2018.12.022
Hossain, M., Muromachi, Y., 2012. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. Accid. Anal. Prev. 45, 373–381. doi:10.1016/j.aap.2011.08.004
Hossain, M., Muromachi, Y., 2011. Understanding Crash Mechanisms and Selecting Interventions to Mitigate Real-Time Hazards on Urban Expressways. Transp. Res. Rec. J. Transp. Res. Board 2213, 53–62. doi:10.3141/2213-08
Hughes, R.G., Council, F.M., 1999. Exploratory Investigation of Relationship Between Congestion and Safety in Terms of Real Time Traffic Performance, in: 78th Annual Meeting of the Transportation Research Board, Washington, DC.
Hydén, C., 1987. The development of a method for traffic safety evaluation:the Swedish traffic conflicts technique. Bull. Lund Inst. od Technol. 70.
Imprialou, M., Quddus, M., 2019. Crash data quality for road safety research: Current state and future directions. Accid. Anal. Prev. 130, 84–90. doi:10.1016/j.aap.2017.02.022
Katrakazas, C., Antoniou, C., Yannis, G., 2019. Time Series Classification Using Imbalanced Learning for Real-Time Safety Assessment, in: Transportation Research Board 98th Annual Meeting.
Katrakazas, C., Quddus, M., Chen, W.H., 2018. A Simulation Study of Predicting Real-Time Conflict-Prone Traffic Conditions. IEEE Trans. Intell. Transp. Syst. 19 10 , 3196–3207. doi:10.1109/TITS.2017.2769158
Katrakazas, C., Quddus, M., Chen, W.H., 2017. A simulation study of predicting conflict-prone traffic conditions in Real-time, in: 96th Annual Meeting of Transportation Research Board (TRB).
Ke, J., Zhang, S., Yang, H., Chen, X. (Michael), 2018. PCA-based missing information imputation for real-time crash likelihood prediction under imbalanced data. Transp. A Transp. Sci. 15 2 , 872–895. doi:10.1080/23249935.2018.1542414
Knoop, V., Hoogendoorn, S.P., Van Zuylen, H., 2009. Empirical differences between time mean speed and space mean speed, in: Traffic and Granular Flow 2007. pp. 351–356. doi:10.1007/978-3-540-77074-9_36
Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of crash precursors on instrumented freeways. Transp. Res. Rec. 1784 1 , 1–8. doi:10.3141/1784-01
Li, J., Guo, J., Wijnands, J.S., Yu, R., Xu, C., Stevenson, M., 2020. Assessing injury severity of secondary incidents using support vector machines. J. Transp. Saf. Secur. doi:10.1080/19439962.2020.1754983
Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. Accid. Anal. Prev. 135, 105371.
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2 3 , 18–22.
Lv, Y., Tang, S., Zhao, H., 2009a. Real-time highway traffic accident prediction based on the k-nearest neighbor method, in: 2009 International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2009. doi:10.1109/ICMTMA.2009.657
Lv, Y., Tang, S., Zhao, H., Li, S., 2009b. Real-time highway accident prediction based on support vector machines, in: 2009 Chinese Control and Decision Conference, CCDC 2009. doi:10.1109/CCDC.2009.5192409
Madanat, S., Liu, P.-C., 1995. A prototype system for real-time incident likelihood prediction. United States. Federal Highway Administration.
Menon, A.K., 2009. Large-scale support vector machines: algorithms and theory. Res. Exam, Univ. California, San Diego.
Nadimi, N., Ragland, D.R., Mohammadian Amiri, A., 2020. An evaluation of time-to-collision as a surrogate safety measure and a proposal of a new method for its application in safety analysis. Transp. Lett. 12 7 , 491–500. doi:10.1080/19427867.2019.1650430

Orsini, F., Gecchele, G., Gastaldi, M., Rossi, R., 2020a. Large-scale road safety evaluation using extreme value theory. IET Intell. Transp. Syst. 14 9 , 1004–1012. doi:10.1049/iet-its.2019.0633
Orsini, F., Gecchele, G., Gastaldi, M., Rossi, R., 2020b. Transferability and seasonality in extreme value theory applications to road safety: a case study in an Italian motorway. Adv. Transp. Stud. 2 Special Issue , 33–46.
Orsini, F., Gecchele, G., Gastaldi, M., Rossi, R., 2019. Collision prediction in roundabouts: a comparative study of extreme value theory approaches. Transp. A Transp. Sci. 15 2 , 556–572. doi:10.1080/23249935.2018.1515271
Osman, O.A., Hajij, M., Bakhit, P.R., Ishak, S., 2019. Prediction of Near-Crashes from Observed Vehicle Kinematics using Machine Learning. Transp. Res. Rec. 2673 12 , 463–473. doi:10.1177/0361198119862629
Paikari, E., Moshirpour, M., Alhajj, R., Far, B.H., 2014. Data integration and clustering for real time crash prediction, in: Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration, IEEE IRI 2014. doi:10.1109/IRI.2014.7051936
Pande, A., Abdel-Aty, M., 2006. Assessment of freeway traffic parameters leading to lane-change related collisions. Accid. Anal. Prev. 38 5 , 936–948. doi:10.1016/j.aap.2006.03.004
Pande, A., Abdel-Aty, M., 2005. A freeway safety strategy for advanced proactive traffic management. J. Intell. Transp. Syst. Technol. Planning, Oper. 9 3 , 145–158. doi:10.1080/15472450500183789
Park, H., Haghani, A., 2016. Real-time prediction of secondary incident occurrences using vehicle probe data. Transp. Res. Part C Emerg. Technol. 70, 69–85. doi:10.1016/j.trc.2015.03.018
Park, H., Haghani, A., Samuel, S., Knodler, M.A., 2018. Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion. Accid. Anal. Prev. 112, 39–49. doi:10.1016/j.aap.2017.11.025
Parsa, A.B., Taghipour, H., Derrible, S., Mohammadian, A.K., 2019. Real-time accident detection: Coping with imbalanced data. Accid. Anal. Prev. 129, 202–210. doi:10.1016/j.aap.2019.05.014
Peng, Y., Li, C., Wang, K., Gao, Z., Yu, R., 2020. Examining imbalanced classification algorithms in predicting real-time traffic crash risk. Accid. Anal. Prev. 144, 105610. doi:10.1016/j.aap.2020.105610
Perkins, S.R., Harris, J.L., 1968. Traffic conflict characteristics-accident potential at intersections. Highw. Res. Rec. 225 .
Qu, X., Wang, Wei, Wang, Wenfu, Liu, P., 2013. Real-time freeway sideswipe crash prediction by support vector machine. IET Intell. Transp. Syst. 7 4 , 445–453. doi:10.1049/iet-its.2011.0230
Roy, A., Hossain, M., Muromachi, Y., 2018. Enhancing the Prediction Performance of Real-Time Crash Prediction Models: A Cell Transmission-Dynamic Bayesian Network Approach. Transp. Res. Rec. 2672 38 , 58–68. doi:10.1177/0361198118797802
Sadeq, H., Sayed, T., 2016. Automated Roundabout Safety Analysis: Diagnosis and Remedy of Safety Problems. J. Transp. Eng. 142 1–8 , Content ID 04016062. doi:10.1061/(ASCE)TE.1943-5436.0000887
Santos, M.S., Soares, J.P., Abreu, P.H., Araujo, H., Santos, J., 2018. Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [Research Frontier]. IEEE Comput. Intell. Mag. 13 4 , 59–76. doi:10.1109/MCI.2018.2866730
Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. Revstat Stat. J.
Shi, Q., Abdel-Aty, M., 2015. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. Transp. Res. Part C Emerg. Technol. doi:10.1016/j.trc.2015.02.022
Simon, R., 2007. Resampling strategies for model assessment and selection, in: Fundamentals of Data Mining in Genomics and Proteomics. doi:10.1007/978-0-387-47509-7-8
Songchitruksa, P., Tarko, A.P., 2006. The extreme value theory approach to safety estimation. Accid. Anal. Prev. 38 4 , 811–822. doi:10.1016/j.aap.2006.02.003
St-Aubin, P., Saunier, N., Miranda-Moreno, L., 2015. Large-scale automated proactive road safety analysis using video data. Transp. Res. Part C Emerg. Technol. 58, 363–379. doi:10.1016/j.trc.2015.04.007
Stehman, S. V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sens. Environ. 62 1 , 77–89. doi:10.1016/S0034-4257(97)00083-7
Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9, 307. doi:10.1186/1471-2105-9-307
Sun, Y., Kamel, M.S., Wong, A.K.C., Wang, Y., 2007. Cost-sensitive boosting for classification of imbalanced data. Pattern Recognit. 40 12 , 3358–3378. doi:10.1016/j.patcog.2007.04.009
Tarko, A.P., 2019. Measuring road safety with surrogate events, Measuring Road Safety with Surrogate Events. Elsevier. doi:10.1016/C2016-0-00255-3
Tarko, A.P., 2018a. Chapter 17. Surrogate Measures of Safety, in: Safe Mobility: Challenges, Methodology and Solutions. pp. 383–405. doi:10.1108/S2044-994120180000011019
Tarko, A.P., 2018b. Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution – A theoretical and numerical exploration. Accid. Anal. Prev. 113, 63–73. doi:10.1016/j.aap.2018.01.008
Tarko, A.P., 2012. Use of crash surrogates and exceedance statistics to estimate road safety. Accid. Anal. Prev. 45, 230–240. doi:10.1016/j.aap.2011.07.008
Tharwat, A., 2018. Classification assessment methods. Appl. Comput. Informatics. doi:10.1016/j.aci.2018.08.003

Theofilatos, A., Chen, C., Antoniou, C., 2019. Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction. Transp. Res. Rec. 2673 8 , 169–178.
Thompson, P., Cai, Y., Reeve, D., Stander, J., 2009. Automated threshold selection methods for extreme wave analysis. Coast. Eng. 56 10 , 1013–1021. doi:10.1016/j.coastaleng.2009.06.003
Vapnik, V.N., 1995. The Nature of Statistical Learning Theory, The Nature of Statistical Learning Theory. Springer. doi:10.1007/978-1-4757-2440-0
Wang, C., Xu, C., Xia, J., Qian, Z., Lu, L., 2018. A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation. Transp. Res. Part C Emerg. Technol. 90, 281–291. doi:10.1016/j.trc.2018.03.011
Wang, J., Luo, T., Fu, T., 2019. Crash prediction based on traffic platoon characteristics using floating car trajectory data and the machine learning approach. Accid. Anal. Prev. 133, 105320. doi:10.1016/j.aap.2019.105320
Wang, L., Abdel-Aty, M., Lee, J., 2017. Safety analytics for integrating crash frequency and real-time risk modeling for expressways. Accid. Anal. Prev. 104, 58–64. doi:10.1016/j.aap.2017.04.009
Wang, L., Abdel-Aty, M., Lee, J., Shi, Q., 2019. Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. Accid. Anal. Prev. 122, 378–384. doi:10.1016/j.aap.2017.06.003
Wang, P., Chan, C.Y., 2017. Vehicle collision prediction at intersections based on comparison of minimal distance between vehicles and dynamic thresholds. IET Intell. Transp. Syst. 11 10 , 676–684. doi:10.1049/iet-its.2017.0065
Wang, Pengfei, Chen, K., Zhu, S., Wang, Peng, Zhang, H., 2020. Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. Resour. Conserv. Recycl. doi:10.1016/j.resconrec.2020.104814
White, H., 2006. Impact evaluation: the experience of the Independent Evaluation Group of the World Bank.
Wu, Y., Abdel-Aty, M., Cai, Q., Lee, J., Park, J., 2018a. Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data. Transp. Res. Part C Emerg. Technol. 87, 11–25. doi:10.1016/j.trc.2017.12.012
Wu, Y., Abdel-Aty, M., Lee, J., 2018b. Crash risk analysis during fog conditions using real-time traffic data. Accid. Anal. Prev. 114, 4–11. doi:10.1016/j.aap.2017.05.004
Xu, C., Tarko, A.P., Wang, W., Liu, P., 2013a. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accid. Anal. Prev. 57, 30–39. doi:10.1016/j.aap.2013.03.035
Xu, C., Wang, W., Liu, P., 2013b. A genetic programming model for real-time crash prediction on freeways. IEEE Trans. Intell. Transp. Syst. 14 2 , 574–586. doi:10.1109/TITS.2012.2226240
Xu, Q.S., Liang, Y.Z., 2001. Monte Carlo cross validation. Chemom. Intell. Lab. Syst. 56 1 , 1–11. doi:10.1016/S0169-7439(00)00122-2
Yasmin, S., Eluru, N., Wang, L., Abdel-Aty, M.A., 2018. A joint framework for static and real-time crash risk analysis. Anal. Methods Accid. Res. doi:10.1016/j.amar.2018.04.001
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. doi:10.1016/j.ssci.2013.10.012
Yu, R., Abdel-Aty, M., 2013a. Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes. Accid. Anal. Prev. 58, 97–105. doi:10.1016/j.aap.2013.04.025
Yu, R., Abdel-Aty, M., 2013b. Utilizing support vector machine in real-time crash risk evaluation. Accid. Anal. Prev. 51, 252–259. doi:10.1016/j.aap.2012.11.027
Yu, R., Abdel-Aty, M., Ahmed, M., 2013. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accid. Anal. Prev. doi:10.1016/j.aap.2012.05.011
Yu, R., Abdel-Aty, M.A., Ahmed, M.M., Wang, X., 2014. Utilizing microscopic traffic and weather data to analyze real-time crash patterns in the context of active traffic management. IEEE Trans. Intell. Transp. Syst. 15 1 , 205–213. doi:10.1109/TITS.2013.2276089
Yuan, J., Abdel-Aty, M., 2018. Approach-level real-time crash risk analysis for signalized intersections. Accid. Anal. Prev. 119, 274–289. doi:10.1016/j.aap.2018.07.031
Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network. Transp. Res. Rec. 2673 4 , 314–326. doi:10.1177/0361198119840611
Yuan, J., Abdel-Aty, M., Wang, L., Lee, J., Yu, R., Wang, X., 2018. Utilizing bluetooth and adaptive signal control data for real-time safety analysis on urban arterials. Transp. Res. Part C Emerg. Technol. 97, 114–127. doi:10.1016/j.trc.2018.10.009
Zhang, S., Abdel-Aty, M., Wu, Y., Zheng, O., 2020. Modeling pedestrians' near-accident events at signalized intersections using gated recurrent unit (GRU). Accid. Anal. Prev. 148, 105844. doi:10.1016/j.aap.2020.105844
Zheng, L., Ismail, K., Meng, X., 2014. Freeway safety estimation using extreme value theory approaches: A comparative study. Accid. Anal. Prev. 62, 32–41. doi:10.1016/j.aap.2013.09.006
Zheng, L., Ismail, K., Sayed, T., Fatema, T., 2018. Bivariate extreme value modeling for road safety estimation. Accid. Anal. Prev. 120, 83–91. doi:https://doi.org/10.1016/j.aap.2018.08.004
Zheng, L., Sayed, T., 2020. A novel approach for real time crash prediction at signalized intersections. Transp. Res. Part C Emerg. Technol. 117. doi:10.1016/j.trc.2020.102683
Zheng, L., Sayed, T., Essa, M., 2019. Bayesian hierarchical modeling of the non-stationary traffic conflict extremes for crash estimation. Anal. Methods Accid. Res. 23. doi:10.1016/j.amar.2019.100100

Zheng, Z., Ahn, S., Monsere, C.M., 2010. Impact of traffic oscillations on freeway crash occurrences. Accid. Anal. Prev. 42 2 , 626–636. doi:10.1016/j.aap.2009.10.009