Corbally *et al.*
European Transport Research Review (2024) 16:14
https://doi.org/10.1186/s12544-024-00632-6

European Transport Research Review

ORIGINAL PAPER

Open Access

Predicting the duration of motorway incidents using machine learning

Robert Corbally[1*], Linhao Yang[1] and Abdollah Malekjafarian[1]

Abstract

Motorway incidents are frequent & varied in nature. Incident management on motorways is critical for both driver safety & road network operation. The expected duration of an incident is a key parameter in the decision-making process for control room operators, however, the actual duration for which an incident will impact the network is never known with true certainty. This paper presents a study which compares the ability of different machine learning algorithms to estimate the duration of motorway incidents on Ireland's M50 motorway, using an extensive historical incident database. Results show that the support vector machine has the best performance in most cases, but a different method may need to be used to improve accuracy in some situations. Results highlight the main challenges in accurately forecasting incident durations in real time & recommendations are made for improving prediction accuracy through systematic recording of various additional incident details.

Keywords Traffic incidents, Incident management, Incident response, Road traffic collisions, Incident duration, Motorway operations, M50 motorway, Machine learning, Regression, Classification

1 Introduction

Up to 1,400 incidents occur on the M50 motorway every year, including various types ranging from road traffic collisions, vehicle breakdowns or overturned vehicles, to animals or pedestrians on the road. It is the most heavily trafficked road in Ireland and once an incident has occurred, control room operators need to immediately instigate an appropriate response to safely and efficiently manage the incident.

The appropriate response is very much dependent on the nature of the incident and the operational response parties that are required to manage the incident [6]. A key parameter in deciding on the appropriate response, is the expected duration of the incident and the impact on traffic during this time [8, 14]. This allows suitable diversion routes or warning messages to be communicated to drivers and helps maintain network efficiency while the incident is being cleared. Therefore, the ability to predict the impact of an incident is of major benefit to road operators. Incident response plans benefit from accurate predictions of duration, especially with regard to the control of variable message signs or signals to give drivers warning of probable delays or safety issues.

Because of advances in Intelligent Transport Systems (ITS) technology, there is a wealth of data which can be used to develop machine learning models [19]. Traffic information can be recorded using cameras [21], inductive loops [1], magnetometers [9], radar systems [28], weigh in motion systems [4], and increasingly, using in-vehicle data [10], amongst many other technologies. All of this information can be leveraged to gain a deeper understanding of traffic conditions and the various factors which influence them [12]. This paper presents a study which uses a comprehensive database of historical incidents on Ireland's M50 motorway, along with traffic flow and weather datasets to assess the ability of different machine learning methods to predict incident durations.

*Correspondence:
Robert Corbally
robert.corbally@ucd.ie
[1] Structural Dynamics and Assessment Laboratory, School of Civil Engineering, University College Dublin, Newstead, Belfield, Dublin 4, Ireland

The paper explores the performance of each method, discusses their advantages and drawbacks and also presents a series of recommendations for systematic recording of relevant incident details to allow more accurate predictions to be made as part of the incident management process.

2 Previous research

Attempts at using machine learning methods to estimate incident duration have been made for more than two decades. Ozbay and Kachroo [15] used decision trees to forecast incident clearance times in Northern Virginia. They developed decision trees that encompass a range of features, including road hazards, property damage, personal injuries, broken trucks, vehicle fires, weather, etc. The significance of independent variables was checked using analysis of variance (ANOVA). Real data was used for testing, and it was found that 60% of the incidents had a prediction error of 10% or less. Smith and Smith [20] assessed three models: stochastic, non-parametric regression, and classification tree. They found the stochastic model unsuitable due to poor fit of Weibull and lognormal distributions. However, the classification tree model, with reliable data and an accuracy rate of 58%, effectively predicted incident duration stages, emphasizing the significance of tow truck response.

Chang and Chang [3] developed a classification tree model for 4908 incidents in Taiwan, which had 96.7% classification accuracy for test data with short incident durations (5–41 min), however, the classification accuracy for medium (42–118 min) and long durations (119–391 min) was extremely low (< 20%). This is because the classification criteria used in the study categorised over 72% of the incidents as short duration, resulting in an unbalanced dataset. Similar to this study, Leahy and Lynch [11] applied multiple machine learning methods to 556 incidents which occurred on Ireland's M50 motorway between 2014 and 2016. This study found that the highest classification accuracy which could be achieved across the various methods tested was 52.7%. It was also confirmed in this study that classification tree methods performed poorly in predicting long duration incidents.

Vlahogianni and Karlaftis [24] analysed 1449 incidents on the Attiki Odos Tollway in Greece, where incident records included incident characteristics, traffic volumes, weather and geometric features. They applied a fuzzy entropy feature selection methodology to determine the redundant factors and ANN models to predict the incident duration time. The results indicate that a model with fewer input variables may be optimal, achieving an accuracy of about 10% when applying survival neural networks to identify the proportion of incidents whose duration could accurately be classified. Dimitriou and Vlahogianni [7] conducted a study using a fuzzy duration model to investigate the same 1,449 incidents. They found that a model with only two input variables (traffic volume and rainfall) achieved the best predictive accuracy with a Minimum Average Percentage Error (MAPE) of 36%. Therefore, they concluded that increasing the number of input variables does not necessarily improve the model performance.

Valenti et al. [23] compared five machine learning methods to examine their ability to predict the duration of incidents. They investigated the use of multiple linear regression, decision trees, ANNs, Support Vector Machine (SVM) and K-Nearest-Neighbour algorithms. The results showed that linear regression is the best approach for short duration incidents, with Relevance Vector Machine (RVM) models achieving the best prediction in the case of medium and medium to long duration incidents. The ANN was shown to be the only model able to predict incidents longer than 90 min. For longer duration incidents, the accuracy of all the proposed models is relatively low, largely due to the relatively small number of severe incidents in the dataset. Yu et al. [25] compared the performance of ANNs and SVMs using data from 235 incidents that occurred on a highway in China between 2012 and 2014. The results were similar to Valenti et al. [23] demonstrating that the SVM model performed better than the ANN model for medium durations. However, it was again reported that the ANN model performed better in predicting long duration incidents. The authors also found significant differences in recorded incident durations due to individual differences in incident management team response to incidents.

Zong et al. [26] analysed police reported traffic incident records from Jilin Province, China, in 2010. In addition to incident and traffic characteristics, they also collected weather factors and road environment factors. They predicted the severity of incidents using SVMs and Ordered Probit models and found that the Ordered Probit model was slightly more accurate than the SVM. The results showed that the presence of hazardous materials, the weather and the location of the incidents had a significant effect in both models.

Park et al. [16] analyzed 13,987 incidents recorded by the Maryland State Highway Administration (SHA) from 2010 to 2011 including incident details, traffic volume, and weather factors. They categorized the incidents into four groups based on clearance time (less than 30 min, 30–60 min, 60–90 min, and over 90 min) and compared four methods, including Classification and Regression Trees (CART), Backpropagation Neural Network (BPNN), and Support Vector Machine (SVM). The study found that the Bayesian Neural Network method performed the best, achieving a Mean Absolute Error (MAE) in the classification accuracy of 0.22 for the test data.

The above research shows that different machine learning methods have different specialisations and overall prediction performance can be improved by combining multiple machine learning methods. However, Li et al. [13] conducted a review of 44 research papers that used machine learning for predicting traffic incident duration. Their meticulous research revealed a significant disproportion in the choice of datasets among these studies. Specifically, only 20% of the reviewed articles employed European traffic data as their dataset, with the majority of studies focusing on traffic situations in the United States. European traffic has its own distinct characteristics, highlighting the need for more research on European traffic conditions. Despite the significant operational need for methods to accurately predict incident durations, studies have not been able to achieve good accuracy. Although existing studies have not achieved sufficient accuracy levels to provide real operational benefits, there have also been no clear recommendations on how accuracy can be improved, particularly in relation to factors affecting incident duration which are not typically recorded or quantified by road operators.

This paper presents a study which incorporates an extensive dataset, with 3 years of incident records from the M50 motorway in Ireland. Weather data, traffic data, and various incident details are used to test the ability of different machine learning models to predict the duration of incidents. A detailed discussion of the results is presented, and a number of practical recommendations are made in relation to how structured recording of relevant details could improve the ability of machine learning algorithms to estimate the expected duration of incidents.

3 Data and methods

3.1 Incident data

The incident dataset used in this study was provided by Transport Infrastructure Ireland (TII) and contains various features for incidents on the M50 motorway, including the date, time, location, category, type, travel direction, closed lane count and emergency services attendance at the incident scene. Time and coordinates cannot be easily used as a direct input to the machine learning models, so the GPS coordinates were used to identify the location of the incidents relative to the nearest junction to provide a systematic way of inputting the incident locations.

Each junction has different characteristics and geometry which were known, based on operational experience, to be an important factor in the management and duration of incidents. Therefore, the distance to the nearest junction to each incident was calculated by measuring the straight distance from the incident to the adjacent junctions. As shown in Fig. 1, and noting that traffic drives on the left in Ireland, the satellite map is used to identify the entry and exit based on the road markings and geometry to determine the geographical extent of the junction. Based on the location, and direction of travel, incidents were classified as (i) before, (ii) within or (iii) after the junction. Thus, three incident location features (junction number, location relative to junction and distance from junction) provide a means of systematically including the relevant information about the incident location and the nearest junction in the machine learning models.

On different days, drivers have various travel patterns, such as commuting to and from work on weekdays, travelling on weekends, and so on. Some routes are predictable, while others are not, and the traffic patterns, demand level and driver behaviour are all influenced by the day type. The UK Department Transport (2016) divided the dates into 13 types and in addition to each day of the week, they distinguished between different types of holidays and different days within holiday periods. The time of day is also an important consideration. Peak periods were defined as 06:30–09:30 am and 3:30–6:30 pm on weekdays for the morning and evening peak periods, respectively.

3.2 Weather data

Weather data was obtained from five weather stations installed along the M50, which record data every 10 min. The measured data included up to 17 different parameters including rainfall/precipitation, road surface temperature, visibility, surface water film thickness amongst others. Due to the limited measuring range of

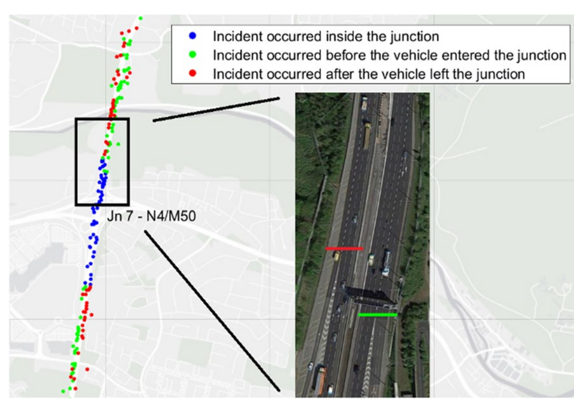

Fig. 1 Incident location characteristics at Junction 7. Source: the authors

the visibility sensors, the visibility data is almost entirely capped at 2000 m. In addition, although the surface water thickness on the road is a useful parameter, its measurement principle limits the measuring distance, making the data only applicable to incidents directly in the vicinity of the sensor. As such, rainfall was used instead of surface water thickness when making predictions. When using the Machine Learning algorithms to predict the duration of incidents, this study used the road surface temperature and the precipitation intensity, taking the average value of each parameter over the duration of the incident.

3.3 Traffic data

Transport Infrastructure Ireland (TII) has installed numerous ITS devices on the M50 to provide real-time information on traffic conditions, as described by De Paor et al. [5]. This includes double inductance loops spaced at 500 m intervals along the length of the M50. These loops record the number of vehicles, average speed, vehicle length, headway and lane occupancy in each lane every 20 s. Figure 2 shows the M50 motorway, which circumnavigates Dublin city. The different junctions along the motorway are labelled, and the location of the double-inductance loops are represented by the blue dots.

Working on behalf of TII, Roughan & O'Donovan (ROD) Consulting Engineers provided the data, including the duration for which each incident affected traffic flows. These durations, based on examination of the data from the inductive loops, were utilised in this study as they are more reflective of the impact of incidents on traffic compared to the durations logged by control room operators. For example, a minor incident at night might not be cleared immediately, but may not affect traffic flow at all, so this type of incident would not be considered in the analysis. The drawback of using this dataset is that a large number of incident records with limited impact on traffic are removed, which makes the average incident duration in this study much longer than typical incidents which are often cleared very quickly with limited impact on traffic. On the other hand, it means that the dataset being used contains high-impact incidents, where the incident response planning and ability to predict the incident duration is much more critical for control room operators.

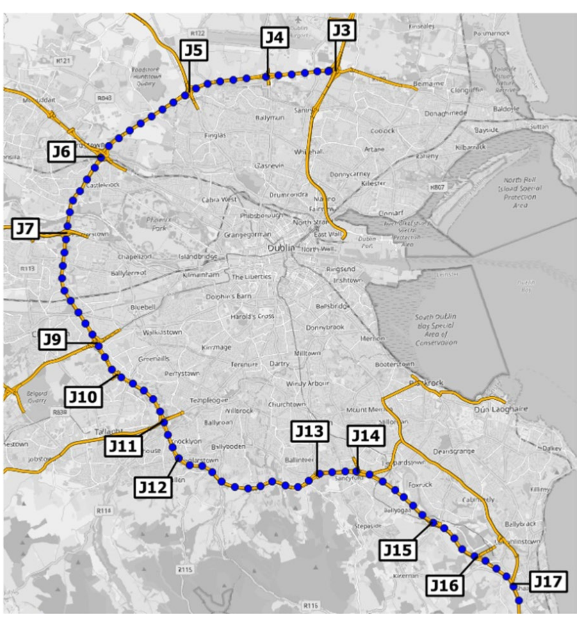

Fig. 2 M50 Motorway in Dublin, showing the location of double-inductance loops for traffic monitoring. Source: the authors

3.4 Machine learning algorithms for estimating incident duration

The following subsections outline the three primary models which were adopted in the various analyses conducted in this study, to examine their capabilities for estimating the duration of incidents by leveraging the various datasets.

3.4.1 Classification and regression trees (CART)

CART is a decision tree model built using data properties for classification and regression analysis [2]. The tree's nodes represent properties, branches represent splits according to the property, and leaf nodes represent the final decision. The CART algorithm starts at the root node and uses the minimum mean square error criterion for regression trees and the Gini index criterion for classification trees to split the tree, dividing the data into two parts at a time until each node becomes a leaf node.

For the CART method's tree structure, nodes closer to the top are usually more important. Node risk is defined as $R_j = P_j \cdot G_j$, where G is the Gini index and P is the node probability. The change in node risk is the risk difference between the parent node and two child nodes. If a tree splits a parent node into two children, the importance of the feature can be thought of as $(R_1 - R_2 - R_3)/N_{branch}$. Feature importance is the sum of all nodes that use the feature.

3.4.2 Support vector machines (SVM)

A Support Vector Machine (SVM) is a binary classification model that separates samples by finding a hyperplane that divides them into two classes while maximizing the distance from the nearest sample point to the hyperplane. For samples that are indistinguishable in a two-dimensional space, the data can be mapped to a higher dimensional space using a transformation, and then distinguished in the new space using a linear method. SVM only cares about data points that are close to the hyperplane, indicating it is not susceptible to disturbance by discrete values. SVM can be used for both classification and regression calculations [18].

3.4.3 Artificial neural networks (ANN)

Artificial neural networks (ANNs) simulate the way neurons in the human brain communicate and collaborate to process information [27]. A traditional ANN model consists of an input layer, an output layer, and one or more hidden layers. The input layer, which is the first layer of the neural network, receives information and passes it on to the next layer without operating on the input information or having any weights. The hidden layers accept signals from the previous layer, weight and sum the input data, compute using an activation function, and output the result to the next layer. The output layer, which is the final layer of the network, receives signals from the last hidden layer, weights and sums it, and outputs the model prediction.

4 Results

The dataset employed for this study consisted of 1391 incidents on the M50 motorway between 2017 to 2019. These incidents represented those which had the full incident details documented and which also had a significant adverse impact on traffic. The mean incident duration in the dataset was 97.04 min, with a median of 85 min, a lower quartile value of 52 min and an upper quartile value of 135 min. The features used in the analysis include incident category, primary incident type (e.g. collision, breakdown etc.), peak hours, day type, closed lane count, rainfall intensity, road surface temperature, police attendance, fire service attendance, ambulance attendance, junction number, location to junction and distance to junction entry/exit. For both regression and classification analysis, Classification and Regression Trees (CART), SVM and ANN are used. For training and validation of each of the models presented in this paper the incident data was split into two sets. To train the models, and optimise the hyperparameters, 80% of the incidents were randomly chosen from the full database and the remaining 20% of incidents were used for validation purposes. The hyperparameters for the different machine learning models, along with the architecture for the ANN models, were automatically varied during the training process, using the classification and regression learner functions in MATLAB. Therefore, the hyperparameters, and relevant architecture for the ANN models, are different for each set of results presented (depending on which gave the optimal result).

4.1 Multiple linear regression

At the outset, an initial test was carried out where a linear regression was applied to examine whether simple linear relationships could be used to assess the influence of different variables on the incident duration. Multiple linear regression is used to analyse the features linearly related to incident duration, with the dependent variable being the logarithm of the duration and the independent variables being normalised to between 0 and 1. Table 1 shows the results, where it can be seen that Closed Lane Count, Location to Junction and Distance to Junction Entry/Exit are not statistically significant (i.e. t-stat between −2 to +2), while the other factors are statistically significant. However, with an R^2 of only 0.185, an F-ratio equal to 41.2 and a p-value equal to 0, the model has a poor explanatory, indicating that the multiple linear regression model cannot accurately capture the incident duration. This is similar to the predictions of Peeta, Ramos, and Gedela (2000) who found an R^2 of 0.234.

Table 1 Coefficients of the multiple linear regression model

	Coefficients (bi)	Std. error	t-Stat	p-value
(Intercept)	3.9266	0.04900	80.134	0
×1: Incident Category	0.22988	0.08386	2.7413	0.00619
×2: Time Period	0.46601	0.03412	13.6580	4.86E−40
×3: Closed Lane Count	0.01313	0.06481	0.2026	0.83951
×4: Rain	0.62985	0.17856	3.5274	0.00043
×5: Police Attendance	0.14502	0.03434	4.2232	2.56E−05
×6: Fire Service or Ambulance	0.23743	0.04566	5.1999	2.28E−07
×7: Location to Junction	−0.06856	0.04360	−1.5725	0.11605
×8: Distance to Junction Entry/Exit	−0.08747	0.08858	−0.9875	0.32356

4.2 Machine learning analysis

4.2.1 Regression analysis

A regression analysis was initially applied to all incident data. The SVM with a Gaussian kernel function was shown to provide the greatest prediction performance, achieving a Mean Absolute Error (MAE) of 40.66 min and MAE% of 42.2%. Figure 3 shows the predicted duration vs. the actual duration for all incidents when using the SVM or ANN. Figure 3a demonstrates that the SVM model overestimates the duration of shorter incidents and is unable to predict incidents longer than 150 min and it is clear that the model only ever makes predictions within a band of between 40 to 150 min. A single hidden layer ANN model using Rectified Linear Unit (ReLU) as the activation function also obtains results with a similar prediction accuracy. Although the ANN model performs less well than the SVM with a MAE of 45.1 min, Fig. 3b shows that it may be more flexible than the SVM and can be suitable for predicting incidents of longer duration, as the model is capable of making predictions of longer duration incidents, however some erroneous predictions of longer durations reduce the overall accuracy of the model.

Table 2 presents the accuracy of the SVM, the ANN and the Regression Tree models, with the accuracy of the predicted durations for collisions, breakdowns and debris/spillage incidents shown separately. The results are presented in terms of the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Mean Average Error (MAE%). The SVM model has the best predictive performance for all three incident types, while the ANN provides the lowest accuracy. It is also noted that the ANN appears to provide much lower accuracy for vehicle breakdowns or incidents involving debris or spillages on the road. This is due to the large variation in the duration of these incidents, purely related to the variable nature of these incidents, which sometimes may be considered a priority and other times may not (whereas collisions are generally considered a priority, irrespective of the impact on traffic conditions). For example, a piece of cardboard debris on the road during heavy, slow-moving traffic, might not be posing a significant risk and the incident may be ongoing for some time, whereas a concrete block which falls from a truck onto the live carriageway would likely result in a partial road closure but would be removed as a matter of urgency. These specific details are not systematically recorded in the incident database, so cannot be used as part of the predictive modelling.

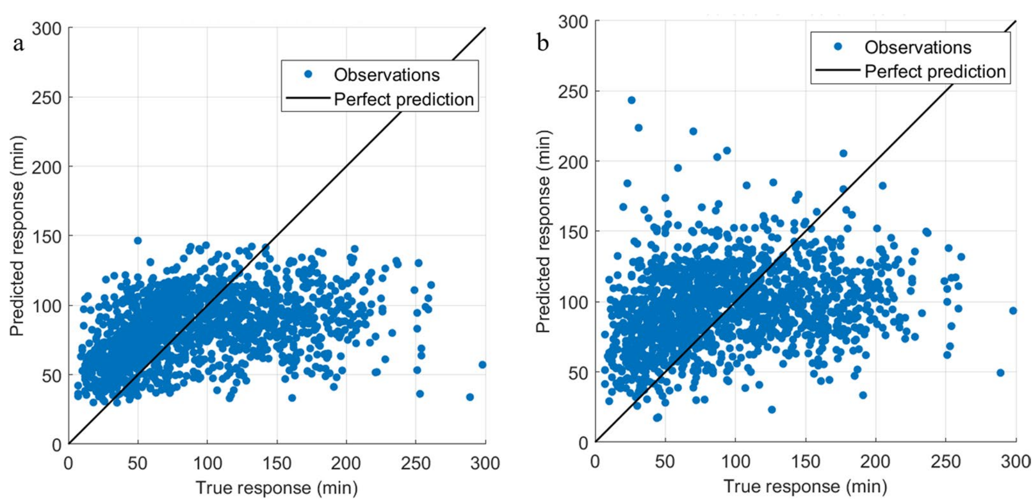

Fig. 3 Predicted vs. true values for **a** SVM and **b** ANN. Source: the authors

Table 2 Regression analysis for specific incident types

Incident type	Collisions			Breakdown			Debris or spillage		
	MAE	**RMSE**	**MAE%**	**MAE**	**RMSE**	**MAE%**	**MAE**	**RMSE**	**MAE%**
SVM	42.65	54.63	38.7	40.75	56.78	45.6	36.98	54.52	53.1
Regression tree	43.51	54.67	39.5	41.90	54.48	46.9	39.20	52.97	56.3
ANN	48.58	62.05	44.1	58.10	80.50	65.0	57.57	80.49	82.6

4.2.2 Classification analysis

In order to facilitate a classification-based assessment, the incident dataset was divided into four parts, with incidents classified based on their duration. This approach aimed to predict which class the incident is expected to fall within, hence providing an indication of the expected duration of the incident. Although it may not be as useful as being able to calculate the exact duration of an incident, it may reduce some of the uncertainty and allow the models to provide a more reliable estimate of the expected timescale for the expected duration. This would provide operators with a general estimate of the expected duration and allow the appropriate response to be determined. To facilitate the classification-based assessment, the dataset was divided as evenly as possible, and Table 3 shows the classification of incidents based on duration.

Classification analyses were first performed on all incident data. The SVM, using a linear kernel function, demonstrated the best prediction performance, with an accuracy of 40.8%. Although this level of accuracy appears to be lower than some of the results reported in previous literature, given the complex nature of the longer duration incidents considered in this study, the accuracy is actually similar to or better than many previous studies. A detailed discussion on the reasons for the accuracy levels achieved is provided in Sect. 5. Fig 4 shows the confusion matrix for the classifications made using the SVM. It can be seen that the model performs better for predicting incidents of less than 52 min duration and those lasting longer than 85 min, highlighting the fact that the data cannot capture some of the unexpected circumstances which typically lead to extended duration of incidents (e.g., recovery trucks might get delayed causing the incident to last much longer). The results using the other models demonstrated a similar trend.

After the classification analysis had been applied to the full incident dataset, the analysis was re-run for different incident types separately. Table 4 shows the results for the scenarios where collisions, vehicle breakdowns and debris/spillages were analysed in isolation, along with the results for the case when all incident types were considered simultaneously. It can be seen from the results that the classification accuracy is noticeably higher for debris or spillages (50.2–52.8%) whereas the classification accuracy for collisions or broken down vehicles remains close to 40%, showing similar accuracy to when all incident types were considered simultaneously. Classification trees provide the best predictions for collisions and breakdowns, while the ANN has the highest accuracy in predicting the debris incidents. It is interesting to note the significant improvement in the classification of debris incidents, when considering them separately. In this case the ANN provides the best results, whereas the SVM model was superior when considering all incident types together. It is also interesting that when collisions and breakdowns are considered separately, although the accuracy is not noticeably improved (in the case of collisions it has actually reduced), the classification tree provides the highest accuracy in both cases. This suggests that the unique characteristics of different incident types may result in different levels of accuracy, but also may require different machine learning models to ensure the best accuracy.

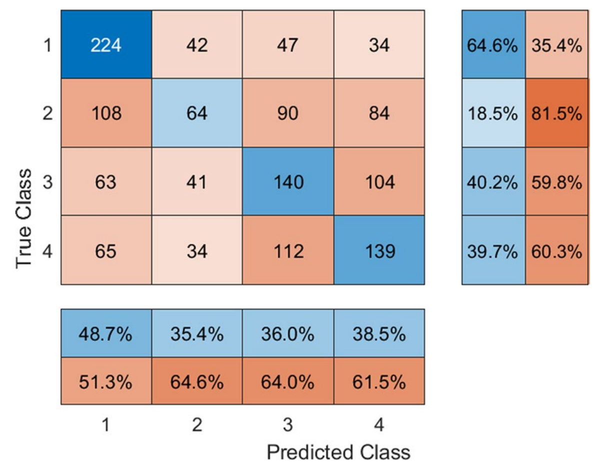

Fig. 4 Confusion matrix for SVM classification predictions, for all incident types. Source: the authors

Table 3 Classification of incident durations

Incident duration	Classification	Sample size
< 52 min	1	347
> = 52 & < 85 min	2	346
> = 85 & < 135 min	3	348
> = 135 min	4	350

Table 4 Classification accuracy for specific incident types

Accuracy	Collisions (%)	Breakdown (%)	Debris or spillage (%)	All incidents (%)
SVM	35.20	39.60	50.20	40.8
Classification tree	38.50	41.60	50.60	39.0
ANN	36.20	39.30	52.80	37.0

4.3 Importance estimates of predictor variables for tree models

In order to examine the importance of the various factors in predicting the duration of the incidents, importance estimates of the predictor variables were calculated according to the node risk of each split, for both regression and classification trees. The relative importance of each input variable is established by summing changes in the node risk (as defined in Sect. 3.4) due to splits on every predictor variable and then dividing the sum by the number of branch nodes. This allows the contribution of each input variable to be quantified and the relative importance to be established. Since the splitting criteria vary, they are normalized to facilitate comparison. Table 5 lists the importance estimates, with a value of 1.0 representing the most important variable.

The results indicate that incident type, and time period during the day are the most important criteria for the regression tree model, with weather variables and attendance of the first service also being relevant considerations. For the classification tree model the incident type and time period are also the most significant factors but the incident category, day type, and closed lane count show greater importance in this model. It is interesting to note that the number of closed lanes is not considered relevant in the regression tree model, which demonstrates a problem with the data recording approach above all else. In reality, the number of lanes affected by an incident will almost always have a direct correlation with the expected duration. However, the way the lane closures are recorded by control room operators only indicates whether each lane was open/closed during the incident. Therefore, a lane which was closed for 5 min of a 2-h incident is not recorded any differently to a lane which was closed for the full 2 h.

Table 5 Importance estimates of predictor variables

Features	Regression trees	Classification trees
Incident category	0.0185	0.0370
Incident type	0.4606	0.5062
Time period	1.0	1.0
Day type	0.0034	0.4431
Closed lane count	0	0.1729
Precipitation	0.2623	0.3557
Surf temperature	0.2501	0.6191
Police attendance	0.0241	0.1088
Fire service	0.2288	0.1876
Ambulance	0	0
Junction number	0.0921	0.3771
Location to junction	0.0334	0.0347
Distance to junction	0.0594	0.2246

5 Discussion

At first glance, the results obtained in this study do not appear to provide particularly high accuracy, and even appear to be less accurate than other studies on the subject. For example, Valenti et al. [23] achieved an MAE of 16 min with the ANN model, which is much better compared to the MAE of 40 min in this study. However, in the Valenti et al. study, the mean incident duration was 45 min, whereas in this study, the mean incident duration is 97 min. As a result, the percentage error is 40%, which is similar to an 35% error achieved by Valenti et al. Similarly, Leahy and Lynch [11] also looked at incident duration on the Irish M50 motorway and achieved a MAE of 24.9 min. However, the mean duration of the incidents considered was approximately 30 min, which is notably lower than the incidents considered in the present study, and equates to an error of approximately 80%. Hence, the results of this study are comparable or better in terms of percentage error to others which have been achieved elsewhere. In addition to this, the fact that this study focuses on longer duration incidents, which tend to be more variable in nature, and hence more difficult to estimate, demonstrates the value of the results presented in this study.

Regarding the classification analysis, Chang and Chang [3] achieved a classification accuracy of 75.1%. This is primarily due to the fact that they categorised 72% of the incidents as short duration, 25.7% as medium duration and only 2.2% as long duration. Their classification was unbalanced compared to this study, and since shorter incidents are usually the most common with less error, it is not surprising that an accuracy rate of 75% was achieved. This study divided incident durations into four even classes with classification accuracy not skewed towards shorter incidents.

Overall, this study and most others have not been able to achieve very accurate results, and the large prediction errors reflect the diversity and unpredictability of real-world incidents and the need to improve accuracy in order to apply this research to highway operations. The results do, however, offer some hope that depending on the type of incidents, different models can be combined to make better predictions.

One shortcoming of this study relates to the fact that more than 60% of the original dataset was removed to focus on incidents which had a high-impact on traffic conditions, potentially resulting in the dataset not being truly representative. Thus, only 1391 incidents between 2017 and 2019 were considered in the analysis.

Irrespective of this issue, it is clear that unpredictability & varied nature of motorway incidents make it difficult to systematically model the expected duration, without knowing the exact influencing factors affecting the duration of a given incident. For example, the availability of vehicle recovery services, the proximity of emergency response vehicles to the incident location, the severity of injuries, number of vehicles involved etc. will all have a large influence on the duration of an incident but are currently not recorded systematically. The primary recommendation of this study would be to develop systematic methods for unbiased recording of additional incident details to allow other important factors to be considered when attempting to estimate the expected duration of an incident.

In this study, three machine learning methods are employed and it was found that certain models perform better in specific scenarios. This is mainly because these three methods have different working principles and modelling characteristics. Classification trees make decisions based on feature splits, while SVMs aim to identify a hyperplane that maximizes the margin between positive and negative samples. ANNs consist of multiple layers and nodes that capture complex nonlinear relationships in the data through activation functions. In addition, these three models have different complexity, hyperparameter selection, and training strategies, which may also lead to different results.

For example, the ANN was seen to provide a much larger spread in the predictions, indicating that the model was more sensitive to the input variables, which in many cases have a significant effect on the incident duration. However, despite the ability of the ANN to fit to trends in the input data, in many cases, this often resulted in poor predictions, because the effect of the input variables and the way that they are recorded is not always consistent. As an example, when more lanes are closed in response to an incident, the incident will usually take longer to clear. However, the data does not contain exact details about how long each lane closure lasted, so in some cases lanes may have been closed for many hours, but in others the lanes may have been closed for less than a minute, but the input variables do not allow these differences to be distinguished from each other. Thus, with more rigorous recording of pertinent incident details, the ANN will have a better chance to provide more accurate predictions.

On the other hand, the SVM showed lower variability in the predicted durations. Shorter duration incidents tended to be over-estimated, with longer duration incidents being under-estimated. For the longer duration incidents, which tended to be under-estimated, there was a much less consistent pattern in the predictions. Again, this is more likely related to the nature of longer-term incidents, and the lack of information which is captured about other contributing factors which actually cause the incidents to last so long. When managing an incident, the cause of the increased duration is usually very apparent, however, the variability and unexpected nature of such causes mean that they are generally not recorded at all, or certainly not in a systematic way which is amenable to training machine learning models.

During collection of the incident data, various observations were made while monitoring incidents from the motorway control room. These observations identified very clear reasons for extended incident durations in many cases. For example, drivers of vehicles involved in incidents often reported medical conditions over the phone which directly influenced the complexity and duration of the response. As another example, the extent/type of debris on the carriageway often had a large influence on the duration of clearance and the impact on the network. The types of vehicles involved in collisions, or the presence of other ongoing incidents elsewhere also had a noticeable influence on response time (e.g. due to the limited availability of recovery trucks for heavy goods vehicles). However, many of these details (medical conditions reported, specific nature of debris, number and types of vehicles involved in collisions etc.) are not systematically recorded. As such, the machine learning models cannot account for the significant influence of these contributing factors. This is further discussed in Sect. 5.1. For this reason, although it was found that certain models were more suited to predicting the incident duration for different scenarios, the primary recommendations for further improvements are in relation to the systematic recording of relevant information to provide improved data sets to allow such models to be comprehensively trained.

It is also worth noting that the incident database, and the other available information used in this study, is already extremely comprehensive and contains significant detail. However, this just highlights the difficulty in capturing and recording the many variables which can influence the life-cycle of an incident, and ultimately affect the overall duration. This is likely a significant reason for the reasonably low accuracy levels of estimated incident duration which are reported in the literature.

5.1 Recommendations for recording of incident details

On the basis of the findings of this study, a number of additional key pieces of information were observed to have a noticeable impact on the duration of incidents, but were not systematically recorded. These additional pieces of information, and the incident types for which they are most relevant, are summarised in Table 6

Table 6 Additional information recommended for collection during incident management

Information	Relevant incident type
Availability / proximity of recovery vehicles or emergency services to incident location	Collision, vehicle breakdown, vehicle fire
Severity of injuries / medical conditions	Collision, medical emergencies, vehicle fire
Number of vehicles involved	Collision, vehicle breakdown
Types of vehicles involved	Collision, vehicle breakdown, vehicle fire
Specific per-lane restriction details	All types
Details of nearby roadworks / traffic restrictions / traffic control measures	All types
Details of the nature of debris on the carriageway	Debris on the carriageway (or incidents which result in debris on the carriageway)
Physical road characteristics including location of nearby junctions, lane configuration, geometry	All types
Traffic conditions at the time of the incident	All types

followed by a discussion of the various pieces of information which are recommended. It should be noted that, in many cases this information was recorded or available in some format, however, it was usually contained in free-text, or in a separate location from the incident log. In order to facilitate the use of machine learning approaches, or any detailed analysis of historical incident records, this information would need be recorded in a systematic way, in a consistent format. In addition, it should be noted that the recommendations are specific to the data set examined and the recommendations are not exhaustive as there may be other key pieces of information which would provide added benefit, which were not apparent or relevant to this study.

Based on the recommendations provided in Table 6, the following observations are made with regard to recording of additional incident details to allow road operators to collect data which can be leveraged to provide more accurate estimates of incident duration:

- Data should be recorded in a systematic way, with limited free-text input from incident responders or control room operators so that the format is consistent across all incident records.
- Details related to the availability and proximity of vehicle recovery services or emergency response vehicles to the incident location (and expected response times), where available, should be recorded at the time when the incident is detected as this has an immediate impact on the response/clearance time.
- The severity of injuries, number and types of vehicles involved should be recorded in a systematic way, as all of these factors can influence the complexity of the response required.
- Exact details of lane closures should be recorded including the start / end chainage of each closure, the time at which each lane was closed / re-opened and the duration and reason for each closure.
- The lane(s) in which the original incident occurred should be recorded, as this is where the immediate impact is observed, and the impact can be very different depending on where the incident occurred.
- For debris incidents, information about the type of debris should be formally recorded in a systematic way to help with understanding the potential risk associated with the debris and to inform the nature of the response type required.
- For vehicle breakdowns, the type of vehicle along with availability and location of recovery vehicles should be recorded as the type of vehicle will determine the recovery vehicle/equipment required.
- Information about the physical characteristics of the road geometry at the incident location, and in the vicinity are also relevant, and depending on the availability of information could be automatically populated/recorded based on the GPS coordinates of the incident location.
- Traffic conditions at the time of the incident should be recorded, if these will not be automatically recorded from ITS equipment / other data sources.
- Similarly, traffic control measures, restrictions, or roadworks which are in place at the time of the incident should be recorded if these are not being recorded separately.

It is acknowledged that there will be many other factors which can influence the potential duration of an incident. In many cases, these things are not easy to systematically record, or even be aware of, however recording as much information as possible, in a systematic way across all incidents, will reduce the number of unknowns and the level of variability of predictions. It is also noted that the nature of road traffic incidents means that the details will likely change during the course of the incident management life-cycle. Therefore, time-stamped records of any changes in the incident details should also be documented to facilitate re-evaluation of the expected incident impact as the incident evolves.

6 Conclusions

This paper incorporates incident, weather and traffic flow datasets from the M50 motorway in Ireland to predict the duration of motorway incidents using different machine learning methods. It is suggested that the accuracy of regression models can be enhanced by combining SVMs and ANNs, with SVMs shown to be superior in predicting short and medium duration incidents, and ANNs providing better accuracy for longer duration incidents. When predicting the duration of specific type of incidents, SVMs demonstrate the greatest predictive performance in regression analysis. Each machine learning method is found to have its own advantages for different durations when using a classification-based approach, and a combination of these three machine learning methods can be employed to improve prediction accuracy.

Multiple linear regression and tree models are employed to examine the importance of influencing factors, with both methods considering the type of incident, time of occurrence and the presence of emergency services at the incident scene as significant factors in determining the incident duration. Weather conditions are also shown to be an influential factor. To improve data availability and systematic recording of incident details, this paper makes a number of recommendations for information which should be recorded to enable more precise estimates of incident duration to enable improved operational responses to real-time incidents. This would ultimately enable better regulation of traffic flow, reducing congestion due to incidents and safeguarding the emergency responders, while limiting the overall impact on the network.

Further development of the techniques presented in this paper has the potential to provide control room operators with more accurate estimates of incident duration which is a key factor in deciding on the appropriate incident management response. Better predictions of incident duration allow suitable diversion routes or warning messages to be communicated to drivers and helps maintain network efficiency while the incident is being cleared.

Acknowledgements
The authors would like to acknowledge Transport Infrastructure Ireland (TII) and Roughan & O'Donovan (ROD) Consulting Engineers for the provision of the data sets used in this study.

Author contributions
LY carried out the data analysis and the application of the machine learning methods along with the initial drafting of the article. RC was responsible for supervision, data collection and preparation, conceptualisation and design of the research along with the final editing and drafting of the manuscript. AM was responsible for supervision and conceptualisation of the research. All authors read and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
The authors do not have permission to share the data used in this study.

Declarations

Competing interests
The authors have no conflict of interests to declare.

Received: 28 February 2023 Accepted: 12 January 2024
Published online: 19 February 2024

References

1. Ali, S. S. M., George, B., Vanajakshi, L., & Venkatraman, J. (2011). A multiple inductive loop vehicle detection system for heterogeneous and lane-less traffic. *IEEE Transactions on Instrumentation and Measurement, 61*(5), 1353–1360.
2. Breiman, L. (2017). *Classification and regression trees*. Routledge.
3. Chang, H.-L., & Chang, T.-P. (2013). Prediction of freeway incident duration based on classification tree analysis. *Journal of the Eastern Asia Society for Transportation Studies, 10*, 1964–1977.
4. Corbally, R., O'Connor, A., & Cahill, F. (2016). Practical applications of weigh-in-motion data. *Paper presented at the Civil Engineering Research in Ireland Conference, Galway, Ireland*.
5. De Paor, C., Corbally, R., Duranovic, M., Feely, L., & O'Sullivan, A. (2018). The role of motorway traffic flow optimisation indicators in enhancing motorway operation services in the Irish road network. *Paper presented at the 25th ITS World Congress, Copenhagen, Denmark*.
6. Dia, H., & Gondwe, W. (2008). Evaluation of incident impacts on integrated motorway and arterial networks using traffic simulation. *Paper presented at the 29th Australasian Transport Research Forum*.
7. Dimitriou, L., & Vlahogianni, E. I. (2015). Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions. *Analytic Methods in Accident Research, 5*, 59–71.
8. Hojati, A. T., Ferreira, L., Washington, S., Charles, P., & Shobeirinejad, A. (2014). Modelling total duration of traffic incidents including incident detection and recovery time. *Accident Analysis & Prevention, 71*, 296–305.
9. Khoury, J. A., Haas, C. T., Mahmassani, H., Logman, H., & Rioux, T. (2003). Performance comparison of automatic vehicle identification and inductive loop traffic detectors for incident detection. *Journal of Transportation Engineering, 129*(6), 600–607.
10. Kong, X., Xia, F., Ning, Z., Rahim, A., Cai, Y., Gao, Z., & Ma, J. (2018). Mobility dataset generation for vehicular social networks based on floating car data. *IEEE Transactions on Vehicular Technology, 67*(5), 3874–3886.
11. Leahy, C., & Lynch, S. (2018). Analysis of traffic incidents using machine learning. *Paper Presented at the Civil Engineering Research in Ireland Conference*, University College Dublin, Ireland
12. Lemonde, C., Arsenio, E., & Henriques, R. (2021). Integrative analysis of multimodal traffic data: Addressing open challenges using big data analytics in the city of Lisbon. *European Transport Research Review, 13*, 1–22.
13. Li, R., Pereira, F. C., & Ben-Akiva, M. E. (2018). Overview of traffic incident duration analysis and prediction. *European Transport Research Review, 10*(2), 1–13.
14. Ma, J., Hu, J., Hale, D. K., & Bared, J. (2016). Dynamic hard shoulder running for traffic incident management. *Transportation Research Record, 2554*(1), 120–128.
15. Ozbay, K., & Kachroo, P. (1999). *Incident management in intelligent transportation systems*. Artech House Publishers.

16. Park, H., Haghani, A., & Zhang, X. (2016). Interpretation of Bayesian neural networks for predicting the duration of detected incidents. *Journal of Intelligent Transportation Systems, 20*(4), 385–400.
17. Peeta, S., Ramos, J. L., & Gedela, S. (2000). *Providing real-time traffic advisory and route guidance to manage borman incidents on-line using the hoosier helper program* (2326–6325). Retrieved from
18. Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101–121). Elsevier.
19. Prabha, R., & Kabadi, M. G. (2016). Overview of data collection methods for intelligent transportation systems. *The International Journal Of Engineering And Science (IJES), 5*(3), 16–20.
20. Smith, K., & Smith, B. L. (2002). Forecasting the clearance time of freeway accidents.
21. Tchuitcheu, W. C., Bobda, C., & Pantho, M. J. H. (2020). Internet of smart-cameras for traffic lights optimization in smart cities. *Internet of Things, 11*, 100207.
22. Transport, D. f. (2016). *Reliability of Journeys on the Highways Agency's Motorway and "A" Road Network*. Retrieved from UK:
23. Valenti, G., Lelli, M., & Cucina, D. (2010). A comparative study of models for the incident duration prediction. *European Transport Research Review, 2*(2), 103–111.
24. Vlahogianni, E. I., & Karlaftis, M. G. (2013). Fuzzy-entropy neural network freeway incident duration modeling with single and competing uncertainties. *Computer-Aided Civil and Infrastructure Engineering, 28*(6), 420–433.
25. Yu, B., Wang, Y., Yao, J., & Wang, J. (2016). A comparison of the performance of ANN and SVM for the prediction of traffic accident duration. *Neural Network World, 26*(3), 271.
26. Zong, F., Zhang, H., Xu, H., Zhu, X., & Wang, L. (2013). Predicting severity and duration of road traffic accident. *Mathematical Problems in Engineering, 2013*.
27. Zou, J., Han, Y., & So, S.-S. (2009). Overview of artificial neural networks. *Artificial Neural Networks: Methods and Applications*, 14–22.
28. Zwahlen, H. T., Russ, A., Oner, E., & Parthasarathy, M. (2005). Evaluation of microwave radar trailers for nonintrusive traffic measurements. *Transportation Research Record, 1917*(1), 127–140.

Publisher's Note