TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–10

DOI: 10.1177/0361198119841571
journals.sagepub.com/home/trr
SAGE

Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction

Athanasios Theofilatos[1,2], Cong Chen[3], and Constantinos Antoniou[1]

Abstract

Although there are numerous studies examining the impact of real-time traffic and weather parameters on crash occurrence on freeways, to the best of the authors' knowledge there are no studies which have compared the prediction performances of machine learning (ML) and deep learning (DL) models. The present study adds to current knowledge by comparing and validating ML and DL methods to predict real-time crash occurrence. To achieve this, real-time traffic and weather data from Attica Tollway in Greece were linked with historical crash data. The total data set was split into training/estimation (75%) and validation (25%) subsets, which were then standardized. First, the ML and DL prediction models were trained/estimated using the training data set. Afterwards, the models were compared on the basis of their performance metrics (accuracy, sensitivity, specificity, and area under curve, or AUC) on the test set. The models considered were *k*-nearest neighbor, Naïve Bayes, decision tree, random forest, support vector machine, shallow neural network, and, lastly, deep neural network. Overall, the DL model seems to be more appropriate, because it outperformed all other candidate models. More specifically, the DL model managed to achieve a balanced performance among all metrics compared with other models (total accuracy = 68.95%, sensitivity = 0.521, specificity = 0.77, AUC = 0.641). It is surprising though that the Naïve Bayes model achieved a good performance despite being far less complex than other models. The study findings are particularly useful, because they provide a first insight into performance of ML and DL models.

Understanding the various factors that cause road crashes and their combined influence is crucial. Although there has been considerable research effort so far, there is still much to be investigated, especially for the purpose of acquiring better knowledge of detailed preaccident conditions for a better proactive safety management on major roads of the transport network. Advances in the field of intelligent transport systems (ITS) and meteorology have enabled the constant and detailed monitoring of real-time traffic and weather conditions and have contributed to the safety assessment of major roads.

As a consequence, during the past decade, interest in effective real-time traffic management strategies has been constantly increasing. This can be achieved through examination of the critical traffic and weather factors that affect crash likelihood on freeways measured on a real-time basis. International literature indicates that much effort has recently been dedicated to the aforementioned relationship (*1–4*). One of the main practical applications is in real-time traffic management schemes such as variable speed limits (*5*).

One of the major limitations of international literature in the field of real-time safety evaluation and analysis is that often a validation framework is missing when predicting crash risk. Furthermore, in the era of machine learning (ML) and deep learning (DL) analytic techniques, there are many relevant techniques that have been rarely or never applied.

Therefore, this study aims to fill this methodological gap and add to the current knowledge by comparing and validating a list of ML and DL methods, to predict crash occurrence. For this study, real-time traffic and weather data from Attica Tollway freeway in Greece were utilized and linked with historical crash data for 2006 to 2011

[1]Transportation Systems Engineering, Technical University of Munich, Munich, Germany
[2]National Technical University of Athens, Athens, Greece
[3]Center for Urban Transportation Research, University of South Florida, Tampa, FL

Corresponding Author:
Address correspondence to Athanasios Theofilatos: a.theofilatos@tum.de

(before the financial crisis). The following ML and DL methods were compared on the basis of their performance metrics (accuracy, sensitivity, specificity, and area under curve, or AUC) on the test set:

- *k*-nearest neighbors (*k*-NN);
- Naïve Bayes (NB);
- Decision trees (DT);
- Random forests (RF);
- Support vector machines (SVM);
- Shallow neural network (SNN–shallow learning); and
- Deep neural network (DNN–deep learning).

Background

Crash risk (or crash likelihood) is usually a typical binary classification problem; the most frequently adopted approach, therefore, is to include data for crash cases, but also for random noncrash cases and apply logistic models or other prediction models, such as SVM, *k*-NN, Bayesian networks, or neural networks (*3*, *6–10*). Very recently, DL techniques have also gained increased attention from research scholars. Yang et al. applied a DL neural network for real-time crash prediction on urban expressways (*11*). The authors linked historical crash data with real-time traffic data from two expressways in Shanghai and they combined and divided data into training and testing sets. The DL models proved to be a promising tool for real-time crash prediction.

Whereas many studies rely on loop detector (LD) data, other data sources such as automatic vehicle identification (AVI) sensors have been used (*12*). The study authors deployed separate models for LD and AVI data. The former model showed that the low average speed and high speed variance (both in the time slice of 5 to 10 min before the crash time) were found to significantly increase crash risk. According to the latter model, only the coefficient of variation in speed had an effect on crash occurrence. The authors conclude that that both LDs and AVI systems can be used for safety applications.

Whereas the majority of related studies concern basic freeway segments, several recent studies have been extended to expressways, weaving segments of freeways, and urban arterials (*2*, *13–16*). With respect to the methods of analysis, alternative modeling approaches have also been recently utilized (*17*, *18*). For instance, Theofilatos et al. considered crashes as rare events and applied alternative methodologies, such as the Firth logistic model, to address this issue (*18*). Similarly, Yasmin et al. proposed an alternative to the case-control binary logit real-time crash risk analysis, through a multinomial logit model (*17*). Their approach also allowed the common unobserved factors influencing crash likelihood to be addressed.

The literature review identified various real-time traffic and weather factors as critical. Overall, some of the major factors that influence crash occurrence were found to be the variance of traffic parameters, such as coefficient of variation of speed and flow (*4*, *9*, *15*, *19*), as well as low visibility and adverse weather (*20–22*). Speed variance is also found to increase the probability of moped and motorcycle crashes (*23*), as well as rear-end crashes (*24*). Interestingly, low driving speeds were often found to be associated with increased crash risk (*12*, *18*, *25*). This could be attributed to low speeds on freeways mostly representing congested traffic.

Although there has been considerable research effort so far, there is still much to be investigated, especially for acquiring better knowledge of detailed precrash conditions for better proactive safety management on major roads of the transport network. Moreover, several issues have been identified and discussed in Roshandel et al., who provided a review and metaanalysis of the factors influencing crash risk and found that only 61% of their reviewing studies provided a validation framework (*26*). Indeed, only a few studies have provided joint model performance metrics (*3*, *27*). Roshandel et al. also found that in real-time prediction literature the prediction accuracy, false positive rate, and false negative rate are usually examined, but only a few studies have utilized all three measures to comprehensively validate their model's performance (*26*). It is also stated in that study that such a comprehensive validation is critical for any prediction model. Moreover, it is reported that balancing prediction accuracy and false positives/negatives is an important issue that has not yet been addressed in the real-time crash risk literature. It is suggested that many studies did not report any measure of their model's performance (*26*).

One of the first road safety studies that applied ML models was that of Abdel-Aty and Haleem (*28*). Despite quite a large number of ML techniques (such as SVM, *k*-NN, RF, and neural networks) having been utilized, their comparative performance for predicting crash likelihood has yet to be explored. Furthermore, there are also some ML models (such as the NB classifier), which have been applied in a few cases (*29*), but not for real-time crash prediction. To the best of the authors' knowledge, Iranitalab and Khattak provide the only extensive comparison among different ML methods, as they explore crash severity prediction (*30*). More specifically, the authors utilized and compared multinomial logit, nearest neighbor classification, SVMs and RFs and found that the correct prediction rates and the proposed approach showed that the nearest neighbors had the best prediction performance in overall and in more severe crashes.

Summing up, several studies have been conducted to predict crash risk when real-time traffic and weather parameters are utilized. However, to the best of the authors' knowledge, there is still very limited research into the compared predictive performance of various ML and DL methods under a validation framework.

Data Collection and Preparation

In the present research scheme, crash and traffic data were collected from Attica Tollway ("Attiki Odos"), which is an urban motorway located in the Greater Athens area, Greece. It is a modern motorway, with a length of 65.2 km and two directionally separated carriageways, each consisting of three lanes and an emergency lane.

Three data sets were used in this analysis: one data set with crash data, one with traffic data, and one with weather data. The required crash data for Attica Tollway were extracted from the Greek crash database SANTRA provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens. Traffic data were collected from the Traffic Management and Motorway Maintenance Department of Attica Tollway. Inductive loops (sensors) were used to provide information about traffic flow (in vehicles per 5-min per lane), occupancy (in %), speed (km/h) and truck proportion (in %). Each traffic parameter is measured in 5-min intervals and each crash was assigned to the closest upstream LD in the crash segment of the road.

Weather data were extracted from the website of the Hydrological Observatory of Athens (www.hoa.ntua.gr). The observatory has more than 10 stations located in the Greater Athens area, measuring various environmental parameters. Thus, each crash was assigned to the closest meteorological station and rainfall, temperature, relative humidity, solar radiation, wind speed, and wind direction were utilized. Each weather measurement is aggregated in 10-min intervals.

Finally, one comprehensive database was created in which traffic and weather characteristics are matched to each crash and noncrash case. Following past literature (*7*, *12*, *31–33*) a matched case study was applied. Therefore, a 2:1 ratio of noncrash to crash cases was created; there were two observations for noncrash cases for the same location (same time of crash 1 week before and 1 week after the actual crash) for each crash observation of the data set. The raw traffic and weather data were further aggregated to obtain averages, standard deviations, and coefficient of variations as in previous literature. In some cases, LDs suffered from problems that might have resulted in unreasonable values for speed, flow, and occupancy. Such unrealistic values (e.g., speed > 0 along with flow $= 0$, etc.) were discarded from the database.

The final data set consisted of 284 crash cases and 592 noncrash cases; all types of vehicle crashes and crash severities were used (slight, severe, and fatal). It is noted that only basic freeway segments (BFS) were included in the analysis.

To carry out the ML and DL techniques, the data set was split into training and test sets. More specifically, 75% of cases were used for calibration and 25% for validation. Afterwards, to account for the different range of independent variables' values, the independent variables of the calibration and validation sets were standardized. This standardization is needed because some ML methods may rely on Euclidean distance measure and consequently the predictability of the models is improved. To the best of the authors' knowledge, this issue is often neglected in road safety studies.

Methodology

In this research, several ML, and one DL, classification methods were used for crash likelihood prediction along with a binary logistic regression that quantifies the impact of the most important variables. This section presents the prediction methods used in the study. Furthermore, the various metrics for comparison of ML and DL while examining the real-time crash likelihood predictive performance are discussed later. In this study, the values of various tuning parameters of the models were determined by testing different values for these quantities and finding the best prediction accuracy and AUC value.

ML Techniques

K-*Nearest Neighbor (*k-*NN).* The k-NN method is an ML prediction method that predicts and classifies an observation by looking at the closest k observations. When implementing the k-NN algorithm, the researcher needs to specify the value of k and the distance function. Lastly, Euclidean distance is used as the distance function. The k-NN approach is briefly summarized in the following steps:

- Step 1: Choose the number, k, of neighbors.
- Step 2: Take the k nearest neighbors of the new data point, according to the Euclidean distance or some other metric.
- Step 3: Among these k neighbors, count the number of data points in each category.
- Step 4: Assign the new data point to the category in which the most neighbors are counted.

Naïve Bayes (NB). The NB classifier technique belongs to the family of simple "probabilistic classifiers," which are based on Bayes' theorem with strong (naïve) independent assumptions between the features. Despite its simplicity, NB can often outperform more sophisticated classification methods (*34*). The Bayes theorem provides a way of calculating posterior probability $P(y|x)$ from $P(y)$, $P(x)$ and $P(x|y)$:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} \tag{1}$$

where

$P(y|x)$ is the posterior probability of class (y, target) given predictor (x, attributes),

$P(y)$ is the prior probability of class,

$P(x|y)$ is the likelihood which is the probability of predictor given class, and

$P(x)$ is the prior probability of predictor.

For a binary class ($y = y_0$ or y_1), a probability threshold t can be set so that data cases where $P(y = y_1|x) > t$ are classified as y_1 and all other cases are classified as y_0.

Decision Trees (DTs). A DT classifier is represented as a tree structure that contains branches splitting the data set from one root node (the topmost node in a tree) to leaf nodes. Each leaf node generally represents a class value which is classified through the splitting path from the root node to the leaf node. This method is often used in road safety (*35–37*) but rarely for real-time crash prediction.

The aforementioned studies (*35–37*) provide a good description of the principles of DTs and the following description is based on them. The basic principle behind tree growing is to recursively partition the target variable to maximize "purity" in the child node. DTs are built recursively with a descending strategy. The root node (which contains all data), is divided into two branches on the basis of an independent variable that creates the best homogeneity. Afterwards, each branch is connected with a child node and the data in each child node are more homogeneous than those in the upper parent node. Lastly, each child node is split recursively until all of them are considered as pure (i.e., when all the cases are of the same class) or their "purity" cannot be further increased. The split criterion of the DT is based on Gini, which represents the diversity of a predictor, and is calculated as

$$\text{Gini} = 1 - \sum_{1}^{n} \rho_i^2 \tag{2}$$

where

i is the category of the target (crash or noncrash),

n is the total number of targets, and

p is the percentage of crash or noncrash cases.

Because the DT in this study is a binary tree, the total number of targets is two.

Random Forests (RF). In this study, the RF model has a dual role; it is used as a preliminary and as a predictive tool. In the former case, it is used to rank the variables according to their relative importance and thus assist in selecting the most appropriate independent variables before applying other statistical or ML models. This is a very common approach in road safety studies having a lot of candidate independent predictors (*15*, *23*, *28*, *38*, *39*).

Overall, an RF is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta k), k = 1,\ldots\}$, where the $\{\Theta k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. For more details and theoretical background about RF models, the reader is encouraged to refer to Breiman (*40*). It is noted that to account for the bias caused by correlated predictor variables (*41*), the conditional variable importance is used. This is another issue often neglected in international literature.

Support Vector Machines (SVM). An SVM constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification or regression. A good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class. The theoretical background of SVMs is well documented in previous studies in road safety (*8*, *42–44*), and so the reader is encouraged to follow the aforementioned studies.

In this study, the radial basis function (RBF) kernel is utilized. More specifically, the RBF kernel on two samples x and x', represented as feature vectors in some input space, is defined as

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{3}$$

where the term $||x\text{-}x'||^2$ is the Euclidean distance between the two feature vectors, and σ is a free parameter.

Shallow Neural Network (SNN–Shallow Learning). The last type of ML model that was applied in this study is a basic neural network (NN), which is a "shallow" NN with only a single layer. NNs contain a series of neurons which are interconnected and process the input. The connections between neurons are weighted. These weights are based on the functions utilized and learned from the data. Activation in one set of neurons and the weights may then feed into the other neurons. In this analysis, the activation function used is the rectifier with dropout.

In simple words, dropout refers to ignoring units (e.g., neurons) during the training phase of a certain set of neurons which is chosen at random. This is done to prevent overfitting.

The hidden units can be denoted by h, the outputs by Y. Different outputs can be denoted by subscripts $i = 1, ..., k$. The paths from each hidden unit to each output are the weights and for the ith output are denoted by w_i. When dealing with a classification problem (as in this study), the following transformation takes place ensuring that the estimates are positive:

$$Y_i = \frac{e^{w_i^T h}}{\sum_i^k e^{w_i^T h}} \tag{4}$$

DL Techniques: Deep Neural Network (DNN–Deep Learning)

The most informative (and perhaps simplest) definition of DNN, is that it is a neural network which has multiple hidden layers. This allows for a more sophisticated build-up from simple elements to more complex ones. Two main complexity aspects exist. The former is simply how many neurons exist in a given layer (i.e., how wide or narrow it is) and the latter how many layers of neurons exist (how deep it is).

In this paper a deep feedforward NN (DFNN) is applied. The DFNN is designed to approximate a function *f*() that maps a set of input variables x to an output variable y. In such models, information flows from the inputs through the hidden layer and finally to the output (without feedback or recursive loops). In this analysis the activation function used is the rectifier with dropout to prevent overfitting.

Performance Metrics

Performance metrics of model validation included accuracy, sensitivity, specificity, and AUC. To calculate sensitivity and specificity values the following measures are calculated:

- True positive (correct identification of a crash case);
- True negative (correct identification of a noncrash case);
- False positive (a noncrash case incorrectly identified as a crash case); and
- False negative (a crash case incorrectly identified as a noncrash case).

Thus the following performance metrics can be calculated:

$$\text{Overall accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{Total crashes}} \tag{5}$$

$$\text{Sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \tag{6}$$

$$\text{Specificity} = \frac{\text{True negative}}{\text{True negative} + \text{False positive}} \tag{7}$$

Ultimately, to combine the false positives and the true positives into one single metric, the two metrics are first computed with many different thresholds (for example 0, 0.01, 0.02, …, 1) for each prediction model and then plotted on a single graph. The resulting curve is the ROC (receiver operating characteristic) curve, but the metric considered is the AUC value of this curve.

Results: Preliminary Analysis

First, an RF model was applied to assess the impact of independent variables according to their relative importance, similarly to the approach of (*15*, *28*, *45*). In this study the conditional importance was used to take into account and overcome potential multicollinearity among independent variables according to Strobl et al. (*41*). Figure 1 illustrates the results of the RF model.

This analysis was proven to be useful, because it provided a first insight into the candidate variables. The following variables were considered for further analysis and are illustrated in Table 1. It is noted that rainfall was also entered as a candidate variable. Afterwards, the candidate variables were entered into the binary logistic model to find the statistically significant ones. The variables identified as statistically significant would then be entered as input to the ML and DL models.

Table 2 illustrates the results of the binary logistic model. The model shows reasonable fit (likelihood ratio test: 114.352, R-square: 0.126). According to the model, only two variables were found to be significant: the standard deviation of speed for the 0 to 15 min time slice before the crash and the total amount of rainfall. More specifically, both variables are found to have a positive relationship with crash likelihood (increase crash risk). This is in line with previous research in the field, such as that by Theofilatos (*15*) and Ahmed et al. (*25*). The reason for indicating only two significant variables may be the limited variance of the data set or potential multicollinearity among variables (which did not allow the use of a high number of predictors at the same time).

As a final methodological step, these two statistically significant variables were entered as input to the ML and DL prediction models. All candidate models were trained on the 75% part of the data set (training set) and their prediction performance was tested on the 25% part of the

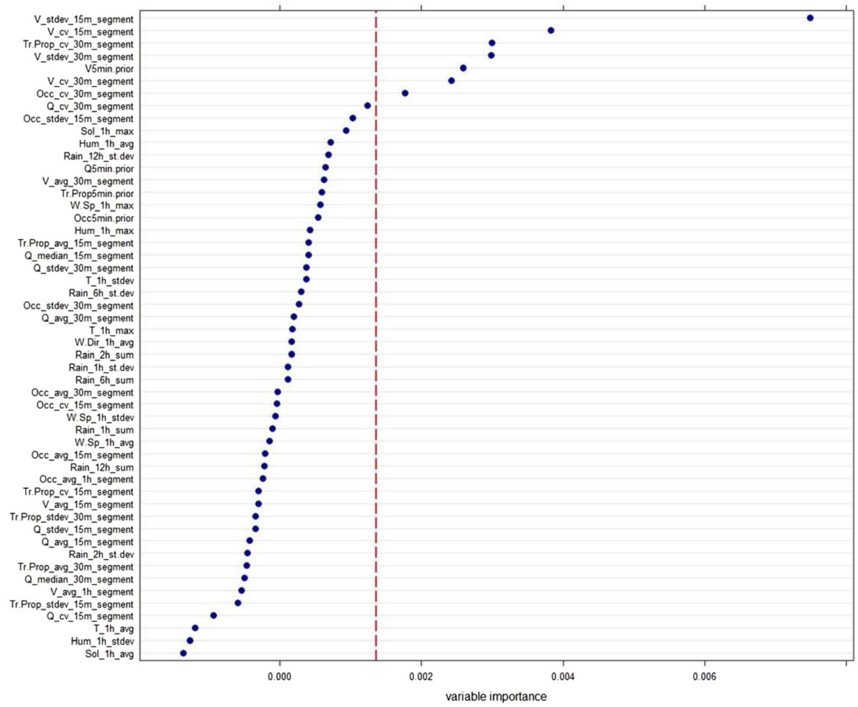

Figure 1. RF model plot.

Table 1. Descriptive Statistics of the Final Selected Variables (before Standardization)

Variable	Description	Type	Mean	Standard deviation
Q_cv_30 m_segment	Coefficient of variation of flow in the time slice 0–30 min before crash occurrence (unitless)	Continuous	0.1178	0.101
V5 min.prior_segment	Average speed in the time slice 0–5 min before crash occurrence (km/h)	Continuous	101.672	16.932
V_stdev_15 m_segment	Standard deviation of speed in the time slice 0–15 min before crash occurrence (km/h)	Continuous	3.119	3.119
V_cv_30 m_segment	Coefficient of variation of speed in the time slice 0–30 min before crash occurrence (unitless)	Continuous	0.036	0.059
V_cv_15 m_segment	Coefficient of variation of speed in the time slice 0–15 min before crash occurrence (unitless)	Continuous	0.032	0.061
V_stdev_30 m_segment	Standard deviation of speed in the time slice 0–30 min before crash occurrence (km/h)	Continuous	3.119	3.687
Occ_cv_30 m_segment	Coefficient of variation of occupancy in the time slice 0–30 min before crash occurrence (unitless)	Continuous	0.129	0.106
Tr.Prop_cv_30 m_segment	Coefficient of variation of proportion of traffic in the time slice 0–30 min before crash occurrence (unitless)	Continuous	0.528	0.563
Rain_12 h_sum	Sum of precipitation in the time since 0–12 h before the crash (mm)	Continuous	0.428	2.714

data set (test set). It is noted that a random split into 75% and 25% was carried out. The following bullets summarize the models' attributes utilized to achieve best results:

- *K*-NN: six neighbors;
- NB: not applicable;
- DT: not applicable;
- RF: 100 trees;
- SVM: radial basis kernel, cost = 1, gamma = 100;
- SNN: one layer with 50 hidden neurons, 100 epochs, activation function = rectifier with dropout;

Table 2. Overview of Binary Logistic Model Results

Variable	Beta coefficient	Standard error	*T*-test	*P*-value
Constant term	–0.747	0.086	–8.73	0.000***
V_stdev_15 m_segment	0.509	0.123	4.13	0.000***
Rain_12 h_sum	0.166	0.096	1.74	0.080**

***Significant at 95% level. **Significant at 90% level.

Table 3. Overview of Overall Prediction Comparison Measures (Test Set)

Method	Overall accuracy	Sensitivity (true positive rate)	Specificity (true negative rate)	AUC
K-NN	62.56%	0.437	0.716	0.575
NB	72.15%	0.225	0.959	0.724
DT	72.15%	0.324	0.912	0.688
RF	62.56%	0.423	0.723	0.573
SVM	67.12%	0.225	0.885	0.574
Binary logistic regression	62.10%	0.521	0.669	0.608
NN with 1 layer (shallow)	57.99%	0.661	0.541	0.628
NN with 4 layers (deep)	68.95%	0.521	0.770	0.641

Note: AUC = area under curve.

- DNN: four layers of hidden neurons (200, 200, 400, and 200 respectively), 500 epochs.

Table 3 illustrates the summary findings of the ML and DL prediction models, in relation to overall accuracy, true positive and negative rate, and AUC value. It is noted that larger values of all these metrics indicate better performance.

A first glance at the table reveals that all candidate models perform relatively well. The highest overall accuracy is achieved by NB and DT models (72.15% in both models), whereas the SNN has a surprisingly low overall accuracy (57.99%) compared with the other ML models. On the other hand, it cannot be ignored that the SNN is the best model in relation to sensitivity (true positive rate detection).

Perhaps one problem of almost all applied models is the detection of true positives (crash cases). The highest sensitivities are achieved by the shallow (0.661) and deep (0.521) NNs. On the other hand, all ML and DL models were capable of predicting the true negatives (noncrash cases). The DT and NB models achieved particularly high specificity (0.912 and 0.959 respectively). However, they had poor sensitivities (0.324 and 0.225 respectively) and as such they are not the preferred models. The SNN had the lowest specificity (0.541), but the best sensitivity (0.661).

In relation to the AUC values, the NB was surprisingly adequate, having the highest AUC value (0.724). Overall, it can be concluded that perhaps the most adequate model was the DL model, because although all other models were good at identifying true negatives, they had very low values of true positive rates. On the other hand, the DL model achieved a balanced and satisfactory performance for all metrics. It also had the second highest true positive detection rate (something important for real-time crash prediction studies) and the second highest AUC value.

Conclusions

Although there are numerous studies which examine and predict the impact of real-time traffic flow and weather parameters on crash occurrence, to the best of the authors' knowledge there were no studies that attempted to compare the prediction performances of well applied ML and DL models. In addition, a common limitation of a lot of previous studies in the field was the lack of a validation framework, as discussed previously by Roshandel et al. (*26*). That paper has some more methodological contributions. For instance, the ML and DL methods are better applied along with standardization of the data, because several such models rely on the Euclidean distance. Consequently, it may be inappropriate to use variables with very different scales.

Therefore, this paper presents a methodological framework for comparing the predictive performance of ML and DL, to examine real-time traffic and weather conditions prone to crashes. With this research purpose, 6 years of crash data and the corresponding real-time traffic and weather data on Attica Tollway ("Attiki Odos") were utilized. First, an RF model and binary logistic model were utilized to provide insight into potentially significant variables among a set of many variables. It was found that the standard deviation of speed and total rainfall were associated with increased crash risk. Afterwards, the ML and DL models were used for predictive purposes by using the significant variables indicated by the aforementioned models as input. That only two variables were found significant can be attributed to limited variance in the set; or perhaps more relevant traffic variables could be used (e.g., median of traffic flow).

The most appropriate model is probably the DL model, because it outperformed all other candidate models. More specifically, the DL model managed to achieve a balanced performance among all metrics, and had the second highest true positive detection rate as well as the second highest AUC value. Lastly, it is surprising that the NB model achieved a very good performance, although being far less complex than the other models.

The authors acknowledge the limitations of the present research. Although ML and DL methods are becoming increasing popular, what still may be an issue is the limitation around the tuning parameters for some of the methods. These methods may be appropriately applied for a particular data training set and validation set, but the issue of transferability arises along with retuning of the parameters, similarly to traditional statistical models. A potential limitation of the applied ML and DL methods is that they can considered as "black box" methods, because they lack a good interpretation when compared with a traditional statistical model, which is explanatory. In that context, unobserved heterogeneity could not be sufficiently explained by this type of model, but this could be counterbalanced by the purpose being not to identify the impact of observed or unobserved factors but to adequately predict crashes. There are also some methods to partially overcome this limitation, such as through a sensitivity analysis, but this was beyond the scope of the present paper. In this analysis, the approach was supplemented by a binary logistic model, which shows the effect and significance of independent predictors on crash risk.

Overall, the findings of the study are particularly useful because they provide a first insight on the compared performance of ML and DL models in the field of road safety. Application of such methods would allow policy makers and freeway operation authorities to identify high risk conditions and apply selected countermeasures. For instance, variable message signs (VMS) or variable speed limits (VSL) could be used to inform/warn drivers after potential risky conditions are identified.

Summing up, the findings suggest that more studies in the future on the field of predicting real-time crash likelihood on freeways should be carried out through the application of DL models to evaluate the results of the present study. Future studies should also attempt to utilize even more complex architectures and structures of DL models.

Acknowledgments

This research was carried out within the framework "Support for Postdoctoral Researchers" of the scientific program "Development of Human Resources, Education and Lifelong Learning" 2014–2020, which is being implemented from the Greek State Scholarships Foundation (IKY) and was co-funded by the European Social Fund and the Greek public. The authors would like to thank the Attica Tollway Operations Authority ("Attiki Odos") for providing the data, especially officers Pantelis Kopelias and Fanis Papadimitriou.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: AT; data collection: AT; analysis and interpretation of results: AT, CC, CA; draft manuscript preparation: AT, CC, CA. All authors reviewed the results and approved the final version of the manuscript.

References

1. Abdel-Aty, M., A. Pande, C. Lee, V. Gayah, and C. Dos Santos. Crash Risk Assessment using Intelligent Transportation Systems Data and Real-Time Intervention Strategies to Improve Safety on Freeways. *Journal of Intelligent Transporation Systems*, Vol. 11, No. 3, 2007, pp. 107–120.
2. Hossain, M., and Y. Muromachi. A Real-Time Crash Prediction Model for the Ramp Vicinities of Urban Expressways. *IATSS Research*, Vol. 37, No. 1, 2013, pp. 1–12.
3. Hossain, M., and Y. Muromachi. A Bayesian Network Based Framework for Real-Time Crash Prediction on the Basic Freeway Segments of Urban Expressways. *Accident Analysis and Prevention*, Vol. 45, 2012, pp. 373–381.
4. Yu, R., X. Wang, and M. Abdel-Aty. A Hybrid Latent Class Analysis Modeling Approach to Analyze Urban Expressway Crash Risk. *Accident Analysis and Prevention*, Vol. 101, 2017, pp. 37–43.
5. Abdel-Aty, M., and L. Wang. Implementation of Variable Speed Limits to Improve Safety of Congested Expressway Weaving Segments in Microsimulation. *Transportation Research Procedia*, Vol. 27, 2017, pp. 577–584.
6. Pande, A., A. Das, M. Abdel-Aty, and H. Hassan. Estimation of Real-Time Crash Risk: Are All Freeways Created Equal? *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2237: 60–66.
7. Sun, J., and J. Sun. A Dynamic Bayesian Network Model for Real-Time Crash Prediction using Traffic Speed Conditions Data. *Transportation Research Part C: Emerging Technologies*, Vol. 54, 2015, pp. 176–186.
8. Yu, R., and M. Abdel-Aty. Utilizing Support Vector Machine in Real-Time Crash Risk Evaluation. *Accident Analysis and Prevention*, Vol. 51, 2013, pp. 252–259.

9. Huang, Z., Z. Gao, R. Yu, X. Wang, and K. Yang. Utilizing Latent Class Logit Model to Predict Crash Risk. *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, China, 2017, pp. 161–165.
10. Katrakazas, C., M. A. Quddus, and W. H. Chen. A Simulation Study of Predicting Conflict-Prone Traffic Conditions in Real-Time. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
11. Yang, K., X. Wang, M.A. Quddus, and R. Yu. Deep Learning for Real-Time Crash Prediction on Urban Expressways. Presented at 97th Annual Meeting of the Transportation Research Board, Washington, D.C., 2018.
12. Abdel-Aty, M., H. M. Hassan, M. Ahmed, and A. S. Al-Ghamdi. Real-Time Prediction of Visibility Related Crashes. *Transportation Research Part C: Emerging Technologies*, Vol. 24, 2012, pp. 288–298.
13. Wang, L., M. Abdel-Aty, Q. Shi, and J. Park. Real-Time Crash Prediction for Expressway Weaving Segments. *Transportation Research Part C: Emerging Technologies*, Vol. 61, 2015, pp. 1–10.
14. Wang, L., Q. Shi, and M. Abdel-Aty. Predicting Crashes on Expressway Ramps with Real-Time Traffic and Weather Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2514: 32–38.
15. Theofilatos, A. Incorporating Real-Time Traffic and Weather Data to Explore Road Accident Likelihood and Severity in Urban Arterials. *Journal of Safety Research*, Vol. 61, 2017, pp. 9–21.
16. Theofilatos, A., G. Yannis, E. I. Vlahogianni, and J. C. Golias. Modeling the Effect of Traffic Regimes on Safety of Urban Arterials: The Case Study of Athens. *Journal of Traffic and Transportation Engineering (English Edition)*, Vol. 4, No. 3, 2017, pp. 240–251.
17. Yasmin, S., N. Eluru, L. Wang, and M. A. Abdel-Aty. A Joint Framework for Static and Real-Time Crash Risk Analysis. *Analytic Methods in Accident Research*, Vol. 18, 2018, pp. 45–56.
18. Theofilatos, A., G. Yannis, P. Kopelias, and F. Papadimitriou. Impact of Real-Time Traffic Characteristics on Crash Occurrence: Preliminary Results of the Case of Rare Events. *Accident Analysis and Prevention*, forthcoming. http://dx.doi.org/10.1016/j.aap.2017.12.018.
19. Yu, R., and M. Abdel-Aty. Multi-Level Bayesian Analyses for Single- and Multi-Vehicle Freeway Crashes. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 97–105.
20. Xu, C., W. Wang, and P. Liu. Identifying Crash-Prone Traffic Conditions under Different Weather on Freeways. *Journal of Safety Research*, Vol. 46, 2013, pp. 135–144.
21. Yu, R., Y. Xiong, and M. Abdel-Aty. A Correlated Random Parameter Approach to Investigate the Effects of Weather Conditions on Crash Risk for a Mountainous Freeway. *Transportation Research Part C: Emerging Technologies*, Vol. 50, 2015, pp. 68–77.
22. Jung, S., X. Qin, and D. Noyce. Rainfall Effect on Single-Vehicle Crash Severities using Polychotomous Response Models. *Accident Analysis and Prevention*, Vol. 42, No. 1, 2010, pp. 213–224.
23. Theofilatos, A., and G. Yannis. Investigation of Powered 2-Wheeler Accident Involvement in Urban Arterials by Considering Real-Time Traffic and Weather Data. *Traffic Injury Prevention*, Vol. 18, No. 3, 2017, pp. 293–298.
24. Dimitriou, L., K. Stylianou, and M. Abdel-Aty. Assessing Rear-End Crash Potential in Urban Locations Based on Vehicle-by-Vehicle Interactions, Geometric Characteristics and Operational Conditions. *Accident Analysis and Prevention*, Vol. 118, 2018, pp. 221–235.
25. Ahmed, M. M., M. Abdel-Aty, and R. Yu. Assessment of Interaction of Crash Occurrence, Mountainous Freeway Geometry, Real-Time Weather and Traffic Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2012. 2280: 51–59.
26. Roshandel, S., Z. Zheng, and S. Washington. Impact of Real-Time Traffic Characteristics on Freeway Crash Occurrence: Systematic Review and Meta-Analysis. *Accident Analysis and Prevention*, Vol. 79, 2015, pp. 198–211.
27. Xu, C., P. Liu, and W. Wang. Evaluation of the Predictability of Real-Time Crash Risk Models. *Accident Analysis and Prevention*, Vol. 94, 2016, pp. 207–215.
28. Abdel-Aty, M., and K. Haleem. Analyzing Angle Crashes at Unsignalized Intersections using Machine Learning Techniques. *Accident Analysis and Prevention*, Vol. 43, No. 1, 2011, pp. 461–470.
29. Chen, C., G. Zhang, J. Yang, J. C. Milton, and A. Dely. An Explanatory Analysis of Driver Injury Severity in Rear-End Crashes using a Decision Table/Naïve Bayes (DTNB) Hybrid Classifier. *Accident Analysis and Prevention*, Vol. 90, 2016, pp. 95–107.
30. Iranitalab, A., and A. Khattak. Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction. *Accident Analysis and Prevention*, Vol. 108, 2017, pp. 27–36.
31. Xu, C., W. Wang, and P. Liu. Identifying Crash-Prone Traffic Conditions under Different Weather on Freeways. *Journal of Safety Research*, Vol. 46, 2013, pp. 135–144.
32. Ahmed, M., and M. Abdel-Aty. The Viability of using Automatic Vehicle Identification Data for Real-Time Crash Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 2, 2012, pp. 459–468.
33. Ahmed, M., and M. Abdel-Aty. A Data Fusion Framework for Real-Time Risk Assessment on Freeways. *Transportation Research Part C: Emerging Technologies*, Vol. 26, 2013, pp. 203–213.
34. Shanthi, S. Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms. *International Journal of Computer Applications*, Vol. 35, No. 12, 2011, pp. 30–37.
35. Hoon, O., W. Rhee, and Y. Yoon. Application of Classification Algorithms for Analysis of Road Safety Risk Factor Dependencies. *Accident Analysis and Prevention*, Vol. 75, 2015, pp. 1–15.
36. Chang, L., and J. Chien. Analysis of Driver Injury Severity in Truck-Involved Accidents using a Non-Parametric Classification Tree Model. *Safety Science*, Vol. 51, No. 1, 2013, pp. 17–22.
37. Griselda, L., D. O. Juan, and A. Joaquín. Using Decision Trees to extract Decision Rules from Police Reports on

Road Accidents. *Procedia – Social and Behavioral Sciences*, Vol. 53, 2012, pp. 106–114.
38. Ahmed, M., and M. Abdel-Aty. The Viability of using Automatic Vehicle Identification Data for Real-Time Crash Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 2, 2012, pp. 459–468.
39. Yu, R., and M. Abdel-Aty. Analyzing Crash Injury Severity for a Mountainous Freeway Incorporating Real-Time Traffic and Weather Data. *Safety Science*, Vol. 63, 2014, pp. 50–56.
40. Breiman, L.E. Random Forests. *Machine Learning*, Vol. 45, 2001, pp. 5–32.
41. Strobl, C., A. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. Conditional Variable Importance for Random Forests. *BMC Bioinformatics*, Vol. 11, 2008, pp. 1–11.
42. Theofilatos, A., G. Yannis, C. Antoniou, and D. Sermpis. Time Series and Support Vector Machines to Predict Powered-Two-Wheeler Accident Risk and Accident Type Propensity: A Combined Approach. *Journal of Transportation Safety and Security*, Vol. 10, No. 5, 2018, pp. 471–490.
43. Chen, C., G. Zhang, Z. Qian, R. A. Tarefder, and Z. Tian. Investigating Driver Injury Severity Patterns in Rollover Crashes using Support Vector Machine Models. *Accident Analysis and Prevention*, Vol. 90, 2016, pp. 128–139.
44. Li, X., D. Lord, Y. Zhang, and Y. Xie. Predicting Motor Vehicle Crashes using Support Vector Machine Models. *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 1611–1618.
45. Abdel-Aty, M., A. A. Ekram, H. Huang, and K. Choi. A Study on Crashes Related to Visibility Obstruction Due to Fog and Smoke. *Accident Analysis and Prevention*, Vol. 43, No. 5, 2011, pp. 1730–1737.

The Standing Committee on Safety Data, Analysis and Evaluation (ANB60) peer-reviewed this paper (19-01388).