Traffic & Transportation

Analysis of Two-Wheeler Casualty Traffic Accidents with Improved Apriori Algorithm Based on the XGBoost Model

Yansong HU[1], Changjun WANG[2], Jinzi ZHENG[3], Yonggang ZHANG[4], Tianya DU[5]

Original Scientific Paper
Submitted: 1 May 2025
Accepted: 25 Sep 2025
Published: 27 May 2026

[1] huyansong12369@163.com, School of Public Security and Traffic Management, Guangdong Police College, Guangzhou, China; School of Traffic Management, People's Public Security University of China, Beijing, China
[2] Corresponding author, wcj121@sina.com, Research Institute for Road Safety of MPS, Beijing, China
[3] jinzi_zheng@sina.com, Research Institute for Road Safety of MPS, Beijing, China
[4] zysteel@163.com, School of Public Security and Traffic Management, Guangdong Police College, Guangzhou, China
[5] dty_525 @163.com, School of Public Security and Traffic Management, Guangdong Police College, Guangzhou, China

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

Accurately identifying the significant factors influencing the severity of two-wheeler casualty accidents and mining association rules between multi-factor combinations and accident outcomes is crucial for implementing targeted prevention strategies. This study integrates machine learning and an improved Apriori algorithm to analyse 5,032 two-wheeler casualty accidents. Initial factors were identified through feature screening, feature fusion and K-means clustering, and then ranked by importance using the XGBoost model, resulting in ten significant factors for association rule mining. Accident severity weights were calculated based on equivalent minor injury values, and an improved Apriori algorithm, incorporating weighted support and constraints on consequent terms, was applied to extract association rules for both major and general accidents. Results show that combinations of factors such as lorries, two-wheeler riders aged 59 years or older, segregated roads, asphalt structures, electric bicycles, summer and lack of helmet use are strongly associated with major accidents. Conversely, combinations involving passenger vehicles, motorcycles, two-wheeler riders aged 30 to 43 years, non-segregated roads, cement structures and spring are associated with general accidents. Among these factors, the collision object has the greatest impact on accident severity. Minimising spatial and temporal conflicts between two-wheelers and other vehicles is essential to reducing two-wheeler casualties.

KEYWORDS

traffic safety; two-wheeler casualty traffic accident; association rule algorithm; XGBoost; weighted support.

1. INTRODUCTION

Road traffic accidents have become a leading cause of unnatural deaths. In 2021, approximately 1.19 million people worldwide lost their lives in road traffic accidents [1], making traffic safety a critical global concern. In China, reducing accident frequency and preventing severe accidents have become the primary objectives of traffic safety initiatives. However, the involvement of diverse vehicle types contributes to the high heterogeneity of traffic accidents, making it difficult to extract actionable insights from the complete volume of accident data for refined safety governance.

Two-wheelers, including motorcycles, mopeds, electric bicycles and bicycles, are widely used for short-distance travel due to their affordability, convenience and practicality [2, 3]. However, their low mass, simple structure, limited passive safety and high susceptibility to loss of control contribute to frequent violations, such as speeding and unsafe lane changes, which increase the risk and severity of traffic accidents [4]. In 2022, fatalities among two-wheeler riders in China accounted for approximately one-quarter of all traffic-related deaths [5]. Therefore, an in-depth analysis of the characteristics of two-wheeler casualty accidents, as well as the relationship between influencing factors and the severity of accidents, is crucial for effective prevention and traffic safety improvements.

Traffic accidents typically result from an interaction of factors related to people, vehicles, roads, the environment and management [6]. These factors have varying degrees of influence on accident severity. Therefore, it is essential to scientifically quantify the importance of different influencing factors, identify and extract the key factors, and construct a representative feature set for accident analysis. This forms the foundation for mining meaningful association rules between multi-factor combinations and accident outcomes. Moreover, traffic accidents exhibit a typical pyramid distribution in severity: the more severe the accident, the less frequently it occurs [7]. As a result, quickly and accurately extracting meaningful association rules from datasets with a low proportion of major accidents holds substantial practical and application value for traffic safety management.

To address these challenges, this study focuses on two-wheeler casualty traffic accidents as the research context. Feature engineering methods are used for data preprocessing and initial factor extraction, and the XGBoost model is applied to identify the key factors influencing accident severity, thereby constructing a feature set for association rule mining. In addition, the Apriori algorithm is improved by incorporating weighted support and constraints on consequent terms, which reduces redundant rule generation and improves the accuracy and computational efficiency of the mining process. This enables the rapid extraction of strong association rules for major casualty accidents, which are a key focus of traffic safety management but occur less frequently. The resulting strong association rules provide a scientific basis for targeted interventions and precision management of two-wheeler traffic accidents.

2. LITERATURE REVIEW

To accurately identify the key factors influencing the severity of two-wheeler casualty traffic accidents and extract their potential association rules, this study reviews existing research on influencing factor analysis and association rule mining in the context of traffic accidents.

2.1 Research on the analysis of factors influencing the accident

As vulnerable road users in traffic systems, two-wheelers are at high risk of accidents due to their limited protection and poor stability. The occurrence and severity of two-wheeler accidents are influenced by a variety of factors, including rider behaviours, road conditions and weather. Albalate et al. identified gender, speeding, road width and alcohol consumption as key determinants of injury severity in powered two-wheeler accidents [8]. Moskal et al. demonstrated that, for both moped and motorcycle riders, being male, non-helmeted, exceeding the legal blood alcohol limit and riding for leisure purposes significantly increase the risk of traffic accidents [9]. In a multivariate probabilistic analysis of two-motorcycle collisions, Schneider IV et al. found that younger age, alcohol consumption, lack of insurance, non-helmeted riding and riding at night substantially elevated accident risk [10]. Hertach et al. revealed that crash time, gender, road characteristics and riding behaviours are closely associated with injury severity [11]. Yang et al. identified collision speed, travel direction and visual obstructions as critical factors contributing to severe e-bike crash outcomes [12]. Panwinkler et al. emphasised that high blood alcohol concentration, excessive cycling speed, limited riding experience, bicycle skidding and downhill segments were major contributors to serious injuries in pedal electric cycle accidents [13]. Additionally, Brown et al. concluded that human-related factors such as observation errors, distraction and vehicle control loss were primary causes of e-bike and bicycle accidents [14]. In summary, previous studies have consistently identified that key factors influencing the severity of two-wheeler traffic accidents include rider characteristics and behaviours, road environment conditions, characteristics of the collision object and temporal variables. Among them, elderly riders, speeding, drunk driving, non-helmeted riders and collisions with large vehicles are widely recognised as critical contributors to increased accident severity.

To effectively identify the influencing factors of two-wheeler traffic accidents and explore their relationship with accident outcomes, both traditional statistical and machine learning methods have been widely adopted [15]. Traditional statistical analysis methods are characterised by simplicity in operation, stability in model parameters and interpretability of results [16]. The importance of influencing factors can be typically evaluated based on the magnitude of correlation or regression coefficients. Discrete choice models based on statistical methods, including logit, probit, as well as their improved and combined variants, have been proven effective for analysing factors and mechanisms underlying accident severity [17-19]. However, traditional statistical methods are limited in handling nonlinear relationships among multiple factors. Moreover, they require large amounts of high-quality, well-distributed data and typically assume data homogeneity [20]. Although data heterogeneity and factor interactions can be partially addressed by adding variables, incorporating stochastic parameters and integrating models [21-23], feature extraction and model selection mainly depend on researchers' expertise [16]. In practice, numerous factors influence traffic accidents, and complex interactions often exist among them.

Data-driven machine learning methods can discover and construct complex relationships within the data, automatically perform feature extraction, and account for intrinsic connections among variables. This reduces reliance on researchers' empirical knowledge and offers significant advantages in dealing with nonlinear and high-dimensional traffic accident datasets [24]. These approaches are widely used to explore the underlying relationships between contributing factors and accident outcomes. Bayesian networks [25], gradient boosting decision trees (GBDT) [26], C4.5 decision trees [27], random forest [28], XGBoost [29] and others can effectively quantify the importance of individual influencing factors on accident severity and rank their importance. However, the results often lack interpretability and do not directly quantify the importance of interactions between variables [30]. To address this, recent studies have applied machine learning models to identify key variables for use in factor interaction (or coupling) analysis [31]. Specifically, methods such as tree-based Bayesian network structures [32] and feature combination approaches [33] have been proposed to reveal complex relationships between variables while reducing dependence on prior expert knowledge. Nevertheless, increasing the number of variables typically leads to higher computational costs and reduced interpretability, which is why most existing studies focus only on pairwise factor interactions.

While prior studies have provided insights into the effects of individual factors or limited factor pairings, the multifaceted nature of traffic accidents demands that decision-makers consider more comprehensive, multi-factor scenarios. Therefore, mining association rules that reflect interactions among multiple factors and their relationship with accident outcomes can offer valuable, scenario-specific guidance for improving the precision and effectiveness of two-wheeler accident prevention and control strategies.

2.2 Research on the association rule mining algorithm

As an unsupervised data mining approach, association rule algorithms do not depend on the stringent assumptions of parametric models and can still perform robustly even when dealing with missing data [34]. By applying user-defined thresholds, these algorithms identify frequent item sets in accident datasets to extract potential associations between influencing factors and accident severity, thus revealing typical accident patterns. Typical association rule algorithms include the Apriori algorithm, FP-Growth algorithm and Eclat algorithm [35]. These algorithms are capable of processing both large-scale and small-scale datasets [36]. They have effectively mined potential association rules in a dataset of 63,325 traffic accidents [37] and have also shown good performance on datasets with as few as 126 major traffic accidents [38]. This flexibility makes it particularly suitable for refined traffic accident analysis. Association rule mining has been widely applied to analyse traffic accidents involving different traffic subjects, such as hazardous material transport vehicles [39], motorcycles [36], buses [40], lorries [41], electric two-wheelers [42] and pedestrians [43]. Based on the association rules between the set of influencing factors and accident outcomes, the severity of accidents can simply be reduced by reducing or eliminating at least one of the chains of factors obtained for each severity level [40]. However, traditional association rule algorithms tend to generate a large number of invalid or redundant rules when applied to multi-featured or high-dimensional accident datasets, which seriously affects the computational efficiency. As the number of influencing features increases, the algorithm's performance deteriorates significantly. To address this, data preprocessing techniques, such as feature screening, feature fusion, K-means clustering and others, are often employed to reduce the data dimensionality. In addition, constraints on the consequent terms of association rules and other engineering techniques [37] can further quickly and efficiently identify the association rules of interest to decision makers. Additionally, numerous studies have focused on enhancing and optimising association rule algorithms through approaches, such as database compression [44], fewer database scans [45] and refinement of pruning strategies to improve efficiency and practical applicability of the algorithms [46].

Traditional association rule algorithms typically assume that all transactions carry equal importance [47], ignoring the variability in decision-makers' attention to different types of accidents. In reality, major accidents, though relatively rare, often receive disproportionate attention from traffic safety authorities. Furthermore, the support-confidence framework used in classical association rule mining often requires repeated manual tuning of thresholds by users [34, 40, 48]. This may even result in the failure to extract meaningful rules for major accidents due to their low frequency in the dataset.

3. METHODS

In order to accurately identify the key factors influencing the severity of two-wheeler casualty traffic accidents and effectively mine their potential association rules, this study proposes a comprehensive analytical framework. Taking two-wheeler casualty traffic accident data as the basis, the framework constructs the initial set of influencing factors through feature engineering techniques and develops an improved Apriori algorithm with weighted support based on the XGBoost model. The entire framework consists of four main components: 1) input the classification of accident severity and the initial set of influencing factors obtained through feature engineering methods, such as feature screening, feature fusion and K-means clustering; 2) use the XGBoost model to quantify the importance of these initial influencing factors and extract the key ones for association rule mining; 3) improve the Apriori algorithm by incorporating weighted support and constraints on consequent terms to facilitate the efficient and accurate extraction of potential association rules; 4) output the ranking of influencing factors by importance and the effective strong association rules, and perform the interpretation. The proposed integrated analysis framework (IAF) is shown in *Figure 1*.

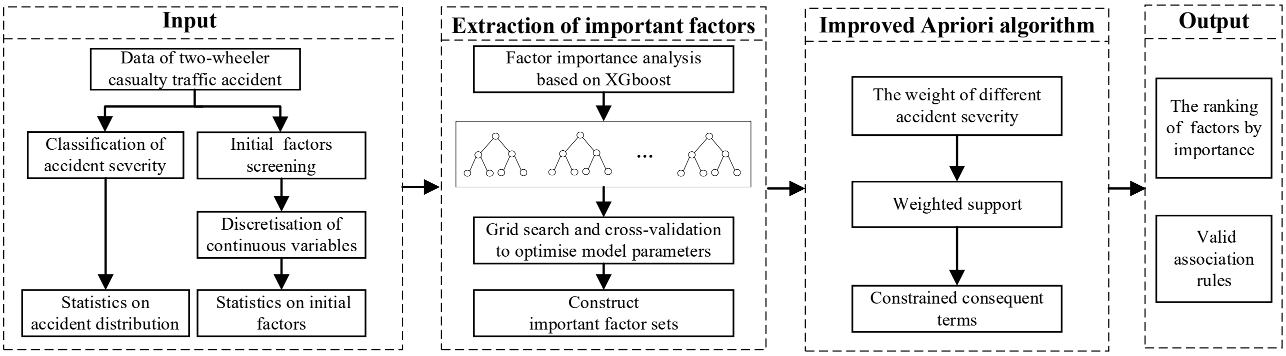

Figure 1 – Integrated analysis framework

3.1 XGBoost for ranking factors

The composition of accident-related factors forms the foundation for association rule mining. To effectively identify the key factors influencing accident severity, reduce data dimensionality and simplify model complexity, the XGBoost model is utilised to quantify the importance of each influencing factor. Subsequently, non-critical factors are eliminated to reduce redundant association rules and increase the computational efficiency of association rule algorithms.

XGBoost is an integrated machine learning model based on the boosting tree algorithm. Built on the gradient boosting decision tree framework, it utilises a loss function expanded using the second-order Taylor series and introduces a regularisation term into the objective function, which effectively prevents overfitting. Through engineering optimisation, XGBoost supports parallel computing, efficiently handles missing data and delivers strong predictive performance with high computational efficiency [49]. During the construction of the optimal decision trees, XGBoost selects features for node splitting based on their importance scores, enabling the ranking of influencing factors.

The importance of features in the XGBoost model can be quantified by the following three metrics:

1) Frequency: The number of times a feature is used to split the data across all trees. This metric is sensitive to the number of categories a feature contains. Multi-category features tend to divide the data space more extensively than features with fewer categories, which may lead to the underestimation of less-categorised but impactful features.
2) Average gain: The average gain in information resulting from splits that use the feature. This metric reflects how much the feature improves the model, but can cause large disparities between high- and low-importance features. In imbalanced datasets, it may also introduce bias in the final ranking.

3) Average coverage: The average number of samples affected by the splits involving the feature. This metric helps reduce bias from differences in feature categorisation and is unaffected by the magnitude of the objective function. It is thus suitable for evaluating features with a small number of categories in classification problems.

Given the low proportion of major accidents in two-wheeler casualty data, the variation in factor categories and imbalanced category distributions, the average coverage metric is adopted to ensure reliable importance ranking. The XGBoost model is further optimised using grid search and cross-validation.

3.2 Improved Apriori algorithm

The Apriori algorithm is a classical method for mining frequent itemsets. It identifies frequent itemsets by iteratively generating candidates, calculating support and pruning, thereby enabling the extraction of potential association rules. The algorithm is widely adopted due to its conceptual simplicity, ease of understanding and better scalability [39]. However, the classical Apriori algorithm relies on a predefined minimum support threshold, which may result in the failure to extract less frequent but practically significant rules in the traffic accident dataset.

Accident-weighted support

Although major traffic accidents occur at a low frequency, their consequences are severe, and their social impact is substantial. Major traffic accidents are the focus of traffic safety management departments and are crucial for preventing and controlling accident risks. To effectively extract the association rules of major accidents involving two-wheelers, this study comprehensively combines casualties from accidents of varying severities. It uses the following equivalencies: 1 fatality equals 3 serious injuries, 1 serious injury equals 3 minor injuries, and calculates the equivalent minor injury values for different accident severities, which are then used to determine the weighting coefficient of each accident.

The numbers of fatalities, serious injuries and minor injuries caused by accidents involving two-wheelers are counted and converted into equivalent minor injuries according to *Equation 1*

$$\bar{M}_i = 9 \cdot F_i + 3 \cdot S_i + M_i \tag{1}$$

where $\bar{M}_i$ denotes the total number of equivalent minor injuries caused by accidents of the severity i ; F_i , S_i and M_i denote the total number of fatalities, serious injuries and minor injuries caused by accidents of the severity i, respectively.

Based on the total number and equivalent minor injuries for accidents of different severities, the average equivalent minor injuries per accident of each severity level are calculated as follows:

$$\bar{W}_i = \frac{\bar{M}_i}{N_i} \tag{2}$$

where $\bar{W}_i$ denotes the average equivalent minor injuries per individual accident of severity i; N_i denotes the total number for accidents of severity i.

Based on the equivalent minor injuries, the weighting coefficients for each severity level of accident can be established as $W = \{w_1, w_2, \cdots, w_n\}$. For each accident in the dataset $D = \{T_1, T_2, \cdots, T_n\}$, which contains n casualty traffic accidents involving two-wheelers, the corresponding coefficient can be determined. The weighted support is then defined as shown in *Equations 3*:

$$W\_Sup(X) = \frac{\sum_{j=1}^{n} w_i \cdot I_j(X)}{\sum_{j=1}^{n} w_j} \tag{3}$$

$$I_j(X) = \begin{cases} 0 & X \notin T_j \\ 1 & X \in T_j \end{cases} \tag{4}$$

where $W\_Sup(X)$ denotes the weighted support of the frequent itemset X, the numerator denotes the sum of the weights of the accidents T_j that contain the itemset X; the denominator denotes the sum of the weights of all accidents within the dataset.

Apriori algorithm with constrained consequent terms

Assuming that X and Y are subsets of the itemset C, where $X \subseteq C, Y \subseteq C$, and $Y \cap Y = \emptyset$, the association rule between the two itemset can be expressed as $Rule\ X \Rightarrow Y$, where X is the antecedent and Y is the consequent. This rule indicates that when X occurs in a transaction, Y is also likely to occur. The generation and evaluation of association rules are primarily achieved using three metrics: weighted support, confidence and lift.

(1) Weighted support (W_Sup) is the probability that a transaction containing both the antecedent X and the consequent Y appears in the entire dataset C, which can be expressed as:

$$W\_Sup(X \Rightarrow Y) = P(XY) = \frac{\sum_{j=1}^{n} w_j \cdot I_j(X \cap Y)}{\sum_{j=1}^{n} w_j} \tag{5}$$

(2) Confidence ($Conf$) is the conditional probability that the consequent Y occurs given that the antecedent X has occurred. A higher confidence value indicates greater reliability of the association rules, which can be expressed as:

$$Conf(X => Y) = P(Y|X) = \frac{W\_Sup(X \Rightarrow Y)}{W\_Sup\,(X)} \tag{6}$$

(3) Lift ($Lift$) refers to the ratio of the conditional probability of the consequent Y occurring given the antecedent X, to the unconditional probability of Y occurring. It reflects the degree of enhancement of the antecedent X to the likelihood of the occurrence of the consequent Y. A higher lift value indicates a more significant enhancement of X to Y, which can be expressed as:

$$Lift(X => Y) = \frac{P(Y|X)}{P(Y)} = \frac{W\_Sup(X \Rightarrow Y)}{W\_Sup(X) \cdot W\_Sup(X)} \tag{7}$$

The minimum weighted support (Min_W_Sup) and minimum confidence (Min_Conf) are used to filter out the rules with lower frequency and weaker validity, respectively. The lift is used to evaluate the rules, and when $Lift > 1$, the association rules are considered meaningful [37, 39].

The traditional Apriori algorithm performs association rule mining without any constraints on the antecedent and consequent terms. This can lead to computational inefficiency and the generation of redundant or irrelevant rules, such as *Rule* {Male, Spring} =>{Sunny}, which lacks practical value in accident analysis. Therefore, to extract association rules between the influencing factors and the severity of accidents, the constraint on the consequent term is set as the severity of accidents. Additionally, a threshold of $Lift > 1$ is used to select the meaningful association rules. This modification enhances the computational efficiency of the Apriori algorithm and reduces the generation of redundant and irrelevant rules. The flowchart of the improved Apriori algorithm is shown in *Figure 2*, and the specific steps are as follows.

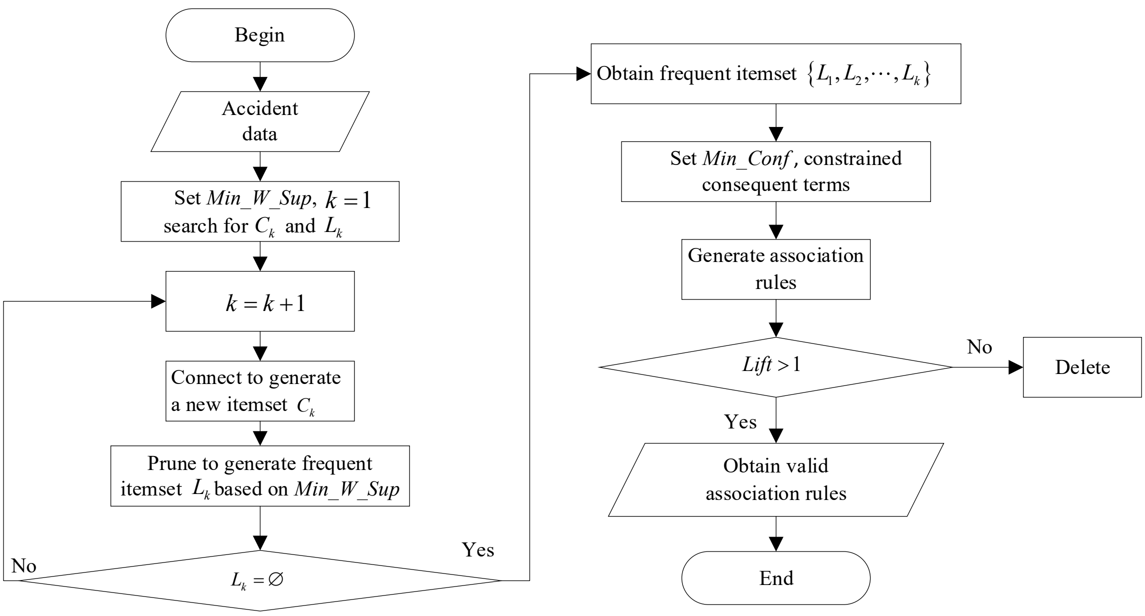

Figure 2 – Flowchart of the improved Apriori algorithm

Step 1: Set the minimum support Min_{W_Sup} threshold and search for the frequent itemset. Iterate through the accident datasets, count the occurrence frequency of different feature values for each factor to obtain the initial itemset C_1, and identify the initial frequent itemset L_1 that meets the Min_{W_Sup} threshold. Based on the frequent itemset L_k, generate a new candidate itemset C_{k+1} through self-joining, and prune it using the Min_{W_Sup} threshold to obtain a higher-dimension frequent itemset L_{k+1}. Continue this process until no larger frequent itemset can be generated.

Step 2: Set the minimum confidence Min_{Conf} threshold and apply constraints on the consequent terms to mine target association rules. Among the frequent itemset $\{L_1, L_2, \cdots, L_k\}$ that meet the Min_{W_Sup} threshold, generate association rules whose consequent terms are limited to accident severity, and retain those rules that meet the Min_{Conf} threshold.

Step 3: Apply the condition $Lift > 1$ to filter meaningful association rules.

4. DATA PROCESS AND DESCRIPTIVE STATISTICS

4.1 Data collection

The dataset used in this study comprises 11,130 road traffic accidents recorded by the public security traffic police in a city in southern China between 2021 and 2023. From this dataset, 5,032 traffic accidents involving two-wheelers and resulting in casualties (including minor injuries, serious injuries or fatalities) were extracted for analysis. The casualty statistics for these accidents are presented in *Table 1*.

Table 1 – Accident statistics

Typology	Number of casualty traffic accidents	Minor injuries	Serious injuries	Fatalities
Total	6,161	5,702	741	829
Involving two-wheelers	5,032	4,726	607	584
Percentage (%)	81.68	82.88	81.92	70.45

According to the data in *Table 1*, casualty accidents involving two-wheelers account for 81.68% of all casualty accidents. These accidents have a high proportion of minor injuries, serious injuries and fatalities, making them the main source of casualties in road traffic accidents.

4.2 Classification of accident severity

In China, the public security traffic police classify the severity of traffic accidents primarily based on the number and severity of casualties involved. Following this principle, traffic accidents in this study are categorised into two levels: general accidents and major accidents. General accidents refer to cases involving 1–2 serious injuries or minor injuries, while major accidents are defined as those resulting in fatalities or three or more serious injuries. The classification criteria and the statistical distribution of accident severity in the dataset are summarised in *Table 2*. Each accident is assigned exclusively to one severity level, and the two levels are mutually exclusive.

Table 2 – Criteria for classifying the severity of accidents and statistical results

Accident severity	Percentage (%)	Guidelines for the classification of severity
General accident	88.65	1-2 serious injuries, or minor injuries
Major accident	11.35	Fatality, or 3 or more serious injuries

The distribution of accidents of different severities by month, day of the week and hour of occurrence is shown in *Figure 3*. March records the highest number of casualty accidents but the lowest proportion of major accidents, while December has the lowest number of casualty accidents. Notably, November has the highest proportion of major accidents at 16.26%. From January to June, the proportion of major accidents remains below the average level, whereas from July to December, it exceeds the average. The number of casualty accidents and the proportion of major accidents are relatively evenly distributed across the days of the week, with no significant difference observed. *Figure 3* shows that the frequency of casualty accidents is higher during morning, midday, and evening peak hours, with the highest number occurring between 7:00 and 8:00 a.m. Although the overall number of casualty accidents is lower during dawn, the proportion of major accidents is notably higher, peaking at 20.34% between 5:00 and 6:00 a.m.

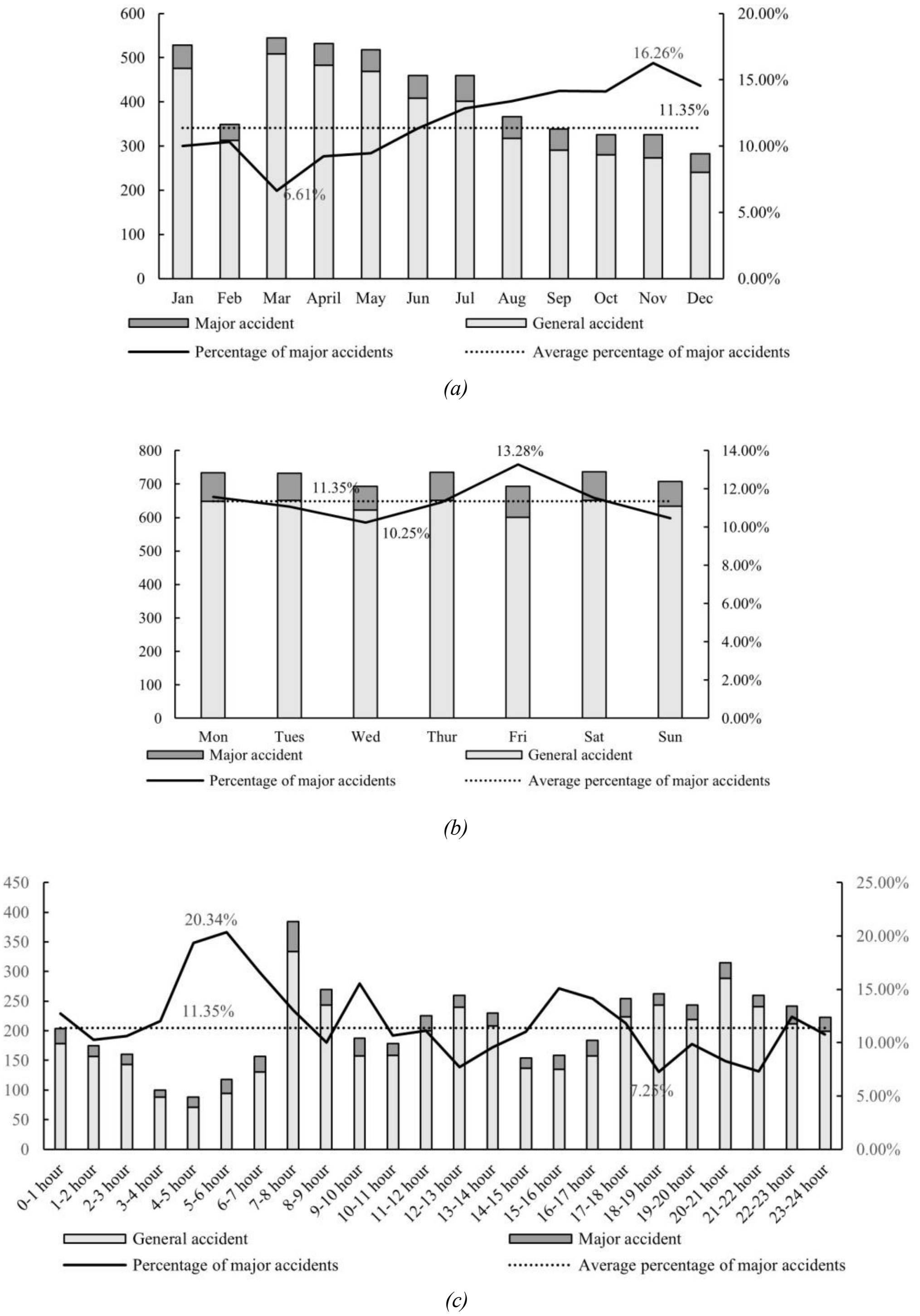

Figure 3 – Distribution of traffic accidents: a) Monthly distribution of traffic accidents; b) Weekly distribution of traffic accidents; c) Hourly distribution of traffic accidents

4.3 Accident influencing factors

With the effective implementation of the public safety industry standards for road traffic accident information recording in China, the comprehensiveness of data recording for road traffic accidents handled through general procedures has been significantly improved, which provides a solid data foundation for mining association rules in two-wheeler casualty accidents. Raw data on two-wheeler casualty accidents are processed through data checking (detecting missing data and identifying outliers), data cleaning (deleting incorrect or inconsistent data) and data standardisation. Based on the structure of two-wheeler traffic accident records, the initial influencing factors were identified with a focus on two-wheelers as the accident subject. Firstly, data irrelevant to the analysis, such as administrative area, case unit, case number, ID number, name, etc., were removed. Then, twelve potential influencing factors related to accident severity were extracted from four categories as follows: time, spatial environment, crash subject and two-wheeler rider attributes. These factors include: month, day of the week, time of day, weather, road structure, junction or section, with segregation or not, type of two-wheeler, collision object, gender, age of the rider and helmet use.

Among the initial influencing factors, the age of two-wheeler riders is a continuous variable, while month, day of the week and time of day are multi-categorical variables. The distribution of limited data samples in continuous or high-dimensional spaces tends to be sparse and uneven, which may lead to insufficient feature learning and overfitting risk [33]. Therefore, it is necessary to discretise the continuous variable and fuse the multi-categorical ones. In order to improve the scientific rigour of the classification, K-means clustering is utilised to discretise the age of two-wheeler riders. The elbow method is employed to determine the optimal number of clusters by identifying the inflexion point where the sum of squared errors sharply decreases. The resulting age classification is shown in *Figure 4*. Additionally, to focus on prevention and control efforts, and in consideration of the needs of the traffic safety management department and expert knowledge, the three time-related factors are processed through feature classification and fusion. The final classification and statistics of the 12 initial influencing factors are summarised in *Table 3*.

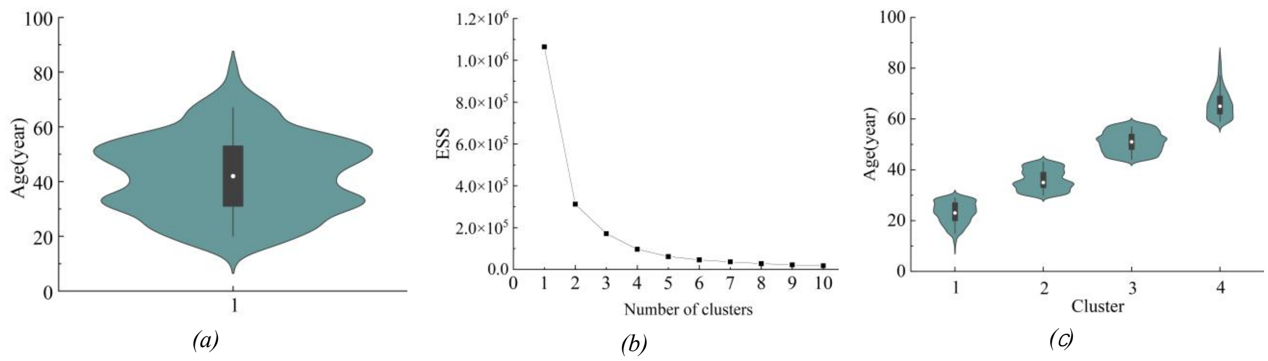

Figure 4 – Classification of age: a) distribution of age; b) elbow diagram of clustering; c) result of age clustering

Table 3 – Classification and distribution of features

Aspect	Factors	Classification of features and percentage (%)
Time	Season	①Spring (Mar-May) (31.70) ; ②Summer (Jun-Aug) (25.56); ③Autumn (Sep-Nov) (19.69); ④Winter (Dec-Feb) (23.05)
	Day of the week	① Weekday (71.28); ② Weekend (28.72)
	Time of day	①Dawn (0:00-6:00) (16.79); ②Morning (6:00-12:00) (27.84); ③ Afternoon (12:00-18:00) (24.66); ④ Evening (18:00-24:00) (30.70)
Spatial environment	Weather	① Sunny (82.97); ② Cloudy (9.46); ③ Rainy (7.57)
	Road structure	① Cement (57.15); ② Asphalt and others (42.85)
	Junction or section	① Road section (82.79); ② Road junction (17.21)
	With segregation or not	① With no segregation (61.98); ② With segregation (38.02)
Crash subject	Type of two-wheeler	① Motorcycle (45.79); ② Moped (24.68) ; ③ Electric bicycle (25.52); ④ Bicycle (4.01)
	Collision object	① Passenger vehicle (46.38); ② Lorry (20.17); ③ Two-wheeler (9.98); ④ Pedestrian (7.25); ⑤ Fixtures and others (16.22)
Two-wheeler riders	Gender	① Male (74.94); ② Female (25.06)
	Age	①Age ≤29 (21.28); ②Age from 30 to 43 (30.43); ③Age from 44 to 58 (35.14); ④ Age ≥59 (13.16)
	Helmet use	① Helmeted (67.51); ② Non-helmeted (30.35); ③ Unknown (2.15)

5. RESULTS AND DISCUSSION

5.1 Ranking of influencing factors by importance

The XGBoost model is optimised using grid search and combined with ten-fold cross-validation, and the final hyperparameters of the XGBoost model are summarised in *Table 4*. The importance scores of the 12 initial influencing factors are obtained by repeating the computation five times to ensure stability and consistency. The ranking results are presented in *Figure 5*.

Table 4 – Hyperparameters of the XGBoost

Parameter	Value	Description
n_estimators	100	Number of trees in the ensemble
max_depth	7	Maximum depth of a tree
learning_rate	0.01	Step size shrinkage used to prevent overfitting
subsample	1	Subsample ratio of the training instances
colsample_bytree	1	Subsample ratio of columns when constructing trees
reg-alpha	0	L1 regularisation term on weights
reg-lambda	1	L2 regularisation term on weights
min_child_weight	1	Minimum sum of instance weights in a child
importance type	cover	Type of feature importance calculation

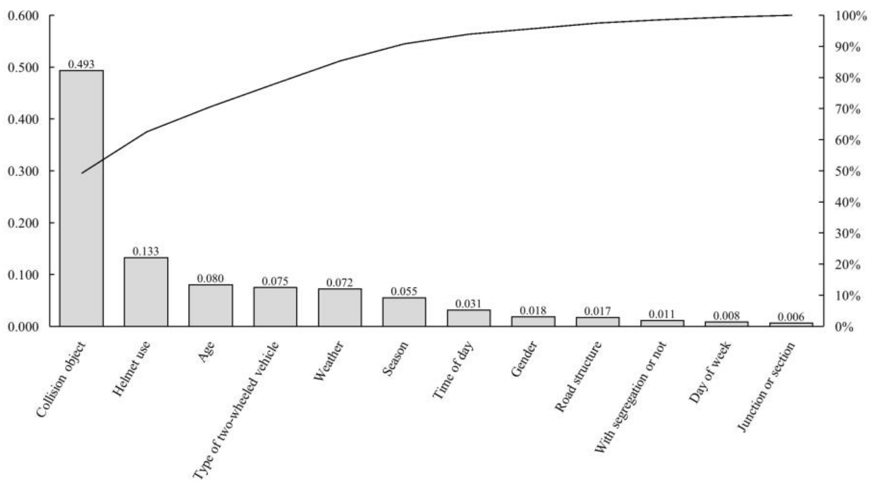

Figure 5 – Factors ranked by importance

Figure 5 shows that the top three factors ranked by importance for the severity of two-wheeler accidents are: collision object, helmet use and age. The collision object has the greatest impact on accident severity, with an importance value of 49.3%. Significant differences in structural dimensions, mass and passive safety performance among various collision objects involved in two-wheeler accidents tend to exacerbate the severity of such accidents [14]. In contrast, day of the week and junction or section contribute less than 1% in importance. Based on the ranking results, factors with importance values below 1% are excluded. The top 10 influencing factors are retained for association rule mining. These top factors account for a cumulative importance of over 98% and include: collision object, helmet use, age, type of two-wheeler, weather, season, time of day, gender, road structure and with separation or not.

5.2 Association rules analysis

Based on the dataset of two-wheeler casualty traffic accidents and *Equation 1*, the total number of equivalent minor injuries is calculated for each accident severity type. Subsequently, the average equivalent minor injuries per accident are derived using *Equation 2*. Based on these values, the weighting coefficient for general accidents is standardised to 1, serving as the baseline. The corresponding weighting coefficient for major accidents is then calculated accordingly, as presented in *Table 5*.

Table 5 – Weighting coefficients of different accidents

Type of accident	Number of traffic accidents	Fatalities	Serious injuries	Minor injuries	Total equivalent minor injuries	Average equivalent minor injuries	Weighting coefficient
Major accident	571	584	15	86	5,387	9.43	6.5
General accident	4,461	0	592	4,640	6,416	1.44	1.0

In order to guarantee the reliability of the association rules, the minimum confidence threshold is set to $Min\_Conf = 0.7$. Based on the selected set of 10 important influencing factors, both the improved Apriori algorithm and the traditional Apriori algorithm are applied for association rule mining under the same hardware and software environment. The results of the association rules are shown in *Table 6*.

Table 6 – Comparison of association rule mining results with different algorithms

Min_W_Sup	Traditional Apriori		Improved Apriori	
	General accident rules	Major accident rules	General accident rules	Major accident rules
0.1	309	0	7	2
0.05	1,098	0	68	21
0.01	10,632	0	2,061	897
0.005	22,237	0	5,899	3,121

From *Table 6*, it can be observed that the number of association rules increases as Min_W_Sup decreases. However, when Min_W_Sup decreases from 0.1 to 0.005, the traditional Apriori algorithm fails to effectively extract association rules for major accidents. In contrast, the improved Apriori algorithm not only can effectively extract the association rules for major accidents but also increases the proportion of these rules as Min_W_Sup decreases.

To investigate the unique and effective strong association rules for traffic accidents of different severities, 89 association rules are selected from *Table 6* under the conditions of $Min\_W\_Sup = 0.05$ and $Min\_Conf = 0.7$. Specifically, there are 21 association rules for major accidents, including 6 for 2-item sets, 10 for 3-item sets and 5 for 4-item sets. For general accidents, there are 68 association rules, including 1 rule for a 1-item set, 5 for 2-item sets, 26 for 3-item sets, 29 for 4-item sets and 7 for 5-item sets. The statistical results for the antecedent item sets are presented in *Figure 6*. The results indicate that two-wheeler casualty traffic accidents are primarily caused by the interaction of multiple influencing factors. This finding is consistent with the findings of He et al. [50], who emphasised the combined effects of various factors in the occurrence of traffic accidents. Combinations involving two to four factors are more likely to result in major accidents, whereas combinations of three to four factors are more commonly associated with general accidents.

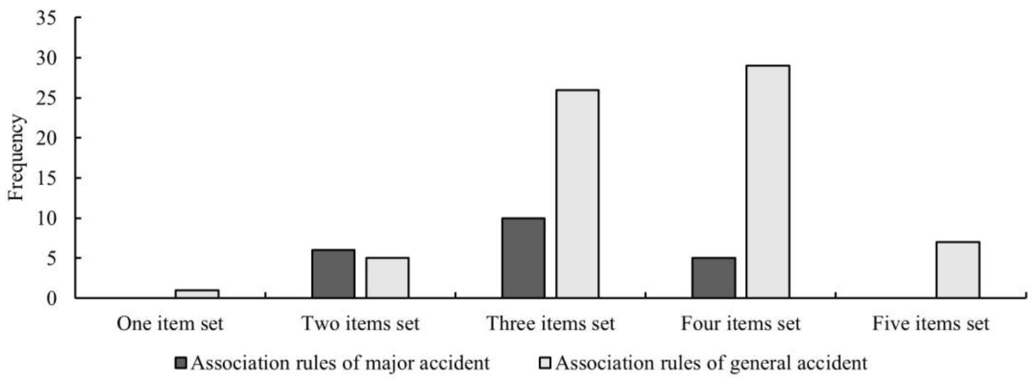

Figure 6 – Statistics of the association rule

To provide a clear and intuitive visualisation of the relationships between antecedents and consequents in the association rules, the matrix of association rules is presented in *Figure 7*.

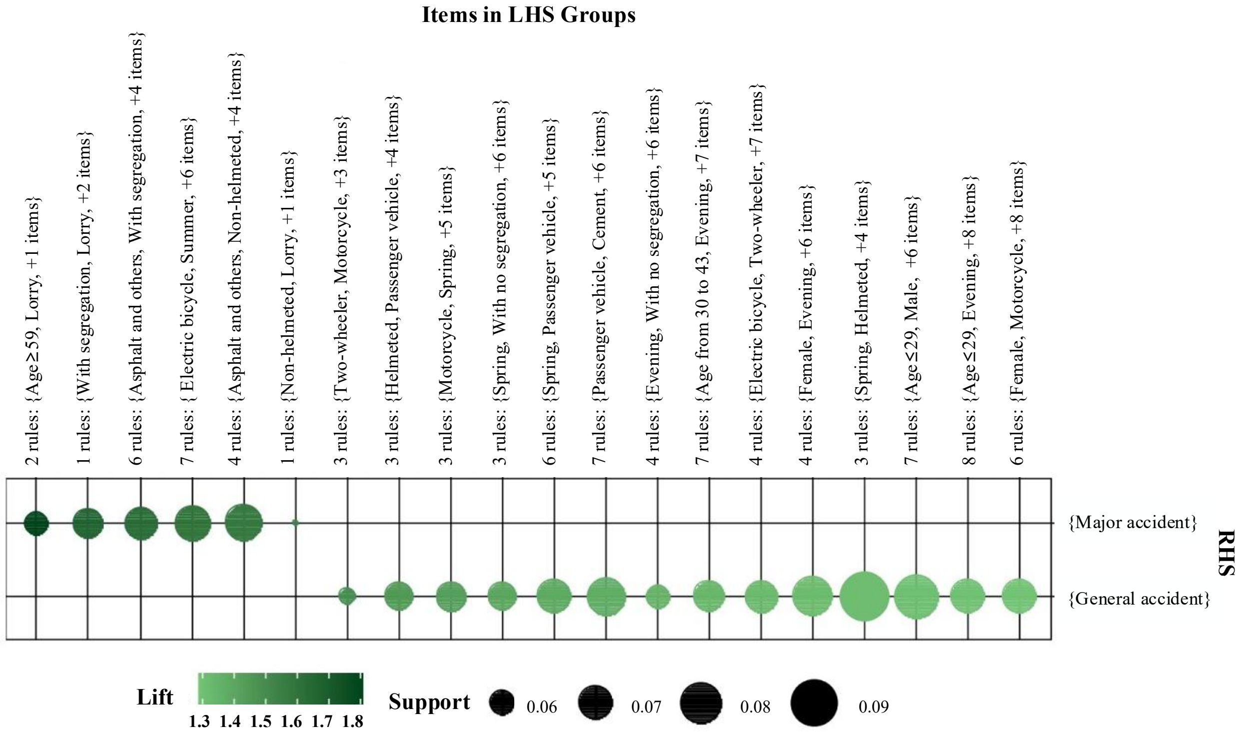

Figure 7 – Matrix of 89 association rules

In *Figure 7*, each dot represents a set of association rules comprising specific antecedent and consequent items. The size of the dot indicates the support, while the colour shade represents the lift, which decreases sequentially from left to right. It can be observed that factors such as lorries, two-wheeler riders aged 59 or older, roads with segregation, asphalt and other road structures, electric bicycles, summer and non-helmeted riders exhibit a stronger association with major accidents. In contrast, passenger vehicles, motorcycles, two-wheeler riders aged between 30 and 43, roads without segregation, cement road structures and spring show stronger associations with general accidents. The effective and strong rules, along with the distinct sets of influencing factors corresponding to different accident severities, can be clearly identified and extracted. These findings provide a practical basis for targeted decision-making in accident prevention and control.

Based on the 89 association rules obtained, the top 10 association rules for major accidents and general accidents, ranked by lift, are extracted and presented in *Tables 7* and *8*.

Table 7 – Top 10 rules ranked by lift of major accident

Rule	*W-Sup*	*Conf*	*Lift*
{age ≥ 59, lorry} => {major accident}	0.064	0.823	1.813
{age ≥ 59, lorry, sunny} => {major accident}	0.054	0.820	1.806
{sunny, male, with segregation, lorry} => {major accident}	0.064	0.755	1.662
{asphalt and others, male, lorry, sunny} => {major accident}	0.062	0.742	1.635
{male, helmeted, with segregation, lorry} => {major accident}	0.050	0.739	1.626
{male, with segregation, lorry} => {major accident}	0.072	0.737	1.623
{sunny, asphalt and others, with segregation, lorry} => {major accident}	0.055	0.737	1.623
{sunny, with segregation, lorry} => {major accident}	0.087	0.736	1.621
{asphalt and others, male, lorry} => {major accident}	0.075	0.733	1.613
{with segregation, lorry} => {major accident}	0.102	0.729	1.604

Table 8 – Top 10 rules ranked by lift values of general accident

Rule	*W-Sup*	*Conf*	*Lift*
{two-wheeler} => {general accident}	0.059	0.814	1.492
{passenger vehicle, cement, motorcycle, sunny} => {general accident}	0.052	0.812	1.488
{two-wheeler, sunny} => {general accident}	0.050	0.808	1.481
{passenger vehicle, helmeted, age from 30 to 43} => {general accident}	0.052	0.792	1.451
{passenger vehicle, cement, male, helmeted, sunny} => {general accident}	0.057	0.792	1.450
{passenger vehicle, helmeted, cement, sunny} => {general accident}	0.081	0.784	1.436
{passenger vehicle, cement, helmeted, sunny, with no segregation} => {general accident}	0.061	0.776	1.421
{passenger vehicle, spring, sunny} => {general accident}	0.071	0.775	1.420
{passenger vehicle, cement, motorcycle} => {general accident}	0.060	0.775	1.420
{passenger vehicle, cement, male, sunny, with no segregation} => {general accident}	0.069	0.763	1.398

Interpreting the first association rule in *Table 7* as an example, in an accident involving a two-wheeler rider aged 59 or older and a lorry, the probability of a major accident is 82.23%, which is 1.813 times higher than the baseline probability of a major accident occurring independently. From *Table 7*, it can be observed that seven influencing factors are involved in the ten strong association rules for major accidents. Notably, the collision object of the lorry appears in all ten rules, totalling ten times. Additionally, factors such as roads with segregation, sunny weather and male riders each appear five times, highlighting their frequent association with serious outcomes. The heightened risk in lorry and two-wheeler collisions can be attributed to the lorry's large structural dimensions, heavy load, high inertia and extensive blind spots [51], which significantly increase the likelihood of serious injuries or fatalities when colliding with smaller, less-protected two-wheelers. This finding aligns with the research of Osman et al., who emphasised that collisions involving lorries are more likely to result in severe injuries [52]. In addition, previous studies have shown that male two-wheeler riders are more prone to engaging in high-speed riding behaviour, and that males are generally at greater risk of cycling accidents [53]. Moreover, on segregated roads during sunny conditions, vehicles typically travel at higher speeds [54], and the increase in kinetic energy at higher speeds further exacerbates the severity of collisions, increasing the risk of serious injuries or fatalities.

To reduce the incidence of major two-wheeler accidents, targeted prevention strategies should prioritise older riders aged 59 and above, who face higher risks due to age-related declines in reaction time and physical resilience [9]. Promoting helmet use among this group can substantially reduce the risk of head injuries, particularly traumatic brain injuries [19]. Infrastructure-based interventions, such as speed control measures and enhanced physical separation between motor vehicles and two-wheelers on segregated roads with asphalt structures, can also mitigate accident severity. In addition, male riders should be identified as a key focus group in traffic safety education campaigns.

According to *Table 8*, ten influencing factors are involved in the ten strong association rules for general accidents. Among these, collision objects of passenger vehicles appear eight times, sunny weather appears seven times, and the road structure of cement appears six times, indicating strong associations with general accidents. Although passenger vehicles often travel at higher speeds compared to lorries, they are equipped with more advanced active safety technologies, which help reduce the likelihood of causing severe casualties in collisions with two-wheelers. In addition to these factors, combinations of seven other factors, including helmeted riders, roads with no segregation, motorcycles, and riders aged 30 to 43, collectively show strong associations with general accidents when paired with the three aforementioned factors. These findings indicate that while these conditions may not typically lead to severe outcomes, they still constitute critical scenarios in which general accidents are likely to occur. Therefore, these factors should be considered in the formulation of targeted strategies aimed at reducing the incidence of general two-wheeler accidents.

6. CONCLUSION

This study identifies and analyses significant factors influencing the severity of two-wheeler traffic accidents by integrating manual empirical techniques with the XGBoost machine learning model. Building on this foundation, an improved Apriori algorithm with weighted support and constraints on consequent terms is developed to mine association rules for multiple factors associated with major accidents and general accidents. These results provide a reference for decision-making in traffic safety management.

Two-wheeler traffic accidents are the primary source of casualties in current road traffic accidents, accounting for over 80% of all injury and fatality cases. Among the various influencing factors, the top three determinants of accident severity are collision object, helmet use and rider age, with collision object having the greatest impact. Combinations of factors, such as collisions with lorries, roads with segregation, male riders and sunny weather, are strongly associated with major accidents. In contrast, combinations of factors, such as collisions with passenger vehicles, cement road structures and sunny weather, show a strong association with general accidents. These findings highlight the importance of spatial and temporal separation between two-wheelers and both lorries and passenger vehicles, as a key strategy for reducing the incidence of both major and general two-wheeler accidents.

From the methodological point of view, integrating manual empirical methods with machine learning methods for extracting influencing factors can significantly improve the scientific validity of the feature set used for mining association rules. The initial influencing factors of concern to decision-makers can be effectively identified through feature screening, feature fusion, K-means clustering and business scenario analysis. The XGBoost model is then applied to rank and extract the most important factors. This hybrid method, which integrates prior knowledge with data-driven techniques, effectively extracts high-impact factors from accident data that are crucial for decision-making. It reduces data complexity and provides a crucial foundation for accurately mining effective association rules.

Another important methodological aspect involves determining the weighting coefficients for various accident severities by calculating the equivalent minor injury values. This enables a differentiated assessment of the level of concern decision-makers assign to different accident severities. Based on this, the Apriori algorithm is improved by incorporating weighted support and constraints on consequent terms, which can address the problem that the traditional Apriori algorithm is difficult to effectively mine association rules for less frequent major accidents.

Nonetheless, the scope of this study is constrained by the limited number of initial influencing factors considered, which may restrict the comprehensiveness of the association rule analysis. Future research could expand the spatial and temporal scope of the dataset and incorporate a broader range of influencing factors to enhance the generalisability of the findings. Furthermore, the current weighting coefficients for accidents are based solely on casualties at the accident severity level. Subsequent studies could explore more nuanced approaches to reflect decision-makers' diverse risk priorities, thereby improving the precision of traffic safety interventions.

ACKNOWLEDGEMENT

The authors appreciate the editors and the anonymous reviewers for their detailed and constructive comments. This work was supported by the Ministry of Public Security of China Research Project Fund (Project number: 2022JSZ16), the Guangdong Police College Research Project Fund (Project number: 2024YBZB03) and the Guangdong Provincial Educational Science Planning Project (Higher Education Special) (Project number: 2025GXJK0472).

REFERENCES

[1] World Health Organization. Global status report on road safety 2021(WHO). *Geneva, Switzerland: WHO Library Cataloguing-in-Publication Data Global*, 2023.

[2] Wang X, et al. Risky behaviors, psychological failures and kinematics in vehicle-to-powered two-wheeler accidents: Results from in-depth Chinese crash data. *Accident Analysis & Prevention*. 2021;156:106150. DOI: 10.1016/j.aap.2021.106150.

[3] Sui B, et l. A clustering approach to developing car-to-two-wheeler test scenarios for the assessment of automated emergency braking in China using in-depth Chinese crash data. *Accident Analysis & Prevention*. 2019;132:105242. DOI: 10.1016/j.aap.2019.07.018.

[4] Karthick L, et al. Vehicle safety system for two wheeler – A critical review. *Materials Today: Proceedings*. 2022; 52:1594-1597. DOI: 10.1016/j.matpr.2021.11.258.

[5] National Bureau of Statistics of China. *China Statistical Year Book*; 2023. https://www.stats.gov.cn/sj/ndsj/2023/indexch.htm [Accessed 12th April 2025].

[6] Tešić M, et al. Identifying the most significant indicators of the total road safety performance index. *Accident Analysis & Prevention*. 2018;113:263-278. DOI: 10.1016/j.aap.2018.02.003.

[7] Alanazi F, et al. Impact of data balancing and feature engineering on accident severity models. *Promet-Traffic&Transportation*. 2025;37(3):665-690. DOI: 10.7307/ptt.v37i3.856.

[8] Albalate D, et al. Motorcycle injury severity in Barcelona: The role of vehicle type and congestion. *Traffic Injury Prevention*. 2010;11(6):623–631. DOI: 10.1080/15389588.2010.506932.

[9] Moskal A, et al. Risk factors for injury accidents among moped and motorcycle riders. *Accident Analysis & Prevention*. 2012;49:5-11. DOI: 10.1016/j.aap.2010.08.021.

[10] Schneider IV, et al. Examination of factors determining fault in two-vehicle motorcycle crashes. *Accident Analysis & Prevention*. 2012; 45:669-676. DOI: 10.1016/j.aap.2011.09.037.

[11] Hertach P, et al. Characteristics of single-vehicle crashes with e-bikes in Switzerland. *Accident Analysis & Prevention*. 2018; 117:232-238. DOI: 10.1016/j.aap.2018.04.021.

[12] Yang N, et al. Analysis of fatal factors influencing accidents involving two-wheel electric vehicle drivers at intersections. *Legal Medicine*. 2020;45:101696. DOI: 10.1016/j.legalmed.2020.101696.

[13] Panwinkler T, et al. Causes of pedelec (pedal electric cycle) single accidents and their influence on injury severity. *Accident Analysis & Prevention*. 2021;154:106082. DOI: 10.1016/j.aap.2021.106082.

[14] Brown L, et al. Investigation of accidents involving powered two wheelers and bicycles–A European in-depth study. *Journal of safety research*. 2021;76:135-145. DOI: 10.1016/j.jsr.2020.12.015.

[15] Ulu M, et al. Bibliometric analysis of traffic accident prediction studies from 2003 to 2023: trends, patterns and future directions. *Promet-Traffic&Transportation*. 2024;36(5):833-851. DOI: 10.7307/ptt.v36i5.576.

[16] Santos K, et al. A literature review of machine learning algorithms for crash injury severity prediction. *Journal of Safety Research*. 2022;80:254-269. DOI: 10.1016/j.jsr.2021.12.007.

[17] Blackman RA, et al. Comparison of moped, scooter and motorcycle crash risk and crash severity. *Accident Analysis & Prevention*. 2013;57:1-9. DOI: 10.1016/j.aap.2013.03.026.

[18] Shaheed MS, et al. A mixed logit analysis of two-vehicle crash severities involving a motorcycle. *Accident Analysis & Prevention*. 2013;61:119-128. DOI: 10.1016/j.aap.2013.05.028.

[19] Santos K, et al. Risk factors associated with the increase of injury severity of powered two wheelers road accidents victims in Portugal. *Traffic Injury Prevention*. 2021;22(8):646-650. DOI: 10.1080/15389588.2021.1987421.

[20] Wei T, et al. Predicting and factor analysis of rider injury severity in two-wheeled motorcycle and vehicle crash accidents based on an interpretable machine learning framework. *Traffic Injury Prevention*. 2024;25(2):194-201. DOI: 10.1080/15389588.2023.2284111.

[21] Chang FR, et al. Crash injury severity analysis of E-Bike Riders: A random parameters generalized ordered probit model with heterogeneity in means. *Safety Science*. 2022;146:105545. DOI: 10.1016/j.ssci.2021.105545.

[22] Zhang XL, et al. Analysis of factors affecting injury severity of shared electric bike riders. *Journal of Southeast University (English Edition)*, 2023; 39(3): 284-291.DOI: 10.3969/j.issn.1003-7985.2023.03.009.

[23] Islam S, et al. A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama. *Accident Analysis & Prevention*. 2017;108:163-171. DOI: 10.1016/j.aap.2017.08.016.

[24] Yang Y, et al. Review of the research toward freeway unconventional traffic accidents. *Journal of Applied Basic and Engineering Sciences*. 2024;32(03):601-626. DOI: 10.16058/j.issn.1005-0930.2024.03.001.

[25] Alkheder S, et al. Risk analysis of traffic accidents' severities: An application of three data mining models. *ISA Transactions*. 2020;106:213-220. DOI: 10.1016/j.isatra.2020.06.018.

[26] Wu WT, et al. Economic development, demographic characteristics, road network and traffic accidents in Zhongshan, China: gradient boosting decision tree model. *Transportmetrica A Transport Science*. 2020; 16(3): 359-387. DOI: 10.1080/23249935.2020.1711543.

[27] Katanalp BY, et al. The novel approaches to classify cyclist accident injury-severity: Hybrid fuzzy decision mechanisms. *Accident Analysis & Prevention*. 2020;144:105590. DOI: 10.1016/j.aap.2020.105590.

[28] Xu RH, et al. Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest. *Safety Science*. 2021;135:105125. DOI: 10.1016/j.ssci.2020.105125.

[29] Parsa AB, et al. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention*. 2020;136:105405. DOI: 10.1016/j.aap.2019.105405.

[30] Jamali DM, et al. Identifying interactions among factors related to death occurred at the scene of traffic accidents: Application of "logic regression" method. *Heliyon*. 2024;10(11):e32469. DOI: 10.1016/j.heliyon.2024.e32469.

[31] Hu LW, et al. Influence factors and coupling relationship of traffic accident injury degree based on a data-driven approach. *Journal of Transportation Systems Engineering and Information Technology*. 2022;22(5):117-124. DOI: 10.16097/j.cnki.1009-6744.2022.05.012.

[32] Guo J, et al. Risk analysis for hazardous chemical vehicle-bridge transportation system: A dynamic Bayesian network model incorporating vehicle dynamics. *Reliability Engineering & System Safety*. 2024;242:109732. DOI: 10.1016/j.ress.2023.109732.

[33] Zhu LP, et al. Analysis of traffic accident causes based on data augmentation and ensemble learning with high-dimensional small-sample data. *Expert Systems with Applications*. 2024;237:121782. DOI: 10.1016/j.eswa.2023.121782.

[34] Das S, et al. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*. 2019;23:30-48. DOI: 10.1080/12265934.2018.1431146.

[35] Wang L, et al. An improved eclat algorithm based association rules mining method for failure status information and remanufacturing machining schemes of retired products. *Procedia CIRP*. 2023;118:572-577. DOI: 10.1016/j.procir.2023.06.098.

[36] Tamakloe R, et al. Factors affecting motorcycle crash casualty severity at signalized and non-signalized intersections in Ghana: Insights from a data mining and binary logit regression approach. *Accident Analysis & Prevention*. 2022;165:106517. DOI: 10.1016/j.aap.2021.1.

[37] Yu S, et al. Identifying Factors that Influence the Patterns of Road Crashes Using Association Rules: A case Study from Wisconsin, United States. *Sustainability*. 2019;11(7):1925. DOI: 10.3390/su11071925.

[38] Xu C, et al. Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. *Journal of Safety Research*. 2018;67:65-75. DOI: 10.1016/j.jsr.2018.09.013.

[39] Hong J, et al. Application of association rules mining algorithm for hazardous materials transportation crashes on expressway. *Accident Analysis & Prevention*. 2020;142:105497. DOI: 10.1016/j.aap.2020.105497.

[40] Samerei SA, et al. Data mining approach to model bus crash severity in Australia. *Journal of Safety Research*. 2021;76:73-82. DOI: 10.1016/j.jsr.2020.12.004.

[41] Hong J, et al. Discovering Insightful Rules among Truck Crash Characteristics using Apriori Algorithm. *Journal of Advanced Transportation*. 2020;2020:1-16. DOI: 10.1155/2020/4323816.

[42] Montella A, et al. A data mining approach to investigate patterns of powered two-wheeler crashes in Spain. *Accident Analysis & Prevention*. 2020;134:105251. DOI: 10.1016/j.aap.2019.07.027.

[43] Das S, et al. Fatal pedestrian crashes at intersections: Trend mining using association rules. *Accident Analysis & Prevention*. 2021;160:106306. DOI: 10.1016/j.aap.2021.106306.

[44] Vo B, et al. A novel approach for mining maximal frequent patterns. *Expert Systems with Applications*. 2017;73:178-186. DOI: 10.1016/j.eswa.2016.12.023.

[45] Liu XB, et al. An improved association rules mining method. *Expert Systems with Applications*. 2012;39(1):1362-1374. DOI: 10.1016/j.eswa.2011.08.018.

[46] Nguyen LTT, et al. Updating mined class association rules for record insertion. *Applied Intelligence*. 2014;42(4):707-721. DOI: 10.1007/s10489-014-0614-1.

[47] Wu WT, et al. A novel software defect prediction approach via weighted classification based on association rule mining. *Engineering Applications of Artificial Intelligence*. 2024;129:107622. DOI: 10.1016/j.engappai.2023.107622.

[48] Besharati MM, et al. Which set of factors contribute to increase the likelihood of pedestrian fatality in road crashes?. *International Journal of Injury Control and Safety Promotion*. 2018;25(3):247-256. DOI: 10.1080/17457300.2017.1363781.

[49] Yang C, et al. The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis. *Accident Analysis & Prevention*. 2021;158:106153. DOI: 10.1016/j.aap.2021.106153.

[50] He M, et al. Road Traffic Accident Prediction based on multi-source data–a systematic review. *Promet-Traffic&Transportation*. 2025;37(2):499-522. DOI: 10.7307/ptt.v37i2.763.

[51] Zhu X, et al. Modeling occupant-level injury severity: An application to large-truck crashes. *Accident Analysis & Prevention*. 2011;43(4):1427-1437. DOI: 10.1016/j.aap.2011.02.021.

[52] Osman M, et al. Analysis of injury severity of large truck crashes in work zones. *Accident Analysis & Prevention*. 2016;97:261-273. DOI: 10.1016/j.aap.2016.10.020.

[53] Beck B, et al. Bicycling crash characteristics: An in-depth crash investigation study. *Accident Analysis & Prevention*. 2016;96:219-227. DOI: 10.1016/j.aap.2016.10.020.

[54] Luburić G, et al. Risk assessment of traffic accidents involving young adult drivers by identification of subjective and objective factors. *Promet-Traffic&Transportation*. 2023,35(1):27-36. DOI: 10.7307/ptt.v35i1.9.