Research Article

Speeding Violation Type Prediction Based on Decision Tree Method: A Case Study in Wujiang, China

Zeyang Cheng ^{1,2,3} Jian Lu ^{1,2,3} Zhenshan Zu,⁴ and Yunxuan Li^{1,2,3}

¹Jiangsu Key Laboratory of Urban ITS, Southeast University, Nanjing 211189, China

²Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Nanjing 211189, China

³School of Transportation, Southeast University, Nanjing 211189, China

⁴Traffic Police Brigade of Wujiang Public Security Bureau, Suzhou, China

Correspondence should be addressed to Jian Lu; lujian_1972@seu.edu.cn

Received 8 April 2019; Accepted 11 June 2019; Published 27 June 2019

Academic Editor: Alain Lambert

Copyright © 2019 Zeyang Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The speeding violation has become a key concern in the traffic safety management, as it increases the risk of traffic crashes, as well as the severity of these crashes. This uncivilized phenomenon is prominent and presents an increasing trend in Wujiang in recent years, which severely endangers the road traffic safety. This study is approved by the Traffic Police Brigade of Wujiang Public Security Bureau and aims to explore the characteristic of the speeding violation behaviour and attempt to make an effective prediction about it. This study proposes a speeding violation type (including type 1 and type 2) prediction method using electronic law enforcement data obtained from the public security administration of Wujiang. Before the prediction, a speeding violation influence factor analysis based on the binary logical regression model is proposed. The binary logical regression analysis identifies that the license plate, season, speeding area, position, and rainfall are the influence factors of Wujiang's speeding violation. Then a decision tree method is used to predict the speeding violation type according to the influence factors, and from which the speeding violation situations can be determined. The prediction results demonstrate that under the hypothetical conditions, the high speeding violation level (i.e., type 2) tends to occur under high rainfall environment, and the foreign license plate and autumn present a larger probability of high speeding violation level than the local license plate and other seasons (i.e., spring, summer, and winter), respectively. Finally, a model comparison between the proposed method and other tree-based approaches is conducted. The comparison results show that the decision tree method outperforms other methods in prediction performance (including accuracy, precision, recall, and classification error), runtime, and ROC curve, which indicates that the decision tree method is feasible in predicting the speeding violation type of Wujiang. Based on the findings, the traffic managers can macroscopically grasp the speeding violation situation of the whole road networks, which can be referred for making the related polices and taking intervention measures.

1. Introduction

Aggressive driving behaviours, driving offenses, and their relations to traffic crashes have been a leading public health problem worldwide. According to the report by Transportation Administration of the Ministry of Public Security in China, a total of 50400 road truck-related crashes occurred in 2016, resulting in 25000 deaths and 46800 injuries. These serious casualties were primarily caused by traffic violations, among which the speeding violation accounts for a big proportion. Generally, the speeding violation is

considered to be a risky traffic behaviour in Wujiang [1], and the consequences of speeding violation are considerable. Relevant statistic illustrated that 31% of road crashes were caused by speeding violation between 2003 and 2007 in the United States [2]. Speeding violation has become a universal phenomenon in the daily life. The survey conducted by National Highway Traffic Safety Administration (NHTSA) of the United States in 2002 reported that 80% of all drivers exceeded the posted speed limit during their driving trips [3]. In addition, large numbers of studies have highlighted that

the speeding violation not only increases the risk of traffic crashes, but also aggravates the severity of them [4–9].

The current study aims to propose an efficient speeding prediction method of road network. Remarkably, this prediction is primarily targeted to the speeding violation type instead of the regular speeding intentions. Based on the influence factors of speeding violation, the prediction can be implemented by the related method. As a critical part of traffic safety management, speeding violation type prediction can macroscopically evaluate the speeding level and speeding distribution under different conditions. Then the policy-makers can take some targeted measures to prevent the speeding violation behaviour. Therefore, the speeding violation type prediction is important to reduce the speeding-related crashes and improve the road traffic safety. Speeding violation influence factor analysis and the prediction methods are two important parts of this study; thereby, the literature review is introduced based on the two regards.

2. Literature Review

2.1. Speeding Violation Influence Factor Analysis. The speeding behaviours are influenced by some biological and social-economic factors (i.e., gender, age, education background, and occupation), and many works have proved it. Tseng [10] analyzed the influence factors contributed to male speeding using a logistic regression model. The results revealed that the male speeding behaviours were significantly associated with age, education, income, yearly driving distance, and driving purposes. Some studies have investigated the factors leading to speeding offenses for truck [11] and taxi drivers [12] in Taiwan. The results showed age, education, mental condition, and driving status were significantly linked to truck speeding offenses. Meanwhile the age, job experience, operating styles, and kilometers driven daily were associated with a taxi driver's speeding. Forward [13] extended the theory of planned behaviour (TPB) and used it to evaluate the intention about the speeding violation behaviour. Finally, the results showed driving experience and age were important factors associated with women's lower speeding intention. Vardaki and Yannis [14] studied the drivers' self-reported behaviours and their attitudes to speeding, drunk driving, and cell phone using by a clustering method. Finally, driver's age, gender, and area of residence were identified as the influence factors of driver's speeding. Roidl et al. [15] introduced a multivariate model for evaluating the speeding driving performance. Then, the gender, age, and anger mood were considered to be the associated factors of speeding violation. The results indicated the drivers who experienced high levels of anger would drive faster and exhibit greater longitudinal and lateral acceleration. Several other works have explored the influence of anger on increased speeding violation behaviour [16–18].

In other aspects, Zhang et al. [19] analyzed 11055 speeding data for the period 2006–2010 in Guangdong Province, China. Finally, they found the private vehicles, the lack of street lighting at night, and poor visibility were the significant

factors related to the high speeding violation. Huang et al. [20] proposed a driver-road-environment identification method for investigating the determinants of taxi speeding violations of Shanghai and New York based on GPS data. The results showed length of segments and good traffic condition were associated with high speeding rates. Moreover, the most frequent speeding behaviour occurring at 5:00 a.m. and 6:00 a.m. in both Shanghai and New York. Fu et al. [21] analyzed the effect of the related factors on speeding behaviour of road intersection. Then the likelihood of speeding behaviour was predicted using a multinomial logistic model. The findings indicated that ages, traffic sign, and driver distraction were significantly associated with the likelihood of speeding behaviour. Several literatures also explored the driving speed choice, as well as the speed limit setting involving the speeding behaviours [22–24].

2.2. Speeding Violation Prediction Method. In the past, TPB model has been widely used in predicting the speeding behaviours. A number of studies conducted in different contexts with different methodologies have proved that TPB model provides a useful frame work towards the speeding violation research [25]. TPB model explains that the behaviour was primarily determined by behavioural intention (i.e., planning to produce or not to produce behaviour) and perceived behavioural control (i.e., the subjective degree of control over performance of the behaviour). A rational behaviour theory and TPB model were proposed by Letirand et al. [26] to predict the drivers' speeding violation behaviour. The study examined the factors (i.e., the amount of explained variance of self-reported speeding behaviour, and the intentions to exceed the speed limit by at least 20 km/h) contributing to improving predictions of speeding behaviour intention. Eyssartier et al. [27] studied the speeding behaviour of motorcyclists based on an TPB model. The results showed that the predictors exceeding the speed limit were different in different types of motorcycles. Elliott [28] identified cognitive predictors of motorcyclists' speeding intentions using a model that comprised selected constructs from TPB theory, identity theory, and social identity theory. The results demonstrated that the significant predictors of different speeding intentions were various, and the proposed model possessed good predictive validity in relation to motorcyclists' speeding intentions. Cestac et al. [29] predicted the influence factors of drivers' speeding intention based on an extended TPB model. A questionnaire contained 3002 drivers aged from 18 to 25 was conducted, and the results showed men had a slightly higher speeding intention than women, and the speeding intention increased with the driving experience and sensation seeking. Atombo et al. [30] predicted drivers' intentions towards speeding using the belief measure of TPB and driver behaviour questionnaire (DBQ). The results indicated that the components of TPB made larger contributions to the prediction of driver' intentions to speeding and overtaking than DBQ. Jovanović et al. [31] examined the determinants of speeding violation and tested the predictive validity of a modified TPB model in relation to speeding violation behaviour.

Other than TPB and its extended methods, some other projections in regard to speeding violation behaviour have also been proposed. Zhao et al. [32] extended a previous mathematical model to quantitatively predict the intentional and unintentional speeding, in which the speeding time and the magnitude of speeding were included. Then the experimental study from a driving simulator was conducted to evaluate the proposed method, and the results showed no significant difference between the modeled predictions and experimental results. Warner and Aberg [33] studied drivers' view on speeding violation and predicted the speeding intention. A total of 162 car owners were selected for the investigation and analysis; then, the results showed that the indicators such as attitude, subjective norms, and perceived behaviours were more accurate in predicting drivers' speeding intentions. Liu et al. [34] predicted the speeding violation behaviour based on GPS data. A speeding prediction algorithm combined the geographic coordinates of the scheduled route and speed limit was proposed, and whether the speeding violation has occurred was determined by comparing the real speed, the current speed limit, and the speeding threshold.

2.3. Objective of the Study. In summary, prior studies have conducted a detailed introduction about the speeding violation influence factor analysis and speeding violation prediction method, respectively. These studies have set a good research basis for the future studies. However, several aspects are still limited. First, most of the predictions were mainly focused on speeding behaviour changes and intentions based on the TPB model and its extended versions. The concentration on speeding violation type prediction is scarce. Moreover, the above literatures were almost the separate study about speeding violation influence factor analysis and speeding prediction, and few works have integrated the two parts as the whole research content. Accordingly, the study conducted in this paper fills these knowledge gaps, and it aims to propose a comprehensive research integrating the speeding violation influence factor analysis and speeding violation type prediction. In this study, the speeding violation influence factor analysis is the foundation, and the speeding violation type prediction is the key part; the speeding violation type prediction can be smoothly implemented based on the result of speeding violation influence factor analysis.

3. Data Source

The raw data used in this study are the electronic enforcement data of Wujiang (i.e., a small-sized city of China, and it owns a population of approximately 830,000, and the motor vehicle number was more than 360,000 in 2017). This data contains speeding collected by the electronic law enforcement equipment (i.e., speeding capture equipment) of Wujiang, which is only officially available. The data were taken from the database of Wujiang Public Security Bureau, and they were collected in 2017. The basic attributes contain vehicle type, license plate, standard speed limits, speeding time (the format is year-month-day-hour-minute), speeding area, position, rainfall (unit: mm /lh), and speeding type. Speeding type

comprised two groups: speeding level between 10% and 20% (i.e., type 1) and speeding level between 20% and 50% (i.e., type 2). The speeding type 1 and type 2 are the existing labels in the raw data of Wujiang, and they are also the two common speeding violation types in China.

The raw data consists of 53283 samples. For the convenience of analysis, some attributes are converted and classified. For example, the vehicle type is classified as car and truck, the license plate is classified as local license plate and foreign license plate (i.e., the license plate is issued from the public security bureau of other cities of China). Likewise, the standard speed limits contain six groups: 40km/h, 60km/h, 70km/h, 80km/h, 90km/h, and 100km/h. The speeding violation time is converted into season (e.g., spring, summer, autumn, and winter), week (i.e., weekday, weekend), day of time (e.g., 0:00-6:59, 7:00-8:59, 9:00-11:59, 12:00-16:59, 17:00-19:59, and 20:00-23:59), which aligns with [19]. The speeding area includes four administrative sections: Songling, Fenhu, Lili, and Taoyuan. Then the violation position is classified as country road, urban road, intersections, and point of interests (POI). Rainfall is classified into four groups according to the division of rainfall by the Chinese Meteorological Department: no rain (rainfall is 0 mm/h), light rain (rainfall is between 0 mm/h-2.5mm/h), moderate rain (rainfall is between 2.6 mm/h-8mm/h), and heavy rain (rainfall is between 8.1 mm/h-15mm/h). Then the descriptive statistic of raw data is shown in Table 1.

4. Method

4.1. Influence Factors Analysis of Speeding Violation. Identifying factors affecting speeding violation is as important as the identification of speeding to cause crashes [35]. Thus, the speeding violation influence factor analysis is not only helpful to the prediction about speeding violation type but also useful for understanding the causes of speeding-related crashes. The related factors in the raw data include vehicle type, license plate, standard speed limits, season, week, time of day, speeding area, violation position, and rainfall. Among these factors, we need to determine which one has an impact on the speeding violation. Thence, the hypothesis testing by the statistical approach is conducted to identify the associated factors of the speeding violation. Notably, the speeding behaviour may be actually influenced by numerous factors, but, in this study, only the factors involved in the raw data of Wujiang is considered. In other words, other external conditions are hypothesized to be fixed. Only in that case, the influence factor analysis about the speeding violation can be meaningful. The effective contributor analysis about speeding violation provides a basis for the speeding violation type prediction.

The predicted target (i.e., the speeding violation type) is a binary variable (i.e., the speeding level between 10% and 20% is type 1, and the speeding level between 20% and 50% is type 2), and it is set as follows: for the presence of speeding violation type resulting from the influence factors, 0="type 1" and 1="type 2". As the outcome measure is either 0 or 1, a binary logistic regression analysis (5%

TABLE 1: Descriptive statistics of the speeding violation samples (N=53283).

Factor	N (%)	Factor	N (%)
<i>Vehicle type</i>		7:00-8:59 (1)	1156 (2.18%)
Truck (0)	25(0.05%)	9:00-11:59 (2)	43310 (81.29%)
Car (1)	52783 (99.95%)	12:00-16:59(3)	8683 (16.30%)
<i>License plate</i>		17:00-19:59(4)	8 (0.00%)
Local license plate (0)	44446 (83.40%)	20:00-23:59(5)	103 (0.19%)
Foreign license plate (1)	8837 (16.60%)	<i>Speeding area</i>	
<i>Standard speed limits (km/h)</i>		Fenhu (0)	373 (0.70%)
40 (0)	2(0.00%)	Songling (1)	49780 (93.43%)
60 (1)	2 (0.00%)	Lili (2)	3106 (5.83 %)
70 (2)	24042(45.12%)	Taoyuan (3)	24 (0.05%)
80 (3)	26652(50.02%)	<i>Position</i>	
90 (4)	1451(2.72%)	Country road (0)	5484 (10.29%)
100 (5)	1134 (2.14%)	Urban road (1)	37209 (69.83%)
<i>Season</i>		Intersection (2)	3658 (6.87%)
Spring (0)	24677 (46.32%)	POI (3)	6932 (13.01%)
Summer (1)	10755 (20.18%)	<i>Rainfall</i>	
Autumn (2)	10274 (19.28%)	No rain (0)	48519 (91.06%)
Winter (3)	7577 (14.22%)	Light rain (1)	4461 (8.38%)
<i>Week</i>		Moderate rain (2)	269(0.50%)
Weekday (0)	39477 (73.70%)	Heavy rain (3)	34 (0.06%)
Weekend (1)	13806 (26.30%)	<i>Speeding type</i>	
<i>Time of day</i>		Type 1 (0)	46757(87.75%)
0:00-6:59 (0)	23 (0.04%)	Type 2 (1)	6526 (12.25%)

TABLE 2: Binary logistic regression analysis results.

Factor	df	P value	OR	95.0% CI for OR	
				Lower	Upper
<i>License plate 1 (base: license plate (0))</i>	1	0.000**	1.238	1.156	1.326
<i>Season (base: season (0))</i>	3	0.000**			
season (1)	1	0.002**	0.876	0.806	0.952
season (2)	1	0.000**	2.309	2.159	2.470
season (3)	1	0.020*	0.896	0.816	0.983
<i>Speeding area (base: speeding area (0))</i>	3	0.000**			
speeding area (1)	1	0.000**	0.516	0.394	0.676
speeding area (2)	1	0.001**	1.643	1.221	2.210
speeding area (3)	1	0.053	2.828	0.986	8.112
<i>Position (base: position (0))</i>	3	0.000**			
position (1)	1	0.000**	0.759	0.691	0.833
position (2)	1	0.000**	1.517	1.312	1.754
position (3)	1	0.267	0.925	0.806	1.062
<i>Rainfall (base: rain (0))</i>	3	0.000**			
rainfall (1)	1	0.000**	1.286	1.174	1.409
rainfall (2)	1	0.137	1.278	0.925	1.767
rainfall (3)	1	0.004**	2.810	1.380	5.725

*p < 0.05.

**p < 0.01.

confidence level) is used to explore the association between the related factors and speeding violation. The binary logistic regression analysis results obtained by SPSS are shown in Table 2.

Several factors, such as the license plate, season, speeding area, position, and rainfall, are identified as the influencing factors of speeding violation type (see Table 2). Then the vehicle type, standard speed limits, and time of day show

TABLE 3: Descriptive statistics of the final speeding violation samples (N=46072).

Variable	Description	N (proportion)	Variable type
Speeding type	Type 1	39752 (86.30%)	Response
	Type 2	6320 (13.70%)	
License plate	Local license plate	39207 (85.10%)	Predictor
	Foreign license plate	6865 (14.90%)	
Season	Spring	225508 (48.90%)	Predictor
	Summer	9611 (20.90%)	
	Autumn	8836 (19.20%)	
	Winter	5117 (11.10%)	
	Fenhu	41 (0.10%)	
Speeding area	Songling	42961(93.20%)	Predictor
	Lili	3070 (6.70%)	
	Country road	36941 (80.20%)	
Position	Urban road	5473 (11.90%)	Predictor
	Intersection	3658 (7.90%)	
	No rain	42219 (91.60%)	
Rainfall (mm/h)	Light rain	3823 (8.30%)	Predictor
	Heavy rain	30 (0.10%)	

no obvious association with speeding violation type. More specifically, among the speeding area groups, speeding area (0) (i.e., Fenhu), speeding area (1) (i.e., Songling), and speeding area (2) (i.e., Lili) are the main influence factors, and speeding area (3) (i.e., Taoyuan) is relatively unrelated to speeding violation type (p value=0.053). Likewise, among the position and rainfall groups, position (0) (i.e., country road), position (1) (i.e., urban road), position (2) (i.e., intersection), and rainfall (0) (i.e., no rain), rainfall (1) (i.e., light rain), rainfall (3) (i.e., heavy rain) are the influence factors, respectively. Position (3) (i.e., the POI, p value=0.267) and rainfall (2) (i.e., the moderate rain, p value=0.137) are relatively unrelated to the speeding violation type. Thereby the final influence labels of the speeding violation type of Wujiang are license plate (include local license plate and foreign license plate), season (include spring, summer, autumn, and winter), speeding area (include Fenhu, Songling, and Lili), position (include country road, urban road, and intersection), and rainfall (include no rain, light rain, and heavy rain).

The following section will predict the speeding violation type based on these influence factors; then, the corresponding policies and measures can be made according to different speeding violation types and their influence factors.

4.2. Speeding Type Prediction

4.2.1. Prediction Labels. The objective of this study is to predict the speeding violation type according to the influenced factors. Thence, the results should include two classifications: type 1 and type 2. The whole labels consist of “type 1”, “type 2”, “local license plate”, “foreign license plate”, “spring”, “summer”, “autumn”, “winter”, “fenhu”, “songling”, “lili”, “country road”, “urban road”, “intersection”, “no rain”, “light rain”, and “heavy rain”. Among these labels, “type 1” and “type 2” are the response labels (i.e., the target prediction labels),

and others are the predictors (see Table 3). Remarkably, the raw data consists of 53283 samples. By the influence factor analysis about the speeding violation type, some unrelated factors such as vehicle type, speeding limits, week, and time of day are removed. As a result, the final data sample is 46072. Table 3 shows the descriptive statistics of the final speeding violation samples. Among those samples, “type 1” (86.30%), “local license plate” (85.10%), “Songling” (93.20%), “country road” (80.20%), and “no rain” (91.630%) occupy absolutely proportionality advantage in their respective classification groups. With regard to seasonal distribution, spring accounts for nearly half of the total samples, while the samples of summer, autumn, and winter are relatively balanced.

4.2.2. Decision Tree Modeling. The decision tree method can predict the binary classification variable with relatively few labels. Our purpose aims to use the related labels to predict the speeding violation type, the target labels in this study is a binary variable (include type 1 and 2), and the predictors contain five kinds of labels (i.e., license plate, season, speeding area, position, and rainfall). This situation fits well with the applicable condition of the decision tree method. Based on this, the decision tree model is selected as the prediction method. Notably, the rainfall (include no rain, light rain, and heavy rain) is restored to the numerical values (i.e., the rainfall values from 0 mm/h to 15 mm/h) to maintain its continuous nature when performing the prediction.

The core idea of decision tree method is to use the information gain to measure the choice of attributes and select the attribute that has the greatest information gain after splitting to divide. This method uses the greedy search from the top down to traverse the possible decision space. Then the gain ratio is selected as the criterion, and the corresponding parameters are set as follows: maximal depth is 20, confidence is 0.25, minimal gain is 0.1, minimal leaf size is 2, minimal

FIGURE 1: Prediction results of the decision tree method.

size for spilt is 4, and number of prepruning alternatives is 3. Additionally, the validation method is a cross-validation way with 5 folds, which protects against overfitting by partitioning the data set into folds and estimating accuracy on each fold. Finally, the prediction results are shown as the dendrogram of Figure 1.

4.2.3. Model Test and Comparison Analysis. In this section, the accuracy (A), class precision (P), and class recall (R) are proposed to test the performance of the decision tree method in predicting the speeding violation type of Wujiang, and these indicators are expressed as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$A = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (3)$$

where TP denotes the true positive sample of the prediction result, FP represents the false positive sample of the prediction result, TN is the true negative sample of the prediction result, and FN denotes the false negative sample of the prediction result. The confusion matrix obtained by the decision tree method is shown in Table 4. From Table 4, A , P , R can be calculated (i.e., $A=92.18\%$, $P=71.93\%$,

TABLE 4: Confusion matrix of the samples.

Classification	True. Type 2	True. Type 1
Predicted. Type 2	4458 (TP)	1740 (FP)
Predicted. Type 1	1862 (FN)	38012 (TN)

$R=70.54\%$). The good performance of the three indicators has preliminarily shown that the decision tree method is a feasible speeding violation type prediction approach.

In order to validate the superiority of the decision tree method in predicting the speeding violation type of Wujiang, two other approaches (i.e., random forest method and gradient boosted trees method) are introduced to be compared. Under the same experimental conditions, the speeding violation types are predicted by the random forest method and gradient boosted trees method. Then the performance and runtime comparison of different methods are shown in Figure 2. By contrast, the accuracy and prediction of the decision tree method are larger than that of the random forest method and the gradient boosted trees method. The recall of the decision tree method is slightly lower than the random forest method but higher than the gradient boosted trees method. Then both the classification error and runtime of decision tree method are the smallest among the three approaches. In general, the decision tree method outperforms other tree-based methods in predictive performance and

FIGURE 2: Prediction comparison of different methods.

FIGURE 3: ROC comparison of different methods.

runtime conducted under the same hardware configuration condition.

In addition, ROC curve is commonly used to evaluate the prediction results of the machine learning methods, especially in the application of the binary classifier. Considering that the speeding violation type prediction of Wujiang is exactly the binary classification problem (i.e., the prediction result is either type 1 or type 2), the ROC comparison between different approaches is performed. This comparison provides a more detailed assessment about the prediction results in addition to the above assessments about the performance and runtime. Figure 3 shows the ROC curve of the three

approaches. As shown in Figure 3, the horizontal and vertical coordinates represent the false positive rate and the true positive rate, respectively. AUC represents the area under ROC. Generally, the closer the ROC curve tilts to vertical axis, the better the prediction result. Similarly, the closer the AUC value is to 1, the better the prediction result. The ROC curve of the decision tree method is above the ROC curve of the random forest method and gradient boosted trees method, and the AUC of decision tree method is also larger than other methods, all of these prove that that the decision tree method is superior to the other two methods in predicting the speeding violation type of Wujiang. These analyses further verified the superiority of the proposed method in this study.

5. Result and Discussion

The current study investigates the influence factor of the speeding violation of Wujiang and then predicts the speeding violation type of the road networks. The license plate, season, speeding area, position, and rainfall are identified as the influence factors of the speeding violation in Wujiang. Then the decision tree method is used to perform the prediction. Finally, a well-layered tree structure is generated to show the prediction results (see Figure 1), and the corresponding discussions are proposed as follows.

(1) The speeding violation level occurring at the country road of Wujiang is more serious than that of the urban road under the same conditions (i.e., the other external conditions are determined). That is to say, the speeding violation type of country road is mainly type 2, whereas it is primarily type 1 on urban road. This result is easier to understand, because some transport infrastructures such as signal light, road separator, and crosswalks at urban road will limit the motor vehicle speed. Then the urban roads are often congested due to the high traffic flow, which reduces the occurrence of speeding violation in some degree. In addition, the large numbers of electronic law enforcement equipment set at different locations of Wujiang can deter the speeding violation. On the contrary, the traffic flow in country road is relatively low, and the transport facilities and electronic law enforcement

equipment on the country road are also scarce. As a result, the speeding violation level of the country road is high.

(2) The results also suggest that the larger the rainfall, the higher the speeding violation level (i.e., type 2 is more than type 1) under the same conditions. For example, when the rainfall is above 5.95mm/h, the speeding violation type 2 (i.e., the high speeding level) accounts for the dominant proportion in the prediction results. While when the rainfall is below 5.95mm/h, most of the prediction results are speeding type 1 (i.e., the low speeding level). In the meantime, when the rainfall is below 0.15 mm/h, the speeding violation is type 1, whereas it contains both type 1 and 2 when the rainfall is above 0.15mm/h. The above analyses present an irregular phenomenon, which means that the drivers do not slow down under heavy rainfall condition, but continue to cause a high speeding violation. This finding may provide an important clue for the speeding violation characteristic of Wujiang, by which the traffic police can pay more attention on the key areas under heavy rainfall conditions.

(3) In addition, the predicted speeding violation types are various according to different labels (i.e., the season, speeding area) under the same rainfall condition. Overall, the speeding violation levels occurring at Lili (it is mainly type 2) are larger than that of Songling (it is type 1); this is because Songling is the core urban area of Wujiang, and the traffic flow, transport infrastructures, and electronic law enforcement equipment of Songling are more perfect than Lili, which reduces the speeding violation level. Then in spring and summer, the speeding violation type is primarily type 1. In winter, the speeding violation types are various according to different rainfalls. For instance, the speeding level is type 2 in winter under the condition of "rainfall > 5.95mm/h", while it is type 1 in winter under the condition of "rainfall ≤ 4.1mm/h". The speeding violation situations are complicated in autumn, as it is influenced by multilabels such as rainfall, speeding area, and license plate. Finally, from the overall predicted results (see Figure 1), the foreign license plate is inclined to cause a high speeding violation level (i.e., type 2), but the local license plate shows a relatively low speeding level (i.e., type 1). This can be ascribed to the fact that the local drivers are more familiar with the road condition and they also know more about the locations of the speeding capture equipment installment, so they rarely cause a high level of speeding in case of a traffic punishment (include a fine or license plate point deduction). On the contrary, the nonlocal drivers cannot necessarily grasp Wujiang's road condition and the setting locations of the speed capture equipment, so the high level of speeding violation will occur among these drivers.

In conclusion, among the speeding violations of Wujiang, country roads are more prone to cause a speeding than urban road under the hypothetical conditions, and the speeding level of the country road is also more serious than that of the urban road. Lili (the remote districts of Wujiang) presents a heavy speeding violation level than Songling (i.e., core urban area of Wujiang). The speeding violation level of Wujiang is increasing with the increasing of rainfall. The license plate and season show a complex relationship with the speeding violation type, and it is not a simple correspondence. However, the overall prediction results indicate the foreign

license plate and autumn presents a high level of speeding violation (it is primarily type 2 in this study) than the local license plate and other seasons (i.e., spring, summer and winter), respectively.

6. Conclusion

The present study is the first attempt to explore the speeding violation influence factor analysis and speeding violation type prediction based on the electronic enforcement data. The results may be targeted towards making some policies to prevent the speeding violations of Wujiang. Before the prediction, a binary logic regression analysis is proposed to identify the determinants of Wujiang's speeding violation. Several factors (i.e., license plate, season, speeding area, position, and rainfall) are significantly associated with speeding violation. Then a machine learning method (i.e., the decision tree method) is conducted for predicting the speeding violation type. Based on the prediction results, the speeding violation level belongs to which type (type 1 or type 2) under different conditions can be identified, and these results are therefore taken into account to formulate appropriate policies for tackling speeding violation issues of Wujiang. Finally, the model test indicates a good performance of the decision tree method (i.e., the accuracy is 92.18%, the precision is 71.93%, and the recall is 70.54%) in predicting the speeding violation. Furthermore, the random forest method and gradient boosted trees method are proposed to compare with the decision tree method in predicting the speeding type of Wujiang. Model comparison results demonstrate that the decision tree method shows a better prediction performance, runtime, and ROC curve in comparison with the random forest method and gradient boosted trees method, which further verifies the feasibility of the proposed method.

The results may have an engineering application value in road traffic safety management. For example, the traffic managers can estimate the speeding violation degree according to the related attributes (e.g., license plate, season, speeding area, position, and rainfall) on the macrolevel, then some intervention policies and measures can be made to decrease the speeding violation from speeding violation quantity and speeding violation level. Meanwhile, the traffic managers can implement different speeding management measures towards various conditions. For instance, they can impose harsher punishments for the high speeding level occurring in the key areas such as the long downhills, road corners, and densely populated areas, especially those crash prone road segments. Because the current speeding violation penalties of Wujiang may be not strict enough (i.e., the speeding violations between 10% and 20% with the punishments including a three-point deduction of license plate and a fine of 50RMB, and the speeding violations between 20% and 50% with the punishments of six-point deduction of license plate and a fine of 200RMB), this encourages the flames of speeding violations to a certain extent. The severe punishments can deter the speeding violation, and it is helpful to reduce the speeding behaviour and the speeding-related crashes. The above measures and policies should also be incorporated in

social campaigns and training courses for targeted drivers or other appropriate traffic safety policies, which is of great significance for the sustainable development of traffic safety management.

However, before the results are applied effectively, several efforts are still needed. First, because of the limitations in data acquisition, the studied data only contains the electronic enforcement data of 2017 from Wujiang's traffic management department. If several years' electronic enforcement data and the related socioeconomic data can be provided, the predicted results may be more microscopic and targeted. Secondly, although the study successfully predicts the speeding violation type using the electronic enforcement data, the corresponding results and conclusions may only be applicable for the situation of Wujiang under the hypothetical conditions. The applicability for other cities is unknown. But it does not hinder the conduction of the whole study, because the modeling framework of the model should be applicable for other cities. Thus, future researches should be conducted to address these issues.

Data Availability

This research has achieved support and cooperation from Wujiang Public Security Bureau. Data were collected by the electronic enforcement equipment of Wujiang. The data is only officially available, and it would be privy. So, the research samples cannot be provided without the permission of Wujiang Public Security Bureau.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 51478110) and the Scientific Research Foundation of Graduate School of Southeast University (No. YBPY1886).

References

- [1] Wujiang Statistical Bureau, *Statistical Bulletin on National Economic and Social Development in Wujiang*, 2017.
- [2] AAA Foundation for Traffic Safety, *Aggressive Driving: Research-Update*, AAA Foundation for Traffic Safety, Washington, DC, USA, 2009.
- [3] D. Royal, "National survey of speeding and unsafe driving attitudes and behaviors: 2002," Volume II-Findings Report, 2003, <https://rosap.nhtl.bts.gov/view/dot/1719>.
- [4] P. J. Cooper, "The relationship between speeding behaviour (as measured by violation convictions) and crash involvement," *Journal of Safety Research*, vol. 28, no. 2, pp. 83–95, 1997.
- [5] J. Mesken, T. Lajunen, and H. Summala, "Interpersonal violations, speeding violations and their relation to accident involvement in Finland," *Ergonomics*, vol. 45, no. 7, pp. 469–483, 2002.
- [6] L. Aarts and I. van Schagen, "Driving speed and the risk of road crashes: a review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.
- [7] V. Viallon and B. Laumon, "Fractions of fatal crashes attributable to speeding: evolution for the period 2001-2010 in France," *Accident Analysis & Prevention*, vol. 52, no. 3, pp. 250–256, 2013.
- [8] S. A. Gargoum and K. El-Basyouny, "Exploring the association between speed and safety: a path analysis approach," *Accident Analysis & Prevention*, vol. 93, pp. 32–40, 2016.
- [9] S. D. Doecke, C. N. Kloeden, J. K. Dutschke, and M. R. Baldock, "Safe speed limits for a safe system: the relationship between speed limit and fatal crash rate for different crash types," *Traffic Injury Prevention*, vol. 19, no. 4, pp. 404–408, 2018.
- [10] C.-M. Tseng, "Speeding violations related to a driver's social-economic demographics and the most frequent driving purpose in Taiwan's male population," *Safety Science*, vol. 57, pp. 236–242, 2013.
- [11] C.-M. Tseng, M.-S. Yeh, L.-Y. Tseng, H.-H. Liu, and M.-C. Lee, "A comprehensive analysis of factors leading to speeding offenses among large-truck drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 38, pp. 171–181, 2016.
- [12] C.-M. Tseng, "Operating styles, working time and daily driving distance in relation to a taxi driver's speeding offenses in Taiwan," *Accident Analysis & Prevention*, vol. 52, pp. 1–8, 2013.
- [13] S. E. Forward, "Intention to speed in a rural area: reasoned but not reasonable," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 13, no. 4, pp. 223–232, 2010.
- [14] S. Vardaki and G. Yannis, "Investigating the self-reported behavior of drivers and their attitudes to traffic violations," *Journal of Safety Research*, vol. 46, pp. 1–11, 2013.
- [15] E. Roidl, F. W. Siebert, M. Oehl, and R. Höger, "Introducing a multivariate model for predicting driving performance: the role of driving anger and personal characteristics," *Journal of Safety Research*, vol. 47, pp. 47–56, 2013.
- [16] J. L. Deffenbacher, D. M. Deffenbacher, R. S. Lynch, and T. L. Richards, "Anger, aggression, and risky behavior: a comparison of high and low anger drivers," *Behaviour Research and Therapy*, vol. 41, no. 6, pp. 701–718, 2003.
- [17] J. Mesken, M. P. Hagenzieker, T. Rothengatter, and D. de Waard, "Frequency, determinants, and consequences of different drivers' emotions: an on-the-road study using self-reports, (observed) behaviour, and physiology," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 10, no. 6, pp. 458–475, 2007.
- [18] A. N. Stephens and J. A. Groeger, "Situational specificity of trait influences on drivers' evaluations and driving behaviour," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 1, pp. 29–39, 2009.
- [19] G. Zhang, K. K. Yau, and X. Gong, "Traffic violations in guangdong province of china: speeding and drunk driving," *Accident Analysis & Prevention*, vol. 64, no. 1, pp. 30–40, 2014.
- [20] Y. Huang, D. J. Sun, and J. Tang, "Taxi driver speeding: who, when, where and how? a comparative study between shanghai and new york city," *Traffic Injury Prevention*, vol. 19, no. 3, pp. 311–316, 2018.
- [21] C. Fu, Y. Pei, Y. Wu, and W. Qi, "The influence of contributory factors on driving violations at intersections: an exploratory analysis," *Advances in Mechanical Engineering*, vol. 5, Article ID 905075, 2013.

- [22] Y. Huang, D. J. Sun, and L. H. Zhang, "Effects of congestion on drivers' speed choice: assessing the mediating role of state aggressiveness based on taxi floating car data," *Accident Analysis & Prevention*, vol. 117, pp. 318–327, 2018.
- [23] H. Liu, L. Zhang, D. Sun, and D. Wang, "Optimize the settings of variable speed limit system to improve the performance of freeway traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3249–3257, 2015.
- [24] D. Sun, C. Zhang, L. Zhang, F. Chen, and Z. R. Peng, "Urban travel behavior analyses and route prediction based on floating car data," *Transportation Letters: The International Journal of Transportation Research*, vol. 6, no. 3, pp. 118–125, 2014.
- [25] D. D. Dinh and H. Kubota, "Speeding behavior on urban residential streets with a 30km/h speed limit under the framework of the theory of planned behavior," *Transport Policy*, vol. 29, pp. 199–208, 2013.
- [26] F. Letirand and P. Delhomme, "Speed behaviour as a choice between observing and exceeding the speed limit," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 6, pp. 481–492, 2005.
- [27] C. Eyssartier, S. Meineri, and N. Gueguen, "Motorcyclists' intention to exceed the speed limit on a 90km/h road: effect of the type of motorcycles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 45, pp. 183–193, 2017.
- [28] M. A. Elliot, "Predicting motorcyclists' intentions to speed: effects of selected cognitions from the theory of planned behaviour, self-identity and social identity," *Accident Analysis & Prevention*, vol. 42, no. 2, pp. 718–725, 2010.
- [29] J. Cestac, F. Paran, and P. Delhomme, "Young drivers' sensation seeking, subjective norms, and perceived behavioral control and their roles in predicting speeding intention: how risk-taking motivations evolve with gender and driving experience," *Safety Science*, vol. 49, no. 3, pp. 424–432, 2011.
- [30] C. Atombo, C. Z. Wu, M. Zhong, and H. Zhang, "Investigating the motivational factors influencing drivers intentions to unsafe driving behaviours: speeding and overtaking violations," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 43, pp. 104–121, 2016.
- [31] D. Jovanović, M. Šraml, B. Matović, and S. Mičić, "An examination of the construct and predictive validity of the self-reported speeding behavior model," *Accident Analysis & Prevention*, vol. 99, pp. 66–76, 2017.
- [32] G. Zhao, C. Wu, and C. Qiao, "A mathematical model for the prediction of speeding with its validation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 828–836, 2013.
- [33] H. Wallén Warner and L. Åberg, "Drivers' beliefs about exceeding the speed limits," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 11, no. 5, pp. 376–389, 2008.
- [34] Y. J. Liu, K. Zhao, Q. Li, and H. W. Xia, "An identification method of violation driving behaviors based on satellite positioning data," *Journal of Highway and Transportation Research and Development*, vol. 34, no. 11, pp. 127–135, 2017.
- [35] M. J. Giles, "Driver speed compliance in Western Australia: a multivariate analysis," *Transport Policy*, vol. 11, no. 3, pp. 228–235, 2004.