

Article

A Multi-Source Data Fusion Framework for Sustainable Drunk-Driving Risk Governance in Megacities: Evidence from Shanghai

Lina Ge, Jiajia Feng *, Jiarong Wu and Zhixuan Jia

School of Vehicle and Traffic Engineering, Taiyuan University of Science and Technology, No. 66 Waliu Road, Wanhailin District, Taiyuan 030024, China

* Correspondence: fengjiajia@tyust.edu.cn

Abstract

Alcohol-impaired driving in megacities is both a road safety problem and a public safety governance challenge. This study proposes a multi-source data fusion framework for identifying the spatio-temporal risk of publicly recorded drunk-driving violations and translating risk estimates into enforcement-oriented priority areas. Using Shanghai as a case study, we integrate 1814 publicly recorded drunk-driving violations with dining and entertainment POIs, 400–800 m annular POI measures, road networks, metro stations, public parking facilities, meteorological conditions, and temporal attributes. Kernel density estimation and ST-DBSCAN are first used to identify observed spatio-temporal clusters. Time-specific grid-based Logit models are then estimated for daytime, evening, and post-midnight periods, with complementary log-log models used for robustness checks. The results show strong nighttime concentration, with T2 showing the highest predicted risk and weekend scenarios further increasing risks in T2 and T3. In-grid dining and entertainment POI variables are mostly negatively associated with risk, whereas surrounding dining POIs show stable positive associations, indicating a spatial mismatch between drinking venues and road spaces where violations are observed. Road exposure, metro accessibility, parking facilities, and weekend context also display time-varying effects. Finally, integrating model-predicted risk grids with observed clusters produces Priority, Warning, and Background areas, supporting evidence-based, time-specific, and location-specific drunk-driving risk governance in megacities.

Keywords: drunk-driving violations; sustainable urban safety; multi-source data fusion; spatio-temporal risk identification; POI spillover; targeted enforcement

Academic Editor: Victoria Gitelman

Received: 23 June 2026

Revised: 10 July 2026

Accepted: 14 July 2026

Published: 16 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Road traffic safety is an important component of urban sustainable development and public safety governance. The World Health Organization has stated that one of the targets of the United Nations Decade of Action for Road Safety 2021–2030 is to reduce road traffic deaths by half by 2030 [1]. Alcohol-impaired driving is both highly dangerous and illegal. It not only increases the probability of traffic crashes but also aggravates crash consequences. China has long treated the governance of drunk and intoxicated driving as a key task in road traffic safety management. In 2023, the Opinions on Handling Criminal

Cases of Drunk Dangerous Driving further emphasized the need to punish drunk dangerous driving offenses in accordance with the law and to protect people's lives, property, and road traffic safety [2]. For megacities such as Shanghai, drunk-driving governance is also closely related to nighttime traffic operations, public safety management, and the allocation of enforcement resources. According to the Shanghai Statistical Yearbook 2025, Shanghai recorded 430,132 traffic crashes in 2024, resulting in 1054 deaths and 10,903 injuries. In the same year, 20.133 million traffic violations were processed, including 12.078 million motor-vehicle violations [3]. At the same time, Shanghai has highly active nighttime activities and nighttime travel. In 2024, the city recorded 4.03 million nighttime active trips, 2906 bars, and a nighttime bus activity area of 1975 km², ranking among the leading cities in China in several indicators related to the nighttime economy and nighttime transport activity [4]. Therefore, identifying the spatio-temporal risk of drunk-driving violations in a high-density, activity-intensive, and polycentric urban space has important practical significance.

Existing traffic safety studies have explained the formation mechanisms of road risk from the perspectives of traffic violations and risky driving behavior. Traffic violations are significantly associated with crash severity, and driver, vehicle, road, and environmental factors can all affect violation and crash risks [5]. Some violations may also evolve into traffic crashes under specific temporal and spatial conditions [6]. In the broader literature on risky driving, speeding, harsh acceleration, harsh deceleration, red-light running, improper lane changing, and aggressive driving are also considered important sources of road safety risk. Wang et al. found, based on severe traffic violation data, that violations exhibit clear temporal variation and spatial differences, and that meteorological conditions and time of day influence the likelihood of violations [7]. Lee et al. showed, in ride-hailing or taxi operating contexts, that working hours and road conditions affect aggressive driving behaviors such as speeding, harsh starting, harsh braking, lane changing, and overtaking, and that prolonged driving gradually increases the risk of such behaviors [8]. Krahé et al., using representative survey data from Germany, found that aggressive driving increased across different periods and was associated with age, gender, annual mileage, and driving stress [9]. These studies indicate that traffic violations and risky driving behaviors are not determined solely by individual factors, but are also jointly influenced by temporal environments, road conditions, and travel activity characteristics.

In terms of influencing factors, traffic safety studies commonly construct explanatory frameworks from the perspectives of road users, vehicles, road conditions, weather environments, and the built environment. Yan et al. found temporal instability in the determinants of alcohol-impaired driving crashes across different days of the week and time conditions [10]. Zhao et al. introduced geographic environment variables, including street-view information, road networks, land use, and points of interest, into the study of urban traffic violations, demonstrating the explanatory role of the urban spatial environment in traffic violations [11]. Li et al. further examined traffic safety risk from the built-environment perspective and showed that road networks, land use, and spatial environmental variables can reflect spatial differences in traffic safety risk [12]. These studies suggest that the formation of drunk-driving violation risk should be understood by considering both road space and urban activity space.

Compared with general traffic violations, drunk-driving violations have stronger behavioral-context characteristics. Existing studies have more often focused on drunk-driving crashes or alcohol-related crashes, examining the effects of alcohol sales and consumption venues, population density, road networks, and land use on crash risk. Treno et al. provided an early spatial panel analysis of the relationship between alcohol outlet density and alcohol-related traffic crashes [13], and alcohol-related POI density has also been found to have a spatial relationship with drunk-driving crashes [14]. Population density,

road networks, land use, and alcohol-related activity venues can affect the risk and consequences of drunk-driving or alcohol-related crashes [15,16]. Based on fine-scale spatial units in Melbourne, Australia, Morrison et al. found that bar density was associated with alcohol-related road crash risk in adjacent areas, while the origins, destinations, and crash locations of alcohol-related trips may not completely overlap [17]. Gruenewald and Johnson also pointed out, from a traffic-flow perspective, that drinking, driving, and crashes are connected through spatial and movement processes [18]. These studies suggest that drunk-driving risk does not necessarily occur only at the locations of drinking venues, but may also appear on surrounding roads, short-distance travel routes, and return corridors after drinking activities.

However, drunk-driving crashes and drunk-driving violations are not identical. Crash data reflect outcomes that have already resulted in collisions or injuries, whereas drunk-driving violation records reflect illegal behaviors that occur before crashes or do not lead to crashes. Sun et al. have begun to examine the spatial heterogeneity of non-crash drunk-driving events, indicating that relying solely on crash data cannot fully characterize the upstream risk of drunk-driving behavior [19]. For traffic management, the identification of violations has stronger preventive significance, because not all drunk-driving violations eventually develop into crashes, although these behaviors already constitute road safety hazards. Therefore, identifying the spatio-temporal distribution and influencing factors of drunk-driving behavior based on public traffic violation records can provide a reference for proactive governance before crashes occur.

At the application level, traffic safety research has gradually shifted from historical hotspot description toward risk identification and governance translation. Machine learning methods can be used to extract key spatio-temporal features of crash risk hotspots on urban road networks [20], and traffic violation data can also be used to identify high-risk areas and common violation types, supporting enforcement priority assessment [21]. Nevertheless, several gaps remain in the existing literature. First, many studies focus on drunk-driving crashes or alcohol-related crashes, while drunk-driving behaviors in public violation records have received relatively limited attention. Second, although alcohol-related venues and built-environment factors have been discussed, the relationship between drinking activity venues and surrounding road spaces remains insufficiently characterized. Third, most hotspot identification studies remain at the level of spatial clustering or risk prediction, and relatively few studies connect clustering results, time-specific risk models, and enforcement-oriented classification.

Based on these considerations, this study takes publicly recorded drunk-driving violations from the Shanghai Traffic Safety Integrated Service Platform as the research object and integrates multi-source data, including road networks, meteorological data, dining and entertainment POIs, metro stations, and public parking facilities, to develop an analytical framework for drunk-driving violation risk identification. First, kernel density estimation and spatio-temporal density-based clustering are used to identify the historical clustering characteristics of drunk-driving violations. Second, using 400 m regular grids as the basic spatial units, time-specific binary risk models are constructed for daytime, evening, and post-midnight periods to analyze the effects of road conditions, urban activity spaces, public transport accessibility, parking facilities, weather, and weekend factors on drunk-driving violation risk. Third, considering the possible influence of dining and entertainment venues on surrounding road spaces, spatial spillover variables of POIs are further constructed. Finally, historical clustering hotspots are spatially overlaid with model-predicted risk grids to generate enforcement-oriented risk classification results. The objective of this study is not merely to describe the spatial distribution of drunk-driv-

ing violations, but to connect drunk-driving risk identification, influencing-factor analysis, and governance zoning, thereby providing a technical pathway for road safety governance and enforcement resource allocation in megacities.

2. Data and Methods

2.1. Study Area and Data Sources

This study takes Shanghai as the study area. The multi-source data used in this study include drunk-driving violation records, road network data, metro station data, meteorological data, and points of interest (POI) data. The drunk-driving violation data were obtained from publicly available records released by the Shanghai Traffic Safety Integrated Service Platform. The study period covers 1 November 2024 to 14 December 2025. The main fields include violation time, violation location, violation type, and administrative district. This study focuses on the drunk-driving violations publicly recorded during the study period, and their spatial distribution is shown in Figure 1.



Figure 1. Spatial distribution of publicly recorded drunk-driving violations in Shanghai.

The road network data were obtained from the Shanghai Public Data Open Platform. The main fields include road name, road class, road code, and road geometry. The metro station data were also obtained from the Shanghai Public Data Open Platform and mainly include station name, line information, and spatial coordinates.

The meteorological data were obtained from the National Historical Weather Data Platform and include hourly weather conditions, air temperature, rainfall, and snowfall in Shanghai during the study period. The POI data were obtained from the Amap platform. The main fields include POI ID, name, address, category, and longitude and latitude coordinates. Considering activity settings that may be related to drunk-driving behavior, this study extracted dining, entertainment, and public parking facility POIs. Among them, entertainment POIs mainly include bars, nightclubs, KTVs, and similar venues.

2.2. Data Preprocessing

To support subsequent spatio-temporal clustering and grid-based risk modeling, the original multi-source data were cleaned, coordinate systems were unified, and spatial

matching was conducted. For the drunk-driving violation data, duplicate records were first removed, resulting in 1814 valid samples. Core fields, including violation time, violation location, and administrative district, were retained, and temporal attributes such as year, month, hour, and day of the week were extracted from the violation time. For violation locations recorded as text addresses, geocoding was used to convert them into longitude and latitude coordinates under the WGS 1984 coordinate system. The coordinates were then projected to WGS 1984 UTM Zone 51N to ensure the accuracy of subsequent spatial distance calculation and grid matching.

For the meteorological data, as no snowfall occurred in Shanghai during the study period, the snowfall field was removed, and hourly air temperature and rainfall were retained. The hourly meteorological information was then matched with violation events according to the occurrence time of each drunk-driving violation, forming a standardized event dataset containing temporal, spatial, and meteorological attributes.

For the road network data, roads were filtered according to road class and motor-vehicle accessibility. Expressways, arterial roads, secondary roads, and major branch roads were retained, while internal minor roads and non-motorized lanes were excluded. The filtered road network was used to calculate the length of road segments included in the analysis within each grid. The proportion of higher-level road length was further extracted to reflect roadway exposure and road hierarchy at the grid level.

For the metro station data, duplicate records were removed based on station names and spatial coordinates to reduce the influence of transfer stations and repeated records on station counts. The processed metro station data were used to calculate the distance from each grid centroid to the nearest metro station, which served as a proxy variable for public transport accessibility.

For the POI data, dining, entertainment, and public parking facility POIs were selected according to the category field. Dining and entertainment POIs were used to describe urban activity spaces related to alcohol consumption, while public parking facility POIs were used to reflect the supply of parking facilities within each grid. Dining and entertainment POIs were further aggregated by the number of POIs within each grid and within the surrounding 400–800 m annular buffer, supporting the subsequent analysis of local effects and surrounding spatial associations.

2.3. Identification of High-Risk Clusters

2.3.1. ST-DBSCAN Algorithm

Drunk-driving violations show both spatial and temporal proximity. Traditional clustering methods consider only spatial proximity and cannot fully capture the spatio-temporal clustering characteristics of drunk-driving events. Therefore, this study uses the spatio-temporal density-based clustering algorithm ST-DBSCAN to identify high-risk clusters of drunk-driving violations. This algorithm incorporates both spatial and temporal neighborhoods within a density-based clustering framework and can identify spatio-temporal hotspots with arbitrary shapes [22,23].

Let the set of drunk-driving violation points be defined as:

$$P = \{p_i = (x_i, y_i, \tau_i) \mid i = 1, 2, \dots, n\} \quad (1)$$

where (x_i, y_i) denotes the planar coordinates of event point p_i in the projected coordinate system, and τ_i denotes the number of minutes within a day corresponding to the event occurrence time.

The event time is converted into a within-day minute variable:

$$\tau_i = 60h_i + m_i, \tau_i \in [0, 1439] \quad (2)$$

where h_i and m_i denote the hour and minute of the event occurrence time, respectively.

The spatial distance between event points p_i and p_j is calculated using the Euclidean distance:

$$d_s(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

Because the high-incidence period of drunk driving usually crosses midnight, temporal distance is represented by the within-day circular distance:

$$d_t(p_i, p_j) = \min\{|\tau_i - \tau_j|, 1440 - |\tau_i - \tau_j|\} \quad (4)$$

Given a spatial neighborhood radius ε_s and a temporal neighborhood threshold ε_t , the spatio-temporal neighborhood of event point p_i is defined as:

$$N_{(\varepsilon_s, \varepsilon_t)}(p_i) = \{p_j \in P \mid d_s(p_i, p_j) \leq \varepsilon_s, d_t(p_i, p_j) \leq \varepsilon_t\} \quad (5)$$

If the number of events in the spatio-temporal neighborhood of p_i is no less than the minimum number of samples, that is, $|N_{(\varepsilon_s, \varepsilon_t)}(p_i)| \geq \text{MinPts}$, then p_i is defined as a core point. Starting from such core points, the algorithm expands through density-connected relationships and finally forms spatio-temporal clusters. Samples that are not assigned to any cluster are treated as noise points.

2.3.2. Parameter Setting for Clustering

The performance of ST-DBSCAN clustering is directly affected by three key parameters: the spatial neighborhood radius, the temporal neighborhood threshold, and the minimum number of samples [24]. To reduce the subjectivity of parameter selection, this study conducted multi-scenario comparisons based on the characteristics of the data.

For the spatial neighborhood radius (ε_s), six radius settings ranging from 200 m to 1000 m were compared using kernel density analysis [25]. The results showed that a smaller spatial radius can preserve finer local differences, but may lead to fragmented clusters, whereas a larger spatial radius can reduce noise points but may merge events located in adjacent road spaces, thereby weakening the identification of micro-scale hotspots. Considering the spatial continuity of event clustering, the scale of urban road spaces, and the connection with the subsequent grid-based modeling, this study selected 400 m as the spatial neighborhood radius of ST-DBSCAN, namely: $\varepsilon_s = 400$ m.

For the temporal neighborhood threshold (ε_t), the temporal distribution of drunk-driving events showed that 20:00 to 02:00 of the following day was the core peak period, accounting for approximately 71.61% of all samples. After further comparing different time windows, a 6 h window was found to effectively cover both the core and marginal peak periods of drunk-driving events. Therefore, 6 h was selected as the temporal neighborhood threshold, namely: $\varepsilon_t = 360$ min.

For the minimum number of samples (MinPts), this study compares three auxiliary indicators under $\text{MinPts} = 3$ to 10: the number of valid clusters, the number of events within clusters, and the number of POIs within 400 m of cluster centers. The parameter definitions are as follows.

Let the k -th cluster be C_k . The number of events within the cluster is $E_k = |C_k|$, The cluster center coordinates are:

$$\bar{x}_k = \frac{1}{E_k} \sum_{p_i \in C_k} x_i, \bar{y}_k = \frac{1}{E_k} \sum_{p_i \in C_k} y_i \quad (6)$$

Let the projected coordinates of the m -th POI be (u_m, v_m) , and let the study area contain M relevant POIs. The number of POIs within 400 m of the cluster center is:

$$\text{POI}_k = \sum_{m=1}^M I \left(\sqrt{(\bar{x}_k - u_m)^2 + (\bar{y}_k - v_m)^2} \leq 400 \right) \quad (7)$$

where $I(\cdot)$ is an indicator function that equals 1 when the condition is satisfied and 0 otherwise.

The comparison results are shown in Table 1. As *MinPts* increases, the number of clusters decreases, but the concentration of events within clusters increases. *MinPts* = 9 achieved a relatively good balance between event concentration and the number of surrounding POIs. Therefore, *MinPts* = 9 was selected as the main analysis parameter. The final core parameters are $\varepsilon_s = 400$ m, $\varepsilon_t = 360$ min, and *MinPts* = 9.

Table 1. Comparison of clustering results under *MinPts* values from 3 to 10.

<i>MinPts</i>	Number of Valid Clusters	Average Violations per Cluster	Average POI Count Within 400 m
3	143	6.6	4
4	94	8.4	4.5
5	70	9.9	3.8
6	57	10.9	3.9
7	39	13.2	4.6
8	35	13.8	5.5
9	25	16.1	6.5
10	20	17.7	4.8

The ST-DBSCAN clustering results were used to identify explicit spatio-temporal concentrations of drunk-driving violations observed during the study period. In the subsequent analysis, this study further combined the grid-based risk model to identify potential high-risk spaces, and used both historical clustering results and model-predicted risks for risk classification.

2.4. Period-Specific Drunk-Driving Risk Models

To examine the effects of alcohol-related activity spaces, road conditions, public transport accessibility, parking facilities, meteorological factors, and temporal attributes on the risk of drunk-driving violations under different temporal contexts, this study uses regular grids as the basic spatial units and develops time-specific binary Logit models as the main models. Given the low-frequency and sparse nature of drunk-driving violations at the grid-date-period level, a complementary log-log model is further employed for robustness testing. In addition, to capture the potential influence of dining and entertainment venues on surrounding road spaces, POI spillover variables within the 400–800 m annular buffer are incorporated into the model.

2.4.1. Grid Construction

To maintain consistency between the spatial unit of the risk model and the clustering radius, 400 m × 400 m regular grids were generated under the projected coordinate system. Water areas and invalid areas without road network coverage were removed, resulting in a valid grid set $G = \{g_1, g_2, \dots, g_j\}$. Drunk-driving violations, road network data, and POI data were then matched to the grid cells. In this study, regular grids serve as the spatial units for risk estimation and governance zoning, while the occurrence locations of drunk-driving violations are still associated with road spaces. Based on this setting, grids without road network coverage were excluded, and the main road length and road hierarchy variables within each grid were incorporated into the subsequent models to strengthen the correspondence between grid-based risk estimates and road spaces.

In the temporal dimension, the whole day was divided into three periods according to the intraday fluctuation pattern of drunk-driving behavior. Specifically, T1 was defined as 6:00–18:00, mainly corresponding to routine daytime travel; T2 was defined as 18:00–

24:00, mainly corresponding to the period when evening dining and entertainment activities are concentrated and some departure trips occur; and T3 was defined as 0:00–6:00, mainly corresponding to return trips after late-night activities. This division distinguishes daytime and nighttime risk differences and further allows comparison between the first half of the nighttime period and the post-midnight period.

Drunk-driving violations occurring during 00:00–06:00 are usually still related to the previous evening's dining, entertainment, and return-trip process. To preserve the continuity of nighttime activity chains in date attribution, T3 was attributed to the previous calendar day. Let the event occurrence time be t_i , and let its assigned date be d_i^* . Then:

$$d_i^* = \begin{cases} \text{date}(t_i) - 1, & 00:00 \leq \text{time}(t_i) < 06:00 \\ \text{date}(t_i), & \text{otherwise} \end{cases} \quad (8)$$

Based on this adjustment, a weekend indicator is constructed. Let W_{ds} indicate whether date d and period s belong to a weekend context. Considering that early Saturday morning activities are directly influenced by Friday evening activities, T2 and T3 periods assigned to Friday, as well as all periods assigned to Saturday and Sunday, are defined as weekend periods. All other periods are defined as weekdays:

$$W_{ds} = \begin{cases} 1, & d^* \in \text{weekend periods} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

A three-dimensional panel dataset consisting of grids, dates, and time periods is then constructed.

2.4.2. Variable Definitions

Studies on traffic violations and traffic crashes generally suggest that road safety risk is jointly influenced by temporal attributes, road conditions, meteorological conditions, and surrounding activity facilities [15,26,27]. Based on data availability and the research objectives of this study, dining POIs, entertainment POIs, surrounding POIs, road length included in the analysis, road hierarchy, metro accessibility, public parking facilities, weekend, rainfall, and average temperature were selected as explanatory variables to analyze their effects on the risk of drunk-driving violations.

The dependent variable was defined according to whether a drunk-driving violation occurred in a given grid cell on a specific date and during a specific time period. Let N_{jds} denote the number of drunk-driving violations observed in grid g_j on date d during period s . The dependent variable is defined as:

$$Y_{jds} = \begin{cases} 1, & N_{jds} > 0 \\ 0, & N_{jds} = 0 \end{cases} \quad (10)$$

where Y_{jds} is a binary variable indicating whether at least one drunk-driving violation occurred in grid g_j on date d during period s .

The core independent variables are the number of entertainment POIs and dining POIs within each grid. Because POI counts are clearly right-skewed, a log transformation is applied to reduce the influence of extreme values:

$$X_j^{\text{ent}} = \ln(1 + \text{Ent}_j) \quad (11)$$

$$X_j^{\text{din}} = \ln(1 + \text{Din}_j) \quad (12)$$

where Ent_j and Din_j denote the numbers of entertainment POIs and dining POIs within grid g_j , respectively. X_j^{ent} and X_j^{din} denote the log-transformed entertainment POI and dining POI variables, respectively.

To further capture the influence of dining and entertainment venues on surrounding road spaces, this study calculated the numbers of entertainment POIs and dining POIs within the 400–800 m annular buffer outside the target grid and applied the same logarithmic transformation. Let Ent_j^{spill} and Din_j^{spill} denote the numbers of entertainment and dining POIs within the 400–800 m annular buffer around grid j , respectively. Then:

$$X_j^{ent,spill} = \ln(1 + Ent_j^{spill}) \quad (13)$$

$$X_j^{din,spill} = \ln(1 + Din_j^{spill}) \quad (14)$$

The road condition variables include the road length included in the analysis and the proportion of higher-level roads. The road length included in the analysis reflects the scale of road space within each grid that can accommodate motor-vehicle travel and the observation of violation records. The proportion of higher-level roads is calculated as the length of expressways, arterial roads, and secondary roads divided by the total road length included in the analysis, and is used to represent the road hierarchy.

Public transport accessibility is measured by the distance from the grid centroid to the nearest metro station. The public parking facility variable is measured by the log-transformed number of public parking facility POIs within each grid. Meteorological variables include rainfall and centered average temperature, and the temporal attribute is represented by the weekend variable. The symbols and definitions of the variables are shown in Table 2.

Table 2. Variable definitions and descriptive statistics.

Category	Variable	Symbol	Description	Mean	Std. Dev.
Dependent variable	Occurrence of drunk-driving violations	Y_{jds}	1 if at least one drunk-driving violation occurred in the grid during the period, 0 otherwise	0.0004	0.0192
Key independent variables	Log of entertainment POIs	X_j^{ent}	Logarithmic transformation of the number of entertainment POIs within the grid	0.2494	0.4652
	Log of dining POIs	X_j^{din}	Logarithmic transformation of the number of dining POIs within the grid	0.8373	0.6836
Spillover variables	Surrounding entertainment POIs	$X_j^{ent,spill}$	Log-transformed number of entertainment POIs within the 400–800 m annular buffer outside the grid	0.6576	0.8761
	Surrounding dining POIs	$X_j^{din,spill}$	Log-transformed number of dining POIs within the 400–800 m annular buffer outside the grid	1.4832	1.1678
Control variables	Road length included in the analysis	RL_j	Total length of expressways, arterial roads, secondary roads, and major branch roads within the grid, measured in km	0.8672	0.8735
	Proportion of higher-level roads	HR_j	Ratio of the length of expressways, arterial roads, and secondary roads to the total road	0.4973	0.4398

		length included in the analysis within the grid		
Distance to the nearest metro station	MD_j	Distance from the grid centroid to the nearest metro station, measured in km	4.0783	6.8090
Public parking facilities	P_j	Log-transformed number of public parking facility POIs within the grid	0.6151	0.6933
Weekend	W_{ds}	1 for T2/T3 on Friday and all day on weekends, 0 otherwise	0.3847	0.4865
Rainfall	R_{ds}	1 if cumulative rainfall > 0, 0 otherwise	0.2812	0.4496
Centered average temperature	T_{ds}^c	Average temperature minus the sample mean	0	9.1314

2.4.3. Logit Model

This study adopts a binary Logit model to estimate the occurrence probability of drunk-driving violations. Logit models have been widely used in traffic safety and crash severity studies to analyze the relationships between influencing factors and event occurrence probabilities [28]. Let $\pi_{jds} = \Pr(Y_{jds} = 1)$ denote the probability that a drunk-driving violation occurs in grid j on date d during period s . The Logit model is specified as:

$$\logit(\pi_{jds}) = \ln\left(\frac{\pi_{jds}}{1 - \pi_{jds}}\right) = \eta_{jds}^{(s)} \quad (15)$$

where $\eta_{jds}^{(s)}$ is the linear predictor. Considering that drunk-driving behavior may be generated by different mechanisms across time periods, T1, T2, and T3 are modeled separately. For any period $s \in \{T_1, T_2, T_3\}$, the linear predictor is specified as:

$$\eta_{jds}^{(s)} = \beta_0^{(s)} + \beta_1^{(s)} X_j^{\text{ent}} + \beta_2^{(s)} X_j^{\text{din}} + \beta_3^{(s)} X_j^{\text{ent,spill}} + \beta_4^{(s)} X_j^{\text{din,spill}} + \beta_5^{(s)} RL_j + \beta_6^{(s)} HR_j + \beta_7^{(s)} MD_j + \beta_8^{(s)} P_j + \beta_9^{(s)} W_{ds} + \beta_{10}^{(s)} R_{ds} + \beta_{11}^{(s)} T_{ds}^c \quad (16)$$

where $\beta_0^{(s)}$ is the intercept for period s , and $\beta_1^{(s)}$ to $\beta_{11}^{(s)}$ are the parameters to be estimated for the explanatory variables in period s . The meanings of the remaining variables are provided in Table 2.

2.4.4. Complementary Log-Log Model

Since grid-level drunk-driving violations are typical low-frequency and sparse events during the observation period, this study further adopts a complementary log-log (cloglog) model to test the robustness of the Logit model results. The cloglog model is commonly used for binary events with low occurrence probabilities and can serve as a complementary specification to the Logit model [29].

Let $\pi_{jds} = \Pr(Y_{jds} = 1)$. The complementary log-log model is expressed as:

$$\text{cloglog}(\pi_{jds}) = \ln[-\ln(1 - \pi_{jds})] = \eta_{jds}^{(s)} \quad (17)$$

The equivalent probability form is:

$$\pi_{jds} = 1 - \exp\left[-\exp\left(\eta_{jds}^{(s)}\right)\right] \quad (18)$$

where $\eta_{jds}^{(s)}$ is consistent with the linear predictor in the baseline Logit model, and the variables are defined in Table 2. By comparing the direction, significance, and period-specific differences in the core variables across the two models, the robustness of the baseline conclusions can be assessed.

2.5. Risk Classification and Spatial Integration

To clarify the transformation from historical clusters and model-predicted risk to enforcement priority classes, the workflow of risk classification and spatial integration is illustrated in Figure 2.

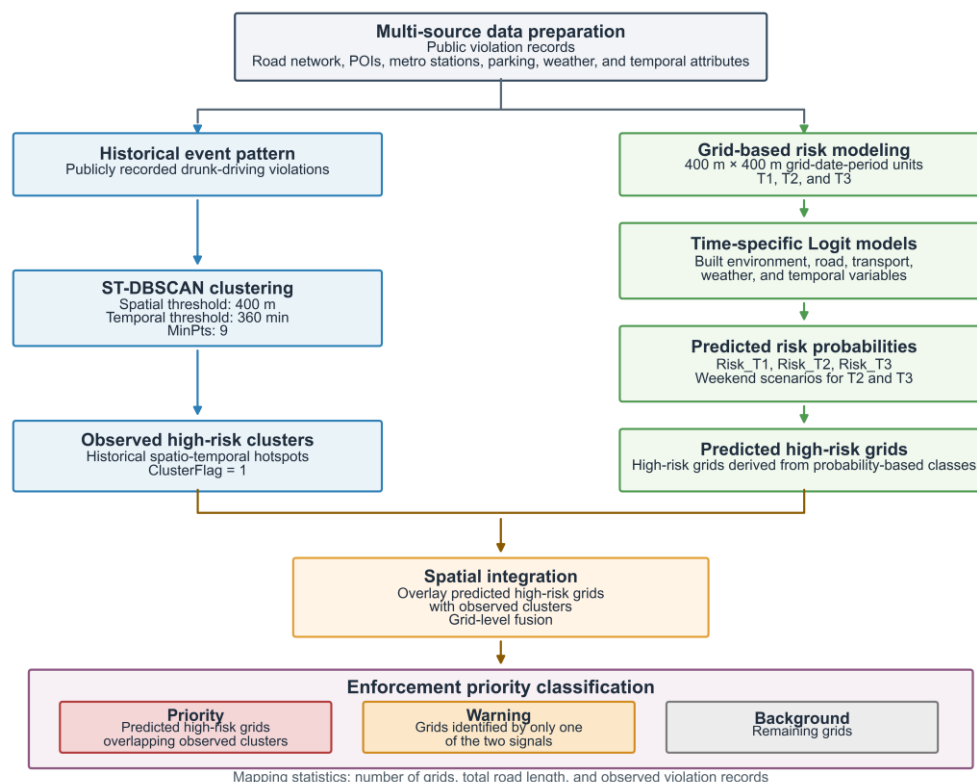


Figure 2. Workflow of risk classification and spatial integration.

3. Results

3.1. Multicollinearity Test

To ensure the stability of parameter estimation in the Logit models, this study conducted a variance inflation factor (VIF) test for the explanatory variables before the regression analysis. The tested variables included entertainment POIs within the grid, dining POIs within the grid, entertainment POIs within the surrounding 400–800 m annular buffer, dining POIs within the surrounding 400–800 m annular buffer, road length included in the analysis, proportion of higher-level roads, distance to the nearest metro station, public parking facilities, weekend, rainfall, and centered average temperature. Previous studies generally suggest that when $0 < \text{VIF} < 10$, there is no serious multicollinearity among independent variables. The test results show that the maximum VIF value is 3.046, indicating that no serious multicollinearity existed among the explanatory variables and that they could be included in the subsequent time-specific risk models.

3.2. Temporal and Spatial Distribution Characteristics of Drunk-Driving Violations

From the temporal distribution, publicly recorded drunk-driving violations in Shanghai show a clear nighttime concentration, as shown in Figure 3. The number of events increases markedly after 18:00 and enters a high-incidence stage during 20:00–23:00. After midnight, the event count remains at its highest level around 00:00 and then gradually declines after 01:00, but it remains relatively high before 02:00. In contrast, the

daytime period from 06:00 to 18:00 has a generally lower number of events, with a particularly flat distribution between 10:00 and 12:00. This pattern is consistent with evening dining, overnight entertainment, and late-night return trips in urban areas.

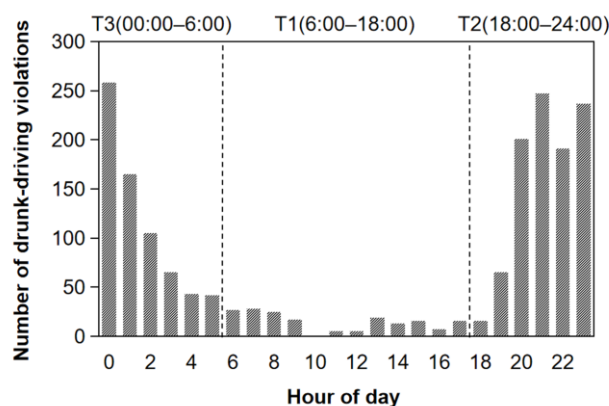


Figure 3. Hourly distribution of publicly recorded drunk-driving violations. Note: The dashed lines indicate the three time periods used in the analysis: T1 (06:00–18:00), T2 (18:00–24:00), and T3 (00:00–06:00). In the panel construction, T3 was attributed to the previous behavioral date.

In terms of spatial distribution, Figure 4 shows that publicly recorded drunk-driving violation records were not evenly distributed across the study area, but instead exhibited a clear pattern of hotspot clustering. Several local high-value areas are observed in the study area. Hotspots are mainly concentrated in the central urban area and nearby areas with dense road networks and active commercial and entertainment activities. Peripheral areas have relatively low overall densities, with only a few scattered high-value nodes. This suggests that drunk-driving risk has significant spatial heterogeneity. Its distribution is shaped not only by the overall urban activity structure but also by stronger local clustering in key areas.

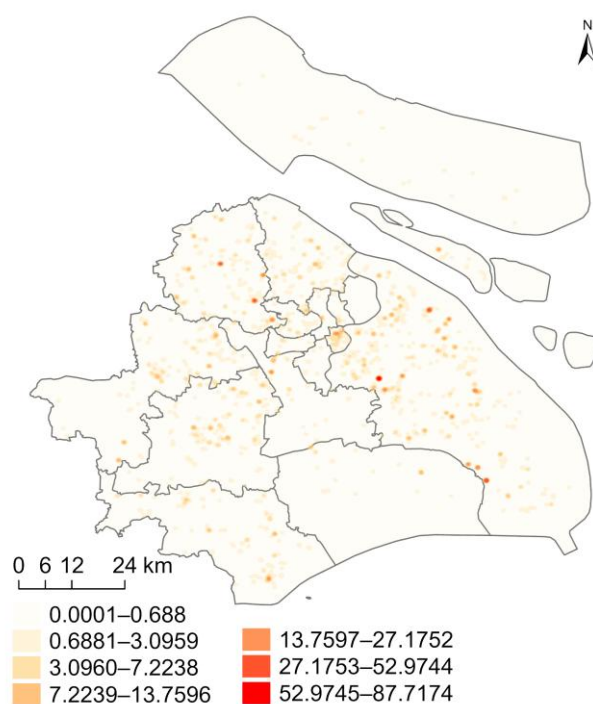


Figure 4. Kernel density distribution of publicly recorded drunk-driving violations in Shanghai.

3.3. Identification Results of High-Risk Clusters

The ST-DBSCAN clustering shows that publicly recorded drunk-driving violation records were not evenly distributed across the study area, but instead exhibited a clear pattern of hotspot clustering. Under this parameter setting, 25 high-risk clusters are identified. Their spatial distribution is shown in Figure 5.



Figure 5. Spatial distribution of high-risk clusters.

The high-risk clusters are not distributed as continuous large areas but appear as multiple discrete local hotspots. Compared with the continuous high-risk pattern revealed by kernel density analysis, the high-risk clusters further condensed the overall hotspots into smaller-scale explicit clustering areas, allowing the risk of drunk-driving behavior to be represented by locatable hotspot clusters. These results help identify explicit clustering areas that had already formed in the historical observation records and provide a basis for subsequent comparison with model-predicted risk grids and risk classification.

3.4. Time-Specific Risk Model Results

To quantify the effects of different variables on the probability of drunk-driving violations, baseline Logit models are estimated separately for T1, T2, and T3. The regression results are shown in Table 3, and the odds ratios (ORs) are shown in Figure 6.

Table 3. Regression results of the time-specific Logit models.

Variable	T1 Coef. (S.E.)	T1 OR	T2 Coef. (S.E.)	T2 OR	T3 Coef. (S.E.)	T3 OR
Entertainment POIs within the grid	−0.489 (0.376)	0.613	−1.152 *** (0.270)	0.316	−0.705 *** (0.242)	0.494
Dining POIs within the grid	−1.135 *** (0.250)	0.321	−1.684 *** (0.198)	0.186	−1.313 *** (0.216)	0.269
Surrounding entertainment POIs	−0.044 ** (0.018)	0.957	−0.061 ** (0.023)	0.941	−0.012 (0.015)	0.988
Surrounding dining POIs	0.016 ** (0.006)	1.016	0.019 *** (0.006)	1.019	0.019 *** (0.004)	1.020

Road length included in the analysis	0.248 ** (0.114)	1.281	0.071 (0.110)	1.074	0.022 (0.107)	1.022
Higher-level road ratio	−0.009 (0.212)	0.991	0.041 (0.131)	1.042	0.078 (0.189)	1.081
Distance to nearest metro station	0.014 (0.011)	1.014	−0.001 (0.006)	0.999	−0.021 ** (0.009)	0.980
Public parking facilities	−0.454 *** (0.164)	0.635	−0.305 *** (0.104)	0.737	−0.234 * (0.137)	0.791
Weekend	0.236 (0.159)	1.266	0.130 * (0.066)	1.138	0.283 *** (0.080)	1.327
Rainfall	0.144 (0.179)	1.155	−0.111 (0.085)	0.895	0.010 (0.108)	1.010
Centered average temperature	−0.004 (0.008)	0.996	−0.003 (0.004)	0.997	−0.009 * (0.005)	0.991
Constant	−8.725 *** (0.252)	—	−6.363 *** (0.127)	—	−7.095 *** (0.184)	—
Observations	1,579,967		1,579,967		1,579,967	
Positive events	179		917		648	
Positive rate	0.0001		0.0006		0.0004	
AIC	3536.552		14,693.240		11,069.455	
Mcfadden Pseudo R ²	0.024		0.053		0.031	

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

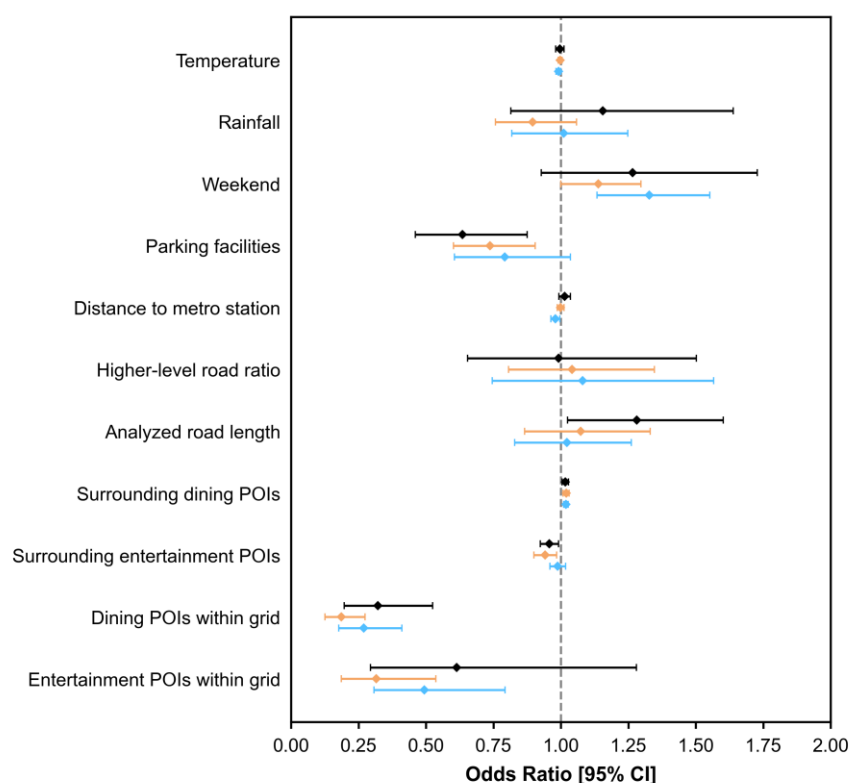


Figure 6. Forest plot of Odds Ratios (ORs) for the Logit models. Note: The dashed line represents OR = 1.

The positive sample rates in all three periods were below 0.1%, indicating that publicly recorded drunk-driving violations were clearly low-frequency and sparse at the grid scale. In terms of model fit, T2 had a relatively larger number of positive samples and a

higher McFadden Pseudo R^2 , suggesting that risk differences were more pronounced during the evening period when dining, entertainment, and departure activities were concentrated.

The POI variables within the grid showed relatively consistent negative associations. Dining POIs were significantly and negatively associated with drunk-driving risk in all three periods, with OR values of 0.321, 0.186, and 0.269, respectively. Entertainment POIs were also significantly negative in T2 and T3, with OR values of 0.316 and 0.494, respectively, but were not significant in T1. These results indicate that publicly recorded drunk-driving violations were not concentrated within the grids where dining and entertainment venues were located. In central urban areas or commercially active districts, dining and entertainment facilities often coexist with better public transport accessibility, more developed road management conditions, and more diverse travel alternatives. Violation records may therefore be more likely to be observed on surrounding roads or return routes after drivers leave activity venues. The negative coefficients of within-grid POIs thus do not indicate a simple inverse relationship between alcohol-related activities and risk; rather, they suggest a possible spatial mismatch between the locations of activity venues and the locations where violation records were observed.

Different from within-grid POIs, the 400–800 m annular POI variables reveal the relationship between surrounding activity spaces and road risk. Surrounding dining POIs were significantly positive in all three periods, with OR values of 1.016, 1.019, and 1.020, respectively, indicating that road spaces around dining-intensive areas were more likely to record drunk-driving violations. This finding is consistent with the travel process of drunk driving: violation records do not necessarily occur within the grid containing dining facilities, but may appear on adjacent road spaces after vehicles leave dining venues. Surrounding entertainment POIs were significantly negative in T1 and T2 and insignificant in T3, suggesting stronger temporal variation in the surrounding associations of entertainment venues. Compared with dining facilities, entertainment venues may involve more dispersed activity times, longer and more variable stay durations, and different departure patterns, so their surrounding road risk did not show a stable positive spillover pattern.

Among the road condition variables, road length included in the analysis was significantly positive in T1, with an OR of 1.281, while it remained positive but insignificant in T2 and T3. This result suggests that, during the daytime period, drunk-driving violation records were more likely to be observed in grids with a larger scale of major road space, possibly because daytime road movements and enforcement observation opportunities are more dependent on the size of the road space. The proportion of higher-level roads was not significant in any of the three periods, indicating that after controlling for road length, POIs, public transport, parking facilities, temporal attributes, and meteorological factors, road hierarchy itself had limited independent explanatory power for drunk-driving risk. By incorporating both road length and road hierarchy, the model controls for road exposure and road composition at the grid scale, thereby avoiding an interpretation of risk distribution based solely on activity facilities.

Public transport accessibility and parking facilities showed different directional patterns. The distance to the nearest metro station was significantly negative only in T3, with an OR of 0.980, indicating that post-midnight drunk-driving risk tended to appear in grids closer to metro stations. Areas surrounding metro stations are often also urban nodes with concentrated commercial activities, nighttime activities, and transport flows. The post-midnight concentration of risk may therefore be related to higher activity intensity and departure travel demand in these areas. Public parking facilities were negative in all three periods; the effect was significant in T1 and T2 and significant at the 10% level in T3, with

OR values of 0.635, 0.737, and 0.791, respectively. This indicates that grids with more parking facilities generally showed lower drunk-driving risk after controlling for other factors. This result may be related to better transport service provision, road order, and spatial management in areas with more complete parking facilities, and may also reflect the combined influence of parking facilities and more mature urban functional areas.

In addition, the robustness test using the complementary log-log model shows that, after replacing the link function, the directions and significance levels of the core variables remain generally consistent. This suggests that the results of the main model are robust.

3.5. Spatial Distribution of Time-Specific Risk and Enforcement Priority Classification

Based on the predicted results of the time-specific Logit models, this study further calculated the predicted probability of drunk-driving violations for each grid under different temporal scenarios and mapped the spatial distribution of time-specific risk. The predicted probability represents the probability that at least one publicly recorded drunk-driving violation is observed in a grid-date-period unit. To reveal local differences in risk distribution under different temporal scenarios, Figures 7–10 classify the predicted probabilities in each scenario using the natural breaks method.

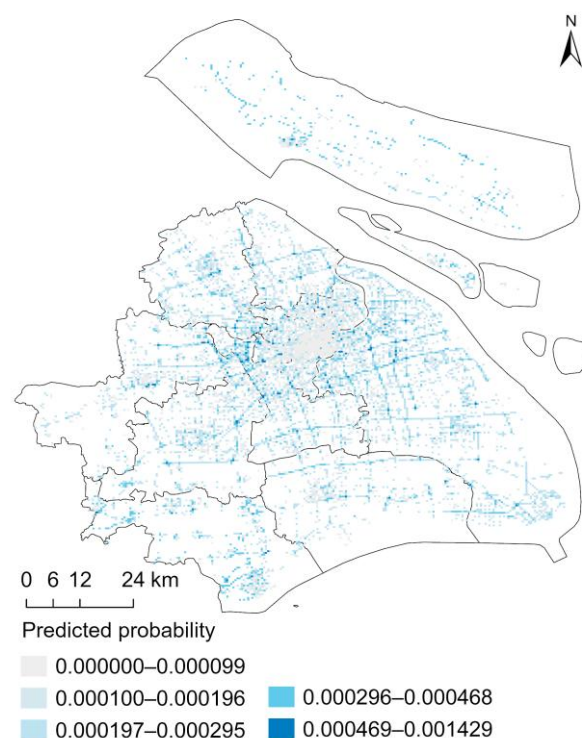


Figure 7. Spatial distribution of predicted drunk-driving violation probability in T1. Note: The predicted probabilities in each scenario were classified separately using the natural breaks method. The color classes indicate relative probability levels within each scenario, while the numerical legend ranges allow comparison of probability levels across scenarios.

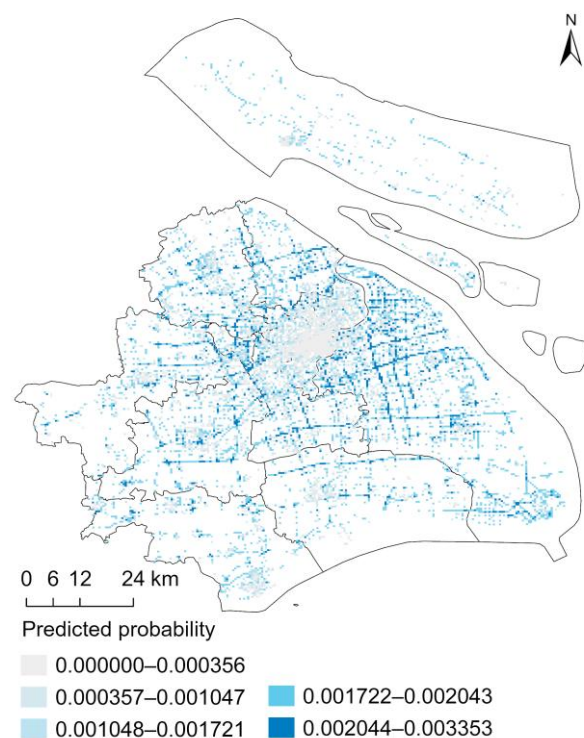


Figure 8. Spatial distribution of predicted drunk-driving violation probability in T2. Note: The predicted probabilities in each scenario were classified separately using the natural breaks method. The color classes indicate relative probability levels within each scenario, while the numerical legend ranges allow comparison of probability levels across scenarios.

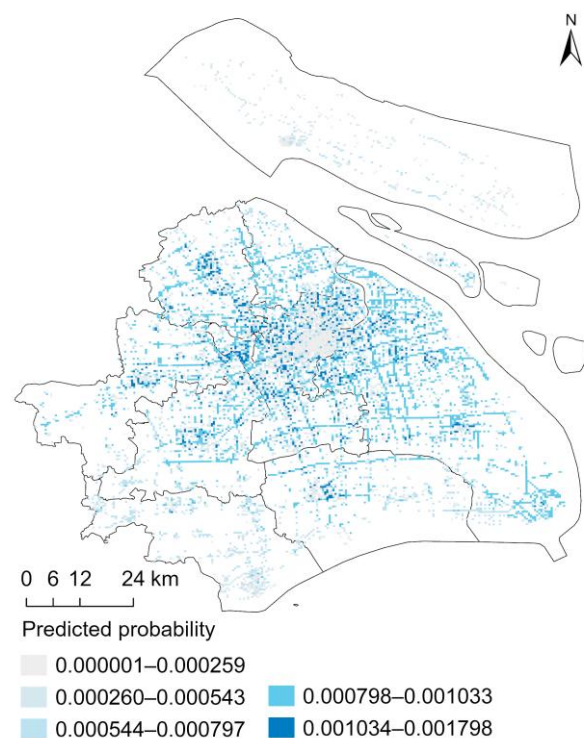


Figure 9. Spatial distribution of predicted drunk-driving violation probability in T3. Note: The predicted probabilities in each scenario were classified separately using the natural breaks method. The color classes indicate relative probability levels within each scenario, while the numerical legend ranges allow comparison of probability levels across scenarios.

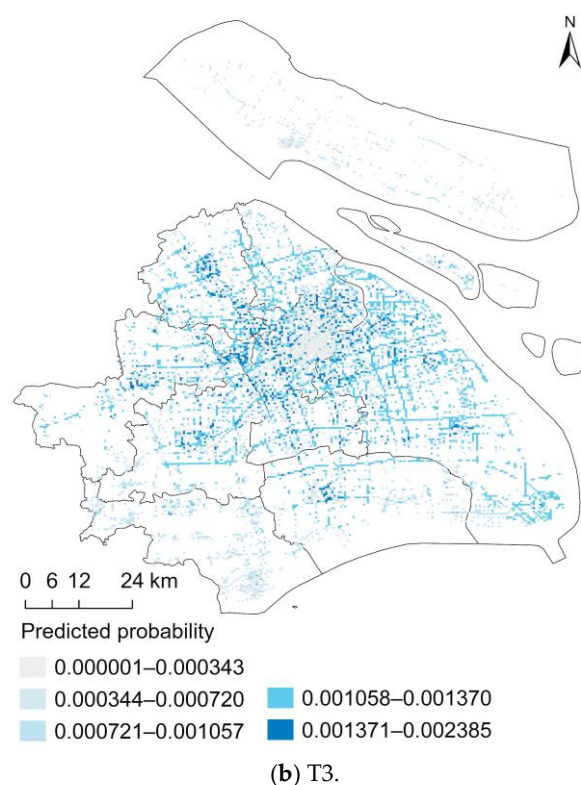
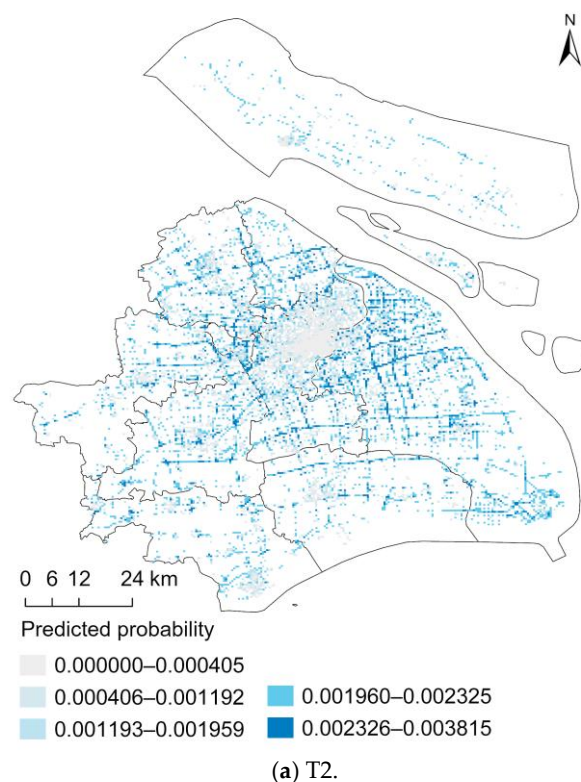


Figure 10. Spatial distribution of predicted drunk-driving violation probability under weekend scenarios: (a) T2; (b) T3. Note: The predicted probabilities in each scenario were classified separately using the natural breaks method. The color classes indicate relative probability levels within each scenario, while the numerical legend ranges allow comparison of probability levels across scenarios.

Overall, the spatial distributions of risk differed markedly across T1, T2, and T3. During T1, the overall predicted probability was relatively low, ranging from 0.000000 to 0.001429. As shown in Figure 7, only a small number of high-value grids appeared, and

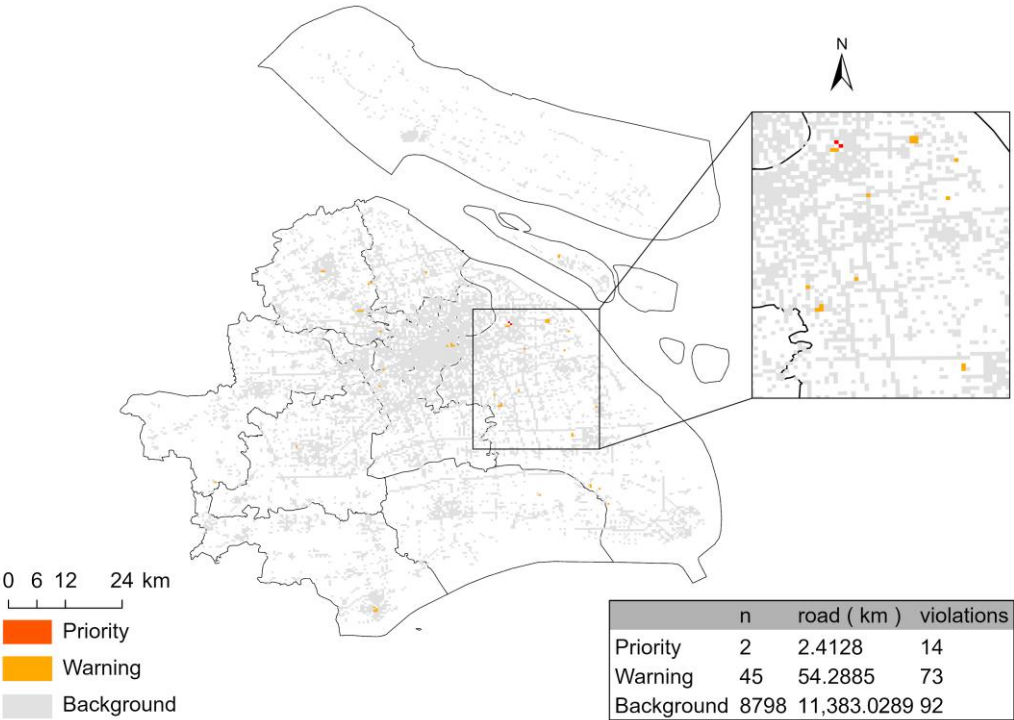
they were spatially scattered. These high-value grids were mainly located in non-central urban areas and around several major road nodes, without forming continuous high-value belts. This pattern is consistent with the daytime routine travel characteristics of T1, indicating that drunk-driving violation risk during the daytime was generally low and was spatially expressed mainly as scattered high-value nodes along local road spaces.

After entering T2, with the beginning of nighttime dining and entertainment activities, the predicted probability increased markedly, and the risk range expanded to 0.000000–0.003353, as shown in Figure 8. Compared with T1, T2 had a substantially larger number of higher-risk grids, and the spatial distribution became more continuous. High-risk areas highly overlapped with core commercial districts, dining streets, and surrounding major road networks, and extended in belt-like patterns along several east–west and north–south major roads. This period corresponds to the stage when evening dining, entertainment, and departure activities are relatively active. The spatial risk pattern therefore shifted from scattered daytime nodes to patch-like and corridor-like distributions shaped jointly by nighttime activity spaces and road traffic spaces.

The risk distribution in T3 differed from that in T2. The predicted probability in T3 ranged from 0.000001 to 0.001798, which was generally lower than that in T2, and the spatial structure of high-risk grids changed to some extent, as shown in Figure 9. Compared with T2, high-value areas in T3 were more concentrated around the central urban area and several key road nodes, while some risk grids extended toward the central urban area along road spaces connecting the central and peripheral areas. This indicates that, compared with the evening dining period, post-midnight drunk-driving violation risk did not diffuse uniformly across the whole city, but was more likely to remain near return routes and traffic transfer nodes after nighttime activities ended.

A further comparison between weekend and weekday scenarios shows that weekends generally increased the risk levels in T2 and T3, as shown in Figure 10. Under the T2 weekend scenario, the predicted probability range increased to 0.000000–0.003815, representing the highest probability range among the five scenarios. Compared with the T2 baseline scenario, the weekend scenario did not change the basic spatial locations of high-value areas but increased the risk levels in the central urban area, peripheral commercial nodes, and along several major roads, causing more grids to enter higher probability intervals. Under the T3 weekend scenario, the predicted probability ranged from 0.000001 to 0.002385, which was higher than that under the T3 baseline scenario. High-value areas remained concentrated around the central urban road spaces, while several peripheral nodes also showed increased risk. Overall, the weekend effect mainly amplified existing nighttime risk spaces rather than generating a completely different spatial pattern.

On the basis of the spatial distribution of predicted risk, this study further integrated model-predicted high-risk grids with historical spatio-temporal high-risk clusters identified by ST-DBSCAN to form an enforcement-oriented priority classification. The high-risk clusters reflect the actual aggregation characteristics of drunk-driving events in both spatial and temporal dimensions, while the time-specific grid regression results reflect the theoretically predicted risk of drunk-driving violations under specific temporal scenarios. Integrating these two types of results allows both predicted risk areas and historically clustered locations to be considered within the same spatial units. Figure 11 presents the enforcement priority classification results for T1, T2, and T3.



(a) T1.



(b) T2.

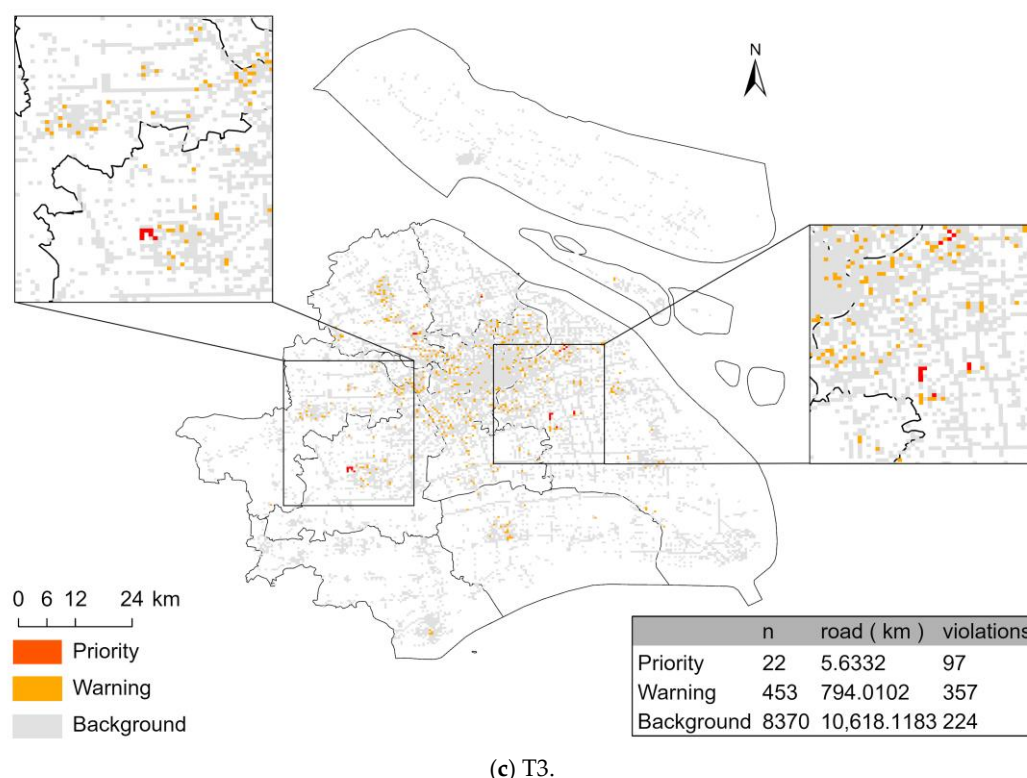


Figure 11. Enforcement priority classification based on risk fusion: (a) T1; (b) T2; (c) T3.

The enforcement priority classification for T1 was dominated by Background areas, as shown in Figure 11a. In this period, only 2 grids were classified as Priority areas, corresponding to 2.4128 km of road length but containing 14 observed violation records. There were 45 Warning grids, corresponding to 54.2885 km of road length and containing 73 observed violation records. By contrast, the Background areas included 8798 grids and 11,383.0289 km of road length, but contained 92 observed violation records. This result indicates that although the overall T1 risk was relatively low, the small number of grids where predicted risk overlapped with historical clusters still showed a high concentration of events and can serve as key reference areas for limited daytime enforcement deployment.

The enforcement priority classification changed markedly in T2, as shown in Figure 11b. The number of Priority grids increased to 47, corresponding to 29.6601 km of road length and containing 360 observed violation records. The number of Warning grids increased to 2573, corresponding to 4602.0989 km of road length and containing 268 observed violation records. Compared with T1, T2 showed not only a substantial increase in Priority grids but also a large expansion of Warning areas within the central urban area and surrounding major road spaces. This indicates that nighttime risk formed a spatial structure consisting of highly concentrated enforcement grids and broader warning grids.

The enforcement priority classification for T3 is shown in Figure 11c. In this period, 22 grids were classified as Priority areas, corresponding to 5.6332 km of road length and containing 97 observed violation records. There were 453 Warning grids, corresponding to 794.0102 km of road length and containing 357 observed violation records. The Background areas included 8370 grids and 10,618.1183 km of road length, containing 224 observed violation records. Compared with T2, the number of Priority grids in T3 decreased, but they remained concentrated around key road nodes. The Warning areas showed a more evident diffusion pattern, indicating that during the post-midnight return period, risk still extended along road spaces beyond local hotspots.

Taken together, Figures 7–11 show clear temporal and spatial heterogeneity in drunk-driving violation risk. T1 was characterized by low predicted probabilities and scattered high-value grids; T2 showed a marked increase in risk level and a more continuous high-value spatial pattern; and T3 exhibited post-midnight concentration around the central urban area and key road nodes. Weekend scenarios further increased the predicted probabilities in T2 and T3, strengthening nighttime and post-midnight risk within existing high-value areas. The risk fusion results further show that Warning areas covered broader potential risk spaces and should be considered in future enforcement deployment. By integrating model-predicted probabilities, historical clustering information, and road-space statistics, the classification results provide a spatial basis for time-specific and area-specific drunk-driving risk governance.

4. Discussion

4.1. Interpretation of the Formation Mechanism of Drunk-Driving Violation Risk

The results indicate that publicly recorded drunk-driving violations in Shanghai exhibit clear activity-chain characteristics and road-space dependence. Drunk-driving violations do not necessarily occur at the locations where drinking activities take place. Instead, they are often observed on departure roads, short-distance transfer routes, and post-midnight return trips after dining and entertainment activities. Therefore, the spatial distribution of drunk-driving violation risk is not a simple mapping of alcohol-related venues, but is jointly shaped by activity venues, road networks, transport nodes, and temporal contexts.

The difference between within-grid POIs and surrounding POIs further supports this interpretation. The model results show that dining and entertainment POIs within the grid were mostly negatively associated with drunk-driving risk, whereas dining POIs within the 400–800 m annular buffer showed significant positive associations across all three periods. This suggests that drunk-driving violation risk was not concentrated within the grids where dining and entertainment venues were located, but was more likely to emerge in the road spaces surrounding these venues. In central urban areas and commercially active districts, dining and entertainment facilities often coexist with higher public transport accessibility, more developed road management conditions, and more diverse travel alternatives. As a result, the observed locations of violation records may shift toward peripheral roads and return routes outside the activity venues. The effect of alcohol-related activity spaces on risk is therefore better understood from the relationship between venues and roads, rather than from venue density alone.

Temporal factors further strengthen this spatial mismatch mechanism. The risk levels in T2 and T3 were clearly higher than those in T1, and the weekend variable also showed stronger effects during the evening and post-midnight periods. This indicates that drunk-driving violation risk is closely related to nighttime dining and entertainment activities, departure trips, and post-midnight return processes. T2 corresponds to evening activities and the initial departure stage, during which risk is more likely to concentrate in activity spaces and surrounding roads. T3 corresponds more closely to return trips after nighttime activities, with risk continuing around key road nodes and transport transition spaces. These results suggest that drunk-driving violation risk is highly dependent on temporal context. Nighttime and weekends do not merely change traffic volume; they also alter the combination of alcohol-related activities and motor-vehicle travel.

Road and transport facility variables further show that drunk-driving violation risk is expressed through road space. Road length included in the analysis was significantly positive in T1, indicating that the scale of road space can affect the likelihood that violation records are observed. The distance to the nearest metro station was significantly negative in T3, suggesting that post-midnight risk is more likely to appear near urban nodes where

commercial activities, nighttime activities, and transport flows are relatively concentrated. The negative association of public parking facilities may reflect the combined influence of parking supply, road order, and spatial management conditions in areas with more developed facilities. Compared with a single road hierarchy variable, the combination of activity spaces, road exposure, and transport node conditions provides a more convincing explanation for the spatial differences in drunk-driving violation risk.

Overall, nighttime dining and entertainment activities provide the behavioral background for risk generation, while road networks and transport nodes constitute the spatial carriers through which risk becomes visible. Weekends and post-midnight periods further strengthen the temporal concentration of risk. The historical high-risk clusters identified by ST-DBSCAN reflect explicit hotspots that have already formed, whereas model-predicted risk reveals potential risk areas shaped by multiple spatial factors. The integration of these two results indicates that drunk-driving violation risk is characterized not only by historical clustering, but also by spatial predictability driven by urban activities and road environments.

4.2. Implications for Risk-Based Governance

The risk identification results provide a spatial basis for shifting drunk-driving governance from experience-based response to risk anticipation. Traditional drunk-driving enforcement often relies on existing violation records, experience with key road sections, and temporary special enforcement campaigns. Such approaches can respond directly to exposed high-incidence areas, but their ability to identify potential risk spaces and temporal differences is relatively limited. By integrating historical spatio-temporal clustering results with model-predicted risk, this study helps identify both explicit hotspots where events have already clustered and potential risk areas with higher predicted probabilities, thereby providing a more refined basis for allocating enforcement resources across different time periods and priority levels.

From the perspective of enforcement priority, Priority areas show both higher predicted risk and historical event clustering. Although the number of such grids is limited, the concentration of observed violation records is high, making them suitable as core areas for targeted enforcement and routine monitoring. These areas can be used to determine relatively stable enforcement deployment locations. This is especially evident during T2, when evening dining, entertainment, and departure activities are concentrated; the stronger overlap between Priority grids and observed violation records indicates that key nighttime areas are highly identifiable. In contrast, Warning areas cover a broader spatial range and reflect a single high-risk signal from either model-predicted risk or historical clustering. These areas should not be regarded in the same way as fixed checkpoints, but they can serve as key areas for mobile patrols, temporary inspections, and coordinated control of surrounding road spaces, helping to reduce spatial omissions caused by reliance on historical hotspots alone.

The time-specific results further suggest that drunk-driving governance should not adopt a uniform spatial strategy throughout the day. During T1, the overall risk level is relatively low, and governance efforts should focus more on a small number of local nodes overlapping with historical clusters, so as to avoid excessive dispersion of enforcement resources in low-risk periods. During T2, the risk level increases markedly, and both Priority and Warning areas expand substantially. This indicates that the evening period is the core window for risk governance, requiring greater attention to roads surrounding dining and entertainment activity spaces, roads extending from commercial nodes, and major departure routes. During T3, the number of Priority grids decreases compared with T2, but Warning areas still show spatial extension, suggesting that governance during the post-midnight return stage should consider both key nodes and connecting road spaces.

Compared with enforcement based solely on fixed points, post-midnight governance is better suited to flexible patrols that combine road corridors, transport transition nodes, and surrounding activity areas.

The weekend scenario results show that temporal context is an important factor in drunk-driving risk governance. Weekends did not completely change the basic locations of high-risk spaces but further increased the probability levels of existing risk areas during evening and post-midnight periods. This means that weekend governance does not require the delineation of an entirely different spatial range. Instead, it should increase enforcement intensity and coverage frequency in existing key and warning areas. In areas where dining and entertainment activities are concentrated, weekend nighttime enforcement can be more targeted by considering parking facilities, metro stations, commercial district exits, and major road intersections.

Drunk-driving governance should also shift from the control of individual venues to the management of surrounding road spaces. The model results show that risk was not necessarily concentrated within the grids where dining and entertainment venues were located but was more likely to emerge on surrounding roads and return routes after vehicles left these venues. Therefore, governance should not be limited to bars, dining streets, or entertainment facility locations themselves. It should also cover peripheral roads, parking entrances and exits, roads connecting commercial areas, and nighttime return corridors. Combining venue information with road space, transport nodes, and temporal context can improve the match between enforcement deployment and the locations where risk actually appears.

From the perspective of sustainable traffic safety governance, the risk classification results can support more efficient allocation of enforcement resources. Concentrating limited police resources in Priority areas and implementing flexible patrols in Warning areas can improve risk coverage without simply expanding the enforcement area. At the same time, identifying Background areas helps avoid excessive resource input in low-risk spaces, making the governance strategy more consistent with refined and evidence-based urban traffic management. For megacities, this risk governance framework based on multi-source data and model prediction can not only improve the efficiency of drunk-driving violation control, but also support nighttime traffic safety, road order maintenance, and the development of a sustainable travel environment.

5. Conclusions

Taking publicly recorded drunk-driving violations in Shanghai as the research object, this study integrated multi-source data, including drunk-driving violation records, dining and entertainment POIs, 400–800 m annular POIs, road networks, metro stations, public parking facilities, meteorological conditions, and temporal attributes. Based on these data, an analytical framework was developed that combines spatio-temporal cluster identification, time-specific risk modeling, and risk-fusion classification. ST-DBSCAN was first used to identify spatio-temporal clusters of drunk-driving violations that had already formed in the public records. Then, using 400 m × 400 m grids as the basic spatial units, binary Logit models were constructed for T1, T2, and T3, and a cloglog model was used for robustness testing. On this basis, model-predicted risk was further integrated with historical spatio-temporal clustering results to generate a classified risk identification result for drunk-driving risk governance.

The results show that the risk of publicly recorded drunk-driving violations in Shanghai exhibits clear temporal differences and spatial heterogeneity. During T1, the overall risk level was relatively low, with only a limited number of high-value grids, mainly appearing as scattered distributions around local road nodes. During T2, the risk

level increased markedly, high-value areas became more continuous, and the spatial pattern was closely related to evening dining and entertainment activities as well as major road spaces. During T3, the overall risk level was lower than that in T2, but risk remained concentrated around the central urban area, key road nodes, and post-midnight return corridors. Weekend scenarios further increased the predicted probabilities in T2 and T3, mainly by strengthening existing nighttime risk spaces rather than generating a completely different spatial pattern. The risk-fusion results show that Priority areas were limited in number but had a high concentration of observed violation records, whereas Warning areas covered broader potential risk spaces. This indicates that historical clustering hotspots and model-predicted risk play complementary roles in identifying drunk-driving violation risk.

The model results further reveal the formation mechanism of drunk-driving violation risk. Dining and entertainment POIs within the grid were mostly negatively associated with risk, whereas dining POIs within the 400–800 m annular buffer showed significant positive associations across all three periods. This indicates that drunk-driving violation risk was not simply concentrated within the grids where alcohol-related venues were located, but was more likely to appear on surrounding roads, departure routes, and post-midnight return spaces. Road length, metro accessibility, public parking facilities, and the weekend attribute also showed varying temporal differences, suggesting that drunk-driving violation risk is jointly influenced by activity spaces, road exposure, transport node conditions, and temporal context. The cloglog robustness test produced results generally consistent with the main model, further supporting the stability of the time-specific Logit model results.

In terms of application value, the multi-source data fusion framework proposed in this study can provide a more refined spatial basis for drunk-driving risk governance in megacities. By distinguishing Priority, Warning, and Background areas, risk governance can gradually shift from relying solely on historical violation records and experience-based judgment toward hierarchical identification based on predicted risk and historical clustering characteristics. This framework can improve risk coverage in key periods and key areas without simply expanding the enforcement range, thereby providing methodological support for nighttime traffic safety, road order maintenance, and the development of a sustainable travel environment.

This study still has several limitations. First, the drunk-driving violation data were derived from public records and may be affected by enforcement intensity and inspection frequency. Second, due to data availability constraints, this study did not incorporate enforcement exposure and traffic exposure variables such as checkpoint deployment, real-time traffic flow, and nighttime travel intensity. Future research could incorporate more specific traffic operation and enforcement data to further improve the explanatory power of risk estimation. Third, this study used regular grids as the units for risk estimation and governance zoning. Although road length and road hierarchy variables were included to strengthen the correspondence between grid-level risk and road space, future studies could further conduct multi-scale validation by incorporating road-segment networks, network distance, and actual travel paths. Finally, with the accumulation of data over longer time periods, future research could further examine seasonal changes, annual stability, and the effects of policy interventions on drunk-driving violation risk, thereby enhancing the applicability of the model in long-term traffic safety governance.

Author Contributions: Conceptualization, L.G. and J.F.; methodology, J.F.; software, J.F.; validation, J.W. and Z.J.; formal analysis, L.G. and J.F.; resources, L.G.; data curation, J.F.; writing—original

draft preparation, J.F.; writing—review and editing, L.G., J.W. and Z.J.; visualization, J.F.; supervision, L.G.; project administration, L.G.; funding acquisition, L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shanxi Provincial Basic Research Program Project [grant number 202403021222195] and the Shanxi Provincial Higher Education Science and Technology Innovation Project [grant number 2023L178].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. Road Traffic Deaths. Available online: <https://www.who.int/data/gho/data/themes/topics/indicator-groups/indicator-group-details/GHO/road-traffic-deaths> (accessed on 10 July 2026).
2. Supreme People's Court; Supreme People's Procuratorate; Ministry of Public Security; Ministry of Justice. Opinions on Handling Criminal Cases of Drunk Dangerous Driving. Available online: https://www.spp.gov.cn/spp/xwfbh/wsfbt/202312/t20231218_637161.shtml (accessed on 10 July 2026).
3. Shanghai Municipal Bureau of Statistics. Shanghai Statistical Yearbook 2025. Available online: <https://tjj.sh.gov.cn/tjnj/index.html> (accessed on 10 July 2026).
4. Shanghai Municipal People's Government. Shanghai's Nighttime Economy Index Remains First Nationwide, and the First Batch of 130 Featured Activities Opens a New Form of Nighttime Tourism in Shanghai. Available online: <https://www.shanghai.gov.cn/nw4411/20250607/26ac836c90b148779b5d8b6796ec58e0.html> (accessed on 10 July 2026).
5. Zhang, G.; Yau, K.K.W.; Chen, G. Risk Factors Associated with Traffic Violations and Accident Severity in China. *Accid. Anal. Prev.* **2013**, *59*, 18–25. <https://doi.org/10.1016/j.aap.2013.05.004>.
6. Fang, T.Y.; Xu, F.X.; Zhu, Q.M.; Zou, Z. Association Rules between Urban Road Traffic Accidents and Violations Considering Temporal and Spatial Constraints. *J. Transp. Syst. Eng. Inf. Technol.* **2025**, *25*, 246–254. <https://doi.org/10.16097/j.cnki.1009-6744.2025.03.022>. (In Chinese)
7. Wang, C.W.; He, J.; Yan, X.T.; Zhang, C.J.; Chen, Y.K.; Ye, Y.T. Temporal-Spatial Evolution Analysis of Severe Traffic Violations Using Three Functional Forms of Models Considering the Diurnal Variation of Meteorology. *Accid. Anal. Prev.* **2022**, *174*, 106731. <https://doi.org/10.1016/j.aap.2022.106731>.
8. Lee, E.H.; Yun, H.; Cho, S.H.; Lee, E. Aggressive Driving in Ride-Hailing: Work Hours and Road Conditions. *J. Transp. Saf. Secur.* **2025**, *18*, 693–716. <https://doi.org/10.1080/19439962.2025.2598268>.
9. Krahé, B.; Kröling, S.; Gehlert, T. Aggressive Driving in Germany: Representative Data on Changes over Time. *Transp. Res. Part F Traffic Psychol. Behav.* **2026**, *116*, 103374. <https://doi.org/10.1016/j.trf.2025.103374>.
10. Yan, X.T.; He, J.; Wu, G.H.; Zhang, C.J.; Wang, C.W. Weekly Variations and Temporal Instability of Determinants Influencing Alcohol-Impaired Driving Crashes: A Random Thresholds Random Parameters Hierarchical Ordered Probit Model. *Anal. Methods Accid. Res.* **2021**, *32*, 100189. <https://doi.org/10.1016/j.amar.2021.100189>.
11. Zhao, Z.Y.; Yu, X.J.; Huang, Y.G.; Wu, S. Analysis of Influencing Factors of Urban Traffic Violations Considering Street-View Information. *J. Huaqiao Univ. Nat. Sci.* **2023**, *44*, 759–768. <https://doi.org/10.11830/ISSN.1000-5013.202307002>. (In Chinese)
12. Li, T.; Liu, S.Q.; Fan, G.Q.; Zhao, H.; Zhang, M.; Fan, J.; Li, C. Spatial Heterogeneity Effect of Built Environment on Traffic Safety Using Geographically Weighted Atrous Convolutions Neural Network. *Accid. Anal. Prev.* **2025**, *213*, 107934. <https://doi.org/10.1016/j.aap.2025.107934>.
13. Treno, A.J.; Johnson, F.W.; Remer, L.G.; Gruenewald, P.J. The Impact of Outlet Densities on Alcohol-Related Crashes: A Spatial Panel Approach. *Accid. Anal. Prev.* **2007**, *39*, 894–901. <https://doi.org/10.1016/j.aap.2006.12.011>.
14. Wang, S.; Chen, Y.; Huang, J.; Liu, Z.; Li, J.; Ma, J. Spatial Relationships between Alcohol Outlet Densities and Drunk Driving Crashes: An Empirical Study of Tianjin in China. *J. Saf. Res.* **2020**, *74*, 17–25. <https://doi.org/10.1016/j.jsr.2020.04.011>.
15. Wang, S.H.; Liu, J.Z.; Chen, N.; Xiao, J.; Wei, P. How Does the Built Environment Affect Drunk-Driving Crashes? A Spatial Heterogeneity Analysis. *Appl. Sci.* **2023**, *13*, 11813. <https://doi.org/10.3390/app132111813>.

16. Yang, M.M.; Bao, Q.; Shen, Y.J.; Qu, Q.K.; Zhang, R.; Han, T.Y.; Zhang, H.S. Determinants Influencing Alcohol-Related Two-Vehicle Crash Severity: A Multivariate Bayesian Hierarchical Random Parameters Correlated Outcomes Logit Model. *Anal. Methods Accid. Res.* **2024**, *44*, 100361. <https://doi.org/10.1016/j.amar.2024.100361>.
17. Morrison, C.; Ponicki, W.R.; Gruenewald, P.J.; Wiebe, D.J.; Smith, K. Spatial Relationships between Alcohol-Related Road Crashes and Retail Alcohol Availability. *Drug Alcohol Depend.* **2016**, *162*, 241–244. <https://doi.org/10.1016/j.drugalcdep.2016.02.033>.
18. Gruenewald, P.J.; Johnson, F.W. Drinking, Driving, and Crashing: A Traffic-Flow Model of Alcohol-Related Motor Vehicle Accidents. *J. Stud. Alcohol Drugs* **2010**, *71*, 237–248. <https://doi.org/10.15288/jsad.2010.71.237>.
19. Sun, Z.Y.; Cui, K.Q.; Wang, J.Y.; Gu, X.; Xing, Y.; Lu, H.; Chen, Y. Investigating the Spatial Heterogeneity of Drunk-Driving Events in Beijing Based on a Hybrid Method with LISA and GeoDetector. *Int. J. Crashworthiness* **2025**, *30*, 80–94. <https://doi.org/10.1080/13588265.2024.2365057>.
20. Wu, P.J.; Chen, T.Y.; Wong, Y.D.; Meng, X.; Wang, X.; Liu, W. Exploring Key Spatio-Temporal Features of Crash Risk Hot Spots on Urban Road Network: A Machine Learning Approach. *Transp. Res. Part A Policy Pract.* **2023**, *173*, 103717. <https://doi.org/10.1016/j.tra.2023.103717>.
21. Ben Laoula, E.M.; Elfahim, O.; El Midaoui, M.; Youssfi, M.; Bouattane, O. Traffic Violations Analysis: Identifying Risky Areas and Common Violations. *Heliyon* **2023**, *9*, e19058. <https://doi.org/10.1016/j.heliyon.2023.e19058>.
22. Birant, D.; Kut, A. ST-DBSCAN: An Algorithm for Clustering Spatial-Temporal Data. *Data Knowl. Eng.* **2007**, *60*, 208–221. <https://doi.org/10.1016/j.datak.2006.01.013>.
23. Yang, M.Y.; Dong, C.J.; Xiong, Z.H.; Zhuang, Y.; Xu, B. Hotspot Segments Identification of Collisions between Electric Vehicles and Pedestrians/Non-Motor Vehicles. *J. South China Univ. Technol. Nat. Sci. Ed.* **2025**, *53*, 112–120. <https://doi.org/10.12141/j.issn.1000-565X.240592>. (In Chinese)
24. Wang, H.D.; Song, Y.L.; Tian, Y.J. Research on a Multi-Attribute Spatiotemporal Data Clustering Algorithm. *J. Chongqing Univ. Posts Telecommun. Nat. Sci. Ed.* **2021**, *33*, 661–668. <https://doi.org/10.3979/j.issn.1673-825X.201909010311>. (In Chinese)
25. Xie, Z.; Yan, J. Kernel Density Estimation of Traffic Accidents in a Network Space. *Comput. Environ. Urban Syst.* **2008**, *32*, 396–406. <https://doi.org/10.1016/j.compenvurbsys.2008.05.001>.
26. Zhang, G.; Yau, K.K.W.; Gong, X. Traffic Violations in Guangdong Province of China: Speeding and Drunk Driving. *Accid. Anal. Prev.* **2014**, *64*, 30–40. <https://doi.org/10.1016/j.aap.2013.11.002>.
27. Wang, J.; Ji, L.; Ma, S.; Sun, X.; Wang, M. Analysis of Factors Influencing the Severity of Vehicle-to-Vehicle Accidents Considering the Built Environment: An Interpretable Machine Learning Model. *Sustainability* **2023**, *15*, 12904. <https://doi.org/10.3390/su151712904>.
28. Savolainen, P.T.; Mannering, F.L.; Lord, D.; Quddus, M.A. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accid. Anal. Prev.* **2011**, *43*, 1666–1676. <https://doi.org/10.1016/j.aap.2011.03.025>.
29. Kitali, A.E.; Alluri, P.; Sando, T.; Haule, H.; Kidando, E.; Lentz, R. Likelihood Estimation of Secondary Crashes Using Bayesian Complementary Log-Log Model. *Accid. Anal. Prev.* **2018**, *119*, 58–67. <https://doi.org/10.1016/j.aap.2018.07.003>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.