Accident Analysis and Prevention 129 (2019) 202–210

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Real-time accident detection: Coping with imbalanced data

Amir Bahador Parsa[a,*], Homa Taghipour[a], Sybil Derrible[b], Abolfazl (Kouros) Mohammadian[c]

[a] Department of Civil and Materials Engineering, University of Illinois at Chicago, 842 W Taylor St, 2054 ERF, Chicago, IL 60607, United States
[b] Department of Civil and Materials Engineering, Institute of Environmental Science and Policy, University of Illinois at Chicago, 842 W Taylor St, 2071 ERF, Chicago, IL 60607, United States
[c] Department of Civil and Materials Engineering, University of Illinois at Chicago, 842 W Taylor St, 2093 ERF, Chicago, IL 60607, United States

ARTICLE INFO

Keywords:
Accident detection
Real-time data
Probabilistic neural network
Support vector machine
Machine learning

ABSTRACT

Detecting accidents is of great importance since they often impose significant delay and inconvenience to road users. This study compares the performance of two popular machine learning models, Support Vector Machine (SVM) and Probabilistic Neural Network (PNN), to detect the occurrence of accidents on the Eisenhower expressway in Chicago. Accordingly, since the detection of accidents should be as rapid as possible, seven models are trained and tested for each machine learning technique, using traffic condition data from 1 to 7 min after the actual occurrence. The main sources of data used in this study consist of weather condition, accident, and loop detector data. Furthermore, to overcome the problem of imbalanced data (i.e., underrepresentation of accidents in the dataset), the Synthetic Minority Oversampling TEchnique (SMOTE) is used. The results show that although SVM achieves overall higher accuracy, PNN outperforms SVM regarding the Detection Rate (DR) (i.e., percentage of correct accident detections). In addition, while both models perform best at 5 min after the occurrence of accidents, models trained at 3 or 4 min after the occurrence of an accident detect accidents more rapidly while performing reasonably well. Lastly, a sensitivity analysis of PNN for Time-To-Detection (TTD) reveals that the speed difference between upstream and downstream of accidents location is particularly significant to detect the occurrence of accidents.

1. Introduction

Traffic conditions of urban expressways are generally expected to follow patterns that do not vary significantly. Nonetheless, several factors can dramatically affect these patterns. For example, weather condition, road work, and the occurrence of an accident can significantly affect traffic conditions. Nowadays, weather condition can be easily forecasted and its impact on traffic condition can be reflected through Advanced Traveler Information Systems (ATIS). Similarly, road works are scheduled, and they can be announced through Variable Message Signs (VMS). The circumstance is different for accidents since predicting their exact time and location is virtually impossible. The next best option is to be able to detect their occurrence rapidly. In other words, the faster accidents are detected and announced to road users, the less delay and inconvenience is imposed on them since they now have the option to take detours. In addition, detecting accurately and rapidly the occurrence of accidents also leads to the swift dispatch of emergency services as well.

Along with various researches regarding analyzing accidents (Jalayer et al., 2018; Mahmoudzadeh et al., 2019; Razi-Ardakani et al., 2018), accident detection has been a well-studied topic in the transportation community at least since the 1970s. Based on traffic flow theory, the California algorithm was developed in 1978, and it detects the occurrence of accidents when traffic-related variables exceed specific thresholds (Payne and Tignor, 1978). Many other algorithms were developed earlier on, including the standard normal deviate (Dudek et al., 1974), Bayesian algorithms (Levin and Krause, 1978), and time series models (Ahmed and Cook, 1982) in which historical traffic information are used to detect accidents. Furthermore, there is a wide range of studies aiming to achieve real-time accident detection through different modern methods, including statistical techniques (Khan and Ritchie, 1998; Zhang, 2005), image processing (Hoose et al., 1992; Zifeng, 1997), pattern recognition (Rong et al., 2013; Zhang and Taylor, 2004), and artificial intelligence (Abdulhai and Ritchie, 1999; Adeli and Karim, 2000; Çetiner et al., 2010; Jin et al., 2001; Lu et al., 2012; Motamed, 2016). In particular, machine learning techniques have proven to be powerful in many transportation applications (Lee et al., 2018; Shabanpour et al., 2017). Among machine learning techniques, Probabilistic Neural Network (PNN) and Support Vector Machines (SVM) are two important techniques that have been used to detect

* Corresponding author.
E-mail addresses: aparsa2@uic.edu (A.B. Parsa), htaghi2@uic.edu (H. Taghipour), derrible@uic.edu (S. Derrible), kouros@uic.edu (A.K. Mohammadian).

https://doi.org/10.1016/j.aap.2019.05.014
Received 28 January 2019; Received in revised form 16 May 2019; Accepted 16 May 2019
Available online 03 June 2019
0001-4575/

accidents (Dong et al., 2015; Jin et al., 2002; Li et al., 2016; Lu et al., 2012; Yu and Abdel-Aty, 2013; Yuan and Cheu, 2003). Yuan and Cheu (2003) have employed SVM to detect accidents in two types of roadways. They notably showed that SVM has lower false alarm rate than other incident detection techniques such as multi-layer feed forward neural network and PNN. In comparison with other families of machine learning techniques such as Neural Network (NN), SVM is famous due to its capability to cope with small sample sizes (Yu and Abdel-Aty, 2013). Nonetheless, SVM cannot cope easily with imbalanced big datasets (You et al., 2017).

In addition, as an emerging field, feature engineering can play an important role to increase model performance, especially when using non-parametric techniques such as many machine learning techniques. Essentially, feature engineering is about creating new and meaningful features (i.e., variables) based on domain knowledge of data to increase model performance. In transportation, many studies have successful applied feature engineering including to create hand-crafted (e.g., Xiao et al., 2017; Zhu et al., 2018) and learned (e.g., Dabiri and Heaslip, 2018; Endo et al., 2016) features (Li et al., 2010).

The required data from Automatic Incident Detection (AID) systems can be collected in several ways such as using loop detectors, probe vehicles, and image processing tools. Each method has its own advantages and disadvantages (Yuan and Cheu, 2003). Among them, the use of loop detectors is most common since, unlike other methods, it is not sensitive to outdoor environment factors such as rain and snow (Rossi et al., 2016), and it has a relatively low cost, good performance, and large knowledge base (Nikolaev et al., 2017). Traffic related data captured by loop detectors has been aggregated into various time intervals by different researchers. For instance, Ahmed et al. (2012) suggested intervals of 5 min for aggregating traffic variables, Ozbayoglu et al. (2017) used the aggregation of volume, speed, and occupancy every 2 min, and Katrakazas et al. (2016) took the weighted average of speed, volume, and travel time in 15-minute intervals.

Another important issue regarding data is the proportion of accident and non-accident records in the dataset. Many researchers follow the traditional ratio of accident to non-accident as 1:4 and therefore take 4 non-accident cases for each accident case (You et al., 2017). Nevertheless, this approach can lead to biases because the number of non-accident cases (i.e., the number of times in which no accident happens) is much larger than the number of accident cases in reality. For example, in a specific section of an expressway, one accident case might happen after thousands of non-accident cases. In addition, inserting a larger number of non-accident cases in the dataset provides more case studies from which machine learning techniques can be trained, and potentially producing more accurate models. From a statistical viewpoint, however, this practice leads to imbalanced data.

Dealing with imbalanced data is a novel and ongoing field of study for which researchers are trying to improve and utilize different techniques. Generally, oversampling and under-sampling are the main methods to deal with imbalanced data (Oqab et al., 2016; Ozbayoglu et al., 2017). That being said, on the one hand, oversampling can lead to overfitting since data points of a minority class are duplicated. On the other hand, under-sampling can cause some important data points of majority class to be excluded from the final sample. To overcome these issues, a powerful method named Synthetic Minority Oversampling TEchnique (SMOTE) was introduced by Chawla et al. (2002) in which new synthetic data points are created by forming a convex combination of neighboring members. In particular, SMOTE and its variants have been shown to be successful when only a few samples are available (Al-azani, 2017; Fernández et al., 2017; Maldonado et al., 2019; Verbiest et al., 2014). One advantage of SMOTE is that synthesizing minority class samples results in larger and less specific decision regions (Han et al., 2005). Moreover, in general, it has been observed that SMOTE is more robust than undersampling approaches, especially when dealing with noisy (Kaur and Gosain, 2018), large, and sparse datasets (Vanhoeyveld and Martens, 2018). Accordingly, several variants of SMOTE have been also introduced and tested by different scholars, such as: borderline SMOTE (Han et al., 2005) that creates new samples from minority classes close to the borderline between the classes; Adaptive Synthetic Sampling (ADASYN) (He et al., 2008); and SVM SMOTE (Tang et al., 2009). Overall, the popularity and success of SMOTE technique and its variants can stem from three factors: simplicity, superior performance, and computational efficiency (He and Garcia, 2008; Sun et al., 2009). To the best knowledge of the authors, SMOTE has not been applied to imbalanced data for accident detection purposes to date.

Overall, the main objectives of this study are twofold. First, it is to test the performance of two powerful supervised machine learning methods—Support Vector Machine (SVM) and Probabilistic Neural Network (PNN)—to detect the occurrence of accidents on an urban expressway by exploiting feature engineering of spatiotemporal data along with applying SMOTE to deal with imbalanced data; Random Forest (RF) was also considered but it did not perform as well (see Supplementary Materials). Second, it is to determine the optimal number of minutes—between 1 to 7—when the accidents can be best detected after their occurrence.

The rest of this article is organized as follows. First, the scope of this study and descriptions of the data are provided. Second, in the methodology section, the modeling approach is presented in detail. Next, the results are reported and interpreted. Finally, a discussion and a conclusion are offered. Overall, this work fits within the general endeavor to make cities smarter and more resilient (Derrible, 2019, 2017; Kermanshah et al., 2014; Kermanshah and Derrible, 2017; Mohareb et al., 2014).

2. Data and application

In this study, a 14-mile stretch of the Eisenhower expressway in the city of Chicago is selected. The Eisenhower expressway is an interstate highway that connects the Chicago loop to the north west side of Chicago, where it meets the I-90. Fig. 1 shows the Eisenhower expressway, and the stretch selected for this study is shown in red. There are 24 operational loop detectors located on this stretch of the Eisenhower. Dividing it into 23 sections gives us an average section length of approximately 1 km (i.e., 0.6 mile). For each of these sections, two loop detectors are located at the beginning and end of the section and they capture the traffic conditions in the upstream and downstream directions. Since the sections' length are relatively short, when an accident happens, and regardless of its location within the section, the changes in the traffic conditions can be rapidly detected by the loop detectors upstream and downstream of the accident location.

This study uses loop detector, accident, and weather condition data from June 2017 to December 2017. In what follows, the three sources of data are described.

2.1. Accident data

For this study, accident data is provided by the Illinois Department of Transportation (IDOT), and it includes 32 accidents in the studied area from June to December 2017. This number of accidents is aligned with the studies dealing with imbalanced data to detect accidents (Jin et al., 2002, 2001; Lu et al., 2012; Ozbayoglu et al., 2017). To properly train machine learning models, a comprehensive dataset is required, which should include almost all possible traffic conditions for both accident and non-accident cases. Although, a 1:4 ratio of accident to non-accident has been selected traditionally, in this study, a lower ratio is preferred to increase the accuracy of the model as discussed in the Introduction. Accordingly, after performing some preliminary analysis, 24 non-accident cases per day (i.e., one case per hour) are selected from each of the 23 sections from June to December 2017. Therefore, after eliminating erroneous non-accident cases, a total number of 85,182 non-accident cases and 32 accident cases constitute the dataset. Since

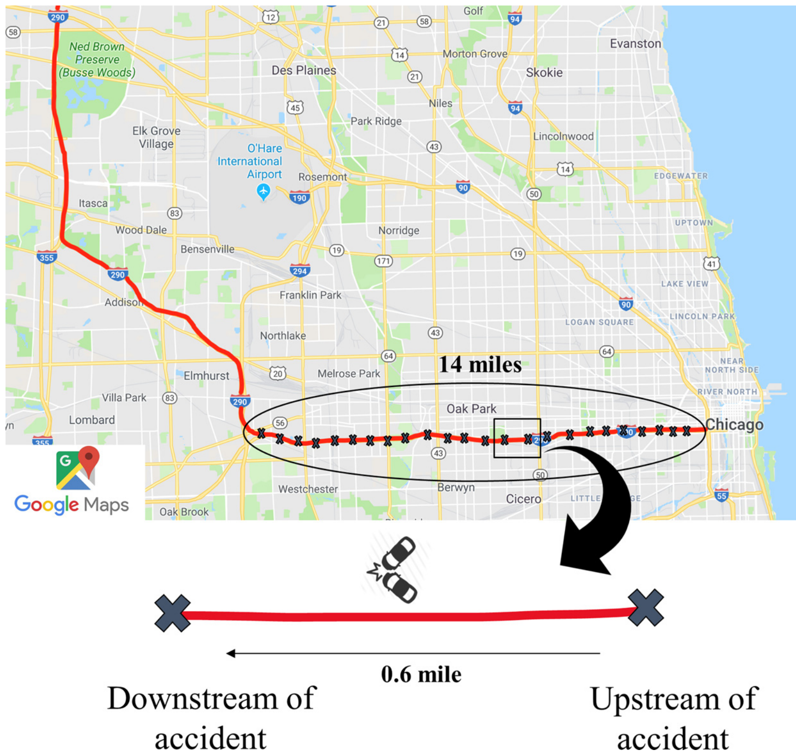

Fig. 1. Eisenhower expressway, Chicago (Google Map).

this dataset is highly imbalanced, new synthetic data points are generated from the 32 accident cases, and the number of accidents is increased to 85,182 cases similar to the number of non-accident cases.

2.2. Loop detector data

One of the main sources of data for accident detection is loop detector data. There are 24 loop detectors located in the studied area that collect volume, occupancy, and vehicle speed every 20 seconds. The loop detector dataset contains many erroneous and missing values, due to detector malfunctioning, roadworks that cover detectors, as well as other factors. Thus, the first step in using loop detector data is to identify erroneous and missing data, and use data imputation and cleansing methods. For data cleansing, several single and combined thresholds in terms of volume, occupancy, and speed are employed to find erroneous or missing data that are then imputed using spatio-temporal data—i.e., using the average of the data points from the previous and the next loop detectors or using the average of the data points from the previous and next timestamps of a given loop detector. In addition, principles of flow conservation are applied to the dataset to find any malfunctioning of consecutive loop detectors (Vanajakshi, 2004).

After applying these data cleansing techniques, traffic-related data is aggregated in intervals of 1 min since 20-second intervals are too short to capture the effect of accidents accurately; i.e., high fluctuations can occur in significantly short time intervals. Nonetheless, aggregating the data to 5-minute or longer intervals would be too long to capture the effect of accidents, especially to ensure a short Time-To-Detection (TTD).

As shown in Fig. 1, when an accident occurs, two loop detectors (upstream and downstream of that accident) are involved. To see how volume, occupancy, and speed are changing per minute during accident versus non-accident occurrences, first, traffic-related variables are collected from 4 min before to 7 min after accident/non-accident cases at the loop detectors in the upstream and downstream directions. Then, to see the impact of accident on traffic-related variables in a given loop detector, the difference of these variables between consecutive timestamps is calculated for that loop detector. In addition, to observe the impact of accident on traffic flow in the upstream and downstream directions of an accident location, the difference between upstream and downstream traffic-related variables is calculated for each timestamp from the moment of an accident/non-accident to 7 min after.

In addition, the standard deviation of the 12 consecutive time intervals (i.e., 4 min before to 7 min after an accident/non-accident) for volume, occupancy, and speed in the upstream and downstream directions are calculated. These variables are calculated because an abrupt change in traffic conditions during accidents is likely to create significant traffic fluctuations that can be captured by a higher standard deviation. Indeed, Fig. 2 compares the average standard deviation of traffic-related variables during accidents (Acc) and non-accidents (N-Acc) in both the upstream and downstream directions, and it shows that the average standard deviation of all traffic related variables for accidents are higher than for non-accidents. Another interesting point that can be derived from Fig. 2 is that the difference between accident and non-accident average standard deviation is higher in the upstream than in the downstream direction. This is likely due to the creation of a shock wave when an accident occurs that affects the traffic conditions upstream more than those downstream. Finally, the difference between accident and non-accident average standard deviation is not as significant for the volume variable in either directions.

2.3. Weather condition data

Another important source of data in this study is weather. The two Chicago airports (Midway and O'Hare) have weather stations that collect and archive weather data. Data from the Chicago Midway Airport is used in this study since it is closer to the studied area.

In total, 94 different variables related to weather conditions are recorded at the weather station, and they are aggregated into four main levels. For this study, an ordinal weather variable is created that goes from 1 for clear or sunny to 4 for harsh stormy, rainy or snowy weather conditions.

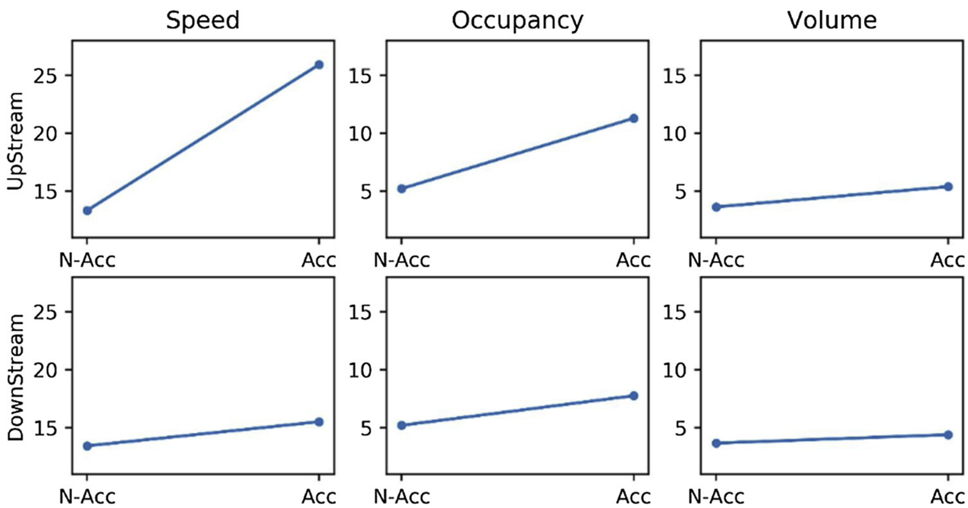

Fig. 2. Comparison of average standard deviation in accident and non-accident conditions.

2.4. Final dataset and pre-processing

The explanatory variables used in this study and their description are displayed in Table 1. In total, seven models are trained separately using SVM and PNN that aim to detect accidents for different TTDs from 1 to 7 min after the occurrence of accident. Therefore, depending on the model, although the type of variables remains the same, the number of variables changes with respect to the TTD used. Moreover, as mentioned above, RF was considered but it did not perform as well as SVM and PNN (see Supplementary Materials).

Among the traffic-related variables, speed and occupancy are found to have more of an impact than volume to detect accidents. Therefore, volume is excluded. Moreover, instead of including the exact value of speed and occupancy, their temporal differences between consecutive timestamps and spatial differences between upstream and downstream directions are used.

3. Methodology

3.1. Synthetic Minority Oversampling TEchnique (SMOTE)

In general, the use of more training data results in a higher performance in machine learning. Accordingly, under-sampling highly imbalanced data causes the loss of a significant portion of information that can impact model accuracy (Han et al., 2005). To overcome this issue, SMOTE oversampling and its variants are used in this study. SMOTE takes each data point of a minority class and produces new data points along the line joining them to their k nearest neighbors. Three types of SMOTE techniques were tested in this study: (1) regular SMOTE (Chawla et al., 2002), (2) borderline SMOTE (Han et al., 2005), and (3) SVM SMOTE (Tang et al., 2009). Eventually, regular SMOTE was selected thanks to its simplicity and the resulting high performance of the trained models. Moreover, in this study, k is assigned to 5; i.e., 5 nearest neighbors. Then, to produce new data points, a random number

Table 1
Description of explanatory variables.

Variables	Description
Speed	
Temporal Difference	
Downstream	Difference between pairs of consecutive speed from 4 minutes before to n minutes after an accident/non-accident*.
Upstream	Difference between pairs of consecutive speed from 4 minutes before to n minutes after an accident/non-accident.
Spatial Difference	
Up-Down	Speed difference between the upstream and downstream directions for each timestamp from an accident/non-accident timestamp to n minutes after.
Occupancy	
Temporal Difference	
Downstream	Difference between pairs of consecutive occupancy from 4 minutes before to n minutes after an accident/non-accident.
Upstream	Difference between pairs of consecutive occupancy from 4 minutes before to n minutes after an accident/non-accident.
Spatial Difference	
Up-Down	Occupancy difference between the upstream and downstream directions for each timestamp from an accident/non-accident timestamp to n minutes after.
Standard Deviation	
Downstream	
Speed	Standard deviation from 4 minutes before to n minutes after an accident/non-accident
Occupancy	.Standard deviation from 4 minutes before to n minutes after an accident/non-accident
Upstream	
Speed	Standard deviation from 4 minutes before to n minutes after an accident/non-accident.
Occupancy	Standard deviation from 4 minutes before to n minutes after an accident/non-accident.
Weather Condition	
Weather	Ordinal variable from 1 for sunny to 4 for stormy weather conditions.
Peak Hour	
Morning	Dummy variable with 1 for weekday morning peak hour and 0 otherwise.
Evening	Dummy variable with 1 for weekday evening peak hour and 0 otherwise.

* n can differ from 1 to 7 based on the model and its required TTD. For each machine learning technique, seven models are trained with respect to different TTDs. For example, to train the model detecting accidents one minute after the occurrence of an accident, n is 1. For the model detecting accidents two minutes after the occurrence of an accident, n is 2, and this procedure is the same for all the models up to a n of 7.

between 0 and 1 is generated and the position of a new data point is defined by multiplying the length of the line between two neighboring data points by this random number.

3.2. Support Vector Machine (SVM)

SVM is a supervised machine learning method mostly used for classification. The method can be employed for high dimensional data and generally leads to accurate classification when coping with small sample size in comparison to the other machine learning methods. The main idea of the SVM method is to generate the optimal separating hyperplane, which can classify the dataset into two classes. In other words, SVM maximizes the margin, which is the distance between the hyperplane and the closest data points of each class (Cortes and Vapnik, 1995). This process is applicable to linearly separable data; however, the method can be extended for the data that is linearly inseparable as well, using data transformation by kernel functions. The kernel functions map the original linearly inseparable data points into a higher dimensional space in which they can be separated linearly. The most popular kernel functions are linear, polynomial, Radial Basis Function (RBF), and sigmoid. In this study, the polynomial kernel function of degree 2 is used, which is shown in Eq. (1) and where X_i and X_j are input vectors.

$$K(X_i, X_j) = (X_i . X_j + 1)^2 \tag{1}$$

Based on the polynomial kernel function, the decision function of SVM (for non-linear classification) follows the form:

$$f(X) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i (X_i . X + 1)^2 + b\right) \tag{2}$$

In Eq. (2), y_i is the classified label, α_i is a lagrange multiplier, X is the input vector to be classified, and b is the intercept of the hyperplane. For more details on the SVM method, Han, et al. (Han et al., 2011) is recommended.

3.3. Probabilistic Neural Network (PNN)

Probabilistic Neural Network (PNN) (also called Naïve Bayes) is a feedforward neural network algorithm mostly used for classification and pattern recognition problems. In this method, at first, the Probability Distribution Function (PDF) of each class of accident and non-accident is estimated from the training samples by using Parzen's method (Parzen, 1962). Then, the Bayesian decision rule is applied on the estimated PDF to assign an input sample into the i_{th} class through Eq. (3) in which h_i is the prior probability that a sample belongs to class i, c_i is the loss resulted from misclassifying the sample, and $f_i(x)$ is the PDF for class i.

$$h_i c_i f_i(x) > h_j c_j f_j(x) \tag{3}$$

Essentially, PNN uses the multivariate kernel estimation with a weight function for the classification. The multivariate kernel estimation with the weight function of Gaussian Kernel for n independent variables can be defined as Eq. (4).

$$\varnothing_i(x_n) = \frac{1}{N_i (2\pi)^{\frac{n}{2}} \sigma^n} \sum_{k=1}^{N_i} EXP\left(\frac{-\|x_n - x_{ik}\|^2}{2\sigma^2}\right) \tag{4}$$

In Eq. (4), x_n is an input vector of n independent variables to which a class should be assigned, x_{ik} is the k^{th} training vector of class i, σ is the standard deviation of kernel function which represents its width, and N_i is number of training vectors of class i. Finally, the classification of the input vector x_n is achieved by applying Bayes decision rule through Eq. (5).

$$C(x) = argmax_i = (\varnothing_i(x_n)) \tag{5}$$

3.4. Model evaluation

Many measures exist to evaluate the performance of a model for classification problems as is the case here. In this study, Accuracy (ACC), Detection Rate (DR), and False Alarm Rate (FAR) are used. The formulas for these three measures are provided in Eqs. (6) to (8):

$$ACC = \frac{Number\ of\ true\ reports}{Total\ number\ of\ cases} \tag{6}$$

$$DR = \frac{Number\ of\ true\ accident\ reports}{Total\ number\ of\ accidents} \tag{7}$$

$$FAR = \frac{Number\ of\ false\ accident\ reports}{Total\ number\ of\ cases} \tag{8}$$

4. Results

All models generated are trained on 65% of the data and tested on the remaining 35%. To cope with imbalanced data, SMOTE and its variants are applied on the training data only. Performance of three SMOTE variants (i.e., regular SMOTE, borderline SMOTE, and SVM SMOTE) in detecting accidents 5 min after their occurrence are compared in Table 2. We can see that the accuracy of regular SMOTE is similar to the two other techniques, but it tends to have a higher detection rate and the false alarm rate tends to be lower. Therefore, regular SMOTE was selected in this study.

In addition, a 5-fold cross-validation procedure is applied on the training data—i.e., the training data is divided into five subsamples randomly, and then four subsamples are kept as the data for training the models and the remaining subsample is used as the validation data. This procedure is repeated five times in such a way that each subsample used exactly once as the validation data, which allows us to measure whether a model is performing well consistently. Furthermore, the SVM and PNN models are trained for seven different TTDs, from 1 min to 7 min after accidents. Fig. 3 shows and compares the results of the final models with respect to three measures of ACC, DR and FAR. It is worth noting that in studies such as this one, the objective is to detect members of a minority class. Therefore, the DR and FAR evaluating measures are more important than ACC since a model could achieve a high accuracy while having a low DR.

In both models, by increasing the TTD from 1 to 5 min after the occurrence of an accident, both ACC and DR increase and reach to their highest value at a TTD of 5 min. Further increasing the TTD to 6 and 7 min then result in a decrease in ACC and DR. FAR also reaches to its lowest value at a TTD of 5 min. From a real-time accident detection viewpoint, a model should detect accidents as rapidly as possible after their occurrence. Put differently, the goal is to have a DR as high as possible and a TTD as low as possible. Based on Fig. 3, the best DR is achieved at a TTD of 5 min. Nonetheless, the DR of models for TTDs of 4 and 3 min remain satisfactory.

Looking into the performance of the two machine learning techniques, PNN achieves best at a TTD of 5 min, with an ACC, DR, and FAR of 99%, 80%, and 0.5% respectively, compared to 99%, 48%, and 0.1% for SVM. Interestingly, although SVM model achieves a higher ACC and a lower FAR on average (i.e., respectively 99% and 0.1% for SVM and

Table 2
Comparison of SMOTE variants.

	ACC		DR		FAR	
	SVM	PNN	SVM	PNN	SVM	PNN
Regular SMOTE	99	99	48	80	0.1	0.5
Borderline SMOTE	99	99	43	72	0.1	0.7
SVM SMOTE	99	98	45	77	0.1	1.1

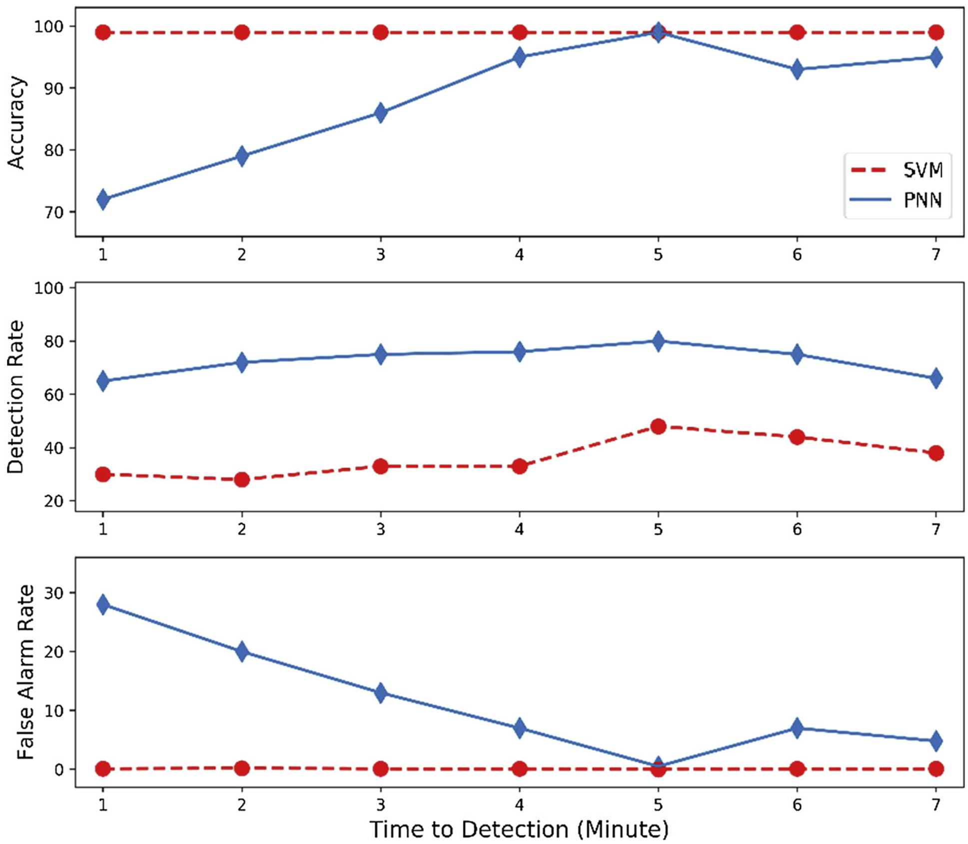

Fig. 3. Results of SVM and PNN models.

88% and 11.5% for PNN), the DR is significantly higher for PNN (i.e., 73% for PNN and 36% for SVM). Regarding the high FAR for PNN at a TTD of 1, 2 and 3 min, it is worth noting that several studies have indicated that even though the number of false alarms might be relatively high in some models, their high DR can still provide useful information on the possibility of occurrence of accidents (Ozbayoglu et al., 2017). This information suggests that, in this context, the use of PNN is preferable.

To explore the results visually, Receiver Operator Characteristic (ROC) curves are plotted and displayed in Fig. 4. ROC curves display the true positive rate (TPR) against the false positive rate (FPR) for both SVM and PNN models in the seven sequential time intervals after accidents. Larger values of the Area Under the ROC Curve (AUC) translate into a better model performance. To this end, the PNN model, which is trained to detect accidents 5 min after the occurrence of an accident, achieves the best result with an AUC of 90%.

In addition, among the traffic-related variables, speed and occupancy are found to be the most significant ones. To further test their significance, they are fed into models in different forms, including as the difference in speed and occupancy between upstream and downstream. Fig. 5 displays the results of a sensitivity analysis of these two variables input in the PNN model with TTD of 5 min to see which one has the greater impact on the probability of accident occurrence.

Based on Fig. 5, both variables have a direct relationship with the probability of accident occurrence, which means that increasing their value leads to an increase in the probability of an accident occurrence. Fig. 5 also shows that the difference in speed between downstream and upstream has a more significant impact with a plateauing effect for percentage changes above 3%.

5. Discussion and conclusion

When an accident occurs, traffic conditions change in the upstream and downstream directions of the accident, and this can impose significant delays to road users. By focusing on a 14-mile stretch of the Eisenhower expressway in the city of Chicago, this study compares the performance of two machine learning models—Support Vector Machine (SVM) and Probabilistic Neural Network (PNN)—to detect the occurrence of accidents by using real-time data. Moreover, one of the novelties of this study is the use of machine learning on different TTDs to detect the occurrence of accidents. Since this dataset is highly imbalanced which consists of 85,182 non-accident and 32 accident cases, SMOTE is also employed.

ACC, DR, and FAR are three performance measures used widely in the literature to evaluate accident detection models. Based on the results, it can be concluded that despite the high ACC and low FAR produced with SVM, PNN performs better due to its higher DR and comparable ACC. In addition, the difference between the downstream and upstream speed is found to have a significant impact to detect accidents. Regarding TTD, it is shown that although both SVM and PNN have the best DR in TTD of 5 min, these models can detect accidents faster in TTD of 4 and even 3 min with only a slight decrease in DR.

When attempting to detect accident occurrence, the number of accidents in a dataset tends to be small, and therefore most studies must cope with highly imbalanced data. To further improve model performance, if available, more spatiotemporal data could be used. Finally, as future work, the performance of deep learning models, such as Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN), could be investigated, especially when used alongside techniques to deal with imbalanced data and that create more data such as SMOTE.

Acknowledgment

The research leading to these findings has received funding from the Illinois Department of Transportation (IDOT) and from the National

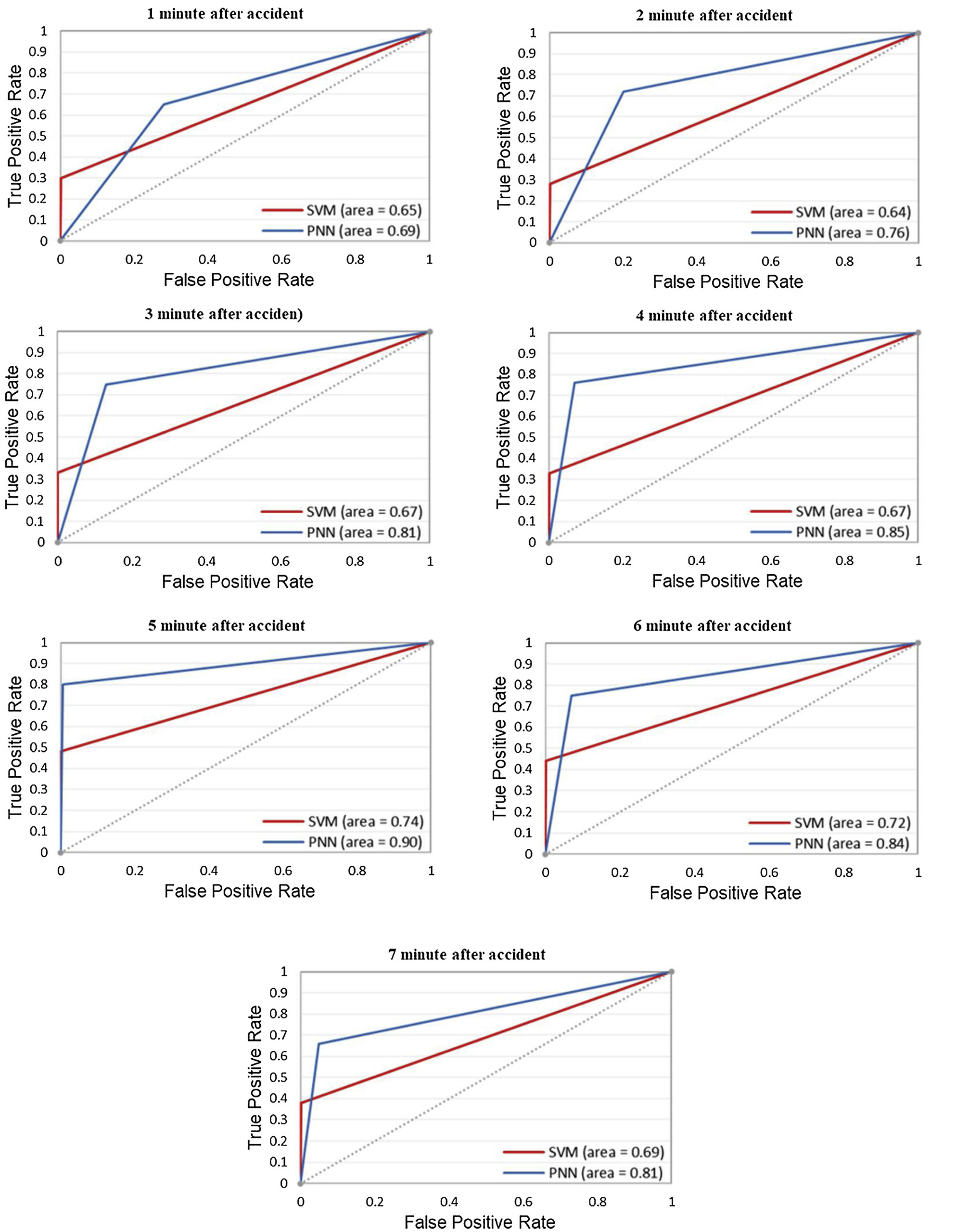

Fig. 4. ROC curves for the SVM and PNN models.

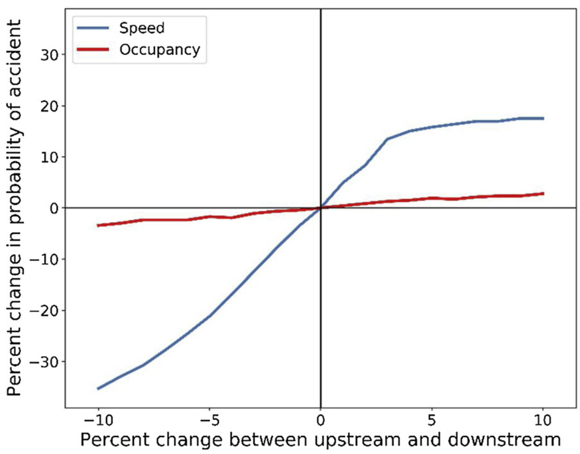

Fig. 5. Sensitivity analysis.

Science Foundation (NSF) CAREER award #155173. The authors would also like to thank IDOT for collecting and archiving the loop detectors data used in this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2019.05.014.

References

Abdulhai, B., Ritchie, S.G., 1999. Enhancing the universality and transferability of freeway incident detection using a Bayesian-based neural network. Transp. Res. Part C Emerg. Technol. 7 (5), 261–280. https://doi.org/10.1016/S0968-090X(99)00022-4.

Adeli, H., Karim, A., 2000. Fuzzy-wavelet RBFNN model for freeway incident detection. J. Transp. Eng. 126 (December (6)), 464–471.

Ahmed, M., Abdel-Aty, M., Yu, R., 2012. Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data. Transp. Res. Rec. J. Transp. Res. Board 2280, 60–67. https://doi.org/10.3141/2280-07.

Ahmed, S.A., Cook, A.R., 1982. Application of time-series analysis techniques to freeway incident detection. Transp. Res. Rec. 841.

Al-azani, S., 2017. ScienceDirect using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short Arabic text. Procedia Comput. Sci. 109, 359–366. https://doi.org/10.1016/j.procs.2017.05.365.

Çetiner, G.B., Sari, M., Borat, O., 2010. A neural network based traffic-flow prediction model. Int. J. Math. Comput. Appl. Res. 15 (2), 269–278. https://doi.org/10.1007/3-540-46016-0_12.

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

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 273–297.

Dabiri, S., Heaslip, K., 2018. Inferring transportation modes from GPS trajectories using a convolutional neural network. Transp. Res. Part C Emerg. Technol. 86 (December), 360–371. https://doi.org/10.1016/j.trc.2017.11.021.

Derrible, S., 2019. Review An approach to designing sustainable urban infrastructure. MRS Energy Sustain. A Rev. J 1–15. https://doi.org/10.1557/mre.2018.14.

Derrible, S., 2017. Urban infrastructure is not a tree: integrating and decentralizing urban infrastructure systems. Environ. Plan. B Urban Anal. City Sci. 44 (3), 553–569. https://doi.org/10.1177/0265813516647063.

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. https://doi.org/10.1016/j.aap.2015.05.018.

Dudek, C.L., Messer, C.J., Nuckles, N.B., 1974. Incident detection on urban freeways. Transp. Res. Rec. 459, 12–24.

Endo, Y., Toda, H., Nishida, K., Kawanobe, A., 2016. Deep feature extraction from trajectories. Pacific-Asia Conf. Knowl. Discov. Data Min 54–66. https://doi.org/10.1007/978-3-319-31750-2.

Fernández, A., Nitesh, R., Herrera, F., 2017. An insight into imbalanced Big Data classification : outcomes and challenges. Complex Intell. Syst. 3 (2), 105–120. https://doi.org/10.1007/s40747-017-0037-9.

Payne, H.J., Tignor, S.C., 1978. Freeway incident detection algorithms based on decision tree with states. Transp. Res. Rec. 30–37.

Han, H., Wang, W., Mao, B., 2005. Borderline-SMOTE : A New Over-Sampling Method in. Int. Conf. Intell. Comput. 878–887.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques.

He, H., Bai, Y., Garcia, E.A., Li, S., 2008. ADASYN: adaptive synthetic sampling approach for imbalanced learning. IEEE Int. Jt. Conf. Neural Networks 3, 1322–1328.

He, H., Garcia, E.A., 2008. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 9, 1263–1284.

Hoose, N., Vicencio, M.A., Zhang, X., 1992. Incident detection in urban roads using computer image processing. Traffic Eng. Control 33.

Jalayer, M., Shabanpour, R., Pour-rouholamin, M., Golshani, N., 2018. Wrong-way driving crashes : A random-parameters ordered probit analysis of injury severity. Accid. Anal. Prev. 117 (April), 128–135. https://doi.org/10.1016/j.aap.2018.04.019.

Jin, X., Cheu, R.L., Srinivasan, D., 2002. Development and adaptation of constructive probabilistic neural network in freeway incident detection. Transp. Res. Part C Emerg. Technol. 10 (2), 121–147. https://doi.org/10.1016/S0968-090X(01)00007-9.

Jin, X., Srinivasan, D., Cheu, R.L., 2001. Classification of freeway traffic patterns for incident detection using constructive probabilistic neural networks. IEEE Trans. Neural Netw. 12 (5), 1173–1187. https://doi.org/10.1109/72.950145.

Katrakazas, C., Quddus, M., Chen, W.-H., 2016. Real-time classification of aggregated traffic conditions using relevance vector machines 2. Transp. Res. Board 95th Annu. Meet 16–3417.

Kaur, P., Gosain, A., 2018. Comparing the behavior of oversampling and undersampling approach of class imbalance learning by combining class imbalance problem with noise. ICT Based Innov. 23–30.

Kermanshah, A., Derrible, S., 2017. Robustness of road systems to extreme flooding: using elements of GIS, travel demand, and network science. Nat. Hazards 86 (1), 151–164. https://doi.org/10.1007/s11069-016-2678-1.

Kermanshah, A., Karduni, A., Peiravian, F., Derrible, S., 2014. Impact analysis of extreme events on flows in spatial networks. IEEE Int. Conf. Big Data (Big Data) 18, 29–34. https://doi.org/10.1109/BigData.2014.7004428.

Khan, S.I., Ritchie, S.G., 1998. Statistical and neural classifiers to detect traffic operational problems on urban arterials. Transp. Res. Part C Emerg. Technol. 6 (5–6), 291–314. https://doi.org/10.1016/S0968-090X(99)00005-4.

Lee, D., Derrible, S., Pereira, F.C., 2018. Comparison of Four Types of Artificial Neural Networks and a Multinomial Logit Model for Travel Mode Choice Modeling.

Levin, M., Krause, G., 1978. Incident detection: a bayesian approach. Transp. Res. Rec. 682, 52–58.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H., 2010. Feature Selection : A Data Perspective. ACM Comput. Surv. 9, 4.

Li, L., He, S., Zhang, J., Ran, B., 2016. Short-term highway traffic flow prediction based on a hybrid strategy considering temporal–spatial information. J. Adv. Transp. 50 (8), 2029–2040. https://doi.org/10.1002/atr.1443.

Lu, J., Chen, S., Wang, W., Van Zuylen, H., 2012. A hybrid model of partial least squares and neural network for traffic incident detection. Expert Syst. Appl. 39 (5), 4775–4784. https://doi.org/10.1016/j.eswa.2011.09.158.

Mahmoudzadeh, A., Razi-Ardakani, H., Kermanshah, M., 2019. Studying crash avoidance maneuvers prior to an impact considering different types of driver's distractions. Transp. Res. Procedia 37 (September), 203–210. https://doi.org/10.1016/J.TRPRO.2018.12.184.

Maldonado, S., López, J., Vairetti, C., 2019. An alternative SMOTE oversampling strategy for high-dimensional datasets. Appl. Soft Comput. J. 76, 380–389. https://doi.org/10.1016/j.asoc.2018.12.024.

Motamed, Moggan, 2016. Developing a Real-time Freeway Incident Detection Model Using Machine Learning Techniques.

Mohareb, E., Derrible, S., Peiravian, F., 2014. Intersections of Sustainability and Jane Jacobs'. Conditions for Diversity : A Look at Four Global Cities 142 (2), 1–20. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000287.

Nikolaev, A.B., Sapego, Y.S., Ivakhnenko, A.M., Mel, E., Stroganov, V.Y., 2017. Analysis of the incident detection technologies and algorithms in intelligent transport systems. Int. J. Appl. Eng. Res. Dev. 12 (15), 4765–4774.

Oqab, R., López, G., Garach, L., 2016. Bayes classifiers for imbalanced traffic accidents datasets. Accid. Anal. Prev. 88, 37–51. https://doi.org/10.1016/j.aap.2015.12.003.

Ozbayoglu, A.M., Kucukayan, G., Dogdu, E., 2017. A Real-Time Autonomous Highway Accident Detection Model Based on Big Data Processing and Computational Intelligence. pp. 1807–1813. https://doi.org/10.1109/BigData.2016.7840798.

Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Stat. 33 (3), 1065–1076.

Razi-Ardakani, H., Mahmoudzadeh, A., Kermanshah, M., 2018. A Nested Logit analysis of the influence of distraction on types of vehicle crashes. Eur. Transp. Res. Rev. 10, 2. https://doi.org/10.1186/s12544-018-0316-6.

Rong, Y.U., Guoxiang, W.A.N.G., Zheng, J., Haiyan, W.A.N.G., 2013. Urban road traffic condition pattern recognition based on support vector machine. Journal Transp. Syst. Eng. Inf. Technol. 13 (1), 130–136.

Rossi, R., Gastaldi, M., Gecchele, G., 2016. Automatic Incident Detection on Freeway Ramp Junctions. A Fuzzy Logic-Based System Using Loop Detector Data. Adv. Concepts, Methodol. Technol. Transp. Logist. 370–383.

Shabanpour, R., Golshani, N., Derrible, S., Mohammadian, A., Miralinaghi, M., 2017. (Kouros). A Cluster-Based Joint Model of Travel Mode and Departure Time Choices. Transp. Res. Board, 96th Annu. Meet. https://doi.org/10.3141/2669-05.

Sun, Y., Wong, A.K., Kamel, M.S., 2009. Classification of imbalanced data: a review. Int. J. Pattern Recognit. Artif. Intell. 23 (04), 687–719.

Tang, Y., Zhang, Y., Chawla, N.V., 2009. SVMs modeling for highly imbalanced classification. IEEE Trans. Syst. Man Cybern. Part B 1 (11), 1–9.

Vanajakshi, L.D., 2004. Estimation and prediction of travel time from loop detector data for intelligent transportation systems applications. OAK Trust (August), 304.

Vanhoeyveld, J., Martens, D., 2018. Imbalanced classification in sparse and large behaviour datasets. Data Min. Knowl. Discov. 32 (1), 25–82.

Verbiest, N., Ramentol, E., Cornelis, C., Herrera, F., 2014. Preprocessing noisy imbalanced datasets using SMOTE enhanced with fuzzy rough prototype selection. Appl. Soft Comput. J. 22, 511–517. https://doi.org/10.1016/j.asoc.2014.05.023.

Xiao, Z., Wang, Y., Fu, K., Wu, F., 2017. Identifying different transportation modes from trajectory data using tree-based ensemble classifiers. ISPRS Int. J. Geo-Information 6, 2. https://doi.org/10.3390/ijgi6020057.
You, J., Wang, J., Guo, J., 2017. Real-time crash prediction on freeways using data mining and emerging techniques. J. Mod. Transp. 25 (2), 116–123. https://doi.org/10.1007/s40534-017-0129-7.
Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. Accid. Anal. Prev. 51, 252–259. https://doi.org/10.1016/j.aap.2012.11.027.
Yuan, F., Cheu, R.L., 2003. Incident detection using support vector machines. Transp. Res. Part C Emerg. Technol. 11 (3–4), 309–328. https://doi.org/10.1016/S0968-090X(03)00020-2.
Zhang, K., 2005. Towards transferable incident detection algorithms. Journal of the Eastern Asia Society for Transportation Studies 6, 2263–2274.
Zhang, K., Taylor, M.A., 2004. A New method for incident detection on urban arterial roads. World Congr. ITS.
Zhu, Q., Zhu, M., Fu, M., 2018. Transportation modes behaviour analysis based on raw GPS dataset Mingzhao Li Zhibiao Huang Qihong Gan Zhenghao Zhou. Int. J. Embed. Syst. 10 (2), 126–136.
Zifeng, J., 1997. Macro and micro freeway automatic incident detection (AID) methods based on image processing. IEEE Conf. 344–349.