Emerging Cutting-Edge Developments in Intelligent Traffic and Transportation Systems
M. Shafik (Ed.)

doi:10.3233/ATDE240037

Predicting the Impact of Traffic Incidents: An Evaluative Analysis

George Rosario JAGADEESH[a,1], Joo Kiat ONG[a] and Chee Chung CHONG[b]
[a] *Data Analytics Strategic Technology Centre, ST Engineering Ltd., Singapore*
[b] *Mobility Road Business Development Centre, ST Engineering Ltd., Singapore*

Abstract. We examine the problem of forecasting the spatial extent of a just-occurred traffic incident's impact and the travel delay induced by it at certain future time points. We present and evaluate a machine learning-based solution for the above problem. The proposed solution is based on a standard classification model that takes in a variety of input features that include the incident attributes and features derived from traffic sensor data. We evaluate several versions of the solution by varying the classification model, the number of impact classes, the type of training data, and the time at which the prediction is made. This is done by conducting a series of experiments using a real-world traffic incident dataset along with the corresponding traffic sensor data. In particular, we investigate the issue of class imbalance in the incident dataset, the disparity in the class-wise prediction accuracies, the benefit of taking the incident's early impact into account, and the relative importance of the input features. The findings of this study are potentially insightful to practitioners and researchers in the field of intelligent traffic management.

Keywords. Traffic incidents, incident impact prediction, traffic prediction, classification models

1. Introduction

Traffic incidents are non-recurring events such as accidents and vehicle breakdowns that cause a temporary reduction in road capacity. Such incidents often lead to congestion and travel delay that evolve over a period following the incident occurrence. The ability to predict the impact of a just-occurred incident is useful for traffic authorities to optimally respond to the incident. Furthermore, forecasts about the impact of an incident are valuable for drivers to plan an optimal route in a dynamically evolving traffic situation. Currently, drivers are alerted through various channels about a delay-causing incident along their route. However, they are generally not provided a forecast of the incident's impact at a future time when they would arrive at the affected area. Such a forecast is necessary to decide if an alternative route should be taken. For instance, if an incident at a location that is 30 minutes ahead on a driver's route has resulted in a long delay currently, but the delay is predicted to be negligible after 30 minutes, then the driver need not change the route. Furthermore, knowing the expected spatial extent of the impacted region helps the driver to decide where to detour from the

[1] Corresponding Author, George Rosario Jagadeesh, Data Analytics Strategic Technology Centre, ST Engineering Ltd., ST Aerospace Aviation Centre, 600 West Camp Road #01-01, Singapore 797654; E-mail: george.jagadeesh@stengg.com.

planned route. Navigation applications with access to incident impact forecasts can lead to significant savings in terms of time and fuel.

Traffic simulation and analytical models have been traditionally used for predicting the impact of incidents [1] [2]. However, such models are limited in their ability to accurately predict the impact of incidents in the real-world. In recent times, the real-time availability of traffic data (including incident data) from various providers have made it feasible to apply data-driven solutions to the incident impact prediction problem. While many researchers have used machine learning models for incident impact prediction, most of them have focused on predicting the impact in terms of the incident duration [3]. This paper, on the other hand, deals with the problem of predicting the spatial extent of an incident's impact on traffic and the incident-induced delay for an individual driver at certain future time points. We refer to the two abovementioned target variables simply as impact extent and delay, respectively.

The main objective and contribution of this paper is to propose and evaluate a machine learning-based solution for predicting the incident impact extent and delay for a given prediction horizon. We model the prediction task as a classification problem where the target variables are expressed as semantically meaningful classes. We evaluate several variations of the proposed solution on a widely used real-world dataset in order to address the following research questions.

i. What are the overall and class-wise prediction accuracies achievable for different number of impact classes?
ii. Given the imbalanced nature of traffic incident datasets where most incidents have a negligible impact, does it help to balance the classes in the dataset used for training the prediction model?
iii. Do the impact prediction accuracies improve if the prediction is delayed by a short period so that the initial effect of the incident is captured in the input features of the model?
iv. Which input features are important for incident impact prediction?

The paper is structured as follows: Section 2 contains a review of related work. Section 3 defines the impact extent and delay and shows how they are calculated from traffic speed data. Section 4 describes the proposed solution for incident impact prediction. Section 5 provides a detailed account of the experimental evaluation and the results obtained. Section 6 concludes the paper and indicates further research directions.

2. Related Work

A few researchers have studied the problem of predicting the impact extent and delay caused by a traffic incident. Pan et al. [4] model the incident impact as a time-varying spatial span (i.e., impact extent) and present a clustering-based method for predicting it. Recently, Sun et al. [5] have proposed a graph neural network that incorporates a graph attention strategy for predicting the impact extent.

Boyles and Waller [6] proposed analytical formulae based on shockwave theory for predicting the incident delay. One of the earliest data-driven methods for predicting the incident delay is by Garib et al. [7] who proposed a regression model trained on historical data. A disadvantage of this model is that it requires incident duration as an input for predicting the delay. Miller and Gupta [8] used several classification models including decision trees, a k-nearest neighbors classifier, and artificial neural networks to predict a cost proportional to the cumulative delay experienced by all affected drivers. The cost of the delay is expressed as a class variable where each class covers a range of cost values.

Some researchers have proposed deep learning models for post-incident prediction of traffic speeds (e.g., [9], [10]). The impact extent and delay for future time points can be calculated from the predicted speeds as explained in Section 3 or by using a method such as [11].

Our review of the literature on the topic of predicting the incident impact extent and delay shows that the effectiveness of state-of-the-art machine learning models for this specific task has not been adequately evaluated. The prediction accuracies reported by the studies leave room for improvement. Furthermore, it is necessary to empirically examine the research questions listed in the previous section.

3. Preliminaries and Definitions

The solution presented in this paper is limited to the problem of predicting the impact of traffic incidents that occur on a continuous stretch of a single freeway that has traffic sensors installed along it. We divide the freeway stretch into virtual segments such that a traffic sensor lies at the midpoint of each segment.

Let us consider a traffic incident that occurs on segment s_a. To be exact, s_a is the segment associated with the sensor that lies immediately upstream of the incident location. Let $S = \{s_a, s_{a+1}, \dots, s_{a+K}\}$ be a contiguous sequence of segments comprising s_a and K segments that are upstream of it. (s_{a+1} is the segment immediately upstream of s_a and so on.) The value of K should be set so that the sequence S is long enough to cover all the segments that could be potentially impacted by the incident. We call S as the *study extent*.

Let $v_{i,j}$ denote the traffic speed of segment s_i at time step t_j. Let $\hat{v}_{i,j}$ denote the normal speed, in non-incident conditions, of segment s_i at time step t_j. (t_j corresponds to a time of the day.) In this work, the normal speed $\hat{v}_{i,j}$ is set equal to the median speed of segment s_i at time step t_j, calculated from historical data. Let l_i denote the length of segment s_i.

A segment is considered *congested* if its speed is less than or equal to a constant proportion of its normal speed. More formally, segment s_i at time step t_j is congested if $v_{i,j} \leq \alpha \hat{v}_{i,j}$, where α is a constant between 0 and 1. (In this study, α is set to 0.7.)

Based on the analysis presented in [4] and [12], the evolution of congestion in incident scenarios is expected to meet certain spatial and temporal constraints. Accordingly, we determine if a segment is *impacted* by an incident by applying the following two rules, in the given order:

i. Spatial consistency rule: A segment s_i in the study extent is considered to be impacted at time step t_j if it is congested at t_j or its immediate upstream segment is congested at t_j. That is, segment s_i is impacted at time step t_j if $v_{i,j} \leq \alpha \hat{v}_{i,j}$ or $v_{i+1,j} \leq \alpha \hat{v}_{i+1,j}$.

ii. Temporal consistency rule: If, after applying the spatial consistency rule, a segment s_i is not impacted at time step t_j but it is impacted at the previous time step t_{j-1} and the subsequent time step t_{j+1}, it should be considered impacted at time step t_j.

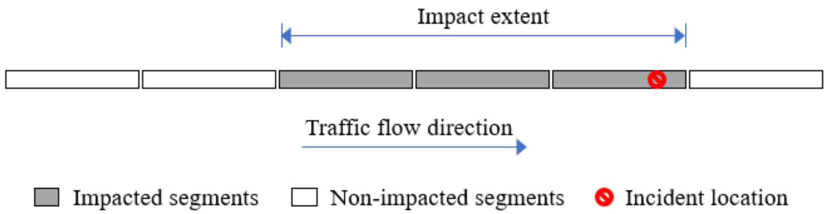


Figure 1. Illustration of the impact extent of a traffic incident

Consistent with the definitions in [4] and [8], we define the *impact extent* at any time step after an incident's occurrence as the length of the contiguous sequence of impacted segments beginning with the segment on which the incident occurred and extending in the upstream direction. Figure 1 illustrates an example case where the impact extent consists of three impacted segments. To be clear, this study considers only the mainline freeway segments and not the ramps connected to the freeway.

We define the incident-induced *delay* at a time step as the extra travel time experienced by an individual driver traversing the impact extent at the time step due to the reduction in speeds caused by a traffic incident with respect to the normal speeds. Most studies (e.g., [8],[11]) calculate the total delay experienced by all the drivers over the entire duration of the incident. We are interested in the delay experienced by an individual driver at a particular time as such a measure is more useful for drivers and navigation applications for route planning.

Let us suppose that the impact extent of an incident occurring on segment s_a corresponds to a contiguous sequence of $M+1$ impacted segments, $\{s_a, s_{a+1}, \ldots, s_{a+M}\}$. The above-defined delay at time step t_j can be calculated as

$$\delta_j = \sum_{k=0}^{M} max\left(\frac{l_{a+k}}{v_{a+k,j}} - \frac{l_{a+k}}{\hat{v}_{a+k,j}}, 0\right) \quad (1)$$

where the variables are as defined earlier.

4. Incident Impact Prediction

4.1 Problem Definition

We aim to forecast the impact of a just-occurred traffic incident at a future point of time in terms of the spatial extent of the incident-induced congestion ("impact extent") and the delay experienced by an individual driver in traversing that extent ("delay"). As other researchers have observed [8], predicting the impact of traffic incidents in the form of exact values is quite difficult. In many scenarios, it is sufficient for the users to know the predicted impact as a range of values. Therefore, we express the target variables, namely, the impact extent and the delay, as class variables where each class corresponds to a range of values. The classes are given meaningful names such as negligible, moderate and long.

In this study, we define the incident impact prediction problem as follows: Given the attributes of an incident that occurred at time step t_j and the traffic variables measured at the incident location, predict the impact extent and the delay, respectively, at time step t_{j+h}, where h is the prediction horizon in number of time steps.

4.2 Input Features

The features used as input to the impact prediction models in this study can be grouped into four types: temporal, spatial, incident and traffic. The input features are listed in Table 1 along with a brief description of each of them. It is worth noting a detail regarding the traffic features: The speed difference and occupancy features are calculated for the two segments in the study extent that are closest to the incident location, and their maximum is taken. The speed change and occupancy change features are calculated based on the speed difference and occupancy values at the current and the previous time steps.

4.3 Prediction Models

As the target variables to be predicted are expressed as class variables, we perform the predictions using standard classification models. We use and evaluate a decision tree classifier and a random forest classifier for predicting the impact extent class and the delay class.

Decision tree-based classification uses a divide and conquer strategy in which the data space is recursively partitioned until all or most data points in each partition have the same class label. A classification criterion such the Gini impurity measure or entropy is used to determine the quality of the partitions. The performance of a decision tree classifier can be optimized by choosing an optimal set of hyperparameters such as the maximum depth of the tree and the classification criterion. In this work, we use a decision tree algorithm named Classification and Regression Trees (CART) [13]. A main advantage of decision tree classifiers is that they are easy to interpret by visualizing the trees. However, they are sometimes prone to overfitting the training data, leading to lower prediction accuracy against new unseen data.

Table 1. Input features used for incident impact prediction in this study

Feature name	Feature type	Description
Time of the day	Temporal	The time elapsed since midnight in hours (continuous).
Weekday	Temporal	1 if the day is a weekday, 0 otherwise.
Location	Spatial	The distance from the start of the freeway stretch to the incident location in km.
Accident	Incident	1 if the incident involves an accident, 0 otherwise.
Injury	Incident	1 if injury is reported, 0 otherwise.
Ambulance	Incident	1 if an ambulance is requested, 0 otherwise.
Hit and run	Incident	1 if it is a hit-and-run accident.
Speed difference (current)	Traffic	The difference between the speed at prediction time and the normal speed.
Speed difference (previous)	Traffic	The difference between the speed at the previous time step and the normal speed.
Occupancy (current)	Traffic	The difference between the occupancy at prediction time and the normal occupancy.
Occupancy (previous)	Traffic	The difference between the occupancy at the previous time step and the normal occupancy.
Speed change	Traffic	The change in speed in the current time step with respect to the previous time step.
Occupancy change	Traffic	The change in occupancy in the current time step with respect to the previous time step.

A random forest model [14] consists of multiple decision trees and outputs the class determined by most of the trees. Each decision tree in the random forest operates on a random sample of the training data (sampled with replacement) and uses a random subset of input features. This helps to create an ensemble of uncorrelated decision trees, which decreases the model's chances of overfitting and improves its ability to generalize to unseen data. Random forest models also have the ability to determine the importance of features. While random forest models are generally preferred to decision trees due to their lower risk of overfitting, they are more complex and difficult to interpret compared to decision trees.

5. Experiments and Results

5.1 Dataset and Data Preparation

We use historical traffic and incident data from the California Performance Management System (PeMS) [15] for the experiments in this study. Specifically, we use 7 months of historical data from September 2022 to March 2023, excluding 7 holidays that fell during this period. We confine this study to a 30-mile (≈ 48 km) stretch of the US-101 (southbound) freeway between postmiles 390 and 420 in the San Francisco Bay Area, shown in Figure 2.

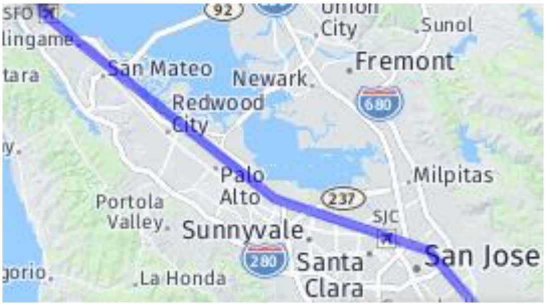


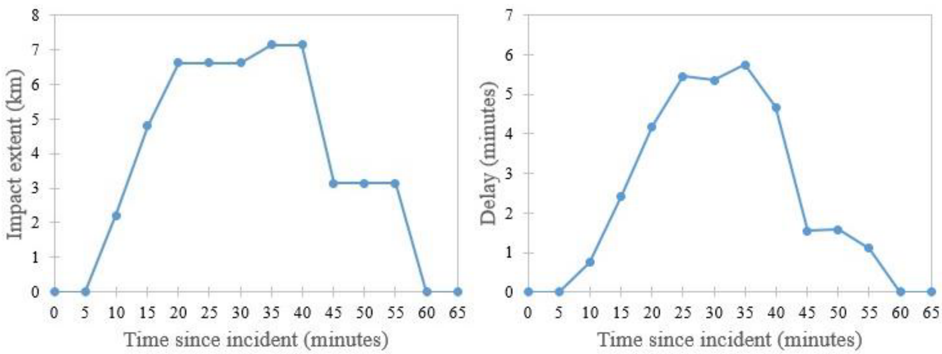
Figure 2. The US-101 freeway stretch considered in this study (Image source: PeMS [15], base map: HERE)

Traffic data (speed and occupancy), aggregated at 5-minute intervals, from 78 sensors installed on the mainline of the freeway stretch are considered. As mentioned in Section 3, we divide the freeway stretch into virtual segments so that each segment is uniquely mapped to a sensor. The average length of the segments is 0.62 km. We calculate historical median speeds for each segment for each time of the day for two groups, namely, weekdays and weekends.

We process the incident data and exclude some incidents based on a few criteria. Incidents that occurred on the first 8 km of the considered stretch are excluded to allow sufficient space for the upstream impact of each incident to be captured. There are many duplicate records in the PeMS database due to multiple agencies handling the same incident. Such duplicates are eliminated. Each incident in the database has a description that indicates the type of the incident. We consider the following 6 types that account for 93% of the total incidents: "Traffic collision - no injuries", "Traffic collision - unknown injuries", "Traffic collision - ambulance requested", "Hit and run - no injuries", "Hit and run - with injuries", and "Traffic hazard". Various minor types (e.g., "wrong way driver") that account for a small number of incidents are filtered out.

Finally, we exclude incidents that occur close to each other, both spatially (within 1.6 km) *and* temporally (within 30 minutes). We do so because it is difficult to distinguish the impact of one incident from another in such situations. After executing the above steps, the incident dataset consists of 966 incidents.

For each incident, we determine the features listed in Table 1. The temporal, spatial and incident features are derived from the incident dataset. The binary values of the four incident features are assigned based on the above-mentioned description of the incident in the dataset. The traffic features are calculated based on the speed and occupancy values of the relevant segments at the prediction time step and the previous time step. For each incident, the ground truth values of the impact extent and delay for different prediction horizons are calculated by applying the method described in Section 3 to the segment speeds at the time step for which the prediction is made. Figure 3 shows the ground truth values of impact extent and delay following the occurrence of a sample incident (traffic collision with unknown injuries) on a weekday morning at postmile 401.5 of the US-101 (southbound) freeway. It can be seen that the impact extent and delay are zero at the time of the incident, but they increase and subside over the next the next 60 minutes.



(a) Impact extent (b) Delay

Figure 3. The ground truth impact extent and delay for a sample incident

5.2 Experimental Setup

We use the implementations of the decision tree and random forest classification models in the Scikit-learn machine learning library. During the training of the model, we use grid search cross-validation to determine the optimal values of two hyperparameters of the models, namely, the maximum depth of the tree and the classification criterion.

Out of the 996 incidents in the incident dataset (in chronological order), we use the first 724 incidents (75%) for training the impact prediction models and the remaining 242 incidents for testing the models. We adopt this approach of chronologically splitting the data into training and test sets as it is consistent with real-world scenarios where models are trained using data available up to a time and applied to data that comes thereafter.

We evaluate the effectiveness of the models for predicting the impact extent and delay at two prediction horizons: 15 minutes and 30 minutes. (We experimented with prediction horizons beyond 30 minutes, but found the results unsatisfactory.) In the first set of experiments, we express the target variable as one of three classes. For impact extent, the three classes are defined as: negligible (impact extent ≤ 0.5 km), moderate (0.5 km < impact length ≤ 3 km) and long (impact extent > 3 km). For delay, the three classes are defined as: negligible (delay ≤ 0.5 minutes), moderate (0.5 minutes < delay ≤ 5 minutes) and long (delay > 5 minutes). In the second set of experiments, the predicted variable is expressed as one of two classes: negligible and significant. The definition of the negligible class is the same as in the first set of experiments, but the moderate and long classes defined above are combined to form the significant class.

An analysis of the ground truth values of the target variables in the incident dataset shows that the distribution of the classes is imbalanced with over 85% of the incidents belonging to the negligible class. We examine if the effect of imbalanced classes on the model performance could be handled by balancing the training dataset through oversampling and undersampling. We investigate if better prediction accuracies could be achieved by delaying the prediction by one time step so that the input traffic features reflect the early impact of the incident. We also examine which of the input features are significant.

We quantify the prediction accuracy of the models using two variants of the F_1 score, namely, the weighted-average (WA) F_1 score and the macro-average (MA) F_1 score. This is done by calculating the F_1 score for each class and taking the weighted average and the macro average, respectively. The WA F_1 score could be misleadingly high for test sets where a vast majority of samples belong to one class that is predicted well, but the other classes are predicted poorly. The MA F_1 score indicates the average F_1 score for all the classes irrespective of how many samples belong to each class.

5.3 Results and Discussion

In the first set of experiments, three classes are considered: negligible, moderate and long. Let us examine a sample case in which the impact extent is predicted for the 15-minutes prediction horizon using the random forest classifier. In this case, the prediction is performed at the time step during which the incident occurs. Table 2 shows the prediction result for this sample case in the form of a confusion matrix. It can be seen that while the model performs well for the negligible class (205 out of 210 samples predicted correctly), it is not quite effective for the other two classes. The F_1 scores for the negligible, moderate and long classes are 0.95, 0.51 and 0.18, respectively. The WA F_1 score and the MA F_1 score for this sample case are 0.88 and 0.55. The substantial difference between the two scores indicates that the model performs quite differently for different classes.

Table 3 shows the overall results for predictions made at the incident time step. Not surprisingly, predictions for the 15-minutes horizon are more accurate than the predictions for the 30-minutes horizon. While the random forest classifier performs marginally better than the decision tree classifier, the MA F_1 score remains quite low. As seen in the sample case above, this is due to the model not performing well for the moderate and long classes, which could be attributed to the imbalanced training dataset with a small number of incident samples belonging to the said classes.

Table 2. The confusion matrix for a sample case (15-minutes-ahead prediction of impact extent made at the incident time step using the random forest classifier)

		Predicted class		
		Negligible	**Moderate**	**Long**
Actual class	**Negligible**	205	5	0
	Moderate	11	11	2
	Long	4	3	1

We attempt to address the above issue by training the models with a balanced training set where each class is made to have an equal number of incident samples through oversampling and undersampling. Table 4 shows the results obtained with the balanced training set for predictions made at the incident time step. A significant improvement in MA F_1 score and a marginal decrease in the WA F_1 score is observed for the random forest classifier. The minor drop in the WA F_1 score is due to a marginal decrease in the model's performance for the negligible class, which account for a vast majority of samples in the test set. This is to be expected because the negligible class has a smaller number of samples in the balanced training set compared to the original training set. Interestingly, using the balanced training set to train the decision tree classifier does not help much.

Table 3. Results for impact extent and delay prediction (3 classes) made at the incident time step, based on the original training set

		Decision tree classifier		**Random forest classifier**	
Target variable	**Prediction horizon (min.)**	**WA F_1 score**	**MA F_1 score**	**WA F_1 score**	**MA F_1 score**
Impact extent	15	0.87	0.49	0.88	0.55
Impact extent	30	0.87	0.34	0.88	0.32
Delay	15	0.90	0.50	0.92	0.52
Delay	30	0.90	0.35	0.90	0.32

Table 4. Results for impact extent and delay prediction (3 classes) made at the incident time step, based on the balanced training set

		Decision tree classifier		**Random forest classifier**	
Target variable	**Prediction horizon (min.)**	**WA F_1 score**	**MA F_1 score**	**WA F_1 score**	**MA F_1 score**
Impact extent	15	0.84	0.51	0.89	0.63
Impact extent	30	0.83	0.42	0.87	0.52
Delay	15	0.84	0.44	0.87	0.47
Delay	30	0.84	0.39	0.87	0.48

Table 5. Results for impact extent and delay prediction (3 classes) delayed by 5 minutes, based on the balanced training set.

		Decision tree classifier		**Random forest classifier**	
Target variable	**Prediction horizon (min.)**	**WA F_1 score**	**MA F_1 score**	**WA F_1 score**	**MA F_1 score**
Impact extent	10	0.88	0.56	0.89	0.63
Impact extent	25	0.89	0.52	0.89	0.53
Delay	10	0.91	0.53	0.92	0.77
Delay	25	0.88	0.42	0.90	0.57

Table 5 shows the results obtained by delaying the prediction by one 5-minute time step. The balanced training set is used for training the models. To be consistent with the previous experiments, we predict the impact extent and delay for time points that are 15 minutes and 30 minutes after the incident time step. However, as the prediction is now made 5 minutes after the incident time step, the corresponding predictions horizons are 10 minutes and 25 minutes, respectively. It can be seen from Table 5 that the delayed prediction leads to improvements in both WA F_1 score and MA F_1 score for both the classifiers. This is not surprising as the input traffic features, based on the speed and occupancy measurements in the time step following the incident occurrence, are likely to be good indicators of the incident's severity.

Despite the improvements observed in Table 4 and Table 5, the difference between the WA F_1 score and the MA F_1 score remains high. We find that this is mainly because the models still struggle to correctly predict the small number of samples in the test set that belong to the moderate and long classes.

In the second set of experiments, the models predict the impact extent and delay as either negligible or significant. Table 6 shows that the MA F_1 scores are substantially better than the corresponding scores obtained with three classes. Similar to the first set of experiments, the overall performance of the random forest classifier is better than the decision tree classifier. In contrast to the three-class scenario, Table 7 shows that in the two-class scenario, training the models with the balanced training set results in a reduction in both the WA F_1 score and the MA F_1 score for both classifiers. Table 8 shows the results obtained by delaying the two-class prediction by one 5-minute time step. The original (imbalanced) training set is used for training the models. The delayed prediction improves the WA F_1 score as well as the MA F_1 score for both the classifiers.

Table 6. Results for impact extent and delay prediction (2 classes) made at the incident time step, based on the original training set

		Decision tree classifier		**Random forest classifier**	
Target variable	**Prediction horizon (min.)**	**WA F_1 score**	**MA F_1 score**	**WA F_1 score**	**MA F_1 score**
Impact extent	15	0.89	0.76	0.91	0.80
Impact extent	30	0.89	0.65	0.92	0.71
Delay	15	0.91	0.76	0.92	0.77
Delay	30	0.91	0.67	0.91	0.63

Table 7. Results for impact extent and delay prediction (2 classes) made at the incident time step, based on the balanced training set

		Decision tree classifier		**Random forest classifier**	
Target variable	**Prediction horizon (min.)**	**WA F_1 score**	**MA F_1 score**	**WA F_1 score**	**MA F_1 score**
Impact extent	15	0.86	0.74	0.88	0.77
Impact extent	30	0.86	0.61	0.86	0.64
Delay	15	0.85	0.68	0.88	0.72
Delay	30	0.86	0.62	0.88	0.62

Table 8. Results for impact extent and delay prediction (2 classes) delayed by 5 minutes, based on the original training set

		Decision tree classifier		**Random forest classifier**	
Target variable	**Prediction horizon (min.)**	**WA F_1 score**	**MA F_1 score**	**WA F_1 score**	**MA F_1 score**
Impact extent	10	0.93	0.85	0.94	0.87
Impact extent	25	0.92	0.72	0.93	0.77
Delay	10	0.94	0.86	0.95	0.88
Delay	25	0.92	0.68	0.93	0.73

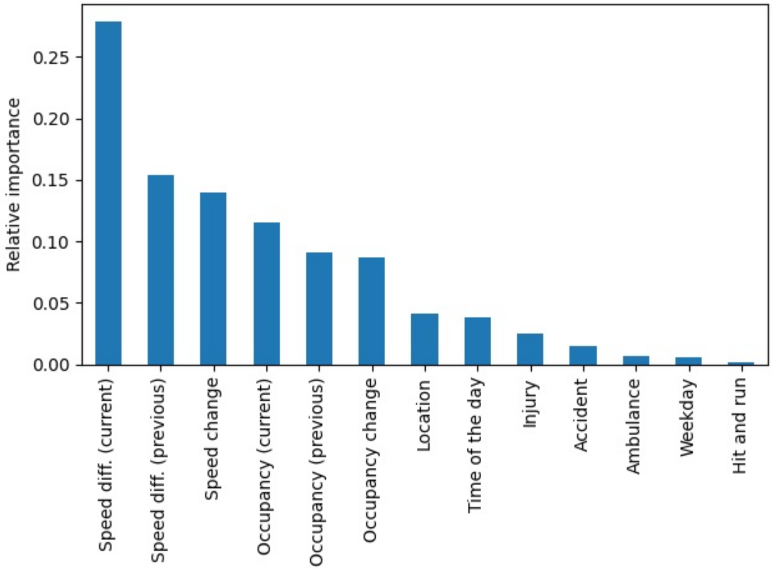


Figure 4. The relative importance of features determined by the random forest classifier for two-class prediction of impact extent

Figure 4 shows the importance of each input feature, determined by the random forest classifier, for predicting the impact extent for the two-class prediction. (The findings are broadly similar for predicting delay and for the three-class prediction.) It can be seen that traffic features are much more important than the other types of features for incident impact prediction. Among the traffic features, speed features are more important than occupancy features. The incident features seem to be of low importance. It may be argued that in reality, the impact of an incident depends more on the level of traffic at the time than the attributes of the incident. For example, an incident that occurs on a freeway during a light-traffic period is unlikely to cause any significant congestion or delay irrespective of the incident's attributes.

6. Conclusions

A forecast of the impact of an ongoing traffic incident in terms of the impact extent and travel delay at a future time is valuable for drivers aiming to minimize their travel times and fuel consumption. In this study, we have proposed a solution using standard machine learning models for predicting the impact extent and delay, respectively, as one of multiple classes. We have used a dataset containing data of 966 real-world traffic incidents on a freeway stretch and the corresponding traffic sensor data to evaluate multiple variations of the proposed solution. We find that a random forest model, which is less prone to overfitting, achieves better prediction accuracy compared to a decision tree model. The prediction models do not perform well for the classes that are severely underrepresented in the training data. Balancing the training data mitigates this issue to an extent for three-class prediction but not for two-class prediction. The results also show that delaying the prediction by one time step so that the initial effect of the incident is captured in the input traffic features improves the prediction accuracy.

When the target variables are expressed as one of two classes, negligible and significant, the prediction models perform significantly better than the three-class scenario. Such a binary classification is useful for identifying false alarms so that drivers need not react to alerts about incidents that are expected to have negligible impact at a future time of interest.

Future extensions of this work could focus on improving the prediction accuracy through the use of a better set of features and a larger, more balanced training dataset. Some intuitively useful incident features such as the number of lanes blocked and the involvement of a tow truck could not be included in this study due to the difficulty in automatically extracting them from free-text incident logs. The effectiveness of such features as well as other feature types such as weather conditions could be evaluated in future. It would also be interesting to study the effect of using a larger training dataset where all the impact classes are adequately represented.

References

[1] Ozbay, K. and Bartin, B., 2003. Incident management simulation. *Simulation*, 79(2), pp.69-82.

[2] Lawson, T.W., Lovell, D.J. and Daganzo, C.F., 1997. Using input-output diagram to determine spatial and temporal extents of a queue upstream of a bottleneck. *Transportation Research Record*, 1572(1), pp.140-147.

[3] Li, R., Pereira, F.C. and Ben-Akiva, M.E., 2018. Overview of traffic incident duration analysis and prediction. *European Transport Research Review*, 10(2), pp.1-13

[4] Pan, B., Demiryurek, U., Shahabi, C. and Gupta, C., 2013, December. Forecasting spatiotemporal impact of traffic incidents on road networks. In *2013 IEEE 13th International Conference on Data Mining* (pp. 587-596). IEEE.

[5] Sun, Y., Fu, K. and Lu, C.T., 2023. DG-Trans: Dual-level Graph Transformer for Spatiotemporal Incident Impact Prediction on Traffic Networks. *arXiv preprint arXiv:2303.12238*.

[6] Boyles, S. and Waller, S.T., 2007. A stochastic delay prediction model for real-time incident management. *ITE Journal*, 77(11), pp.18-24.

[7] Garib, A., Radwan, A.E. and Al-Deek, H., 1997. Estimating magnitude and duration of incident delays. *Journal of Transportation Engineering*, 123(6), pp.459-466.

[8] Miller, M. and Gupta, C., 2012, August. Mining traffic incidents to forecast impact. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing* (pp. 33-40).

[9] Yu, R., Li, Y., Shahabi, C., Demiryurek, U. and Liu, Y., 2017, June. Deep learning: A generic approach for extreme condition traffic forecasting. In *Proceedings of the 2017 SIAM international Conference on Data Mining* (pp. 777-785). Society for Industrial and Applied Mathematics.

[10] Xie, Q., Guo, T., Chen, Y., Xiao, Y., Wang, X. and Zhao, B.Y., 2020, October. Deep graph convolutional networks for incident-driven traffic speed prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 1665-1674).

[11] Chung, Y. and Recker, W.W., 2012. A methodological approach for estimating temporal and spatial extent of delays caused by freeway accidents. *IEEE Transactions on Intelligent Transportation Systems*, 13(3), pp.1454-1461.

[12] Yang, H., Bartin, B. and Ozbay, K., 2013. Use of sensor data to identify secondary crashes on freeways. *Transportation Research Record*, 2396(1), pp.82-92.

[13] Breiman, L., Friedman, J.H., Olshen, R. and Stone, C.J., 1984. *Classification and Regression Trees*. CRC Press.

[14] Breiman, L., 2001. Random forests. *Machine Learning*, 45, pp.5-32.

[15] California Department of Transportation, 2023. Freeway Performance Measurement System (PeMS), Available at: http://pems.dot.ca.gov (Accessed: 27 July 2023).