sustainability

MDPI

Article

Traffic Accident Severity Prediction Based on Random Forest

Miaomiao Yan [1,2,3] and Yindong Shen [1,2,*]

[1] School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China; ymmwhu@163.com
[2] Key Laboratory of Image Processing and Intelligent Scheduling (Huazhong University of Science and Technology), Ministry of Education, Wuhan 430074, China
[3] Department of Mathematics, Huzhou University, Huzhou 313000, China
* Correspondence: yindong@hust.edu.cn

Abstract: The prediction of traffic accident severity is essential for traffic safety management and control. To achieve high prediction accuracy and model interpretability, we propose a hybrid model that integrates random forest (RF) and Bayesian optimization (BO). In the proposed model, BO-RF, RF is adopted as a basic predictive model and BO is used to tune the parameters of RF. Experimental results show that BO-RF achieves higher accuracy than conventional algorithms. Moreover, BO-RF provides interpretable results by relative importance and a partial dependence plot. We can identify important influential factors for traffic accident severity by relative importance. Further, we can investigate how the influential factors affect traffic accident severity by the partial dependence plot. These results provide insights to mitigate the severity of traffic accident consequences and contribute to the sustainable development of transportation.

Keywords: traffic accident severity; random forest; Bayesian optimization; road traffic safety; road safety

Citation: Yan, M.; Shen, Y. Traffic Accident Severity Prediction Based on Random Forest. *Sustainability* **2022**, *14*, 1729. https://doi.org/10.3390/su14031729

Academic Editors: Armando Cartenì and Matjaž Šraml

Received: 10 December 2021
Accepted: 27 January 2022
Published: 2 February 2022

1. Introduction

It is essential to precisely predict traffic accident severity for traffic safety management and control. Traffic accidents are a major threat that has induced enormous human injury and economic losses. The World Health Organization indicates that over 1.2 million lives have been lost and 50 million people have been injured owing to road traffic accidents [1]. The prediction of traffic accident severity provides emergency responders with crucial information for estimating the potential impacts and implementing timely accident management strategies [2]. However, traffic accidents are the result of complex interactions between humans, vehicles, roads, and the environment. Traffic accident data, which are composed of several to dozens of features, have the characteristics of high dimensions, nonlinearity, and multicollinearity. Hence, developing an accurate method for the prediction of traffic accident severity is a challenging task.

A large quantity of work has been devoted to traffic accident prediction. There are mainly two categories of models for predicting traffic accident severity: statistical models and artificial intelligence models. Statistical models have been widely used since the early 1990s, e.g., logit models and probit models [3–5]. Statistical models usually have strict assumptions about explanatory variables and response variables. The assumptions allow the results of the statistical models to be easily interpreted. However, violating these assumptions may lead to erroneous results [6]. Because the assumptions are not always satisfied, the application of statistical models is limited.

On the contrary, artificial intelligence models, which have no assumptions, are more flexible than statistical models. They can deal with complex nonlinear relationships and usually have higher prediction accuracy than statistical methods. Hence, artificial intelligence models have gained more and more popularity in recent years. For instance,

Moghaddam et al. estimated crash severity using artificial neural networks and identified significant crash-related factors in urban highways [7]. Taamneh et al. compared the performance of Decision Tree (J48), Multilayer Perceptron, and Naive Bayes in predicting accident severity [8]. Zheng et al. applied a convolutional neural network to predict traffic accident severity [9]. However, most artificial intelligence models are a black box and their results lack reasonable interpretation.

Random forest (RF) is an ensemble model based on decision trees. It can handle nonlinear, high-dimensional variables and is robust to outliers and noise [10]. Moreover, RF provides the relative importance of variables and partial dependence plots. Thus, it is easy to interpret results from RF. RF has been widely used for classification and regression in the transportation field, e.g., travel mode choice identification, road traffic condition prediction, and incident duration prediction [11–13].

The performance of RF is largely affected by the setting of hyperparameters. To enhance the performance of RF, it is important to find the optimal parameter values. Existing studies mostly use grid search to look for values in the parameter space. However, this requires a large number of evaluations when the parameter space is high-dimensional. Thus, grid search is computationally intensive. Alternatively, Bayesian optimization (BO) usually requires less computation. BO is powerful for selecting high-quality parameters of machine learning problems. It is suitable for the optimization of objective functions that are characterized by either non-existent analytical expressions or expensive evaluation.

In this study, we aim to predict the severity of traffic accidents on urban roads. Due to the advantage of RF in terms of both prediction accuracy and interpretation power, we choose RF as the basic predictive model. Further, to improve model performance, BO is used to adjust the parameters of RF. This hybrid approach is denoted as BO-RF.

It should be noted that we are concerned with traffic accidents on urban roads. Urban roads have the characteristics of heavy traffic flow and a complex traffic environment. It is more difficult to predict traffic accidents and ensure traffic safety on urban roads relative to the expressway. To ensure the right level of road traffic safety, the first task that can be performed is designing a safe road infrastructure. Roundabout intersections are a very good example of point road infrastructure. The advantages of this type of intersection (e.g., a small number of collision points compared to other types of intersections, speed reduction when crossing the intersection, low loss of time for drivers at inlets, etc.) contribute significantly to ensuring the appropriate level of road safety at a given point of the transport network [14–16]. On the other hand, the road infrastructure, e.g., intersections, bus stops, and junctions, reflect the characteristics of human activities and the built environment. Hence, the road infrastructure is closely related to traffic accidents. It is necessary to consider information about intersections, bus stops, and other points of interest (POI) in the prediction model. Hence, we incorporate POI as well as time, weather, and other information to construct features for the proposed BO-RF. Experimental results on a real traffic accident dataset show that BO-RF achieves higher prediction accuracy than the commonly used machine learning models. Moreover, it can identify the important influencing factors and display the relationship between influential factors and traffic accident severity. This will facilitate the analysis of the traffic accident severity and the design of proactive measures. Thus, it can help to enhance road traffic safety and contribute to the sustainable development of transportation.

The rest of this paper is arranged as follows. The second section introduces the accident data used in this study, followed by a specific description of the BO-RF model. Section 4 analyzes the experimental results. Section 5 presents the discussions. Conclusions are outlined at the end.

2. Data

The analysis of this study is based on the dataset US-Accidents [17–19]. This dataset contains approximately 2.25 million samples relating to traffic accidents in the United States from February 2016 to March 2019. Due to limited computational resources, the accident

data in Montgomery county of Pennsylvania state were selected as the target data. Each accident record is described by a wide range of data attributes, including the accident location, weather, time, POI, et al.

Raw data need to be pre-processed, including removing variables with too many missing values, filling variables, and coding variables. After pre-processing the data, a total of 30426 accident records were obtained. Each record consisted of 15 feature variables and one response variable. These variables are listed in Table 1. The 15 feature variables are Start_lat, Start_lng, Distance (mi), Month, Day, Hour, Weekday, Pressure (in), Temperature (F), Humidity (%), Visibility (mi), Traffic_signal, Junction, Crossing, Stop. They can be divided into four categories: traffic attributes, temporal attributes, weather attributes, and POI attributes. Table 2 presents a summary of the qualitative variables, including Start_lat, Start_lng, Distance (mi), Temperature (F), Humidity (%), Pressure (in), and Visibility (mi). Table 3 gives a summary of the quantitative variables including Crossing, Junction, Stop, and Traffic_signal.

Table 1. Variable description.

Category	Variable	Description
Traffic Attributes	Start_lat	Latitude in GPS Coordinates of the Start Point
	Start_lng	Latitude in GPS Coordinates of the Endpoint
	Distance (mi)	The Length of the Road Extent Affected by the Accident
Temporal Attributes	Month	1–12 [for January–December]
	Day	1,2, . . . ,31 [day]
	Hour	0,2, . . . ,23 [hour]
	Weekday	0–6 [MondayvSunday]
Weather Attributes	Pressure (in)	The Air Pressure (in inches)
	Temperature (F)	Temperature (in Fahrenheit)
	Humidity (%)	The Humidity (in percentage)
	Visibility (mi)	Visibility (in miles).
POI	Traffic_Signal	A POI annotation indicating the presence of traffic_signal in a nearby location: 0 = no; 1 = yes
	Junction	A POI annotation indicating the presence of junction in a nearby location: 0 = no; 1 = yes
	Crossing	A POI annotation indicating the presence of crossing in a nearby location: 0 = no; 1 = yes
	Stop	A POI annotation indicating the presence of a stop in a nearby location: 0 = no; 1 = yes
Response Variable	The Severity Level of Traffic Accident	1 = slight; 2 = serious; 3 = fatal

Table 2. Summary of quantitative variables.

	Mean	Std.	Min.	Max.
Start_lat	40.15	0.08	39.98	40.44
Start_lng	−75.29	0.14	−75.69	−75.02
Distance (mi)	−0.23	1.55	0	43.82
Temperature (F)	56.91	18.41	−6.00	97.00
Humidity (%)	70.43	20.35	18.00	100.00
Pressure (in)	30.02	0.25	28.85	30.84
Visibility (mi)	8.77	2.52	0.10	10.00

Table 3. Summary of qualitative variables.

	Severity Level: 1		Severity Level: 2		Severity Level: 3	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Crossing	3474	12.47%	19	0.81%	16	7.17%
Junction	251	0.90%	163	6.95%	26	11.66%
Stop	492	1.77%	1	0.04%	1	0.45%
Traffic_Signal	8105	29.09%	201	8.58%	41	18.39%

The response variable is traffic accident severity. It is classified into three ordered levels. Level 1 denotes a slight accident, which accounts for 91.56% of all accident records. Level 2 indicates serious accidents, which account for 7.7% of total accidents. Level 3 denotes fatal accidents, which have the lowest percentage of 0.73%.

3. Methodology

3.1. Random Forest

Random forest was proposed by Breiman in 2001 [10]. It aggregates the results from multiple decision trees to obtain a stronger learner. RF has good resistance to noise and does not easily fall into overfitting. It mainly consists of two ideas: bagging and random feature selection. In bagging, for a given training dataset with sample size n, the k new training set, whose sample size is n, is sampled from the original training set uniformly and with replacement. Then, k base trees are trained using the k new training set. Moreover, to obtain diversity between base trees, a feature subset is randomly selected to grow each tree in RF. When generating a single decision tree, splitting cannot be stopped unless the tree reaches its maximum depth. The procedure of RF is as follows:

- Using the bootstrap sampling technique, k training sets are produced from the original dataset.
- Each tree is trained on a different training subset. Noted in an individual decision tree, a randomly selected feature subset is used for splitting nodes.
- Step 2 is repeated with k iterations to obtain k decision trees.
- New inputs are fed into RF, and each tree gives a prediction result. The final classification result is determined by majority voting from all the decision trees.

Based on what is described in reference [20], we chose three main parameters that affect the tuning performance of random forest, including the total number of trees (n_estimators), the number of features used for each node segmentation (max_feature), and the maximum tree depth (max_depth).

3.2. Bayesian Optimization

For the RF model, several hyperparameters that need to be preset before the learning process have a marked impact on the model performance. Therefore, hyperparameter optimization is important to obtain optimal performance.

In this study, the objective function to be optimized is the performance of RF on a validation set. This can be regarded as the chosen parameters' function, which lacks an analytical expression. This means that the gradient descent method is infeasible. In addition, computation for evaluating a set of hyperparameter configurations that requires retraining the prediction model is very demanding. Bayesian optimization is a powerful sequential optimization tool for selecting high-quality parameters of machine learning problems. It is suitable for the optimization of black box functions whose evaluations are expensive. Unlike grid or random search, where every evaluation is independent of previous evaluations, BO uses the history of function evaluations to determine the next point to sample [21]. This greatly reduces the search time and enhances the optimization efficiency.

At each iteration of the optimization process, BO estimates the posterior distribution of the objective function using a surrogate model based on Bayes's theorem. Hence, the information of the previous sample can be fully utilized and many unnecessary evaluations of the objective function are avoided. Then, according to the distribution, an acquisition function is generated to evaluate the expected utility of assessing a candidate point. The acquisition function should offer a trade-off between exploration and exploitation, with little evaluation cost. The next evaluation point is the one that maximizes the acquisition function. This process terminates when enough data are obtained.

The basic procedure of BO is [22]:

- For $n = 1, 2, \ldots$, do

- select the next point to assess by maximizing the acquisition function $x_{n+1} = \underset{x}{\text{argmax}}\alpha(x; D_n)$
- obtain y_{n+1} by querying the objective function.
- increase data $D_{n+1} = \{D_n, (x_{n+1}, y_{n+1})\}$
- end for

In this study, BO is used to tune the hyperparameters of RF to obtain optimal performance. We adopt the tree parzen estimator (TPE) and expected improvement (EI) [23] as the surrogate model and acquisition function, respectively.

3.3. Relative Importance and Partial Dependence

Generally, it is useful to explore the influence of predictor variables on the response variable. The RF model can distinguish the contribution of a predictor variable to response prediction, while still maintaining relatively high accuracy. The importance value of the variable X_k in the tree T_m and RF model is defined in (1) and (2), respectively [24].

$$I_k^2(T_m) = \sum_{j=1}^{J-1} \tau_j^2 I_j(X_k) \tag{1}$$

$$I_k^2 = \frac{1}{M} \sum_{m=1}^{M} I_k^2(T_m) \tag{2}$$

where T_m denotes the mth decision tree, which has J leaf nodes (m = 1, 2, . . . , M). $I_j(X_k)$ indicates whether the feature X_k is used to split node j in T_m. τ_j^2 is the performance improvement as a result of using the variable X_k to branch at node j.

Moreover, RF provides a partial dependence plot, which can be used to investigate the effect of one or two features on the traffic accident severity. Suppose that X_S is a subset of the variable; the partial dependence on X_S can be defined as follows [25]:

$$F(X_S) = E_{X_C}[F(X_S)] = \frac{1}{n} \sum_{i=1}^{n} F(X_S, X_{iC}) \tag{3}$$

where X_C, X_S are complementary sets, and X_{iC} is the value of X_C for training ($i = 1, 2, \cdots, n$).

3.4. Performance Measure

In this study, predicting traffic accident severity is a multiclass classification problem. Since the traffic accident data are highly imbalanced, the commonly used accuracy-based evaluation index is not sufficient to determine the optimal classifier. Indices including precision, recall, and F1 score, defined in (4), (5), and (6), are used to evaluate the proposed multiclass classification model.

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 \cdot \frac{precison \cdot recall}{precison + recall} \tag{6}$$

where true positive (TP) denotes positive samples predicted by the model as positive classes, and false positive (FP) denotes negative samples predicted by the model as a positive class; false negative (FN) denotes positive samples predicted by the model as a negative class. In addition, the F1 score and the area under the receiver operating characteristic curve (AUC) are also used to evaluate the overall classification ability of the predictive model. A

larger precision, recall, F1 score, or AUC value corresponds to better prediction ability of the model.

3.5. The Proposed BO-RF Approach

In this study, we integrate RF and BO for the prediction of traffic accident severity. In the proposed approach, BO-RF, we use RF as a basic algorithm to fit the relationship between candidate influential variables and traffic accident severity. Further, the parameters of RF are optimized by BO. The flow chart of BO-RF is shown in Figure 1.

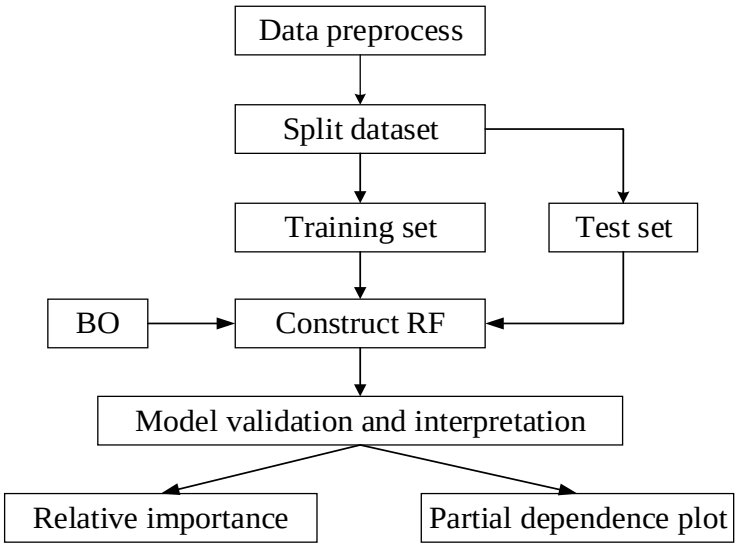

Figure 1. The flow chart of BO-RF.

Step 1. Preprocess data. The raw traffic accident data are processed before model training. Because some features' missing ratio exceeds 10%, we delete these features, including wind_chill (F), wind_speed (mph), and precipitation (in). Then, missing data are imputed with the mean or mode of the corresponding variables. Categorical variables are converted into dummy variables. Lastly, we obtain a dataset including 30,426 records.

Step 2. Split the dataset into a training set and test set according to ratio 80:20. Here, 80% of the records are used for model training and 20% is used for validation.

Step 3. Use RF as a basic forecasting model to predict traffic accident severity. The RF model is trained on the training set. The input is feature vectors (Start_lat, Start_lng, Distance (mi), Month, Day, Hour, Weekday, Pressure (in), Temperature (F), Humidity (%), Visibility (mi), Traffic_signal, Junction, Crossing, Stop). The output is traffic accident severity.

Step 4. Adopt BO to adjust the hyperparameters of RF. The hyperparameters include n_estimators, max_depth, and max_feature. The objective function of BO is the prediction result of *k*-fold cross-validation for RF. The input variables of the objective function are the hyperparameters of RF: n_estimators, max_depth, and max_feature. The output variable is the F1 score of *k*-fold cross-validation on the training set.

Step 5. Input the test data into the trained RF with the optimal hyperparameters obtained in Step 4. The performance of BO-RF is compared with the commonly used machine learning models, such as artificial neural network (ANN), k-nearest neighbor (KNN), and support vector machine (SVM).

Step 6. Model interpretation. To explore the significant factors of traffic accident severity, the relative importance of each feature is calculated by using the RF with the optimal parameters. A feature with higher relative importance implies that it has a stronger effect on traffic accident severity. Further, the relationship between input features and traffic accident severity is investigated by the partial dependence plot. These results provide insights for enhancing road traffic safety.

4. Results

The proposed model was completed in Python. For model training, we used the Scikit-learn library and Hyperopt package for model optimization.

4.1. Model Optimization

To obtain a good prediction result, it is essential to optimize hyperparameters of the RF model before prediction. In this study, we use BO to find the optimal combination of three parameters, namely the number of trees (n_estimators), the tree complexity (max_depth), and the maximum subfeatures (max_features). Moreover, 10-fold cross-validation is carried out on the training set to avoid overfitting. We randomly divide the dataset into 10 subsets of equal size, where 9 of the 10 subsets are used as training data. The remaining subset is regarded as validation data to measure the performance of the fitted model. This process is repeated 10 times, once for each subset as validation data. The final output is the average of the 10 test results.

We complete BO for RF using the Hyperopt package in Python. Hyperopt consists of three main components: the objective function, search space, and optimal algorithm. In this research, the objective function of BO is the mean F1 score of 10-fold cross-validation for RF; the tree parzen estimator is the algorithm. Regarding the search space, we set n_estimators$\in$(1,1000), max_depth$\in$(1,50), and max_features$\in$(1,15). The best prediction performance of RF is derived when n_estimators, max_depth, and max_features are set as 359, 42, and 12, respectively. The results are presented in Table 4.

Table 4. Hyperparameters optimization of RF.

Hyperparameters	Type	Search Space	Optimal Values
n_estimators	Discrete	(1,1000)	359
max_depth	Discrete	(1,50)	42
max_features	Discrete	(1,15)	12

4.2. Model Comparison

To examine the performance of the BO-RF model for traffic accident severity prediction, we compare it with several conventional algorithms, including ANN, KNN, and SVM, with a radial basis function kernel, and benchmark RF. The latter four models are all trained with the default values of the hyperparameters specified in the Scikit-learn package. All the models are implemented in Python.

Table 5 shows the comparison results measured by precision (macro P), recall (macro R), F1 score (macro F1), and AUC. It can be seen that BO-RF achieves the highest recall value of 0.53, F1 score of 0.57, AUC value of 0.9625, and the second-highest precision value of 0.66. RF obtains the highest precision value of 0.7 and the second-highest recall value of 0.47, F1 score of 0.54, and AUC value of 0.958.

Table 5. Comparison of the forecasting performance of different models.

	Precision	Recall	F1 Score	AUC
ANN	0.43	0.37	0.4	0.921
KNN	0.38	0.27	0.28	0.629
SVM	0.42	0.3	0.32	0.8488
RF	0.7	0.47	0.54	0.958
BO-RF	0.66	0.53	0.57	0.9625

According to the F1 score, the improvement of BO-RF over ANN, KNN, SVM, and RF is 42.5%, 103.57%, 78.13%, and 5.56%, respectively. Regarding AUC, the improvement of BO-RF over ANN, KNN, SVM, and RF is 4.51%, 53.02%, 13.4%, and 0.47%, respectively. In addition, KNN has the worst performance for accident severity prediction, with the lowest F1 score and AUC value. This indicates that BO-RF has substantial strengths over others.

Further, we discuss the effect of dataset segmentation. The forecasting results are presented in Tables 6 and 7. When 70% of the dataset is used as the training set, BO-RF achieves the highest recall value of 0.51, F1 score of 0.56, AUC value 0.956, and the second-highest precision value of 0.64. RF obtains the highest precision value of 0.67 and the second-highest recall value of 0.45, F1 score of 0.51, and AUC value 0.9481. When 60% of the dataset is used as the training set, BO-RF achieves the highest recall value of 0.5, F1 score of 0.54, and AUC value 0.9579, and the third-highest precision value of 0.63. RF obtains the highest precision value of 0.69 and the second-highest recall value of 0.44, F1 score of 0.49, and AUC value 0.9494. Hence, the performance of the proposed model is almost independent of the dataset partitioning. BO-RF always outperforms ANN, KNN, SVM, and RF in terms of precision, recall, F1 score, and AUC, except that its precision is slightly smaller than RF.

Table 6. Prediction results with 70% as the training set and 30% as the test set.

	Precision	Recall	F1 Score	AUC
ANN	0.47	0.39	0.41	0.9115
KNN	0.46	0.28	0.29	0.616
SVM	0.42	0.3	0.32	0.8938
RF	0.67	0.45	0.51	0.9481
BO-RF	0.64	0.51	0.56	0.956

Table 7. Prediction results with 60% as the training set and 40% as the test set.

	Precision	Recall	F1 Score	AUC
ANN	0.47	0.4	0.43	0.9052
KNN	0.47	0.27	0.28	0.613
SVM	0.66	0.3	0.32	0.8346
RF	0.69	0.44	0.49	0.9494
BO-RF	0.63	0.5	0.54	0.9579

4.3. Model Interpretation

The input features with their importance are presented in Table 8. We can see that each feature has a different impact on traffic accident severity. According to the categories of input features, the traffic, temporal, weather, and POI characteristics account for 68.89%, 13.78%, 13.69%, and 3.64% contributions to the accident severity prediction, respectively. Regarding traffic-related attributes, Start_lat and Start_lng are the most important factors of traffic accident severity, with the relative importance of 24.37% and 24.2%, respectively. Hence, accident locations are the critical factor related to traffic accident severity. With the relative importance of 20.32%, Distance (mi) is the third most influential factor. Hour is the fourth most important variable, with a relative importance of 4.65%. Next, Pressure (in), Temperature (F), and Humidity (%) are the following most influential variables, with the relative importance of 4.55%, 4.15%, and 3.68%, respectively. We can find that the characteristics of a traffic accident are the most contributing factor influencing accident severity, whereas POI is the least important influencing factor.

Further, the effect of the input features on traffic accident severity can be investigated by a partial independence plot in RF. As fatal accidents will bring serious consequences, we present the partial dependence plot of level 3. Figure 2 displays the relationship between traffic accident severity and potentially influential factors. It can be seen from the figure that most influential factors have a nonlinear effect on traffic accident severity. A detailed analysis is presented in Section 5.

Table 8. Relative importance of input features for traffic accident severity.

Category	Variable	Rank	Relative Importance (%)
Traffic	Start_lat	1	24.37
	Start_lng	2	24.2
	Distance (mi)	3	20.32
Temporal	Month	10	2.59
	Day	8	3.51
	Hour	4	4.65
	Weekday	9	3.03
Weather	Pressure (in)	5	4.55
	Temperature (F)	6	4.15
	Humidity (%)	7	3.68
	Visibility (mi)	12	1.31
POI	Traffic_signal	11	2.25
	Junction	13	0.86
	Crossing	14	0.41
	Stop	15	0.12

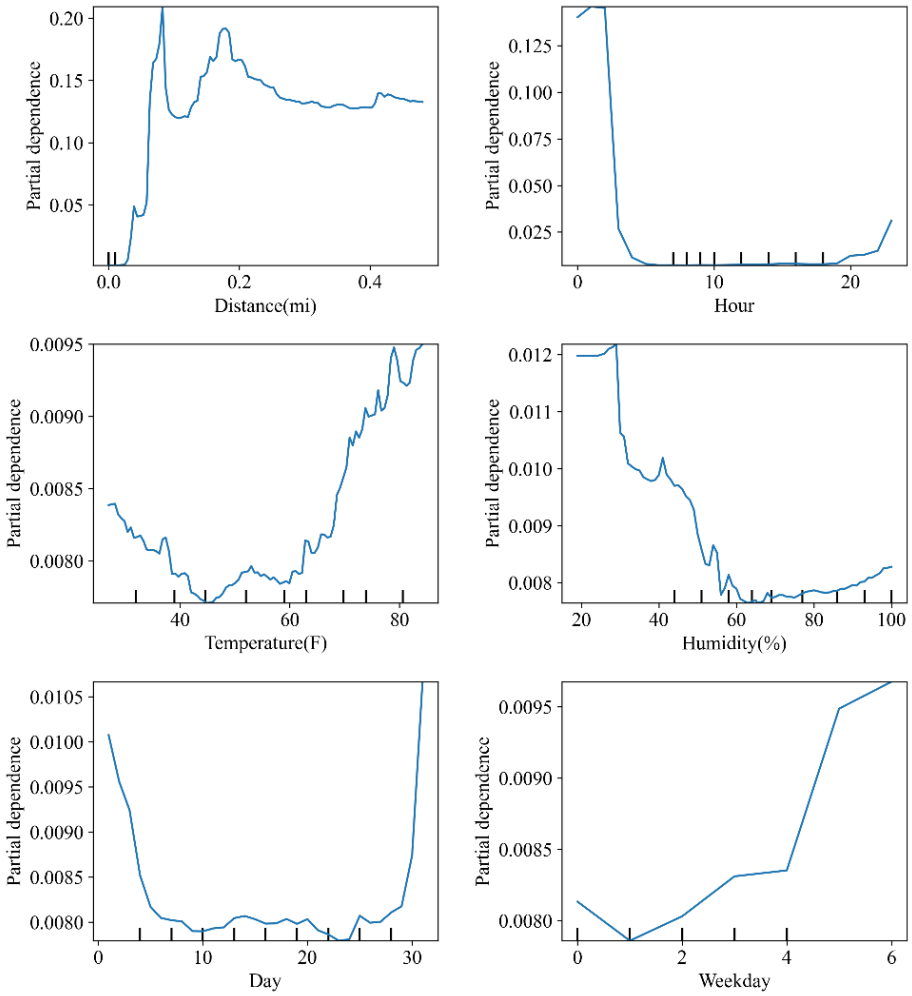

Figure 2. *Cont*.

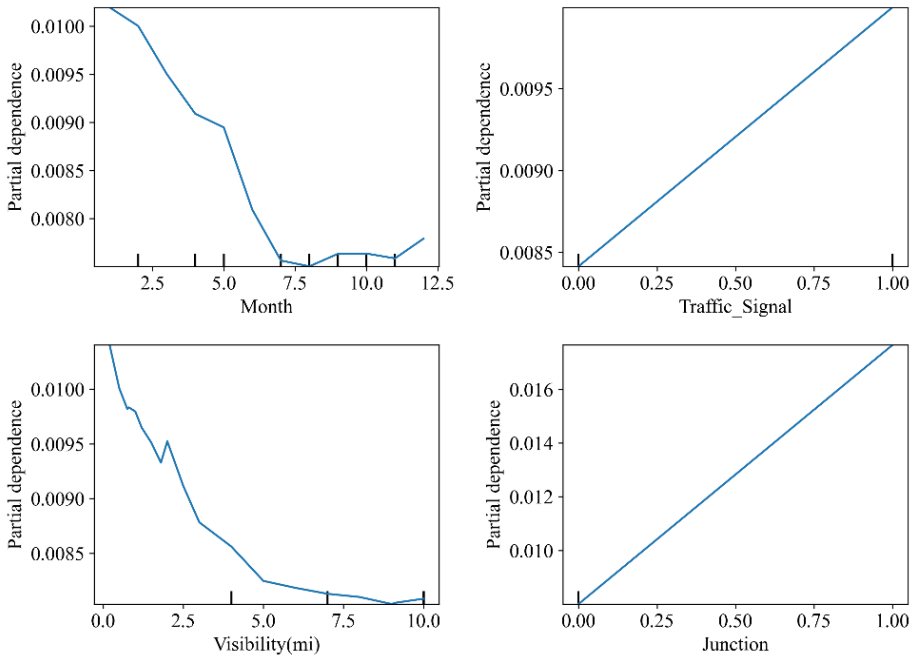

Figure 2. Partial dependences of key influential factors on fatal accident probability.

5. Discussions

This study predicts traffic accident severity by a hybrid model, BO-RF. We compare the model with four base learner models: ANN, KNN, SVM, and RF. The results in Table 5 show that RF performs better than ANN, KNN, and SVM in terms of precision, recall, F1 score, and AUC. This indicates the superiority of RF over ANN, KNN, and SVM in predicting traffic accident severity. Moreover, BO-RF performs better than the benchmark RF in terms of recall, F1 score, and AUC. This suggests that the BO can significantly improve the prediction performance of RF.

We also tested the performance of BO-RF by dividing the dataset into different proportions. As seen in Tables 6 and 7, BO-RF always outperforms ANN, KNN, SVM, and RF in terms of precision, recall, F1 score, and AUC, except that its precision is slightly smaller than that of RF. This means that the partitioning of the dataset hardly affects the performance of the model. BO-RF is robust.

The RF model not only achieves higher accuracy but also has good interpretability. Table 8 lists the relative importance of input features for traffic accident severity. We can find that Start_lat and Start_lng are the most important explanatory factors of traffic accident severity. Moreover, Distance (mi), Hour, Pressure (in), Temperature (F), Humidity (%), Day, Weekday, and Month are the next most significant factors affecting traffic accident severity. A similar finding is reported in a previous study, which indicates that the accident location, day of the week, and lighting conditions significantly influence accident severity [26].

The partial independence plot in Figure 2 further reveals the relationship between the input features and fatal traffic accidents. The probability of fatal accidents sharply rises from 0 to 0.2 with the Distance (mi) increasing from 0 to 0.05 miles (0 to 0.08 km); the probability of fatal accidents fluctuates within the range of 0.05–0.2 miles (0.08–0.32 km). When the Distance (mi) is larger than 0.2 miles (0.32 km), the probability of fatal accidents slowly reduces to 0.13. These results are in line with common sense: the more serious the accident, the larger the distance it affects.

The partial dependence value exhibits a downward trend when the Temperature (F) increases from 30 °F to 45 °F (−6.67 °C to 7.22 °C). When the Temperature (F) is larger than 45 °F (7.22 °C), the partial dependence value rises gradually from 0.007 to 0.0095 with the increase in Temperature (F). This suggests that 40 °F–60 °F (4.44 °C–15.56 °C) is comfortable for driving. The partial dependence of Humidity (%) is almost unchanged in the range of 20–30%. After this range, the probability of fatal accidents decreases sharply

from 0.012 to 0.007. When Humidity (%) is larger than 60%, the probability of fatal accidents increases slightly. A possible explanation is that the human body feels comfortable when the humidity is in the range of 50–80%. Outside of this range, the human body feels uncomfortable. Hence, the probability of fatal accidents increases. The partial dependence decreases in a nearly linear manner while Visibility (mi) rises from 0 to 10 miles (0 to 16.09 km). This is in line with common sense: the lower the visibility, the more likely a serious traffic accident will occur.

The partial dependence of Hour reaches the maximum value at approximately 2 o'clock. When the Hour increases from 2 o'clock to 5 o'clock, the partial dependence decreases from the maximum value to 0. Then, the partial dependence remains unchanged in the range of 5 o'clock–19 o'clock. After this range, the partial dependence increases slightly. This indicates that fatal traffic accidents are prone to occur in the small hours. A possible reason is that light is poor or drivers are more likely to lose concentration due to drowsiness [27,28]. When Day increases from 1 to 5, the partial dependence decreases from 0.01 to 0.008. The partial dependence increases from 0.008 to 0.0105 when Day increases from 28 to 31. This means that fatal accidents are more likely to occur at the end of the month or the beginning of the month. When Weekday increases from 1 (Tuesday) to 6 (Sunday), the partial dependence increases. In particular, when the Weekday increases from 4 (Friday) to 6 (Sunday), the partial dependence increases largely. This means that the probability of fatal accidents on weekends is much higher than that on weekdays. The partial dependence decreases from 0.01 to 0.0075 when Month increases from 1 to 7. However, within the range of 7–12, the effect of Month is trivial. This means that the effect of the Month occurs in the first half of the year, especially in winter. This is consistent with previous research that finds that the most severe accidents tend to occur in December, January, and February [29].

The partial dependence increases when Traffic_Signal or Junction increases from 0 to 1. This means that traffic accidents tend to happen near traffic signals and junctions [30]. A possible reason is that intersection traffic is characterized by a high mixture of vehicles, non-motorized vehicles, and pedestrians.

In summary, the hybrid BO-RF not only improves the prediction performance but also provides interpretable results. Based on the conclusions, some countermeasures can be developed, such as paying attention to traffic black spots, designing appropriate traffic facilities, and ensuring sufficient light.

6. Conclusions

Traffic accident severity prediction is important for accident management. In this study, we propose a hybrid model, BO-RF, for the prediction of traffic accident severity on urban roads. Experimental results show that BO-RF not only has higher prediction accuracy than traditional machine learning models but also provides interpretable results. With the precise prediction of traffic accident severity, traffic operators can deploy timely measures to reduce the side effects of the accident, such as providing timely medical assistance to the persons injured in the traffic accident, thus reducing casualties.

Moreover, the prominent influencing factors on traffic accident severity can be identified by the relative importance of the proposed model. The ways in which they influence traffic accident severity can be investigated by the partial dependence plot. The results provide reference suggestions for taking measures to mitigate the severity of accident consequences and improving traffic safety.

Author Contributions: Conceptualization, M.Y.; methodology, M.Y.; software, M.Y.; formal analysis, M.Y.; writing—original draft preparation, M.Y.; writing—review and editing, Y.S.; supervision, Y.S.; funding acquisition, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China under Grant No. 71571076 and by the Major Program of National Social Science Foundation of China (Grant No. 13&ZD175).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The travel time data used in this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Save LIVES-A road safety technical package. Geneva: World Health Organization; 2017. Available online: https://www.who.int/publications/i/item/save-lives-a-road-safety-technical-package. (accessed on 21 January 2022).
2. Gan, J.; Li, L.; Zhang, D.; Yi, Z.; Xiang, Q. An Alternative Method for Traffic Accident Severity Prediction: Using Deep Forests Algorithm. *J. Adv. Transp.* **2020**, *2020*, 1257627. [CrossRef]
3. Shiran, G.; Imaninasab, R.; Khayamim, R. Crash Severity Analysis of Highways Based on Multinomial Logistic Regression Model, Decision Tree Techniques, and Artificial Neural Network: A Modeling Comparison. *Sustainability* **2021**, *13*, 5670. [CrossRef]
4. Abdel-Aty, M. Analysis of driver injury severity levels at multiple locations using ordered probit models. *J. Saf. Res.* **2003**, *34*, 597–603. [CrossRef] [PubMed]
5. Sze, N.N.; Wong, S.C. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accid. Anal. Prev.* **2007**, *39*, 1267–1278. [CrossRef]
6. Savolainen, P.T.; Mannering, F.; Lord, D.; Quddus, M.A. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accid. Anal. Prev.* **2011**, *43*, 1666–1676. [CrossRef] [PubMed]
7. Moghaddam, F.R.; Afandizadeh, S.; Ziyadi, M. Prediction of accident severity using artificial neural networks. *Int. J. Civ. Eng.* **2011**, *9*, 41–48.
8. Taamneh, M.; Alkheder, S.; Taamneh, S. Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *J. Transp. Saf. Secur.* **2017**, *9*, 146–166. [CrossRef]
9. Zheng, M.; Li, T.; Zhu, R.; Chen, J.; Ma, Z.F.; Tang, M.J.; Cui, Z.Q.; Wang, Z. Traffic Accident's Severity Prediction: A Deep-Learning Approach-Based CNN Network. *IEEE Access* **2019**, *7*, 39897–39910. [CrossRef]
10. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
11. Lu, Z.; Long, Z.; Xia, J.; An, C. A Random Forest Model for Travel Mode Identification Based on Mobile Phone Signaling Data. *Sustainability* **2019**, *11*, 5950. [CrossRef]
12. Evans, J.; Waterson, B.; Hamilton, A. Forecasting road traffic conditions using a context-based random forest algorithm. *Transp. Plan. Technol.* **2019**, *42*, 554–572. [CrossRef]
13. Hamad, K.; Al-Ruzouq, R.; Zeiada, W.; Abu Dabous, S.; Khalil, M.A. Predicting incident duration using random forests. *Transp. A-Transp. Sci.* **2020**, *16*, 1269–1293. [CrossRef]
14. Macioszek, E. Roundabout Entry Capacity Calculation—A Case Study Based on Roundabouts in Tokyo, Japan, and Tokyo Surroundings. *Sustainability* **2020**, *12*, 1533. [CrossRef]
15. Severino, A.; Pappalardo, G.; Curto, S.; Trubia, S.; Olayode, I.O. Safety Evaluation of Flower Roundabout Considering Autonomous Vehicles Operation. *Sustainability* **2021**, *13*, 10120. [CrossRef]
16. Macioszek, E. The Comparison of Models for Critical Headways Estimation at Roundabouts. In Proceedings of the 13th Scientific and Technical Conference on Transport Systems. Theory and Practice (TSTP), Katowice, Poland, 19–21 September 2016.
17. US-Accidents. Available online: https://smoosavi.org/datasets/us_accidents (accessed on 13 November 2021).
18. Moosavi, S.; Samavatian, M.H.; Parthasarathy, S.; Ramnath, R. A countrywide traffic accident dataset. *arXiv Preprint* **2019**, arXiv:1906.05409. Available online: https://arxiv.org/abs/1906.05409 (accessed on 20 December 2021).
19. Moosavi, S.; Samavatian, M.H.; Parthasarathy, S.; Teodorescu, R.; Ramnath, R. Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights. In Proceedings of the 27th Acm Sigspatial International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 5–8 November 2019; pp. 33–42.
20. Cheng, L.; De Vos, J.; Zhao, P.; Yang, M.; Witlox, F. Examining non-linear built environment effects on elderly's walking: A random forest approach. *Transp. Res. Part D-Transp. Environ.* **2020**, *88*, 102552. [CrossRef]
21. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]
22. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; de Freitas, N. Taking the Human out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148–175. [CrossRef]
23. Bergstra, J.S.; Bardenet, R.; Bengio, Y.; Kegl, B. *Algorithms for Hyper-Parameter Optimization*; Curran Associates Inc.: New York, NY, USA, 2011.
24. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
25. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009.

26. Zhu, L.; Lu, L.; Zhang, W.; Zhao, Y.; Song, M. Analysis of Accident Severity for Curved Roadways Based on Bayesian Networks. *Sustainability* **2019**, *11*, 2223. [CrossRef]
27. Pillajo-Quijia, G.; Arenas-Ramírez, B.; González-Fernández, C.; Aparicio-Izquierdo, F. Influential Factors on Injury Severity for Drivers of Light Trucks and Vans with Machine Learning Methods. *Sustainability* **2020**, *12*, 1324. [CrossRef]
28. Miqdady, T.; de Oña, J. Identifying the Factors That Increase the Probability of an Injury or Fatal Traffic Crash in an Urban Context in Jordan. *Sustainability* **2020**, *12*, 7464. [CrossRef]
29. Zhang, C.; He, J.; Wang, Y.; Yan, X.; Zhang, C.; Chen, Y.; Liu, Z.; Zhou, B. A Crash Severity Prediction Method Based on Improved Neural Network and Factor Analysis. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 1–13. [CrossRef]
30. Jaber, A.; Juhász, J.; Csonka, B. An Analysis of Factors Affecting the Severity of Cycling Crashes Using Binary Regression Model. *Sustainability* **2021**, *13*, 6945. [CrossRef]