

## Non-motorized Vehicle Traffic Accidents in China: Analysing Road Users' Precrash Behaviors and Implications for Road Safety



Xiaoxia Xiong<sup>1,2\*</sup>, Shuichao Zhang<sup>3</sup>, Lin Guo<sup>3</sup>

<sup>1</sup> School of Automotive and Traffic Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, China

<sup>2</sup> Key Laboratory of Ministry of Public Security for Road Traffic Safety, Wuxi 214151, China

<sup>3</sup> School of Civil and Transportation Engineering, Ningbo University of Technology, 201 Fenghua Road, Ningbo 315211, China

Corresponding Author Email: [XXiong623@ujs.edu.cn](mailto:XXiong623@ujs.edu.cn)

<https://doi.org/10.18280/ijss.110112>

### ABSTRACT

**Received:** 27 July 2020

**Accepted:** 23 December 2020

#### Keywords:

*e-bike, bicycle, precrash behavior, road safety*

The paper aims to explore underlying patterns of non-motorized vehicle (NM, including both regular bicycles and e-bikes) traffic accident occurrences based on precrash behaviors. A quarter-year data of NM accidents was collected by Yinzhou Traffic Police Department of Ningbo, China. Descriptive statistics and Rough Set theory were used to examine rules within different types of NM accidents from temporal, spatial, and behavioral aspects. Some main findings include: behavior patterns of different parties involved vary across different accident types, levels of roads, and intersections; motorized vehicle's illegal turning as well as NM's reverse riding are the two key behaviors that deserve concern across all levels of roads and intersection; in addition, for higher level urban roads more attention should be focused on lane violations of motorized vehicles, and for branch roads and intersections prevention efforts could be directed to motorized vehicles' illegal turning around and NM's red-light running respectively. Results from this paper could facilitate related staff formulating more targeted policies to make roadways safer.

## 1. INTRODUCTION

Non-motorized vehicle transportation has been one of the main modes of urban road traffic in China since decades ago. In recent years, a growing number of people choose electric bicycles (e-bikes) for short to medium distances travels thanks to their advantages of convenience, fast speed (compared to conventional bicycles), and low cost (compared to cars) [1-4]. Note that e-bikes with maximum design speed under 25km/h and vehicle mass (including battery) under 55kg are legally classified as non-motorized vehicle according to regulations of China [5], and are required to ride on non-motorized vehicle lanes of urban roads. With the rise of using e-bikes, mixed traffic conflicts between motorized vehicles, non-motorized vehicles (including both conventional bicycles and e-bikes), and pedestrians have become more complicated and raised more concerns regarding their safety impacts. According to national bureau of statistics of China, injuries and fatalities associated with non-motorized vehicles are observed to have an increasing trend over the years, with fatalities rising from 1,600 in 2010 to 3,741 in 2018 (doubled in less than ten year) [6]. Consequently, more attention and efforts are needed for non-motorized vehicle traffic safety improvement.

Safety research on non-motorized vehicles has been conducted from three main perspectives. The first relates to injury patterns of non-motorized vehicle riders among regular bicycles and e-bikes. Lawinger and Bastian [7] and Schepers et al. [8] showed that e-bike riders are more prone to severe crashes compared to regular bicycles. Weber et al. [9] found diverging results when evaluating injury severity of e-bikers

compared to bicyclists in Switzerland, and claimed no clear conclusion can be drawn on differences in injury severity between them. Otte et al. [10] compared crashes between e-bikers and bicyclists in Germany, but no significant differences were found in either crash propensity or injury severity. Based on hospital injury records in Switzerland, Papoutsis et al. [11] conducted an analysis on e-bike accidents at an aggregate level, and found that patients were predominantly male with age around 47, and the main causes of injury were self-accident. Hu et al. [12] explored risk factors (road user and environmental as well as safety factors) affecting injury severity of non-motorized vehicle crashes at a disaggregate level using crash data from a hospital in Hefei, China. Based on the developed multiple-factor non-conditional logistic regression model, they found traffic rule violation is one of the primary factors contributing to injury severity; however, rule violation type was not specified in their study.

The second perspective evaluates factors affecting cyclists' crashes based on questionnaire surveys. Weinert et al. [13] conducted a survey on safety perceptions of e-bikers in Shijiazhuang, China, and found e-bikers were generally satisfied with traveling and females tended to feel safer in crossing intersections with e-bikes compared to bicycles. Yao and Wu [14] conducted a self-reported questionnaire on e-bikers in two large cities in China. They found that more males were at-fault in e-bike crashes compared to females, and e-bike riders were less prone to crashes when he/she had a driver's license. Risk perception and safety attitudes were also identified as significant factors affecting e-bikers' crashes.

Guo et al. [15] obtained cyclists' crash related data via a telephone interview based on crash police reports, with non-crash samples collected by a random questionnaire survey. Factors including gender, age, education level, driver license, car in household, experiences in using e-bike, law compliance, and aggressive driving were found to significantly impact on e-bike crashes. E-bike license plate use was also found to have a strong negative correlation with e-bike crashes. Findings from such studies could provide insight on cyclists' characteristics impacting crashes, but could not inform on specific cyclist behaviors that could lead to accidents.

The third perspective examines contributing factors to cyclists' crashes based on on-site investigation or police crash reports. There have been extensive studies on factors affecting bicycle crashes from the aspects of road users, vehicles, infrastructure, environment, and exposure [16-19], the results of which indicate that road user behaviors could be critical in crash development. However, most of the studies were limited to bicycle-motorized vehicle crash type only, without consideration of e-bikes and their related e-bike-motor vehicle, e-bike-e-bike, and e-bike-pedestrian crashes. In the regard of e-bike related crash analysis, Wu et al. [20] observed red-light running behavior of e-bikers and bicyclists at intersections in Beijing, China, and employed logistic regression analysis to identify factors contributing to a rider running a red light. Du et al. [21] carried out a cross-sectional observational study at randomly selected intersections in Suzhou, China, which described illegal riding behaviors among e-bikers such as riding in a motor vehicle lane and mobile phone use. However, no relationship was established between the observed illegal behaviors and crash outcome in these studies. Bai et al. [22] manually identified non-motorized traffic conflicts/ incidents at signalized intersections by reviewing videos, and compared risk-taking behaviors of e-bikers and bicyclers causing incidents. However, their retrieved incidents were judged based on observed evasive actions (such as braking and swerving) by person, which could be subjective and limited by experience/knowledge of the person; also, only two types of behaviors (turning without yielding and red-light running) were considered. Wang et al. [23] developed a mixed logit model for fault assignment (i.e. e-bike riders solely at fault and drivers solely at fault) of e-bike related fatal crashes in Taixing, China based on factors related to person, environment, and precrash maneuvers. Their modelling results indicate that the precrash maneuvers of both e-bike riders and drivers are significant to fault assignment. However, the maneuvers analyzed in their study are limited to directional operations of e-bikes/vehicles (i.e., going straight and left/right turning, etc.), while discussion on a higher behavioral level is absent.

Many studies in the field of motorized vehicle accidents have shown that precrash driver behaviors are important factors contributing to accidents thanks to large scale naturalistic driving data collected from motorized vehicles [24-26], while due to the lack of such data no extensive study has been focused on different types of non-motorized vehicle accidents (i.e., the roadway accidents that involve non-motorized vehicles, including both bicycles and e-bikes) from the perspective of road users' precrash behaviors. Thus, more research efforts are needed for behavioral level analysis of non-motorized vehicle accidents of different types to understand their safety implications and develop safety countermeasures.

In this paper, in order to explore behavior characteristics of road users involved in non-motorized vehicle accidents, data

on driver/rider/pedestrian behaviors prior to non-motorized vehicle accident were obtained based on witness on site or statements of the two parties involved. For the rest of the paper (organized as shown in Figure 1), the collected data for analysis will be first introduced in Section 2. General temporal and spatial characteristics of NM accidents will be examined in Section 3. Descriptive statistics of road user behaviors will then be analyzed to provide an overall understanding of causes NM accidents in Section 4. Rough set, a useful methodology to discover hidden knowledge under data, will finally be utilized to reveal potential rules/laws within NM accidents from temporal, spatial, and behavioral aspects in Section 5. Section 6 concludes the paper.

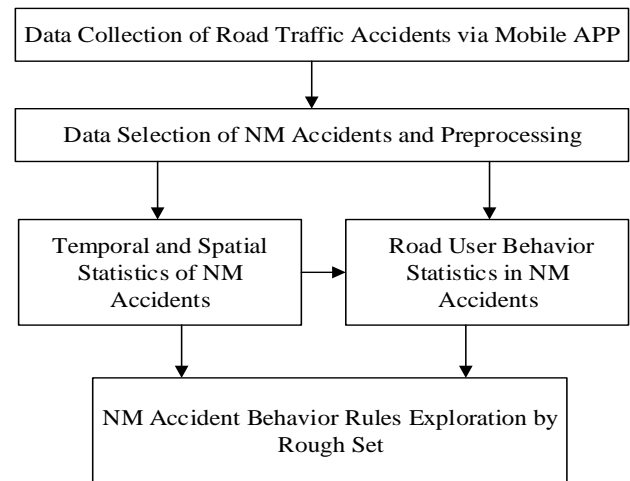


Figure 1. Research methodology and organization

## 2. DATA ASSEMBLY

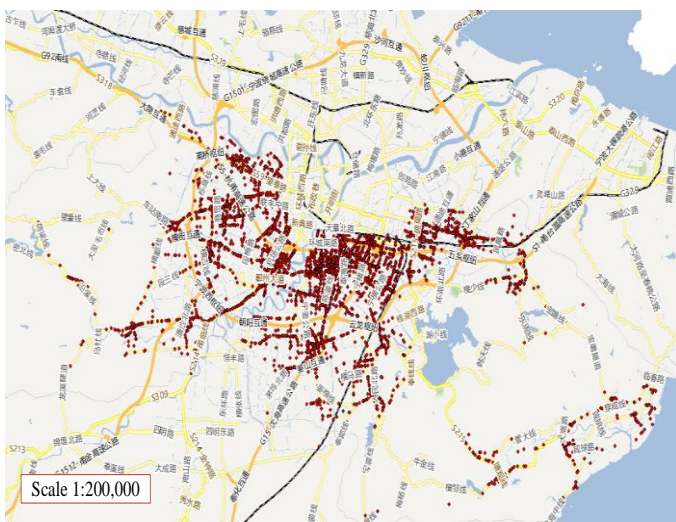
Data were collected in Yinzhou district of Ningbo by Yinzhou Traffic Police Office in cooperation with Ningbo University of Technology (NBUT) in the 4<sup>th</sup> quarter of 2016 (from Oct. 1<sup>st</sup> to Dec 31<sup>th</sup>, 2016). Ningbo, a sub-provincial city under Zhejiang province, is one of the largest cities along the eastern coast of China, with e-bike ownership reaching 26.5 thousand in 2015 [15]. Yinzhou, one of the main districts of Ningbo, covers a total area of 1,346 km<sup>2</sup> and ranks first in all districts of Ningbo in terms of local GDP. Traffic accident information, including both motorized vehicle accidents and non-motorized vehicle accidents, were recorded through a mobile APP by on-site traffic policeman and uploaded to a remote server through the mobile Internet [27]. Collected attributes cover conventional accident data including latitude and longitude coordinates of accident, accident occurrence time, accident type, and weather/environment, as well as detailed precrash behaviors by parties involved (see Table 1), making behavioral analysis feasible for our study. Note that “non-motorized vehicles” includes both regular bicycles and e-bikes in the study.

A total of 37,654 traffic accidents were recorded, within which 133 records have GPS outside of the region of study (generally caused by misuse of mobile phone or GPS positioning failure) and were deleted from the database. To meet the objective of our study, only accidents with non-motorized vehicles involved were extracted, yielding a total of 7,725 NM accident records (with 1,159 at intersection and 6,566 on road sections). To explore the location characteristics

of NM accidents, ArcGIS “Near” tool [28] was utilized to retrieve the road level attribute from OpenStreetMap (OSM) [29] for each accident record happened on road sections (i.e., through mapping accident point and its nearest road link). Note that based on the road type attribute extracted from OSM, road levels in the paper were re-categorized into main road, secondary road, and branch road according to design standards of urban roads in China [30]. In the paper, ‘primary’ and ‘trunk’ from OSM were coded into main road, which mainly connects main districts of city with transportation as its primary function; ‘secondary’ was kept the same as secondary road, which mainly collects and distributes traffic with function somewhere between main road and branch road; ‘tertiary’ and ‘residential’ were coded into branch road, which mainly connects secondary road and local communities with service as its primary function; ‘unclassified’ was manually checked on road map and recoded into corresponding road level. Unreasonable location records (e.g., with the distance from the accident coordinate point to the nearest road link over 50 meters, generally due to GPS positioning failures or featuring NM accidents on roads within communities, which is not in the scope of this study) were manually checked and further dropped, and finally a total of 6,855 valid NM records were retrieved for analysis here (see Figure 2).

**Table 1.** Traffic accident data collection by mobile APP in Yinzhou, Ningbo

Accident Attributes Collected by Mobile APP	
Location	GPS (Longitude and Latitude)
Date & Time	Year-Month-Day; Hour-Minute-Second
Location	Road Section, Intersection, Overpass, Parking Lot
Accident Type	Motorized Vehicle and Motorized Vehicle (M-M), Motorized Vehicle and Non-motorized vehicle (M-NM), Motorized Vehicle and Pedestrian (M-P), Non-motorized vehicle and Non-motorized vehicle (NM-NM), Non-motorized vehicle and Pedestrian (NM-P), Single vehicle accident, Multi-vehicle accident
Pre-crash Behavior	Normal driving/ riding/ walking, red-light violation, turning without yielding, lane violation, etc. (see Section 4 for more details)
Weather Environment	Sunny, cloudy, rainy, snowy, frozen Normal, night, poor lighting, etc.

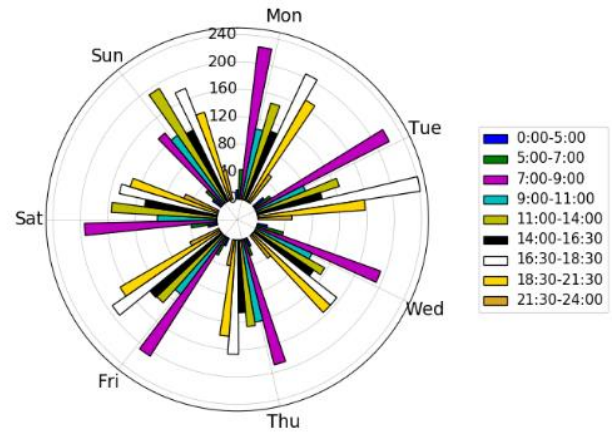


**Figure 2.** NM traffic accidents records distribution in Yinzhou, Ningbo, 2016

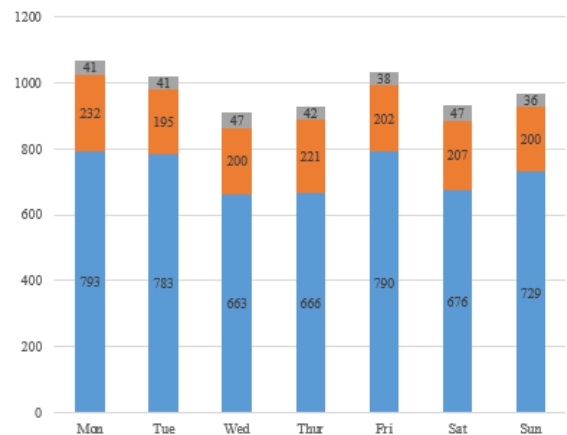
### 3. TEMPORAL AND SPATIAL CHARACTERISTICS OF NM TRAFFIC ACCIDENTS

In this section, a general analysis of NM traffic accidents would be carried out from two aspects, including their temporal distribution and spatial distribution characteristics.

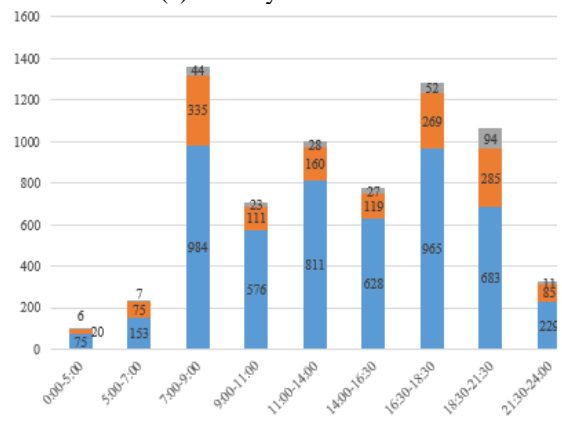
#### 3.1 Temporal distribution



**Figure 3.** Daily and hourly distribution of NM traffic accidents in Yinzhou, Ningbo, 2016



(a) Weekly distribution



(b) Hourly distribution

**Figure 4.** Weekly and hourly distribution by NM accident type

Figures 3 and 4 presents the distribution of NM accident counts by week and hour. It could be noted that the overall number of NM accidents does not vary significantly across weekdays, indicating a diversity of trip purposes (not limited to general commuting trips) by NM mode in the study area. In working days, 7:00-9:00 and 16:30-18:30 (which usually features commuting trips) shows generally higher accident counts than other time slots; while no such pattern are observed at weekend, except that higher proportion of NM accidents are still observed during 7:00-9:00 on Saturday, probably due to 5.5 or 6 working days for some occupation/enterprises. It could also be noted that 11:00-14:00 (which usually features lunch trips) rises to the top and second time slot in ranking of NM accident occurrences on Sunday and Saturday respectively, possibly due to more lunch trips by NM at weekends than on weekdays.

Temporal distribution was further explored for different types of NM accidents. Note that the types of NM accidents are categorized according to parties involved in the accident, i.e. M-NM, NM-NM, and NM-P, as listed in Table 1. Results show that M-NM type generally feature a larger number of accidents at the beginning and end of working days (i.e., Monday and Friday), while NM-NM and NM-P types do not show such significant pattern across working days. The overall temporal distribution of all types of NM accidents generally conforms to the morning and evening peak characteristics of traffic flow on urban roads. Specifically, M-NM accidents peaks at 7:00-9:00 and 16:30-18:30, which are typical peak hours for commute trips. Compared to M-NM accidents, the evening peak period of NM-NM accidents seem to extend from 18:30 to 21:30, during which usually features dinner/evening entertainment trips. NM-P accidents are significantly higher at 18:30-21:30 than other time slots, which pattern is not observed in either M-NM or NM-NM accidents. Such temporal characteristics of NM accidents indicate that targeted measures should be different in terms of time for M-NM, NM-NM, and NM-P accidents, e.g., special attention should be paid to 18:30-21:30 time slot when treating NM-P accidents.

### 3.2 Spatial distribution

Among all retrieved NM accident records of Yinzhou in Oct-Dec 2016, 83.1% of NM accidents occurred on road sections, remaining 16.9% at intersections. Detailed distribution of accident locations is presented in Figure 5. It could be noted that secondary roads and branch roads dominate over 70% of all NM accidents, with each accounting for approximately 35%. Taking into account that the ratio of road length for different road classes in Yinzhou is approximately 1:0.73:1.62 for main road : secondary road : branch road according to statistics from Ningbo Housing and Urban-Rural Development Bureau [31], the average ratio of accident rates per unit length are 1:1.74:0.76 for the three road classes, indicating higher NM accident rate was observed on secondary roads compared to other road levels in the research area.

From location distribution of NM accidents of different types in Figure 6, it could be noted that larger amount of NM-NM accidents was observed on higher level roads in the dataset. However, no conclusion could be drawn on the

relationship between the number of NM-NM accidents and road level. It could be that lower level of road (such as branch road) also has comparable or even higher number of NM-NM accidents, but the severity of these accidents may be slight (the severity of collision generally decreases when travel speeds are smaller, i.e., typically on lower level roads) and thus were not reported to police office (and thus were not included in our dataset). Also note that M-NM accidents dominate intersection NM accidents (over 80%), which is in accordance with one would expect, as intersections generally feature more conflict points for motorized and non-motorized traffic flows compared to road sections (where M and NM traffic generally travel on separate lanes). Such result indicates that intersections are the location worth focusing on for reducing M-NM accidents.

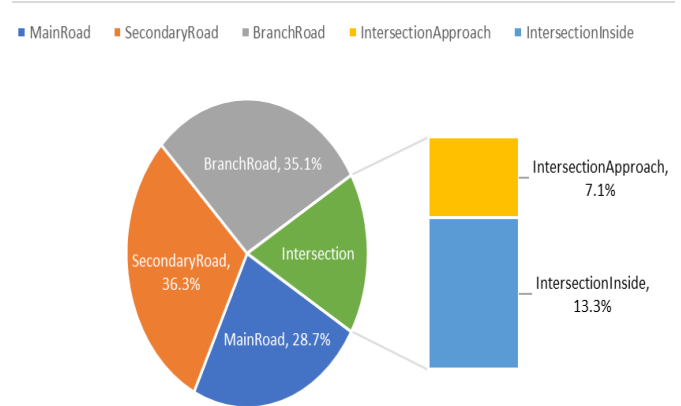


Figure 5. Location distribution

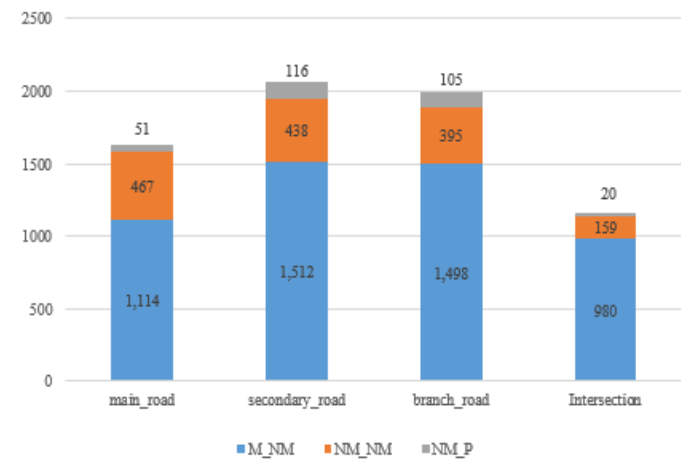


Figure 6. Location distribution by NM accident type

## 4. BEHAVIORAL CHARACTERISTICS OF NM TRAFFIC ACCIDENTS

Based on the behavior information recorded of different parties involved a NM accident (see Table 1), the paper attempts to explore the role of road user behaviors in NM accident occurrences. Behaviors prior to accident occurrence were coded as presented in Table 2 according to motor driver, NM rider, and pedestrian perspectives, respectively.

**Table 2.** Coding of road user behaviors

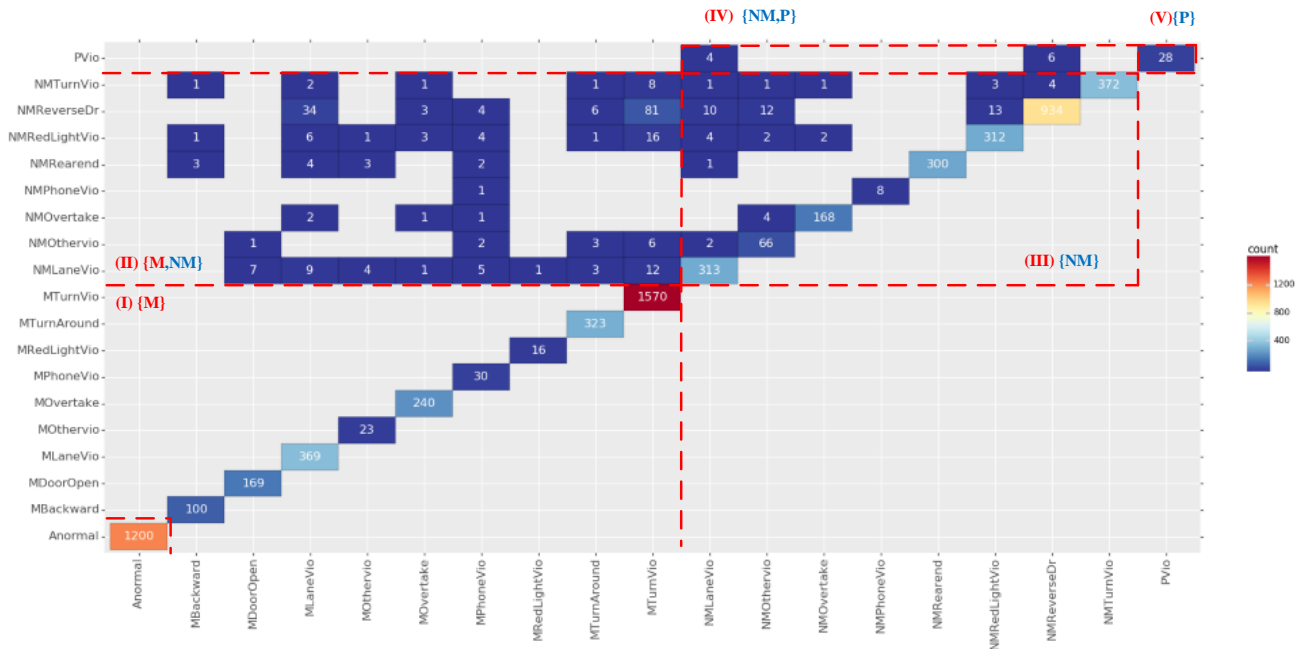
Driver Behavior Set	
Mnormal	Motorized vehicle driving normally at road sections or intersections
MRedLightVio	Motorized vehicle running at red lights
MTurnVio	Motorized vehicle turning without yielding to NM with higher right of way (e.g., NM going straight)
MLaneVio	Motorized vehicle occupying NM lanes
MPhoneVio	Driver of motorized vehicle using mobile phone while driving
MDoorOpen	Passenger of motorized vehicle opening door without watching coming NM
MOvertake	Motorized vehicle overtaking leading vehicle
MBackward	Motorized vehicle reversing or slipping backward
MTurnAround	Motorized vehicle illegally turning around (e.g., at double yellow line)
Mothervio	Other illegal behaviors of motorized vehicle (e.g., illegal cargo or parking)
NM Rider Behavior Set	
NMnormal	Non-motorized vehicle driving normally at road sections or intersections
NMRedLightVio	Non-motorized vehicle running at red lights
NMTurnVio	Non-motorized vehicle turning without yielding to vehicles with higher right of way
NMReverseDr	Non-motorized vehicle riding at a reverse/opposite direction (i.e., facing oncoming traffic)
NMLaneVio	Motorized vehicle illegally occupying motorized vehicle lanes
NMPhoneVio	Rider of non-motorized vehicle using mobile phone while riding
NMRearend	Non-motorized vehicle hitting the back of front vehicle
NMOvertake	Non-motorized vehicle overtaking leading vehicle
NMOthervio	Other illegal behaviors of non-motorized vehicle (e.g., illegal carrying passengers)
Pedestrian Behavior Set	
Pnormal	Pedestrian walking normally at road sections or intersections
PVio	Pedestrian violating laws, e.g., crossing road at red light, climbing separating facility

Note that for each NM accident, behaviors of the two parties involved in the accident were recorded separately as {Duty1, Duty2} in the database, where Duty 1 and 2 refer to the behaviors of party 1 and 2 prior to accident occurrence respectively. For the following analysis, based on the coding scheme in Table 2, for one party being a driver, its corresponding “Duty” would be one behavior from “Driver Behavior Set”, while for a NM rider and a pedestrian it would be from “NM Rider Behavior Set” and “Pedestrian Behavior Set” respectively.

**4.1 Interrelationship of behaviors prior to NM accident occurrence**

To explore potential relationship between different behaviors prior to accident occurrence, the distribution of {Duty1, Duty2} sets observed in the collected dataset was summarized and presented in Figure 7, which could be divided into five regions as follows:

- (I) {M} refers to accident caused by behavior of motorized vehicle only (i.e., behavior of NM is NMnormal);
- (II) {M, NM} refers to accident caused by behaviors of both motorized vehicle and non-motorized vehicle;
- (III) {NM} refers to accident caused by behavior of non-motorized vehicle only (i.e., for those on the diagonal in this region, behavior of motorized vehicle is Mnormal for M-NM accidents, and behavior of another NM is NMnormal or being the same for NM-NM accidents, and behavior of pedestrian is Pnormal for NM-P accidents);
- (IV) {NM, P} refers to accident caused by behaviors of both non-motorized vehicle and pedestrian;
- (V) {P} refers to accident caused by behavior of pedestrian only (i.e., behavior of NM is NMnormal).



**Figure 7.** Behavior distribution of NM accidents

Note that in Figure 7, “Anormal” (i.e., all normal) represents an accident record with both Duty1 and Duty 2 being behaviors from {Mnormal, NMnormal, Pnormal} set.

It could be seen from Figure 7 that a majority of NM accidents happened without faults by any party involved (i.e., “Anormal”), which could be due to unobserved attributes such as facility condition and built environment, or simply random nature within accident occurrence. Except for all normal behavior cases, MTurnVio only and NMReverseDr only are the most two frequently observed behaviors in all NM accidents. Specifically, MTurnVio-NMReverseDr and MLaneVio- NMReverseDr are the most two frequently observed behavior sets for M-NM accidents. In addition, the observed numbers of accidents in the “MTurnVio” column (i.e., “8, 81, 16,6,12” from up to down in Figure 7) are generally higher compared to those in the same row (except on diagonal), indicating a higher chance in NM accident when an illegal/risk behavior is coupled with MTurnVio behavior. Similarly, the observed numbers of accidents in the “NMReverseDr” row (i.e., “34, 3, 4, 6 ,81, 10, 12, 13” from left to right in Figure 7) are generally higher compared to those in the same column (except on diagonal), indicating other illegal/risk behaviors are more likely to result in NM accidents when coupled with NMReverseDr behavior. No clear relationship could be found for other observed behavior sets. Such results imply that more improvement efforts could be devoted to reducing motorized vehicle’s illegal turning behavior as well as non-motorized vehicle’s reverse riding behavior.

#### 4.2 Behavior distribution across different accident types by location

Behavior analysis was further carried out across different accident types at different locations, with results presented in Figure 8-10. Note that when calculating percentages within each figure, behaviors were listed for both parties of each accident, possibly making an accident record counted twice in numerator, with denominator being the total number of observed accidents within a certain accident type at road section/ intersection (e.g., in a particular accident involving a motorized vehicle and a non-motorized vehicle, the motorized vehicle run at red light while the non-motorized vehicle occupied motorway at the same time, making this accident

record appear twice in both “MRedLightVio” and “NMLaneVio” categories, while it was only counted once in denominator for calculating category percentages). It should also be noted that for better presentation of the distribution of behaviors with fault, the numbers of normal driving/riding/walking behaviors (which account for a large proportion of accidents, see Figure 7) are not presented in the following figures.

From Figure 8, it could be noted that MTurnVio dominates the behavior list leading to M-NM accidents, with its observed number higher at intersection than on road section, as generally more turning manoeuvres are required at intersections. The number of NMRedLightVio behaviors significantly rises at intersection compared to road sections, which is the same with one would expect as signal lights are more frequently installed at intersections and NM riders’ non-compliance rate could be high. MTurnAround and MLaneVio are observed at an approximately 5% higher rate in road section accidents than intersection accidents, probably due to larger difficulty in managing these violations on road sections compared to intersections.

Figure 9 shows that NMReverseDr is the most frequently observed behavior leading to NM-NM accidents, with its observed number approximately the same across road sections and intersections. Similar to M-NM accidents, MTurnVio and RedLightVio are more often observed at intersection compared to road sections. Furthermore, NMOvertake behaviors are observed at a higher rate in road section NM-NM accidents compared to intersection NM-NM accidents, probably because these behaviors generally happen more frequently on road sections.

It could be noted from Figure 10 that pedestrian violations are more frequently observed at intersection accidents than road section accidents, as generally more pedestrian violations such as crossing at red light would happen at signalized intersections. Compared to NM-NM type in Figure 9, NMReverseDr has an over 5% higher observation rate in road section accidents than intersection accidents for NM-P type, probably because when compared to road sections pedestrians could pay more attention to reversing non-motorized vehicles at intersections and avoid accidents. Other illegal/ risk non-motorized behaviors are also more frequently observed in road section accidents than intersection accidents except for NMRedLightVio.

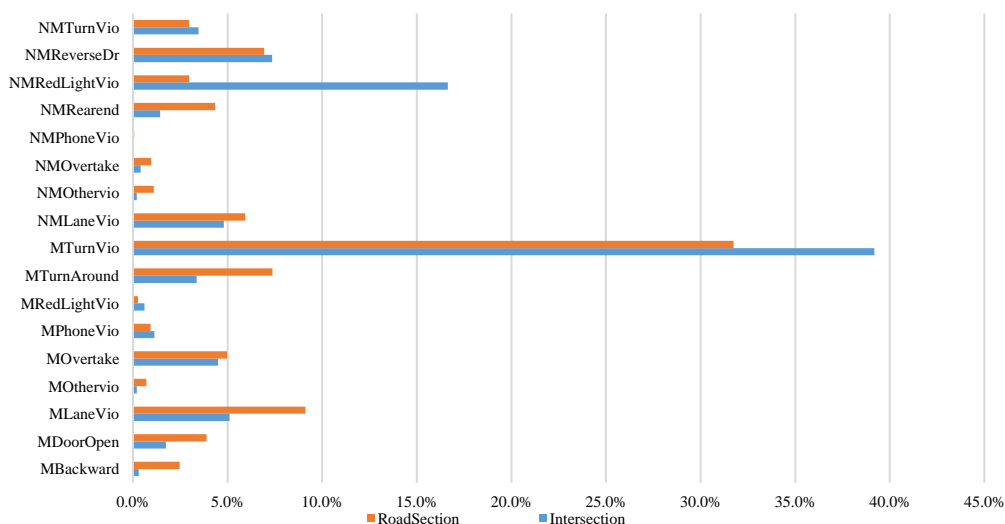


Figure 8. M-NM accidents

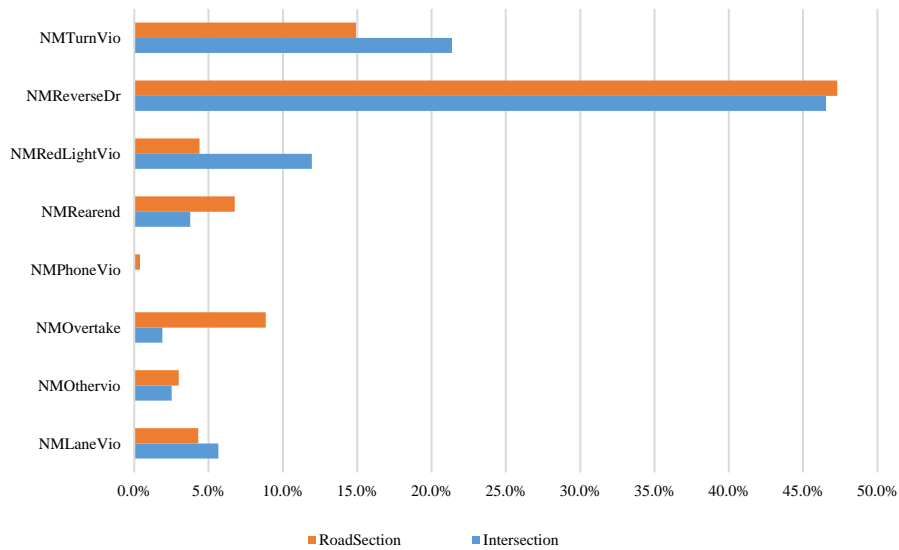


Figure 9. NM-NM accidents

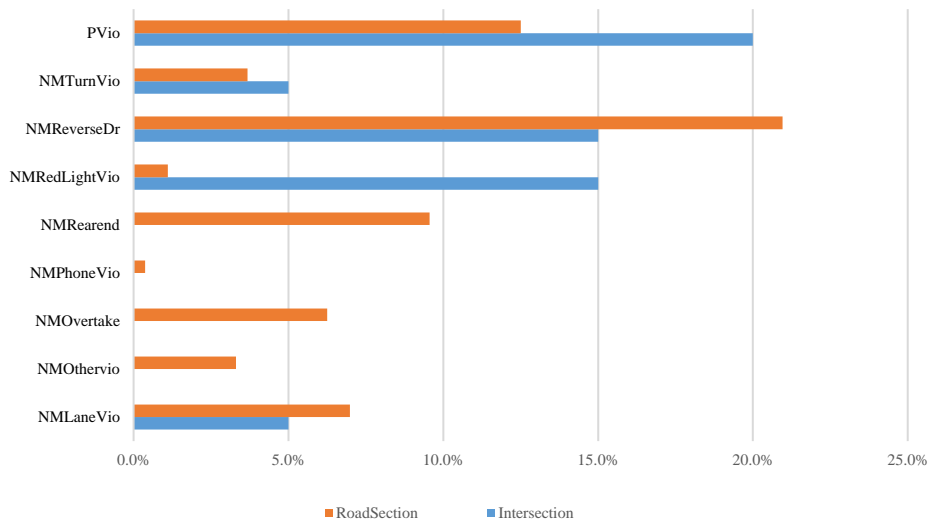
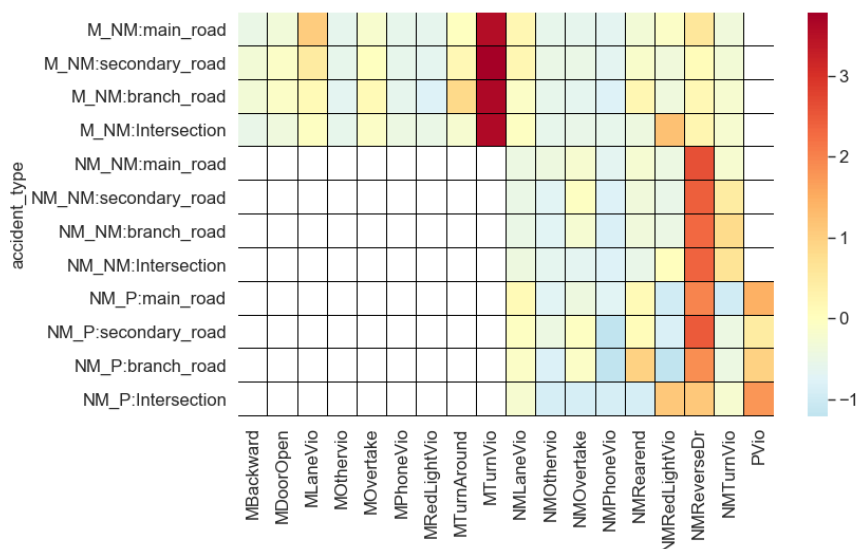


Figure 10. M-P accidents



Note: Rectangular in white represents no accidents could be observed because behavior and accident type does not match (e.g., MBackward could not be observed in a NM-NM or NM-P accident).

Figure 11. Behavior distribution across different accident types and locations

To further explore potential differential relationship between road user behavior and accident type across different levels of roads and intersections, their observation distribution was summarized as a heatmap presented in Figure 11. Note that for better presentation, observation frequencies of each row (i.e., each location type) were standardized (with mean set to zero and variance to one) using `sklearn.preprocessing.scale` function in Python.

It could be noted from Figure 11 that in accordance with Figure 8, MTurnVio is the most frequently observed behaviors in M-NM accidents across all levels of roads and intersections. Similarly, NMReverseDr is the most frequently observed behaviors in NM-NM accidents across all levels of roads and intersections. No behaviors are observed at a significantly higher rate for NM-P accidents, except that NMReverseDr is mostly observed in NM-P accidents on secondary roads while PVio is more frequently observed at intersections. Also, in accordance with Figure 8-10, the observed number of NMRedlightVio caused accidents is generally higher at intersections compared to all levels of road sections across all three types of accidents. This does not necessarily mean that red light violation behaviors could cause more accidents at intersections than on road sections, but could be probably due to fewer signal control are implemented on road sections. Red light violation could be an important factor causing accidents at both intersections and road sections for all accident types. Results suggest for a certain accident type, differentials in behavior-accident occurrence pattern across different levels of roads and intersections are generally not obvious, and thus are difficult to quantify directly.

## 5. ANALYSIS OF NM ACCIDENT BEHAVIORAL RULES BY ROUGH SET

Although statistical data in the last two section could facilitate general understanding of NM accidents in terms of temporal, spatial, and behavioral characteristics, no clue could be obtained as to the interrelationship of these features in NM accident occurrence. In this section, Rough Set is employed to explore the underlying patterns of NM accidents from all of these aspects.

### 5.1 Rough Set Theory

Rough Set (RS) Theory, proposed by Pawlak [32], is recognized as an effective mathematical approach for exploring data patterns/ discovering knowledge based on rule induction method (expressed as IF-THEN statement, or called decision rules). RS targets at classification of indiscernible objects (i.e., cannot be grouped/ classified with certainty) using vague information, and does not require assumptions/constraints on data nature (e.g., independence hypothesis in Bayesian analysis) or model structures (like membership functions in fuzzy theory and other parametric methods). Previous work show that RS could still perform well in discovering relationships when incomplete, imprecise, or even inadequate data exist [33]. Such capability fits well with our study here, as the behaviors recorded in the mobile APP by policeman could be subjective and involve impreciseness. Besides, RS has been used in analysing the process of traffic accident occurrences by Wong and Chung [34] and achieves good results, however which is focused on accidents of motorized vehicle rather than non-motorized vehicles. Thus,

RS is employed here to uncover potential patterns in NM accidents.

Similar to independent-dependent variable pairs in statistics, a set of condition attributes and a decision attribute are defined in RS theory. Specifically, a lower and an upper approximation are defined for each category of the decision attribute  $Y$  (which has a total of  $n$  categories, leading to  $n$  non-overlapping classes for the whole universe  $U = \bigcup_{i=1}^n Y_i$ ), where both lower and upper approximations are formed by elementary sets  $X$  (see Eqns. (1) and (2)), objects within which are indiscernible given the specified set of condition attributes. Thus, elementary sets  $X$  actually represent the smallest partitions of objects, with objects from different  $X$  discernible and those from the same  $X$  indiscernible.

$$\underline{AY}_i = \bigcup X \{X \in A^* \text{ and } X \subseteq Y_i\} \quad (1)$$

$$\overline{AY}_i = \bigcup X \{X \in A^* \text{ and } X \cap Y_i \neq \emptyset\} \quad (2)$$

where,  $A^*$  represents the family of all elementary sets;  $i=1,2,\dots,n$  represents the categories of decision attribute  $Y$ . As implied in the equations above, lower approximation consists of objects that could be fully identified by the elementary sets in  $A^*$ , while upper approximation consists of those cannot be fully identified by the elementary sets in  $A^*$ . As a result, by varying the set of condition attributes, lower and upper approximations of different categories of the decision attribute would be changed, and result in different performances in distinguishing objects. Condition attributes with poor performance in distinguishing objects would be considered redundant and excluded from the condition attribute set (typically called attribute reduction in RS theory). Combinations of values of condition and decision attributes (typically called decision rules; a rule exists if and only if at least one object with such attribute values exists) could be finally retrieved based on rule generation algorithms [35]. More details in theory and application of RS could be found at [32-36] and would not be presented here due to space limit.

### 5.2 Behavior and location rule generation

Based on the accident attributes collected in the dataset (see Table 1), five condition attributes are coded for RS analysis on accident location (decision attribute), including WeekDay, TimeDay, Lighting, WetRoad, and Duty/ Behavior Set (as explained in Table 3).

Based on the retrieved dataset, genetic algorithm-based attribute reduction and rule generation were carried out using Rosetta software [37]. Results showed that all five variables are indispensable in discovering knowledge and thus were all kept for generating rules, and a total of 2,315 rules were finally generated. Among all rules obtained, 19 significant rules could be extracted (as presented in Table 4) based on the following two filter criteria:

1) Rule RHS Support  $\geq 5$ . Rule RHS support is defined as the number of cases that fully satisfy the IF-THEN condition; note that other studies have shown that a rule with RHS at 3 to 4 could already be regarded as significant in accident chain exploration [34] and construction project classification [36].

2) Rule RHS accuracy  $\geq 70\%$ . Rule RHS accuracy is defined as the number of RHS support divided by the number of cases that fully conform to the IF condition.



**Table 3.** Attribute coding for RS analysis

Condition Attributes	
<b>WkDay</b>	Mon, Tue, Wed, Thur, Fri, Sat, Sun
<b>TimeDay</b>	0:00-5:00, 5:00-7:00, 7:00-9:00, 9:00-11:00, 11:00-14:00, 14:00-16:30, 16:30-18:30, 18:30-21:30, 21:30-24:00
<b>Lighting</b>	Normal, Poor (Poor lighting in Table 1)
<b>RoadSurface</b>	Normal, Wet (Snow and Rainy weather in Table 1)
<b>BehaviorSet</b>	{Duty1, Duty2} (note that two sets with same two behaviors in different order are treated as identical in analysis)
Decision Attribute	
<b>Accident Location</b>	Main Road, Secondary Road, Branch Road, Intersection

**Table 4.** Rule generation for NM accident locations

Location	Rule Description						RHS Accuracy (%)
	WkDay	Time Day	Lighting	Road Surface	Duty/ Behavior Set		
Main Road <sup>(35)</sup>	Mon	18:30-21:30	Normal	Normal	NMReverseDr	NMnormal	76.5% <sup>(13)</sup>
	Tue	7:00-9:00	Normal	Normal	MDoorOpen	NMnormal	83.3% <sup>(5)</sup>
	Thur	16:30-18:30	Normal	Normal	MOvertake	NMnormal	71.4% <sup>(5)</sup>
	Fri	11:00-14:00	Normal	Normal	MLaneVio	NMnormal	77.8% <sup>(7)</sup>
	<b>Sun</b>	<b>11:00-14:00</b>	<b>Normal</b>	<b>Normal</b>	<b>NMReverseDr</b>	<b>NMRedLightVio</b>	<b>100.0%</b> <sup>(5)</sup>
Secondary Road <sup>(39)</sup>	Mon	16:30-18:30	Normal	Normal	MLaneVio	NMnormal	76.9% <sup>(10)</sup>
	<b>Tue</b>	<b>16:30-18:30</b>	<b>Normal</b>	<b>Wet</b>	<b>MBackward</b>	<b>NMnormal</b>	<b>100.0%</b> <sup>(6)</sup>
	Tue	18:30-21:30	Normal	Wet	NMReverseDr	NMnormal	77.8% <sup>(7)</sup>
	<b>Wed</b>	<b>18:30-21:30</b>	<b>Normal</b>	<b>Normal</b>	<b>Mnormal</b>	<b>NMLaneVio</b>	<b>85.7%</b> <sup>(6)</sup>
	Wed	18:30-21:30	Normal	Normal	Mnormal	NMRearend	71.4% <sup>(5)</sup>
Branch Road <sup>(27)</sup>	Thur	7:00-9:00	Normal	Normal	NMnormal	NMnormal	71.4% <sup>(5)</sup>
	Wed	7:00-9:00	Normal	Normal	MTurnAround	NMnormal	70.0% <sup>(7)</sup>
	<b>Wed</b>	<b>16:30-18:30</b>	<b>Normal</b>	<b>Normal</b>	<b>Mnormal</b>	<b>NMRearend</b>	<b>85.7%</b> <sup>(6)</sup>
	Sun	9:00-11:00	Normal	Normal	Mnormal	NMnormal	75.0% <sup>(9)</sup>
	Sun	11:00-14:00	Normal	Normal	MDoorOpen	NMnormal	71.4% <sup>(5)</sup>
Intersection <sup>(25)</sup>	Thur	7:00-9:00	Normal	Normal	Mnormal	NMRedLightVio	75.0% <sup>(6)</sup>
	Thur	16:30-18:30	Normal	Normal	Mnormal	NMRedLightVio	70.0% <sup>(7)</sup>
	Fri	7:00-9:00	Normal	Normal	Mnormal	NMRedLightVio	71.4% <sup>(5)</sup>
	<b>Sun</b>	<b>11:00-14:00</b>	<b>Normal</b>	<b>Normal</b>	<b>Mnormal</b>	<b>NMRedLightVio</b>	<b>87.5%</b> <sup>(7)</sup>

<sup>a</sup> The value in the parenthesis represents rule strength (i.e., RHS support); Boldness indicates rules with accuracy over 85%; Shaded context indicates faults by motorized vehicle.

Note that the main purpose of establishing RS here is not for accurate classification (which could be vital to other application scenarios such as online driving risk state classification), but to examine potential patterns that could affect accident locations, and to further enlighten on probable measures in preventing these accidents. And thus the established RS rules are considered acceptable for the subsequent qualitative analysis. More detailed data collection (with more attributes such as age, gender of riders, etc.) in the future is expected to improve classification accuracy.

Although most patterns of NM accidents are unique, some significant accident patterns of M-NM and NM-NM accidents could be identified and worth noting based on the above RS generated rules:

1) A strong pattern in precrash behaviors is found for main roads NM accidents, where rules for M-NM type generally feature accidents caused by motorized vehicles' precrash behaviors (i.e., {MDoorOpen, NMnormal}, {MOvertake, NMnormal}, {MLaneVio, NMnormal} behavior sets) while NM-NM type rules feature reverse riding of non-motorized vehicles (i.e., {NMReverseDr, NMnormal}, {NMReverseDr, NMRedLightVio} behavior sets).

2) Secondary roads show a strong pattern in time of NM accident occurrences (16:30-21:30 on working days) from RS generated rules; in addition, when motorized vehicle and non-motorized vehicle's behaviors involve reverse driving (i.e., MBackward and NMReverseDr), wet road surface is also found to be influential in generating rules of secondary road accidents.

3) Both branch roads and intersections show stronger rules in M-NM accidents than NM-NM accidents, with non-motorized vehicle's red-light running being the most important precrash behavior in rules generated for intersection accidents, which is in accordance with results from Section 4.2.

It should be noted that no strong rules could be identified for NM-P accidents, probably due to its relatively small proportion of observations. Other patterns of rules could be noted from the aspects of time, environment, and behavior:

1) Temporal pattern is not clear for different levels of roads and intersection rules except for secondary roads (i.e., 16:30-21:30 on working days as discussed above).

2) Environment factors (including lighting and road surface condition) do not show a significant role in deducing strong rules on accident locations, except for wet road surface showing up in the strong rules for secondary road.

3) Within a certain location type, no notable patterns could be found in motorized vehicle behaviors for NM accidents. When compared across different level of roads, MLanevio tend to induce more NM accidents on higher level roads (showing up in both main road and secondary road rules), while MTurnAround shows up more often on branch road (which is in accordance with the behaviour distribution heatmap in Figure 2), suggesting MLanevio could play a more significant role in NM accidents on higher level roads, while MTurnAround exerts more influence on branch roads.

Above results suggest that for improving NM safety on main roads, more efforts should be devoted to reducing lane violations of motorized vehicles (such as illegal door opening

and illegal lane occupying) and reverse riding behaviors of non-motorized vehicles; 16:30-21:30 of working days as well as rain/snow weather could be key surveillance time periods for the safety of secondary roads; red-light running behavior of non-motorized vehicles is worth of great attention at intersections. Moreover, for illegal behavior governance of motorized vehicles, their illegal lane occupying should be more strictly restricted on higher level roads, while for branch roads their illegal turning around should be the targeting behavior that needs to be corrected.

Note that although motorized vehicle's turning without yielding is identified as the most frequent illegal behavior across all levels of roads and intersection for M-NM accidents (as shown in Figure 11), it does not show up in the strong rules obtained from RS analysis here. This is in line with one would expect, as RS rules are generated to examine differential patterns in accidents across locations, while motorized vehicle's illegal turning dominates all location categories (i.e., does not have significant variation) and could not contribute to rule identification. From this perspective, the RS-based analysis here provides additional/underlying information on NM accident occurrences that are unable to be extracted from descriptive statistics.

## 6. CONCLUSION

Non-motorized vehicle is one of the main modes of urban road traffic in China, which also accounts for a large proportion of fatalities and injuries in urban road traffic accidents. However, research on behavioral aspects of non-motorized vehicle accidents is still insufficient in literature. The road traffic accident data collected via mobile APP in Yinzhou, Ningbo 2016 covers details on precrash behaviors in NM accidents apart from conventional accident time and location information, which provides a good foundation for behavioral research in our study. Descriptive statistics and Rough Set were employed to explore the temporal, spatial, and behavioral aspects of non-motorized traffic accidents. Results show that behavior patterns of different parties involved vary across different accident types, levels of roads, and intersections, indicating the need to consider all types of locations and road users (drivers, cyclists, and pedestrians) when developing countermeasures to improve road safety. Among all recorded precrash behaviors, motorized vehicle's illegal turning as well as NM's reverse riding are the two key behaviors that deserve concern across all levels of roads and intersection. Besides, for higher level urban roads more attention should be focused on lane violations of motorized vehicles, and for branch roads and intersections prevention efforts could be directed to motorized vehicles' illegal turning around and NM's red-light running respectively. Findings from this study could provide specific focusing areas for preventing NM accidents, and facilitate urban traffic managers/ traffic police officers formulating more targeted policies to make roadways safer.

However, there also exist several limitations of this study. As only accidents reported to police were recorded in the dataset, likely slight accidents (the least severe accidents or accidents without property damage) were not included in analysis, which could underestimate NM accidents on lower level roads (with lower travel speed). Also note that the purpose of the paper is not for accurate classification but to explore potential patterns in NM accidents concerning

behavioral characteristics and locations, and thus the limited collected attributes are considered acceptable for current study; however, more research efforts should be devoted to more extensive data collection (like collecting more data attributes, such as age and gender of riders as well as road facility conditions) and in-depth data analysis on NM traffic accidents in the future. In addition, in the current stage, NM accident types are divided based on the parties involved in the accident (i.e., M-NM, NM-NM, and NM-P), while severity of accident outcome (i.e., injury or fatality) could provide more insights on countermeasures and should be considered for further analysis. Moreover, non-motorized vehicles in our study include both electric-bikes and conventional bicycles (undifferentiated in dataset), while their differential behaviors could have different influential pattern in accident occurrences and should be explored in the future.

## ACKNOWLEDGMENT

The research is supported by National Natural Science Foundation of China (52002154) and the Open Project of Key Laboratory of Ministry of Public Security for Road Traffic Safety (2020ZDSYSKFKT11). The authors also thanks Traffic Police Bureau of Ningbo Municipal Public Security Bureau for collecting the traffic accident source data.

## REFERENCES

- [1] Fyhri, A., Fearnley, N. (2015). Effects of e-bikes on bicycle use and mode share. *Transportation Research Part D*, 36: 45-52. <https://doi.org/10.1016/j.trd.2015.02.005>
- [2] Cairns, S., Behrendt, F., Raffo, D., Beaumont, C., Kiefer, C. (2017). Electrically-assisted bikes: Potential impacts on travel behaviour. *Transportation Research Part A: Policy and Practice*, 103: 327-342. <https://doi.org/10.1016/j.tra.2017.03.007>
- [3] Cairns, S., Behrendt, F., Raffo, D., Beaumont, C., Kiefer, C. (2017). Electrically-assisted bikes: Potential impacts on travel behaviour. *Transportation Research Part A: Policy and Practice*, 103: 327-342. <https://doi.org/10.1016/j.tra.2017.03.007>
- [4] Li, H., Zhong, X., Zhang, W., Li, S., Xing, Y. (2020). An algorithm for e-bike equivalents at signalized intersections based on traffic conflict events number. *Transportation Research Part A: Policy and Practice*, 134: 78-95. <https://doi.org/10.1016/j.tra.2020.02.007>
- [5] GB17761. (2018). *National E-Bike Compelling Standards*, Chinese Standards Publisher.
- [6] China Road Traffic Accidents Statistics. (2019). National Bureau of Statistics, Beijing, China. <http://www.stats.gov.cn/english/>.
- [7] Lawinger, T., Bastian, T. (2013). Neue Formen der Zweiradmobilität. Eine empirische Tiefenanalyse von Pedelec-Unfällen in Baden-Württemberg. *Zeitschrift für Verkehrssicherheit*, 59(2): 99-106. <http://worldcat.org/issn/00443654>
- [8] Schepers, J.P., Fishman, E., Den Hertog, P., Wolt, K.K., Schwab, A.L. (2014). The safety of electrically assisted bicycles compared to classic bicycles. *Accident Analysis & Prevention*, 73: 174-180. <https://doi.org/10.1016/j.aap.2014.09.010>

- [9] Weber, T., Scaramuzza, G., Schmitt, K.U. (2014). Evaluation of e-bike accidents in Switzerland. *Accident Analysis & Prevention*, 73: 47-52. <https://doi.org/10.1016/j.aap.2014.07.020>
- [10] Otte, D., Facius, T., Mueller, C. (2014). Pedelecs im Unfallgeschehen und Vergleich zu konventionellen nicht motorisierten Zweirädern. *VKU Verkehrsunfall und Fahrzeugtechnik*, 52(2): 48-60. <http://worldcat.org/issn/07242050>
- [11] Papoutsis, S., Martinolli, L., Braun, C.T., Exadaktylos, A. K. (2014). E-bike injuries: Experience from an urban emergency department—a retrospective study from Switzerland. *Emergency Medicine International*, 2014. <https://doi.org/10.1155/2014/850236>
- [12] Hu, F., Lv, D., Zhu, J., Fang, J. (2014). Related risk factors for injury severity of e-bike and bicycle crashes in Hefei. *Traffic Injury Prevention*, 15(3): 319-323. <https://doi.org/10.1080/15389588.2013.817669>
- [13] Weinert, J.X., Ma, C., Yang, X., Cherry, C.R. (2007). Electric two-wheelers in China: effect on travel behavior, mode shift, and user safety perceptions in a medium-sized city. *Transportation Research Record*, 2038(1): 62-68. <https://doi.org/10.3141/2038-08>
- [14] Yao, L., Wu, C. (2012). Traffic safety for electric bike riders in China: attitudes, risk perception, and aberrant riding behaviors. *Transportation Research Record*, 2314(1): 49-56. <https://doi.org/10.3141/2314-07>
- [15] Guo, Y., Zhou, J., Wu, Y., Chen, J. (2017). Evaluation of factors affecting e-bike involved crash and e-bike license plate use in China using a bivariate probit model. *Journal of Advanced Transportation*, 2017. <https://doi.org/10.1155/2017/2142659>
- [16] Reynolds, C.C., Harris, M.A., Teschke, K., Crompton, P. A., Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature. *Environmental Health*, 8(1): 1-19. <https://doi.org/10.1186/1476-069X-8-47>
- [17] Handy, S., Van Wee, B., Kroesen, M. (2014). Promoting cycling for transport: Research needs and challenges. *Transport Reviews*, 34(1): 4-24. <https://doi.org/10.1080/01441647.2013.860204>
- [18] Hamann, C.J., Peek-Asa, C., Lynch, C.F., Ramirez, M., Hanley, P. (2015). Epidemiology and spatial examination of bicycle-motor vehicle crashes in Iowa, 2001–2011. *Journal of Transport & Health*, 2(2): 178-188. <https://doi.org/10.1016/j.jth.2014.08.006>
- [19] Prati, G., Marín Puchades, V., De Angelis, M., Fraboni, F., Pietrantonio, L. (2018). Factors contributing to bicycle-motorised vehicle collisions: A systematic literature review. *Transport Reviews*, 38(2): 184-208. <https://doi.org/10.1080/01441647.2017.1314391>
- [20] Wu, C., Yao, L., Zhang, K. (2012). The red-light running behavior of electric bike riders and cyclists at urban intersections in China: an observational study. *Accident Analysis & Prevention*, 49: 186-192. <https://doi.org/10.1016/j.aap.2011.06.001>
- [21] Du, W., Yang, J., Powis, B., Zheng, X., Ozanne-Smith, J., Bilston, L., Wu, M. (2013). Understanding on-road practices of electric bike riders: An observational study in a developed city of China. *Accident Analysis & Prevention*, 59: 319-326. <https://doi.org/10.1016/j.aap.2013.06.011>
- [22] Bai, L., Liu, P., Chen, Y., Zhang, X., Wang, W. (2013). Comparative analysis of the safety effects of electric bikes at signalized intersections. *Transportation Research Part D: Transport and Environment*, 20: 48-54. <https://doi.org/10.1016/j.trd.2013.02.001>
- [23] Wang, C., Xu, C., Xia, J., Qian, Z. (2017). Modeling faults among e-bike-related fatal crashes in China. *Traffic Injury Prevention*, 18(2): 175-181. <https://doi.org/10.1080/15389588.2016.1228922>
- [24] Talbot, R., Fagerlind, H., Morris, A. (2013). Exploring inattention and distraction in the SafetyNet Accident Causation Database. *Accident Analysis & Prevention*, 60: 445-455. <https://doi.org/10.1016/j.aap.2012.03.031>
- [25] Zhang, G., Yau, K. K., Zhang, X., Li, Y. (2016). Traffic accidents involving fatigue driving and their extent of casualties. *Accident Analysis & Prevention*, 87: 34-42. <https://doi.org/10.1016/j.aap.2015.10.033>
- [26] Xiong, X., Chen, L., Liang, J. (2018). Analysis of roadway traffic accidents based on rough sets and Bayesian networks. *Promet-Traffic & Transportation*, 30(1): 71-81. <https://doi.org/10.7307/ptt.v30i1.2502>
- [27] Guo, L., Zhou, J., Dong, S., Zhang, S.C. (2018). Analysis of urban road traffic accidents based on improved K-means algorithm. *China J. Highway Trans*, 31(4): 270-279.
- [28] Wu, X. (2007). *ArcGIS 9 Geographic Information System Application and Practice*. Tsinghua University Press. ISBN 978-7-30215-134-0.
- [29] Ramm, F., Topf, J., Chilton, S. (2011). *OpenStreetMap: using and enhancing the free map of the world*. Cambridge: UIT Cambridge.
- [30] CJJ 37. (2016). *Code for design of urban road engineering*, Chinese Standards Publisher.
- [31] *Urban Road Length 2017*. (2017). Ningbo Housing and Urban-Rural Development Bureau, Ningbo, China. <http://www.nbjs.gov.cn/>.
- [32] Pawlak, Z. (1982). Rough sets. *International Journal of Parallel Programming*, 11(5): 341-356. <https://doi.org/10.1007/BF01001956>
- [33] Mak, B., Munakata, T. (2002). Rule extraction from expert heuristics: A comparative study of rough sets with neural networks and ID3. *European Journal of Operational Research*, 136(1): 212-229. [https://doi.org/10.1016/S0377-2217\(01\)00062-5](https://doi.org/10.1016/S0377-2217(01)00062-5)
- [34] Wong, J.T., Chung, Y.S. (2007). Rough set approach for accident chains exploration. *Accident Analysis & Prevention*, 39(3): 629-637. <https://doi.org/10.1016/j.aap.2006.10.009>
- [35] Greco, S., Matarazzo, B., Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1): 1-47. [https://doi.org/10.1016/S0377-2217\(00\)00167-3](https://doi.org/10.1016/S0377-2217(00)00167-3)
- [36] Akbari, S., Khanzadi, M., Gholamian, M.R. (2018). Building a rough sets-based prediction model for classifying large-scale construction projects based on sustainable success index. *Engineering, Construction and Architectural Management*. <https://doi.org/10.1108/ECAM-05-2016-0110>
- [37] Komorowski, J., Øhrn, A., Skowron, A. (2002). The ROSETTA Rough Set Software System, In *Handbook of Data Mining and Knowledge Discovery*, W. Klösgen and J. Zytkow (eds.), ch. D.2.3, Oxford University Press.

## NOMENCLATURE

$Y$  Decision attribute  
 $X$  Elementary set

$\bar{A}$  Upper approximation  
 $\underline{A}$  Lower approximation  
 $A^*$  Family of all elementary sets