Exploring Factors Affecting the Yellow-Light Running Behavior of Electric Bike Riders at Urban Intersections in China

Jing Cai,1 Jianyou Zhao,1 Jing Liu,1,2 Ke Shen,1 Xun Li,1 and Yuntao Ye1

1School of Automobile, Chang’an University, Xi’an 710064, Shaan Xi, China
2School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 230601, China

Correspondence should be addressed to Jianyou Zhao; jyzhao@chd.edu.cn

Received 13 November 2019; Revised 30 January 2020; Accepted 25 February 2020; Published 10 June 2020

1.Introduction

Electric bike as one of the flexible transportation modes is popular in China and other Asian countries, which constitutes about 34% proportion among all travel modes in China [1]. Due to the convenience in congestion traffic, energy efficiency, and high manoeuvrability [2], the electric bike (e-bike) has experienced a tremendous growth in China and its total number was more than 250 million according to the China Bicycle Industry Information Center in 2018 [3]. This trend is likely to continue with the soaring prices of fuel and the traffic jam due to the growing ownership of motor vehicles. More e-bikes traffic accidents may be incurred by the surging number of e-bikes which has attracted some safety concerns around the world. Riders are considered as vulnerable road users since they are not protected by any metal structures of vehicles in traffic crashes [4]. In 2015, e-bike accidents accounted for more than 70% of nonmotorized traffic accidents which involved 14471 casualties [5]. The total number of road e-bike traffic accidents from 2016 to 2017 was 25990, which resulted in 4070 deaths and 28509 injuries [6]. Wu et al. revealed that over 60% of fatal crashes involving two-wheelers resulted from violation of signal rules [7]. According to the Chinese road rules, nonmotorized traffic including regular bicycles and electric bikes should obey the same signal as motor vehicles at signalized intersections. Because of lower speed and inappropriate signal controlling, e-bike riders may do an inappropriate go-stop decision when facing signal changes. Due to the insufficient clearance time, riders may meet the opposite vehicles, thus causing a right-angle collision when doing a go-decision at the onset of yellow-light. Therefore, a study focusing on e-bike riders’ crossing behavior at intersection in yellow-light interval is imperative.

Although the growing violation in e-bikes has attracted widespread interest of researchers to model the signal violation behavior, few studies have investigated e-bike riders yellow-light running behavior. Therefore, this study mainly reviewed the red-light running behavior of e-bike riders and the yellow-light running behaviors of motor vehicle drivers.
Several researchers have studied associated factors, which may influence the e-bike riders’ red-light running (RLR) behavior. Wu et al. [7–9] focused on the rider characteristics such as gender and age group, which were proven to have significant impact on the RLR behavior. Wu et al. studied that the young and middle-aged riders were more likely to run against a red-light than the old ones and males were more likely to act in a risk-taking manner than females [7]. Some studies examined the effects of environment factors on the RLR behavior. Yan et al. found that the type of day (weekday, weekend, and holiday) and period per day (peak and off-peak hours) had effects on RLR violation rate. And the result showed that the RLR rate of e-bike riders was high in off-peak hours but low on weekends and holidays [10]. Yu et al. found that riders were more likely to stop at the intersection with pedestrian countdown signal devices [11]. The study conducted by Zhang and Wu suggested that the sunshields installed at intersection can reduce RLR violation rates of e-bikes on both sunny and cloudy days [8]. Bai et al. found that the type of vehicle had a significant effect on red-light running behavior occurrences [9].

Most of yellow-light running (YLR) behavior studies focused on motor vehicles. Some useful reviews of existing research studies on vehicle YLR behavior can be seen as follows. A number of studies have conducted to explore effects of driver’s personal attributes on YLR behavior violation. The finding of the study conducted by Papaioannou indicated that drivers’ age and gender significantly impacted their YLR violation behavior [12]. Consistent conclusions were found by Rakha and Haque et al. [13, 14]. Many scholars studied the effects of vehicle operation characteristics on yellow-light running behavior. Köll et al. concluded that drivers were likely to pass through intersection rather than stop with a higher approaching speed and shorter distance [15]. Similar conclusions were found by Papaioannou, Bharat, Hurwitz, and Ding et al. [12, 16–18]. Elmitiny et al. [19] and Pathivada, and Perumal [16] focused on the exposures affecting the violation behavior and they found that vehicle type had a statistically influence on the YLR behavior.

The existing signal violation research studies have similarities and differences in the contributing factors that are explained in terms of the following aspects: (a) General summarization: drivers’ characteristics including gender and age were proved to have significant impact on the e-bike riders’ red-light running and vehicle drivers’ yellow-light running violation behaviors, and in both violations, vehicle type is a significant exposure. (b) Differences of exposures: compared with the vehicle YLR studies, the researchers studying e-bike RLR focused on the environment exposures (e.g., the type of day, the weather, and facilities in the intersection). However, the vehicle operation characteristics (e.g., the approaching speed and the distance to stop line) were the crucial factors which have been proved to affect the YLR violation in many vehicle YLR research studies. However, till now, rare studies addressed that the operation characteristics of the e-bike impacted the riders’ crossing behavior in yellow-light interval. The crucial factors of approaching speed and distance to the stop line are taken into consideration about how they affect the intersection crossing behavior of riders in this study.

Modeling techniques were explored by many researchers to explain the e-bike rider’s and vehicle driver’s signal violation behavior at the signalized intersections. The logistic regression model was the most commonly used model to explain the signal violation behavior in studies conducted by Wu, Zhang, Tang, and Bharat et al. [7, 8, 11, 16]. Yan et al. used the Poisson model to describe the e-bike cyclist’s and bicyclist’s RLR behavior [10]. Elmitiny et al. proposed a decision tree model to classify the driver’s stop-go decision behavior in yellow-light interval [19]. Hurwitz et al. [17] and Tang et al. [20] developed a fuzzy logic model to explain the probability of a driver’s go-stop decision at the intersection when facing the signal changes. In this study, the principle component analysis logistics model is proposed to investigate e-bike riders’ yellow-light running behavior. This approach can be used to describe how various factors affect the violation behavior and to eliminate the multicollinearity in the observed data, further improving the measurement accuracy. PCA logistics models have been widely applied in biometrics [21], engineering application [22], economics [23], and management [24] fields to determine causality from collected data. Results of these studies indicated that the PCA model had high model accuracy.

In summary, the objective of this paper is twofold. The first aim is to examine the effects of a set of contributing factors on yellow-light running behavior, which included e-bike rider characteristics (gender, age group, and the type of vehicle), the e-bike riders’ operation characteristics (approaching speed, distance to the stop line), and the characteristics of intersection (the width and the facility of intersection). The second aim is to compare the PCA logistics model and base logistics model, further revealing the contribution of variables excluded or omitted by the base model and improving the model performance in e-bike riders’ yellow-light running behavior analysis. The results of this paper would explain why e-bikers infringe the traffic signal at intersections and might help propose some suggestions on enhancing the safety of e-bike riders, which is a major issue in China and other developing countries.

2. Methodology

2.1. Intersection Selected. The signal intersection chosen to conduct this observation was in Xi’an, the provincial capital city of Shaan Xi province, China, where the total number of electric bikes has been estimated to be more than 3 million [25]. The selected intersection shared the following characteristics:

(1) The electric bike traffic flow is smooth and not queueing
(2) An exclusive nonmotorized lane exists
(3) Before the onset of yellow-light, the signal device has 3-second green-light countdown flashing
To clearly record the approaching behaviors of e-bike at intersection by using a UAV camera, the landscape trees do not exist on the side of nonmotorized lane.

Details and characteristics of the selected intersection are shown in Table 1.

2.2. Data Collection Using UAV. In this study, the field observation approach was used to record the yellow-light running behavior of e-bike riders, which has been commonly used to investigate the red-light running violation [7, 8, 10, 11] and yellow-light running behavior of road users at urban intersections [16, 20, 26]. An unmanned aerial vehicle which has been applied in traffic behavior analysis [27–29] recorded e-bike riders’ crossing behaviors, and synchronized cameras were used to collected riders’ individual characteristics. The location of two cameras is listed in Figure 1. The unmanned aerial vehicle, also called drone, was operated to hover for 50 m altitude over the nonmotorized lane located in upstream intersection. The fight altitude ensured that it would not be visible and audible for the riders and that the rider’s entire crossing process could be recorded. To avoid being spotted by subjective riders and consequently causing changes in crossing behaviors, the synchronized camera installed was hidden behind telegraph poles and pointed towards the riders to observe their detail characteristics. This field survey was conducted during everyday’s peak hour periods (8:00 a.m.-9:00 a.m.; 5:30 p.m.-6:30 p.m.) in good weather conditions for 3 weeks in April 2019.

2.3. Data Extraction. All road users’ crossing behaviors were recorded on the camera videos, but only the behaviors of e-bike riders were extracted when they entered the scope of the UAV camera in 6 seconds prior to the onset yellow-light to the end of yellow-light. We restricted the coding process to only include e-bikes straightly passing through at intersections. Right-turn e-bikes were ignored because they could not subject to the traffic signal controlling on the basis of the road law in China. Meanwhile, left-turn e-bikes were also excluded due to having dedicated left signal light for them. After the coding process, video analysis software was applied to extract trajectory information. The software could mark the objective e-bike and automatically track its position at the frame rate of 25 frames per second. The Cartesian coordinate was established with the cross point of the pavement and the stop bar as the origin which can be regarded as a reference to extract the X and Y of e-bikes’ position (see in Figure 2). Hence, the approaching speed and the distance to the stop line of e-bikes could be obtained. Then, the riders’ individual characteristics coded by the synchronized camera including gender, estimated age group, and vehicle type were extracted and are listed in Table 2.

In order to avoid data recording mistake, two trained graduate students independently extracted the data from collected videos. The recoding reliability was calculated by Cohen’s Kappa for categorical variables and intraclass correlation for continuous variables. All the coefficients ranged from 0.79 to 0.99, which ensured the reliability of extracting process.

2.4. Modeling Rider Behavior. Some limitations may exist in predicting the result by the linear regression model due to the binary response variables frequently involved in traffic behavior. The logistics regression has been widely applied and proven to be successful to model traffic safety research, such as evaluating the contributing factors for vehicle accident [31], especially in examination of risk factors involved in red-light running and yellow-light running behavior [7, 8, 11, 16]. In our model, \( Y = 1 \) denoted yellow-light running behavior of e-bike riders and \( Y = 0 \) denoted that riders stopped their riding behavior in the yellow-light interval. The YLR behavior model was formulated as follows:

\[
\log \left( \frac{p_i}{1 - p_i} \right) = \alpha + \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_n x_n, \tag{1}
\]

where \( p_i \) is the probability that events occurred, \( \alpha \) is a constant term, and \( \beta_i \) are the corresponding coefficient of \( x_i \) estimated by the method of maximum likelihood.

The probability of the rider \( i \) running against the yellow-light was obtained in the following equation:

\[
p_i(y_i = 1 | x_{i0}, x_{i1}, \ldots, x_{in}) = \frac{\exp(\alpha + \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_n x_n)}{1 + \exp(\alpha + \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_n x_n)} \tag{2}
\]

2.5. Method of Estimating Variables’ Multicollinearity. The logistics model presented in equations (1) and (2) assumes that explanatory variables are independent. However, the observed variables had a multicollinearity problem. Multicollinearity may result in prejudices in the estimation of the model and the interpretation of its parameters. In this paper, we used two variable selection methods to eliminate the multicollinearity problem.

2.5.1. Stepwise Regression Selection Model. The stepwise regression has been commonly used to deal with multicollinearity in logistics regression process. The stepwise regression selection method was used to extract possible combination of explanatory variables affecting the rider behavior at the yellow-light onset, and the final combination of variables was decided based on whether those were significant at a 95% confidence level using a T-test in the SPSS 22.0 software.

2.5.2. Principle Component Analysis Selection Model. In order to obtain an accurate estimation of explanatory variables under multicollinearity, principal component analysis (PCA) was proposed to extract features of explanatory variable. The Kaiser–Meyer–Olkin (KMO) test is shown whether the sample data met the requirements for PCA. Since PCA is suitable for continuous variables, categorical principal component analysis (CATPCA) was used in this study to transform the categorical variables into numerical
The CATPCA process is briefly described as follows:

Assuming that measurements of \( n \) individuals on \( m \) variables scores were given in an \( n \times m \) scores matrix \( H \) where each variable was dedicated by \( X_j \), \( j = 1, \ldots, m \), that was the \( j^{th} \) column of \( H \), \( X_j \) was measured in nominal or ordinal level. Equation (3) transformed the score into category quantification:

### Table 1: Characteristics of study intersections.

<table>
<thead>
<tr>
<th>Intersection</th>
<th>Wenyi North rd.–Huangheng South rd. (W-H intersection)</th>
<th>Hanguang North rd.–Xiaozhai West rd. (H-X intersection)</th>
<th>Taibai North rd.–Keji rd. (T-K intersection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of intersection</td>
<td>Four-arm</td>
<td>Four-arm</td>
<td>Four-arm</td>
</tr>
<tr>
<td>Approaches(^a)</td>
<td>SB</td>
<td>SB</td>
<td>E-WB</td>
</tr>
<tr>
<td>Width of intersection(^b) (m)</td>
<td>60</td>
<td>50</td>
<td>75</td>
</tr>
<tr>
<td>Width of nonmotorized lane (m)</td>
<td>2.7</td>
<td>3.0</td>
<td>2.7</td>
</tr>
<tr>
<td>Length of yellow-light (s)</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Flashing countdown time</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

\(^a\)NB: south-bound approach; E-WB: east-west approach. \(^b\)Width of intersection: the distance between the stop line in the entrance of the intersection and the other invisible stop line in the exit of the intersection which is the extension cord of the opposite nonmotoried lane’s stop line.

![Figure 1: Bird’s eye view and two photos of camera view of W-H intersection.](image1)

![Figure 2: Image of software for data extraction.](image2)
The loss function was defined as
\[ F(Q, A, B) = n^{-1} \sum_j tr\left( qa_j^T - B\right)^T\left( qa_j^T - B\right), \]
where \( Q \) is the matrix of category quantifications. \( A \) is the \( m \times p \) matrix of the component loadings where the \( j^{th} \) column is denoted by \( a_j \). \( B \) is the \( n \times p \) matrix of object scores, which are the individuals on the principal component. \( tr \) is the trace function.

The score matrix \( H \) is replaced by the matrix \( Q \) that has the categorical variables into numerical values [23]. Then, the PCA analysis was conducted by software SAS 9.4 to replace the original corrected variables by uncorrelated principal components to regress the logistics model [21, 33, 34].

Two logistics models based on different methods (e.g., stepwise regression for the base logistics model and principle component analysis for the PCA logistics model) were employed to deal with multicollinearity. The results about e-bike riders’ YLR behaviors were estimated and compared.

Furthermore, the Hosmer–Lemeshow test and some goodness-of-fit measure were used for model comparison.

3. Results

3.1. Descriptive Statistics. In 28 h high-resolution videos, a total of 396 valid clearing intersection behaviors in the yellow-light interval events were observed, among which 248 riders had yellow-light running behaviors. YLR riders were divided into different characteristic groups, as summarized in Table 3.

### Table 2: Definition of variables coded.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Type of variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (GEN)</td>
<td>0 for female; 1 for male</td>
<td>Categorical</td>
</tr>
<tr>
<td>Estimated age group* (AGE)</td>
<td>1 for middle-age group (30–50)</td>
<td>Categorical</td>
</tr>
<tr>
<td></td>
<td>2 for old group (&gt; 50)</td>
<td></td>
</tr>
<tr>
<td>Vehicle type* (VT)</td>
<td>0 for bicycle-style electric bike</td>
<td>Categorical</td>
</tr>
<tr>
<td></td>
<td>1 for scooter-style electric bike</td>
<td></td>
</tr>
<tr>
<td>Physical barrier (PB)</td>
<td>0 for without physical barriers between motorized vehicle lane</td>
<td>Categorical</td>
</tr>
<tr>
<td></td>
<td>1 for with physical barriers between motorized vehicle lane</td>
<td></td>
</tr>
<tr>
<td>Approaching speed (AS)</td>
<td>The approaching speed of vehicle at the onset of the yellow</td>
<td>Continuous</td>
</tr>
<tr>
<td>Distance to the stop bar (DTS)</td>
<td>Vehicle’s distance to the stop bar at the onset of the yellow indication</td>
<td>Continuous</td>
</tr>
<tr>
<td>Width of intersection (WI)</td>
<td>The distance between the stop line in the entrance of the intersection and the other invisible stop line in the exit of the intersection which is the extension cord of the opposite nonmotorized lane’s stop line</td>
<td>Continuous</td>
</tr>
</tbody>
</table>

*The rider’s age information extracted from the recorded videos using the estimated age group could be more effective, and the group-dividing method was reported by Wu et al. [7].

Bicycle-style e-bike can be solely electric-powered or require pedal assistance, while the power of scooter-style e-bike only comes from the electromotor. The speed of the scooter-style e-bike is faster than the bicycle-style ones, and the features of each type are shown in Figure 3 [30].
The overall proportion of riders who cross against yellow-light was 62.6%, and it varied with the intersections (from 50% to 72.7%). The majority of observed riders were male (216) and were aged under 50 years (332). Scooter-style e-bike accounted for more than 57% of the total vehicles in all intersections. Chi-square test was used to analyze the number of YLR behavior riders in different gender, age, and vehicle type groups. It was observed that male riders were prone to YLR behaviors (161 vs. 87, \( p = 0.01 \)). In addition, riders who drove scooter-style e-bikes were more likely to go against yellow-light than bicycle-style e-bike riders (181 vs. 76, \( p < 0.01 \)). A higher proportion of yellow-light runners were observed in young and middle-age riders than in old ones (73.6% and 57.4% vs. 48.4%); however, marginal difference can be observed from the result of the chi-square test for the number of YLR riders in the age group (\( p = 0.167 \)).

Table 4 lists the vehicle’s approaching speed and distance to the stop line extracted from the acquired trajectory data. The data were tested by the Levene test (\( F = 3.307, p < 0.05 \)) prior to Student’s t-test. The result indicated that the mean approaching speed of YLR riders at the onset of yellow-light was 18.9 km, significantly higher than the non-YLR ones (\( t = -5.248, p < 0.05 \)). Furthermore, the distance to the stop line was different between the YLR group and non-YLR group (\( F = 8.814, p < 0.05; t = 4.02, p < 0.01 \)). The YLR riders were closer to the stop line than non-YLR riders (8.0 m vs. 14.9 m).

### 3.2. Characteristic Parameters Analysis of the YLR Behavior Model

In order to further study the correlation of each driver’s own characteristics in different stages, the strength of a relationship between variables should be quantitatively measured. Pearson correlation coefficients were evaluated by continuous variables, and the categorical variables were estimated by Cramer’s V coefficient by the chi-square test. The results are listed in Table 5.

AS and VT and GEN and VT had significant positive correlation with each other, the value of correlation coefficient reaching 0.767 and 0.612, respectively. The positive correlation coefficient illustrated that scooter-style e-bike riders had a higher approaching speed to clear the intersection at the onset of yellow-light and male riders were more likely to ride a scooter-style e-bike than bicycle-style. In addition, gender had a significantly positive influence on the approaching speed and the distance to the stop line (\( r = 0.506 \) and \( r = 0.350 \), respectively). That is to say, at the onset of yellow-light, in comparison with female riders, males rode faster and were closer to the stop line. The value of correlation coefficients between AS and AGE and AS and DTS were negative, reaching −0.495 and −0.05, respectively. The negative correlation between AS and these two parameters showed that the faster the approaching speed riders had, the younger they were, and that riders with faster approaching speed could be closer to the stop line. Furthermore, the DTS had slight negative correlation between WI and PB (\( r = -0.266 \), \( r = -0.370 \) respectively), which indicated that the riders were far away from the stop line at the onset of yellow-light when the crossing distance was longer or when there existed physical barriers separating the motorized vehicle lane and nonmotorized vehicle lane.

The variance inflation factory (VIF) was calculated to confirm multicollinearity among the explanatory variables in the following equation:

\[
VIF_i = 1 - \frac{1}{R_i^2},
\]

where \( R_i^2 \) is the model determination coefficient.

The results were as follows: VIF = \[12.629, 11.396, 1.355, 1.540, 12.087, 1.637, 2.946\]. Three VIF values were greater than 5 (12.629, 11.396, and 12.087), which indicated that the multicollinearity problem existed in explanatory variables.

### 3.3. Principle Component Analysis of Explanatory Variables

Principal component analysis was used to eliminate the multicollinearity problem of explanatory variables mentioned above. Four explanatory variables, physical barriers, gender, age group and vehicle type, are the categorical variables in this study. Therefore, we utilized CATPCA to transform categorical variables into continuous ones. The Kaiser–Meyer–Olkin (KMO) test was used to calculate the four transformed variables and three original variables, and the value of the test was 0.7 which suggested that the sample

<table>
<thead>
<tr>
<th>Intersection site</th>
<th>W-H intersection</th>
<th>H-X intersection</th>
<th>T-K intersection</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>62.6% (52/83)</td>
<td>75% (33/44)</td>
<td>85.4% (76/89)</td>
<td>74.5% (161/216)</td>
</tr>
<tr>
<td>Female</td>
<td>34.8% (24/69)</td>
<td>45.8% (11/24)</td>
<td>59.8% (52/87)</td>
<td>48.3% (87/180)</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>59% (36/61)</td>
<td>82.9% (29/35)</td>
<td>82.1% (55/67)</td>
<td>73.6% (120/163)</td>
</tr>
<tr>
<td>Middle-age</td>
<td>42.9% (21/49)</td>
<td>42.9% (12/28)</td>
<td>69.6% (64/92)</td>
<td>57.4% (97/169)</td>
</tr>
<tr>
<td>Old</td>
<td>45.2% (19/42)</td>
<td>60% (3/5)</td>
<td>52.9% (9/17)</td>
<td>48.4% (31/64)</td>
</tr>
<tr>
<td>Vehicle type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle-style electric-bike</td>
<td>29.7% (19/64)</td>
<td>42.9% (12/28)</td>
<td>47.4% (36/76)</td>
<td>39.9% (67/168)</td>
</tr>
<tr>
<td>Scooter-style electric-bike</td>
<td>64.8% (57/88)</td>
<td>80% (32/40)</td>
<td>92% (92/100)</td>
<td>79.4% (181/228)</td>
</tr>
<tr>
<td>Overall</td>
<td>50% (76/152)</td>
<td>64.7% (44/68)</td>
<td>72.7% (128/176)</td>
<td>62.6% (248/396)</td>
</tr>
</tbody>
</table>
Eigenvector

3.4. Model Estimation. The two binomial logistics models of rides’ yellow-light running behavior which was based on the field observation were established by the forward stepwise (likelihood ratio) and PCA methods, respectively.

3.4.1. Base Logistics Model. Based on the previous study, seven explanatory variables were selected to explore riders’ yellow-light running behavior. Among those variables, only the estimated age group is a three-category variable. Two dummy variables and a consultative valuable were set for the base logistics model, which are shown in Table 7. Through the forward stepwise (likelihood ratio) method, the yellow-light running behavior using the base logistics model was established. The result is reported in Table 8. Estimation step terminated at step 6 because the change in parameter estimations is less than 0.01.

The probability prediction of YLR behavior function was given by

\[ \hat{p} = \frac{\exp \left( -1.040 + 0.191AS - 0.2DTS + 1.826\text{GEN} \right)}{1 + \exp \left( -1.040 + 0.191AS - 0.2DTS + 1.826\text{GEN} \right)} \]  

(8)

The Hosmer–Lemeshow test and the value of prediction accuracy are summarized in Table 9, respectively, to access goodness of fit in the proposed model.

3.5. PCA Logistics Model. As mentioned earlier, we applied principle component analysis to eliminate the multi-collinearity of explanatory variables. The result of the PCA logistics model, obtained from the output of the logistics model with the uncorrelated factors, is listed in Table 10. Hence, the PCA logistics model is given as follows:

\[ \log(\pi_{i+1}) = 0.704 + 1.027 \text{factor}_1 + 0.868 \text{factor}_2 \]  

(9)
Substituting equations (6) and (7) into equation (8), the probability prediction of YLR behavior PCA logit function is given by

\[
p = \frac{\exp \left( 0.704 + 0.39 \text{AS} - 0.267 \text{DTS} + 0.38 \text{WI} + 0.47 \text{PB} + 0.206 \text{GEN} - 0.285 \text{AGE} + 0.339 \text{VT} \right)}{1 + \exp \left( 0.704 + 0.39 \text{AS} - 0.267 \text{DTS} + 0.38 \text{WI} + 0.47 \text{PB} + 0.206 \text{GEN} - 0.285 \text{AGE} + 0.339 \text{VT} \right)}
\]

Table 9 summarizes the result of goodness of fit for the base logistics and PCA logistics models. Prediction accuracy of the PCA logistics model was 77.8%, and the value of chi-square calculated by the Hosmer–Lemeshow test was 5.72, while the two values of base model were 74.7% and 6.398, respectively.

4. Discussion

Electric bike is a critic transportation mode in China and other developing Asian countries, while it is involved in massive casualties at intersections. The objective of this study was to explore e-bike riders’ behavior when facing the yellow-light interval. The results indicated that 62.6% (248/396) of e-bike riders arriving during the yellow-light period run against the traffic signal at intersection. This finding was higher than that of Bharat [16]. The difference in ratio of the number of yellow-light running behaviors may be caused by different collecting data time.

Two types of yellow-light running behavior analytical models for e-bike riders were developed, with different multicollinearity eliminating method (forward stepwise and PCA). Table 9 lists goodness-of-fit measures for the base logistics model and PCA logistics model, respectively. Results revealed that the principle component analysis model improved overall fit as measured by the Hosmer–Lemeshow test and percentage correct compared to the base model. As shown in Table 9, the PCA logistics model’s chi-square value was smaller than that of the base model’s which indicated that the variance between model prediction and origin data was small \(p > 0.1\). In addition, the
that a shorter distance to the stop line reduced the likelihood of stopping in the yellow-light period. With respect to the appropriate measures aiming at improving the current situation, speed limit measure should urgently be carried out in e-bike vehicle and strict law enforcement related to riders violate in yellow-light signal should be considered.

4.2. Effect of e-Bike Riders’ Characteristics. According to the model in equation (8), gender was found to be a significant variable to estimate the YLR behavior. Male riders had 1.299 times yellow-light running violation than female riders at signal intersection (the odds ratio = $e^{0.206}$), which meant that males had more propensity to running against yellow-light than females. Consistent result showed that the female drivers had lower rates of signal infringement than male drivers [36]. Also, Wu et al. analyzed information recorded by video cameras, which implied that males were more likely to cross the intersection in a risk-taking manner among 451 e-bikers and bicyclists [7]. The previous study conducted by Parker et al. found that males reacted with more incon siderate driving and impatient driving than females when they faced impede progress such as the signal light changing into yellow-light or red-light [37], which may explain the gender difference in the yellow-light running behavior.

The estimated age group was found to be significant for prediction of the YLR behavior and had a negative effect on it. The young group had higher violation behaviors than middle-aged and old riders. This tied in with the findings of Chung and Wong [38] and Zamani-Alavijeh et al. [39] that compared to other drivers, the young driver was more likely to engage in risky driving behavior and involved in severe accidents. However, the result was not consistent with the result in which the age group failed to be a significant variable for predicting signal violation proposed by Wu et al. [7]. One possible reason was that researchers used the stepwise regression method in the logistics model, and the method of stepwise regression could obtain an optimal combination of variables, but this combination of variables may eliminate some variables due to the small interpretation.

4.1. Effect of e-Bike Riders’ Operation Characteristics. The result of the PCA model indicated that the approaching speed had a significant positive impact on the YLR decision. The odds ratio was 1.477($e^{0.39}$) which suggested that the rider, who traveled 1 km/h faster than the other ones, was 1.477 times likely to do a go-decision against yellow-light signal. The result was consistent with the result in a motor vehicle observation study reported by Papaioannou [12]. One possible reason would be that the riders traveling in a fast approaching speed may think they can pass the stop line in the remanent yellow-light time and quickly cleared the intersection to avoid waiting for the next green light signal. Therefore, effectively lowering riders’ approaching speed at the approach to the intersection may result in a significantly less YLR violation.

The effect of distance to the stop line (DTS) on the YLR violation was significantly negative (−0.276). The odds ratio value was 0.765 which indicated that the riders, who were 1 m farther than the others, was 0.765 times more liable to run against yellow-light, that was to say riders who are closer to the stop line at the onset of yellow-light will be more likely to cross the intersection than the one who was far away from the stop bar. The GHM model proposed by Gazis et al. proved that a smaller distance to the stop distance than the vehicle’s minimum stopping line could lead to a rear-end accident [35]. Riders may feel they cannot safely stop before the stop line so that they prefer continuously running against the yellow-light. This result was in line with the previous study [16], proving that a shorter distance to the stop line reduced the percentage correct also proved that the PCA logistics model was statistically superior, which had a higher percentage correct of 3.1%. The results suggested that stepwise regression could obtain an optimal combination of variables, but those combinations of variables may eliminate some variables due to the small interpretation which should not be omitted in the YLR behavior model. Therefore, the following analysis was mainly based on PCA logistics model.

**Table 9: Goodness-of-fit statistics for logistics regression.**

<table>
<thead>
<tr>
<th></th>
<th>Base logistics model</th>
<th>PCA logistics model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosmer–Lemeshow test</td>
<td>Chi-square</td>
<td>6.398</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>0.603</td>
</tr>
<tr>
<td>Percentage correct</td>
<td></td>
<td>74.7</td>
</tr>
</tbody>
</table>

**Table 10: PCA logistics regression values of the predicting variables.**

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>Wald df</th>
<th>Sig.</th>
<th>95% CI for Exp (β)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>1.027</td>
<td>0.272</td>
<td>14.209</td>
<td>1</td>
<td>0.000</td>
<td>2.791</td>
<td>4.760</td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.868</td>
<td>0.268</td>
<td>10.456</td>
<td>1</td>
<td>0.001</td>
<td>2.382</td>
<td>4.032</td>
</tr>
<tr>
<td>Constant</td>
<td>0.704</td>
<td>0.251</td>
<td>7.850</td>
<td>1</td>
<td>0.002</td>
<td>2.021</td>
<td></td>
</tr>
</tbody>
</table>

**Table 11: Summary of factors that affect the YLR behavior.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>β</th>
<th>SE</th>
<th>Wald df</th>
<th>Sig.</th>
<th>95% CI for Exp (β)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>1.299</td>
<td>0.272</td>
<td>1</td>
<td>0.000</td>
<td>2.791</td>
<td>4.760</td>
</tr>
<tr>
<td>Age</td>
<td>Young</td>
<td>0.765</td>
<td>0.251</td>
<td>7.850</td>
<td>1</td>
<td>2.021</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>Slow</td>
<td>1.477</td>
<td>0.268</td>
<td>10.456</td>
<td>1</td>
<td>2.382</td>
<td>4.032</td>
</tr>
</tbody>
</table>

**Table 12: Goodness-of-fit statistics for PCA logistics regression.**

<table>
<thead>
<tr>
<th></th>
<th>Hosmer–Lemeshow test</th>
<th>BASE logistics model</th>
<th>PCA logistics model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square</td>
<td>6.398</td>
<td>5.720</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Sig.</td>
<td>0.603</td>
<td>0.679</td>
<td></td>
</tr>
<tr>
<td>Percentage correct</td>
<td>74.7</td>
<td>77.8</td>
<td></td>
</tr>
</tbody>
</table>
be carried out in practice to avoid yellow-light running behavior so that riders would take YLR behavior as an illegal behavior rather than a normal riding behavior.

4.3. Effect of the Characteristics of Intersection. The width of intersection had a significant positive impact on the yellow-light running behavior. The odds ratio was 1.462 ($e^{0.47}$) which suggested that 1 m longer width may result in the rider having 1.462 times to go against the yellow-light signal. One possible reason was that the gap in the cross traffic is larger in a big intersection than that in a small one; therefore, riders consider that they may have a chance to safely cross the intersection in an enough gap.

The probability of riders running against yellow-light in an intersection with physical barriers was 1.6 times higher than in intersections without physical barriers (the odds ratio = $e^{0.47}$). One possible explanation was that the existence of a barrier between motorized vehicle lane and nonmotorized vehicle lane might increase speed among riders due to the separation from motorized traffic; thus, riders may run against the yellow-light in fast speed.

4.4. Limitation. There are several limitations in the present study. First, the data collection in this study is performed in a single city and in the four-leg intersection which may not be representative of Chinese complicated intersection environments. More data should be collected at multiple intersections in other cities to validate the findings in further study. Second, the various types of traffic light devices and phase of traffic light may impact e-bike riders’ behavior, which were not considered in present research. Future research studies including traffic light devices and phase are required to better understand how these variables influence the YLR behavior. Third, the current study focused on individual e-bike rider’s yellow-light running behavior while ignoring the impacts of platoons. E-bike riders’ yellow-light running behavior could be affected by the behaviors of other people, and future research is required to take these effects into consideration.

5. Conclusion

This study sought to examine the effects of factors on e-bike riders’ yellow-light running behavior. 396 samples were collected using field observation in Xi’an, China, for model development. Except for riders’ attributes, the type of e-bikes, operation characteristics related to approaching speed, distance to stop line, and characteristics of intersection including the width of intersection and the existence of physical barriers are also considered as explanatory variables. Given that riders have two choices either to pass or to stop at the intersection when facing the yellow-light, two logistics regression models, PCA logistics model and base logistics model, have been developed to explain this behavior.

The empirical analysis revealed a number of findings. First, results of the PCA logistics model showed that seven factors had significant effect on YLR behavior, while results of the base logistics model showed that just two factors significantly affected YLR violation. The comparison between the results of two established models revealed that omission of effective variables would result in lower prediction accuracy of the model and misunderstanding in riders’ YLR behavior. Second, the characteristics of operation were found to have effect on e-bike riders YLR behavior. It was observed that the probability of stopping decreased with the increase in the approaching speed of the e-bike and riders’ stopping probability decreased with the increase in their distance from the stop line when green traffic light changed to yellow. Third, e-bike riders’ attributes were important factors affecting YLR behaviors. Female riders demonstrated a more obedient behavior to signals than male riders. Riders in the old age group were more likely to stop at the onset of yellow-light compared to other age groups. Also, scooter-style e-bike riders were less likely to stop at the intersection. Last but not the least, e-bike riders’ violation behavior was affected by the intersection characteristics. The existence of physical barrier decreased the riders’ stopping probability. And an intersection with a shorter crossing distance may increase the stopping probability of e-bike riders.

Based on the findings of this study, some countermeasures to improve e-bike riders’ yellow-light running violation should be adopted. Adjustment of the signal time may be necessary to reflect the e-bike operation characteristics. In addition, YLR infringement education must be given repeatedly in a rider’s daily life. Moreover, a license system should be required to e-bike riders; thus, penalization for those YLR riders can link YLR violation with their individual credits under a stricter enforcement at intersection. Improvements for the existing road infrastructure as well as some other countermeasures are required for crossing safety of e-bike riders.

Data Availability

The data used to support the findings of this study have not been made available because we have no right to share it.

Conflicts of Interest

The authors declare that there are no conflicts of interest with respect to the research, authorship and/or publication of this article.

Acknowledgments

This work was jointly supported by the National Key R & D Plan of China (No. 2018YFC0807500), the National Nature Science Foundation of China (No. 51878066), and Funds for Central Universities and Colleges of Chang’an University (No. 300102229201).

References


