Traffic & Transportation

Factors Influencing Drivers' Tolerance for Large Vehicle Proportions

Hanbin WANG[1], Jianxiao MA[2], Wenyun TANG[3], Chenyang YANG[4]

Original Scientific Paper
Submitted: 7 Nov 2024
Accepted: 24 Jan 2025

[1] hanbinwang2022@163.com, College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing, China
[2] majx@njfu.edu.cn, College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing, China
[3] Corresponding author, tangwy@njfu.edu.cn, College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing, China
[4] 19852832575@163.com, College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing, China

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

ABSTRACT

To enhance road traffic safety, this study develops a classification model to assess drivers' tolerance of large vehicle proportions under varying road conditions. It explores how personal and socioeconomic factors influence this tolerance. Six road scenarios were designed with differences in large vehicle proportion, driving time, vehicle types and road types. Behavioural and willingness surveys collected drivers' demographics and choices. K-means clustering segmented drivers into "low", "medium" and "high" tolerance groups, accounting for 6.77%, 51.13% and 42.10%, respectively. Based on clustering results, an ordered logistic regression model further analysed factors influencing large vehicle tolerance. Tolerance correlated positively with household vehicle usage, annual household income, driving duration and weekly driving frequency, and negatively with urban GDP, vehicle ownership and peak congestion index, with age showing no significant effect. Additionally, linear trends were observed for urban peak congestion index, urban vehicle ownership, age and driving duration. In contrast, urban GDP, weekly driving frequency, annual household income and household vehicle usage showed curvilinear effects with gradually diminishing rates of change.

KEYWORDS

urban traffic; drivers' tolerance to vehicles; ordered logistic regression; drivers; K-means clustering algorithm.

1. INTRODUCTION

With the continuous development of traffic infrastructure, the proportion of large vehicles (such as buses and heavy trucks) in road traffic is steadily increasing, especially on urban expressways, motorways and areas around logistics hubs. In China, the number of civilian large passenger vehicles and heavy-duty freight vehicles grew from 6.73 million in 2014 to 10.25 million in 2023, reflecting a growth rate of 52.4% [1]. The average share of trucks on most motorways is around 25%, while in certain regions, this share can reach as high as 60% [2]. Due to their substantial size, lower speeds and potential to hinder visibility, large vehicles have a significant impact on the driving behaviour and route choices of smaller vehicles. In real-world driving situations, drivers tend to prefer routes with fewer large vehicles to avoid the inconveniences and safety risks associated with mixed traffic. Therefore, a deeper examination of drivers' tolerance towards large vehicle proportions under different conditions, as well as the factors influencing this tolerance, is essential.

Despite the increasing research on traffic flow and large vehicles in recent years, studies specifically addressing small vehicle drivers' tolerance to the proportion of large vehicles remain scarce. Existing studies have predominantly focused on macro-level issues, with less attention given to individual differences in driving behaviour. For instance, prior research has concentrated on the effects of different road sections, types of large vehicles, or specific time periods on traffic flow, overlooking how small vehicle drivers' tolerance to large vehicle proportions varies across different road conditions and traffic patterns. Previous research methods have mainly relied on traffic flow models, road capacity analysis and large-scale traffic surveys, with limited focus on the factors influencing individual drivers' tolerance. Therefore, this study aims to fill this gap by exploring small vehicle drivers' tolerance to large vehicle proportions and its influencing factors under varying conditions.

This study aims to investigate the differences in small vehicle drivers' tolerance towards large vehicle proportions, influenced by multiple factors such as road conditions, traffic patterns and driving experience. Specifically, we will examine how road type, type of large vehicles and time affect drivers' attitudes and decisions regarding large vehicle proportions. Furthermore, this study will explore tolerance differences across various regions and traffic environments, revealing how these differences affect road management and policy development. By conducting this research, we aim to deepen our understanding of driving behaviour and route choice decisions, ultimately contributing to traffic flow optimisation and providing scientific support for improving road service levels. This study will also offer theoretical support for traffic flow optimisation and road safety while providing practical insights for the formulation of relevant policies, ultimately advancing traffic management practices.

2. LITERATURE REVIEW

Numerous studies have thoroughly examined the diverse impacts of large vehicle proportions on road capacity and traffic conditions. The use of connected and automated heavy vehicles (CAHV) in construction zones has significantly reduced delays and queue lengths, particularly in areas with a high proportion of large vehicles. Improving the cooperative efficiency of large vehicles through the application of intelligent technologies can enhance overall traffic flow and road capacity in these regions [3]. A refined method has been proposed for adjusting saturated flow by examining the interactions between lane width, the proportion of large vehicles and the share of left-turning vehicles, leading to improved accuracy in road saturation flow predictions [4]. It was observed that an increase in the proportion of large vehicles may lead to a 5% to 11% reduction in average saturation headway at intersections under oversaturated conditions. Furthermore, regression models have been developed to estimate the reduction in intersection capacity across varying proportions of large vehicles, providing valuable insights for traffic management strategies [5]. Advanced deep learning models incorporating multi-source data fusion have proven effective in predicting traffic flow patterns for various vehicle types. These methods, particularly for predicting large vehicle flow over extended periods, contribute to a better understanding of how large vehicle proportions influence traffic conditions and allow for improved responses to their effects [6]. Although these studies have highlighted the effects of large vehicle proportions on road efficiency, there is still a lack of systematic analysis regarding drivers' psychological and behavioural responses to such changes.

The increasing proportion of large vehicles is closely linked to road traffic safety risks. Previous research has shown that a high mixing ratio of large vehicles not only increases the likelihood of traffic accidents but also exacerbates their severity. Studies have identified that the lower braking capabilities of large vehicles contribute to an increased likelihood of rear-end collisions within traffic flow [7]. An innovative method has been proposed to assess traffic conflict risks by incorporating vehicle weight as a key influencing factor. Results indicate that as the proportion of large vehicles rises, the risk of traffic conflicts increases significantly [8]. Spatial analysis methods have also been applied to identify high-incidence locations for large vehicle accidents by examining accident frequency, the proportion of large vehicles and accident severity. These findings provide important insights for traffic management agencies in developing targeted safety improvement measures [9]. Concerning accident severity, low-speed driving of large vehicles on inclines has been linked to severe rear-end collisions, further increasing accident severity [10]. Further investigations have analysed the severity of single-vehicle heavy truck accidents by considering various influencing factors such as driver characteristics, vehicle technology and road conditions. Recommendations have been made to address these risks through educational, technological and engineering measures [11]. In addition, evidence indicates that large pickup trucks, characterised by their elevated and blunt front-end design, pose a significantly higher risk of fatal injuries to pedestrians, particularly when compared to other vehicle types [12]. Analysis of large vehicle accident data has revealed that the fatality rate in large vehicle accidents is significantly higher than that of small vehicles. These findings underscore the elevated risks associated with large vehicles and the importance of addressing them in traffic safety management strategies [13]. Although these studies reveal the negative effects of large vehicle proportions on traffic safety, further exploration is needed on how to effectively mitigate these risks through optimising driver route choices and adjusting traffic strategies.

To gain a deeper understanding of drivers' tolerance for the proportion of large vehicles and the factors influencing it, the choice of appropriate analytical methods is essential. For instance, researchers have applied logistic regression to explore factors such as individual attributes, travel modes and perceptions of safety that influence commuters' choices [14]. Binary logistic regression has been used to analyse the severity of accidents involving large trucks, highlighting factors such as driver fatigue and lane deviations [15]. Multinomial logistic regression has been employed to examine travel pattern choices, identifying primary travel modes and their effects on distance and reliance [16]. Ordered logistic regression has proven effective in analysing injury severity in large truck rollover accidents, showcasing its flexibility in handling ordered outcome variables [17]. Additionally, during the pandemic, logistic regression was utilised to study the influence of perceived safety on public transportation use, revealing shifts in travel preferences [18]. A study combined classification tree models and logistic regression to analyse the acceleration and climbing characteristics of pure electric large vehicles. This analysis led to the proposal of new design standards for motorway gradients and lengths to better accommodate the specific requirements of electric large vehicles [19]. Although logistic regression models perform well in addressing binary choice problems and effectively explaining the relationship between individual characteristics and behavioural choices, their data classification is relatively simplistic, making it difficult to fully capture the multi-layered characteristics of drivers' tolerance. Therefore, this study employs a K-means clustering model to categorise drivers into three groups based on their tolerance of the proportion of large vehicles: "low tolerance", "medium tolerance" and "high tolerance". Further analysis using an ordered logistic regression model enables effective examination of the associations between continuous variables and categorical outcomes. This methodological combination addresses the limitations of traditional models, providing new perspectives for understanding and predicting driver behaviour.

This study utilises K-means clustering and ordered logistic regression models to investigate drivers' tolerance for the proportion of large vehicles, addressing critical gaps overlooked in previous research. The research objectives are articulated along two dimensions. From the perspective of the research subject, this study examines how drivers' tolerance for the proportion of large vehicles varies across different driving scenarios, aiming to fill a significant gap in this area of research. From the perspective of results analysis, the study employs ordered logistic regression models to systematically evaluate the influence of various factors on drivers' tolerance and the magnitude of their effects. Specifically, it explores how factors such as road types, time distribution and vehicle classifications impact tolerance levels. By clarifying the core factors shaping drivers' tolerance and uncovering their underlying mechanisms, this study provides a robust scientific basis for optimising road design and traffic management strategies, ultimately enhancing traffic efficiency and overall safety.

3. RESEARCH DESIGN

3.1 Questionnaire design

The questionnaire consists of two parts: the basic information section and the scenario experimental design. The basic information section collects respondents' personal information, driving habits and socio-economic attributes as influencing factors. The scenario experimental design examines the effects of four factors – the proportion of large vehicles, types of large vehicles, travel time and road types – on drivers' tolerance for large vehicles. Each factor's parameter values (levels) are shown in *Table 1* below.

Table 1 – Experimental factors and their levels

Variable level	Proportion of large vehicles	Type of large vehicles	Travel time	Road type
1	Small	Coach	27 minutes	Regular road
2	Moderate	Bus	25 minutes	Motorway
3	Large	Truck	23 minutes	Expressway
4				Main road
5				Secondary road
6				Branch road

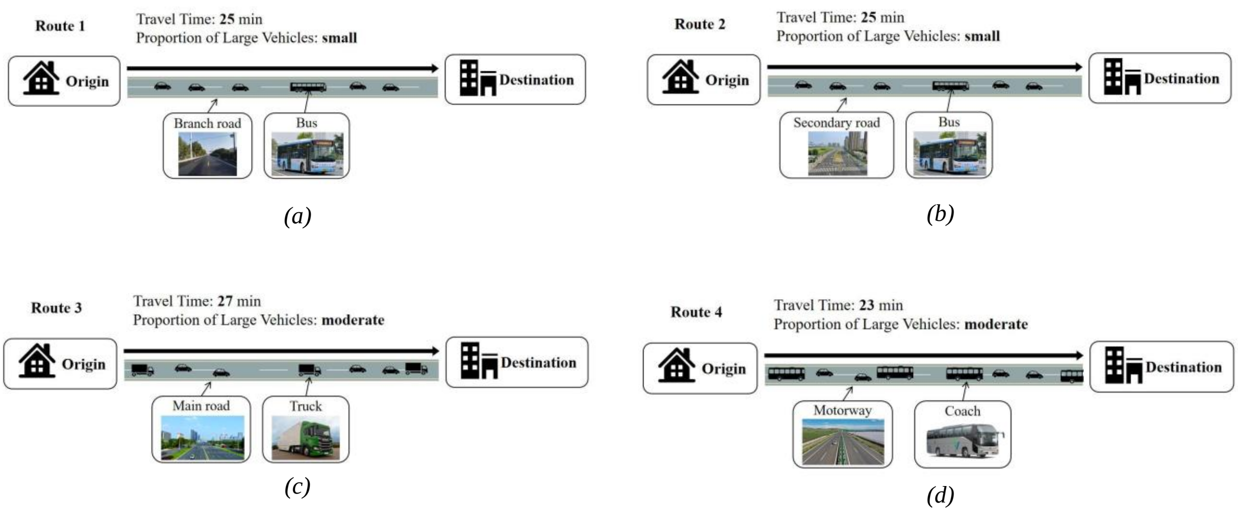
Considering the differences in respondents' perceptions of the proportion of large vehicles categorised as 'small', 'moderate' and 'large', this study adopts a combined scenario design, using a mix of images and text to illustrate the proportion of large vehicles under different road conditions. Detailed demographic and personal characteristics of the respondents are provided in the descriptive statistics section. The types of large vehicles (coach, bus, truck) were selected based on their prevalence and significant impact on traffic flow, as documented in previous traffic studies [20]. According to the Guideline on the design of traffic operation of urban roads (GB/T 36670-2018) [21] and the Technical Standard of Highway Engineering (JTG B01-2014) [22], roads are classified into six categories: regular road, motorway, expressway, main road, secondary road and branch road. Finally, the time intervals (27, 25 and 23 minutes) were determined based on route choice studies, simulating realistic peak-hour traffic scenarios to facilitate respondent decision-making [23]. The specific levels of each influencing factor are depicted through images to enhance respondents' understanding and perception of the presented road conditions.

The experiment involves four factors, each with levels of 3, 3, 3 and 6, respectively. A full factorial design would require 162 trials, whereas an orthogonal experimental combination requires 12 trials. However, considering the efficiency and complexity arising from the mixed levels of the road type factor, a mixed-level uniform design table U6*(64) was adopted, resulting in an optimised experimental design requiring only 6 trials. The uniform design has been demonstrated to be effective in capturing the key interactions between factors while significantly reducing the number of trials, especially in cases involving mixed-level factors [24]. In this study, the selected design ensures an even distribution of experimental points across the factor space, optimally balancing experimental efficiency and accuracy, yielding the uniform experimental results shown in *Table 2*.

When completing the questionnaire, respondents will be presented with road conditions for each travel scenario through a combination of images and text. Each scenario is characterised by four key attributes: the proportion and types of large vehicles, travel time and road types. Respondents are required to assess their tolerance to six different road scenarios using a five-point Likert scale. *Figure 1* illustrates the 6 scenarios.

Table 2 – Uniformly designed experimental scenarios

Scenario	Proportion of large vehicles	Type of large vehicles	Travel time	Road type
1	Small	Bus	25 minutes	Branch road
2	Small	Bus	23 minutes	Secondary road
3	Moderate	truck	27 minutes	Main road
4	Moderate	Coach	23 minutes	Motorway
5	Large	Coach	27 minutes	Expressway
6	Large	truck	25 minutes	Regular road



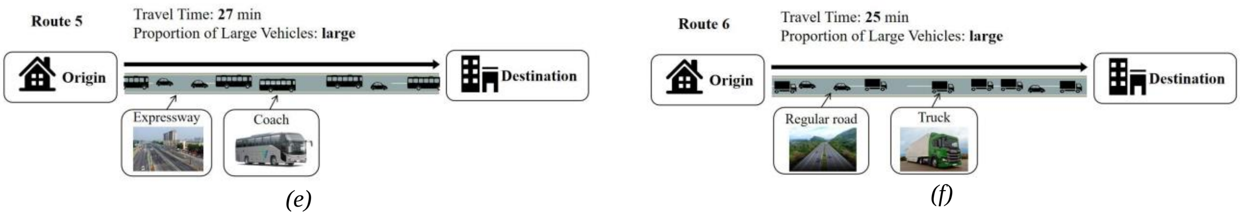


Figure 1 – Uniformly designed road scenarios:
a) Route 1; b) Route 2; c) Route 3; d) Route 4; c) Route 5; d) Route 6

3.2 Survey implementation

Data were collected through a combination of online and offline surveys. The first part of the questionnaire requires drivers to provide basic information based on their circumstances, while the second part presents six scenarios of different road attributes through a combination of images and text. The survey was conducted from 16 April to 22 April 2022, focusing on licensed drivers in China. Prior to administering the tolerance survey on the proportion of large vehicles, respondents were required to confirm possession of a valid driver's licence. Respondents who confirmed their status as licenced drivers proceeded with the survey, while those without a valid licence were excluded from participation. This ensures that participants possess practical driving experience, which is crucial for the validity of the results. By focusing on licensed drivers, we ensure that respondents have the necessary knowledge and experience to accurately assess their tolerance to the scenarios, thereby improving the reliability and relevance of the findings. A total of 1,662 responses were collected, of which 1,596 were valid, resulting in a response rate of 96.03%.

4. DATA PROCESSING AND VALIDATION

4.1 Descriptive statistics

The basic information from the survey results is presented in *Table 3*.

Table 3 – Personal attributes

Personal factors	Variable	Percentage
Occupation	Employed	63%
	Unemployed	37%
Age (years)	18~35	28%
	35~45	38%
	45~60	31%
	>60	3%
Gender	Male	54%
	Female	46%
Household annual income (RMB/year)	<100,000	40%
	100,000~500,000	48%
	500,000~1,000,000	10%
	>1,000,000	2%
Weekly driving frequency (driving occurrences/week)	0	23%
	1~5	39%
	6~15	29%
	16~30	7%
	> 30	2%

Personal factors	Variable	Percentage
Driving duration (years)	0	16%
	1~3	29%
	3~6	30%
	6~10	15%
	> 10	10%
Household vehicle usage (vehicles)	0	58%
	1	34%
	2~4	7%
Peak congestion index	Smooth (1~1.2)	31%
	Basically smooth (1.2~1.4)	11%
	Light congestion (1.4~1.6)	19%
	Moderate congestion (1.6~1.8)	25%
	Severe congestion (> 1.8)	14%
Population (persons)	1~2 million	30%
	2~5 million	12%
	5~10 million	27%
	10~20 million	18%
	> 20 million	13%
GDP (RMB/year)	< 100 billion	24%
	100~1000 billion	30%
	1000~2000 billion	27%
	> 2000 billion	19%
Vehicle ownership (vehicles)	Below 1 million	40%
	1~2 million	24%
	2~3 million	26%
	> 3 million	10%

Gender and occupation were encoded using a binary coding system (0-1), transforming them into binary classification variables. Other factors were classified based on their magnitude to form ordinal variables, with codes incrementing from 1 to 5.

In the surveyed population, drivers aged between 36 and 45 years accounted for 38.0%, while the proportion of elderly drivers aged over 60 was the lowest, showing the predominance of middle-aged and young drivers among road traffic participants. The majority of the sample had driving experiences of 3 to 6 years (30.1%), with only 9.8% having over 10 years of experience. Additionally, 91.4% of respondents reported driving 15 times or fewer per week, which aligns with the commuting and travel needs of drivers who typically drive 1 to 2 times daily. Socioeconomic data include information about the population, GDP, vehicle ownership and peak congestion index of the cities where respondents reside. The distribution of urban populations and GDP among the samples is relatively even, with the majority of vehicle ownership in cities being below one million, accounting for 39.8%. The most common urban peak congestion index is categorised as generally smooth (1-1.2), making up 31.2% of responses. This reflects the representation of drivers from various types of cities across China.

4.2 Validity and reliability testing

Validity testing is used to assess the reliability and validity of the results from the questionnaire survey. Reliability analysis commonly employs Cronbach's α and composite reliability (CR) to measure the correlation and consistency among questionnaire items. Validity analysis generally utilises average variance extracted (AVE) to evaluate the measurement validity of the questionnaire. The results of the validity and reliability testing for the latent variables related to drivers' basic information and perceptions are presented in *Table 4*.

Table 4 presents the reliability and validity results for these latent variables. The "Driving factors" refer to factors directly associated with driving behaviour and vehicle usage, measured through three indicators: respondents' weekly driving frequency, driving duration and household vehicle usage. The "Urban factors" capture the influence of urban traffic conditions and socioeconomic attributes, measured by four indicators: the peak congestion index, the population, GDP and vehicle ownership in the city where the respondents reside. The "Tolerance score" measures the respondents' tolerance of road attributes, assessed based on their evaluations of six road scenarios in the questionnaire.

Table 4 – Reliability and validity tests for latent variables

Latent variable	Items	Cronbach's α	AVE	CR
Driving factors	Weekly driving frequency	0.658	0.593	0.811
	Driving duration			
	Household vehicle usage			
Urban factors	Peak congestion index	0.965	0.891	0.970
	Population			
	GDP			
	Vehicle ownership			
Tolerance score	Tolerance for Road 1	0.820	0.521	0.867
	Tolerance for Road 2			
	Tolerance for Road 3			
	Tolerance for Road 4			
	Tolerance for Road 5			
	Tolerance for Road 6			

The overall reliability, indicated by Cronbach's α, was 0.788, which is greater than the acceptable threshold of 0.7, suggesting a good internal consistency of the questionnaire. The Cronbach's α values for the latent variables ranged from 0.658 to 0.965, all exceeding 0.6, indicating that the questionnaire design meets the established requirements. The composite reliability (CR) values were between 0.811 and 0.970, all above 0.8, demonstrating strong consistency among the latent variables. The KMO and Bartlett's sphericity tests yielded a KMO value of 0.821, exceeding the standard value of 0.5, and Bartlett's test achieved statistical significance ($p < 0.01$), confirming the adequacy of the sample. The AVE values for the latent variables ranged from 0.521 to 0.891, indicating good content validity of the questionnaire.

4.3 Correlation analysis

To identify the factors influencing drivers' tolerance of the proportion of large vehicles, correlation analysis was conducted on eleven selected influencing factors. Given that age and occupation are binary categorical variables while the other factors are ordinal variables, Spearman's correlation coefficient was used to analyse the relationships between ordinal and categorical variables, as well as among ordinal variables. Chi-square tests were employed to analyse the correlations between categorical variables. The results are shown in *Figure 2*.

The correlation analysis revealed significant correlations between occupation and gender with annual household income, household vehicle usage, driving duration and weekly driving frequency. Additionally, a strong correlation was observed between urban population and urban GDP, urban vehicle ownership and urban peak congestion index, with correlation coefficients exceeding 0.7. The occupation and gender factors exhibit considerable overlap with other household characteristics and driving behaviour variables, while the urban population factor is highly correlated with other urban-related factors. Including these variables in the model would not enhance its explanatory power; instead, it would exacerbate multicollinearity, which could undermine the independent contribution of each factor. To improve the model's robustness, minimise redundancy and ensure the independent effect of each factor on tolerance, these variables were excluded.

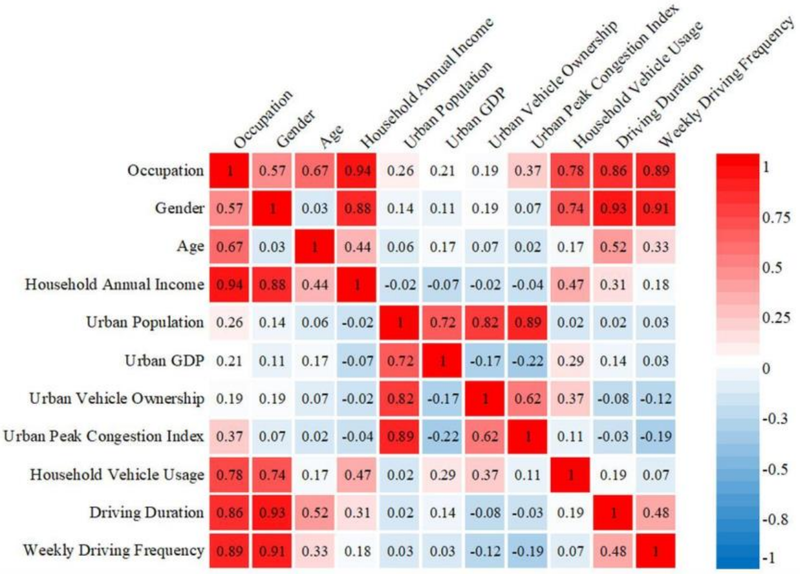


Figure 2 – Correlation coefficient of factors influencing the tolerance of large vehicle proportions

5. CLASSIFICATION OF LARGE VEHICLE PROPORTION TOLERANCE

5.1 K-means clustering algorithm

The K-means clustering algorithm is notable for its simplicity, effectiveness, rapid processing speed and widespread application in data mining. This study develops a classification model for determining tolerance levels of large vehicle proportions based on drivers' perceptions, employing the K-means clustering algorithm.

K-means clustering partitions data into a predetermined number of clusters, denoted as k, while minimising the sum of squared errors (SSE). The similarity between two objects is evaluated based on the distance between them; the smaller the distance, the greater the similarity. The core idea of executing the K-means algorithm is to minimise the squared error sum for the clustering partition $\{E_1|l = 1,2, \dots, k\}$ derived from the dataset $D = \{X_1, X_2, \dots X_m\}$. The formula for calculating SSE is as follows:

$$SSE = \sum_{l=1}^{k} \sum_{x \in E_l} dist(\mu_l, x)^2 \tag{1}$$

where k represents the number of clusters; E_l is the lth cluster; x is the sample; and μ_l is the centroid of cluster E_l.

To validate the stability and feasibility of the K-means algorithm, this study incorporates several other commonly used clustering algorithms for comparative analysis, including hierarchical clustering and the Gaussian mixture model (GMM). These comparisons aid in determining the most appropriate clustering method for the dataset, ensuring the robustness of the results. The performance of each algorithm is assessed using several evaluation metrics, including the silhouette coefficient and the Davies-Bouldin index (DB index). The silhouette coefficient measures the similarity of each point to its own cluster relative to other clusters, while the Davies-Bouldin index assesses the average similarity between each cluster and its most similar counterpart.

As shown in *Table 5*, the K-means method exhibits the highest silhouette coefficient (0.69) and the lowest DB index (0.95), indicating clear separation between the clusters. In contrast, the hierarchical clustering and GMM methods show lower silhouette coefficients and higher DB indices, suggesting weaker cohesion and less distinct separation between clusters. These results indicate that the K-means method is the most suitable clustering algorithm for this dataset, ensuring the validity and interpretability of the clustering results and providing a robust foundation for further analysis.

Table 5 – Comparison of different clustering methods

Clustering method	Silhouette coefficient	DB index	Number of clusters (k)
K-means	0.69	0.95	3
Hierarchical clustering	0.55	2.02	3
GMM	0.62	1.45	3

5.2 Driver's tolerance of large vehicle proportions

This study utilises the silhouette coefficient to determine the optimal number of clusters for the K-means algorithm, as illustrated in *Figure 3*. The silhouette coefficient is a widely adopted metric for evaluating clustering quality, particularly in selecting the most appropriate number of clusters. By calculating the silhouette values for various cluster numbers, the optimal number is identified, which maximises intra-cluster cohesion and inter-cluster separation. A higher silhouette value is indicative of more effective clustering performance, suggesting clearer and better-separated clusters.

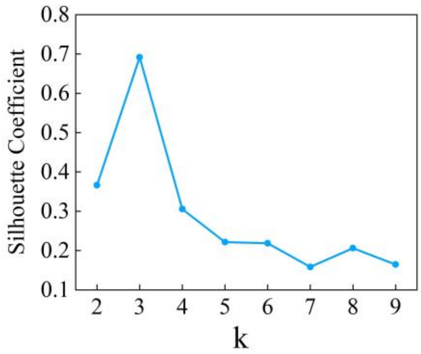


Figure 3 – Silhouette coefficient diagram

The results shown in *Figure 3* indicate that the silhouette coefficient reaches its maximum value when K=3, signifying optimal clustering performance. Consequently, the K-means clustering algorithm categorises drivers' tolerance levels into three distinct classes.

The clustering results reveal that among the 1,596 drivers, the three categories account for 6.77%, 51.13% and 42.10% of the total population, respectively. Drivers in Category 1 exhibit low tolerance levels across six scenarios, indicating sensitivity to the proportion of large vehicles and a preference for routes with fewer or no large vehicles to ensure driving safety. Conversely, drivers in Category 3 show high tolerance across all six roads, suggesting that this group is less influenced by the proportion of large vehicles and has a broader tolerance range. Based on this, and in conjunction with relevant literature [25][26], the three categories are defined as follows:

Low Tolerance: Drivers exhibit a narrow range of tolerance toward large vehicle proportions, preferring routes with fewer or no large vehicles when selecting paths.

Moderate Tolerance: Drivers have a moderate tolerance level of large vehicle proportions, allowing for the presence of large vehicles on the road but maintaining a certain ratio with smaller vehicles.

High Tolerance: Drivers are less influenced by large vehicles, with most able to accept their presence during driving.

This study utilises the Kolmogorov-Smirnov (KS) test and the Kruskal-Wallis (KW) test to further verify significant differences in tolerance levels of large vehicle proportions among different categories [27][28]. Results from the KS test indicate that the distributions of scores for the six roads are non-normally distributed, and the KW test shows significant differences in tolerance of large vehicle proportions scores among categories for the six roads ($p < 0.05$), thereby validating the rationale of the clustering.

To ensure the accuracy of the clustering results for "tolerance levels of large vehicle proportions", this study employs an experimental design with partially fixed control variables, namely fixing the types of large vehicles, temporal distribution and road types while systematically varying the proportion of large vehicles. Repeated clustering analyses were conducted to observe the consistency of category divisions, as shown in *Table 6*. The silhouette coefficient remains consistent under these four experimental conditions. The minor fluctuations in the silhouette coefficient values (ranging from 0.62 to 0.67) further validate the consistency and reliability of the clustering results.

Table 6 – Comparison of clustering results under fixed variable conditions

Experimental design	Number of categories	Silhouette coefficient	Sample proportion
Original clustering result (retain all features)	3	0.67	100%
Fixed road type (road type: motorway)	3	0.62	32%
Fixed travel time (travel time: 25 minutes)	3	0.64	38%
Fixed type of large vehicles (type of large vehicle: bus)	3	0.66	41%

6. ANALYSIS OF INFLUENCING FACTORS ON TOLERANCE LEVELS OF LARGE VEHICLE PROPORTIONS

6.1 Ordinal logistic regression model

This study employs an ordinal logistic regression model to investigate the extent to which various influencing factors affect the tolerance levels of large vehicle proportions. The model is represented by *Equation 2*.

$$P(Y \leq i|X_1, X_2, \dots, X_j) = \frac{exp(a_i + b_1X_1 + b_2X_2 + \cdots b_jX_j)}{1 + exp(a_i + b_1X_1 + b_2X_2 + \cdots b_jX_j)} \tag{2}$$

In the equation, i denotes the type of tolerance level for large vehicle proportions; $P(Y \leq i|X)$ represents the probability that the tolerance level is less than or equal to i; $X_1, X_2, \dots, X_j$ encompasses j influencing factors, including annual household income, household vehicle ownership, urban GDP, urban vehicle ownership, urban peak congestion index, age, driving duration and weekly driving frequency; a_i is the corresponding intercept of the model; and $b_1, b_2, \dots, b_j$ represents the regression coefficients.

6.2 Regression result analysis

Based on the results of the correlation analysis, four distinct ordered logistic regression models were constructed: Model (1) retains all factors; Model (2) excludes gender and occupation factors based on correlation analysis; Model (3) removes demographic factors also based on correlation analysis; and Model (4) simultaneously excludes gender, occupation and demographic factors while considering the results of the correlation analysis. The fit of the models was compared using the -2 log-likelihood criterion, with the results shown in *Table 7*. The findings indicate that Model (4) exhibits the best fit, demonstrating an enhancement in the model's explanatory power and robustness after the exclusion of gender, occupation and demographic factors during the model construction process based on the correlation analysis.

Table 7 – Comparison of fitting effects of different ordered logistic regression models

Model	Description	-2 Log-likelihood
Model (1)	Retain all factors	1192.94
Model (2)	Remove gender and occupation	1155.77
Model (3)	Remove population	1195.57
Model (4)	Remove gender, occupation and population	1147.74

Table 8 presents the results of the ordered logistic regression model after removing gender, occupation and demographic factors, highlighting significant differences in the impact of various factors on drivers' tolerance of large vehicle proportions.

Table 8 – Ordinal logistic regression analysis of the tolerance of large vehicle proportion

	Variable	Coefficient	p-value
Threshold	Low tolerance	-8.973	0
	Moderate tolerance	-1.106	0
Influencing factors	Household annual income	5.18	0
	Household vehicle usage	5.122	0
	Urban GDP	-7.91	0
	Urban vehicle ownership	-1.441	0.001
	Urban peak congestion index	-1.918	0
	Age	0.067	0.84
	Driving duration	0.704	0.045
	Weekly driving frequency	2.074	0

Household annual income and household vehicle usage

Both household annual income and household vehicle usage show a positive correlation with the tolerance of the large vehicle proportion, significant at the 0.01 level. This may be attributed to the fact that drivers with higher household incomes and more personal vehicles tend to have a more positive attitude toward travel. They are more lenient in assessing road risks, leading to a higher tolerance of the large vehicle proportion.

Urban GDP, urban vehicle ownership and urban peak congestion index

Urban GDP, urban vehicle ownership, and the urban peak congestion index exhibit a negative correlation with the tolerance of the large vehicle proportion, significant at the 0.01 level. As urban GDP increases, urban vehicle ownership rises, leading to increased traffic flow and heightened peak congestion, which exacerbates traffic congestion issues. Compared to smaller vehicles, large vehicles occupy more space on the road, increasing the risk of collisions with surrounding smaller vehicles and having a greater impact on their movement, thereby decreasing drivers' tolerance of large vehicles.

For drivers, the growth of urban GDP typically accelerates the pace of life, increasing dependency on urban space and the demand for travel efficiency. Consequently, drivers may find large vehicles inconvenient and disruptive, further reducing their tolerance of these vehicles.

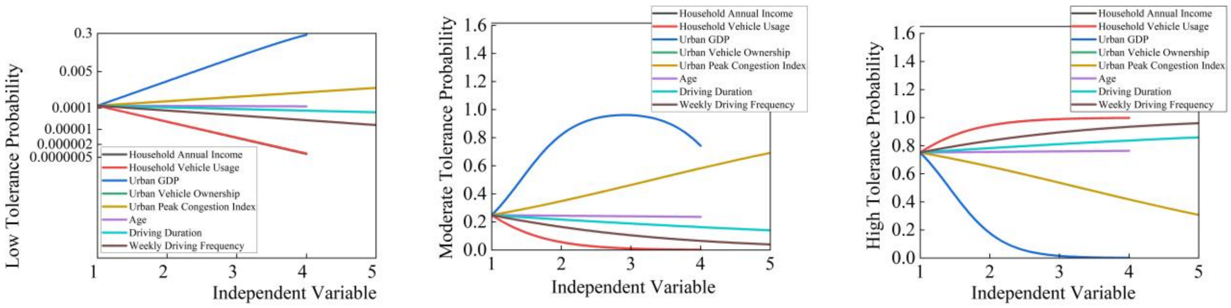
Driving duration and weekly driving frequency

Age, driving duration and weekly driving frequency all show a positive correlation with the tolerance of the large vehicle proportion, with weekly driving frequency significant at the 0.01 level and driving duration significant at the 0.05 level. As drivers accumulate more time and increase their weekly driving frequency, their driving skills become more refined. This allows them to better handle complex traffic situations, thereby enhancing their tolerance of the large vehicle proportion and placing greater emphasis on road accessibility and driving time costs.

7. MARGINAL BENEFIT ANALYSIS OF LARGE VEHICLE PROPORTION TOLERANCE

To explore the changing trends of various factors influencing the tolerance of large vehicle proportions, the marginal effects of these factors are analysed based on the results of the ordered logistic regression analysis, as shown in *Figure 4*. Due to the similar coefficients between household annual income, household vehicle usage, urban peak congestion index and urban vehicle ownership parameters, the corresponding curves in the line graph exhibit overlapping phenomena.

The results indicate that the speed of influence from each factor on the tolerance of large vehicle proportions varies. Based on the probability of changes in large vehicle tolerance, the influencing factors can be roughly categorised into four types. Urban GDP, urban peak congestion index and urban vehicle ownership show a negative correlation with large vehicle tolerance; thus, the probability of low tolerance increases as the independent variables rise. As GDP increases, the probability of moderate tolerance first increases and then decreases, while the probability of high tolerance decreases to zero, with a slowing rate of change. Due to differences in coefficient magnitudes, urban vehicle ownership and urban peak, the congestion index exhibits variations in the probabilities of moderate and high tolerance compared to GDP. As the urban peak congestion index and urban vehicle ownership increase, the probability of moderate tolerance rises while that of high tolerance declines, maintaining a constant rate of change. Household annual income, household vehicle usage, age, driving duration and weekly driving frequency are positively correlated with large vehicle tolerance. As the independent variables increase, the probability of low tolerance approaches zero, while the probabilities of moderate tolerance decrease and high tolerance probabilities increase. Specifically, as age and driving duration increase, the probability of moderate tolerance shows a linear downward trend, while the probability of high tolerance shows a linear upward trend. Increases in household vehicle usage, household annual income and weekly driving frequency slow the rate of decrease in the probability of moderate tolerance and also slow the rate of increase in the probability of high tolerance.



(a) *(b)* *(c)*

Figure 4 – Analysis of the magnitude of changes in factors influencing the tolerance of large vehicle proportions: a) Low tolerance; b) Moderate tolerance; c) High tolerance

Furthermore, the interaction effects of two factors on the tolerance of large vehicle proportions are analysed, as shown in *Figure 4*. Based on the marginal effects analysis of individual factors, the factors with the highest coefficients in each group (household annual income, urban GDP and weekly driving frequency) were selected for interaction analysis.

The results in *Figure 5* indicate that under the interaction of household annual income and weekly driving frequency, the probabilities of low and moderate tolerance are relatively low, while the probability of high tolerance is relatively high. As household annual income and weekly driving frequency increase, the probabilities of low and moderate tolerance decrease towards zero, while the probability of high tolerance approaches one, which aligns with the positive correlation between household annual income, weekly driving frequency and tolerance.

The interaction analysis graphs of urban GDP with household annual income and weekly driving frequency exhibit similar trends, with the probability of low tolerance approaching zero as GDP decreases and both weekly driving frequency and household annual income increase.

The marginal effects of the dual factors on the tolerance of the large vehicle ratio differ from those of the single factors. As shown in *Figure 4*, as urban GDP increases, the probability of moderate tolerance of the large vehicle ratio first increases and then decreases. In the interaction between weekly driving frequency and household annual income with urban GDP, both weekly driving frequency and household annual income are positively correlated with the tolerance of the large vehicle ratio. Therefore, when both weekly driving frequency and household annual income increase, the probability of moderate tolerance of the large vehicle ratio changes from initially increasing and then decreasing to consistently increasing as urban GDP rises. Similarly, when household annual income and weekly driving frequency are at lower values, the decreasing trend of high tolerance probability slows down with increasing urban GDP. Conversely, when household annual income and weekly driving frequency are at higher values, the decline in high tolerance probability accelerates as urban GDP increases.

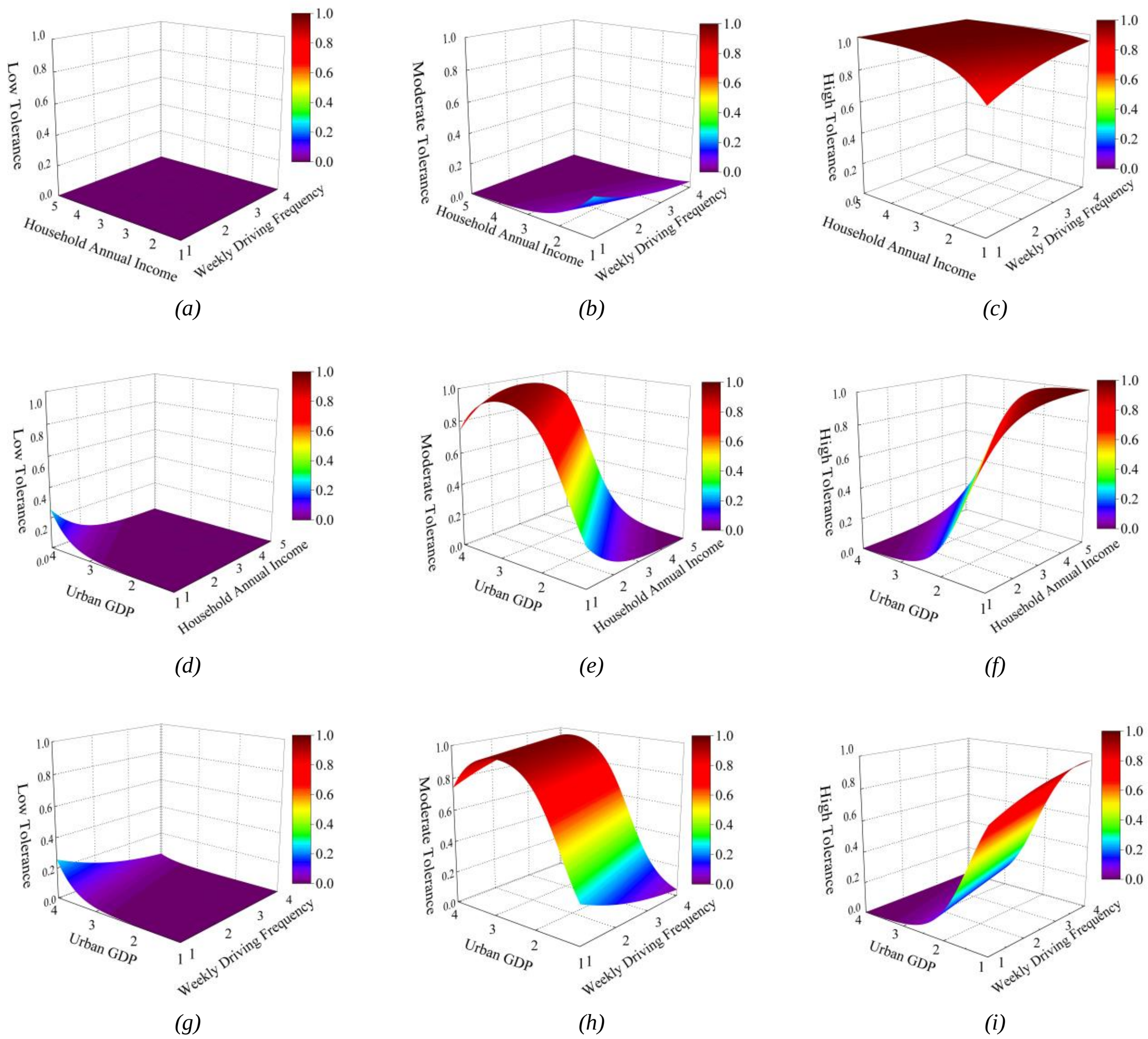


Figure 5 – Two-factor cross-analysis of the tolerance of large vehicles:
a) Low tolerance (income & frequency); b) Moderate tolerance (income & frequency); c) High tolerance (income & frequency)
d) Low tolerance (GDP & income); e) Moderate tolerance (GDP & income); f) High tolerance (GDP & income)
g) Low tolerance (GDP & frequency); h) Moderate tolerance (GDP & frequency); i) High tolerance (GDP & frequency)

8. CONCLUSION

This study constructs a model to classify drivers' tolerance of large vehicle proportions by examining how various factors, including personal information, driving habits, socioeconomic attributes and road traffic environment, influence this tolerance. The findings are as follows.

1) By employing the K-means clustering algorithm, drivers' tolerance of large vehicle proportions was categorised into three levels: "low tolerance", "medium tolerance" and "high tolerance", with respective proportions of 6.77%, 51.13% and 42.10%. These findings indicate that the majority of drivers exhibit medium to high tolerance toward large vehicles, reflecting a general acceptance of higher proportions of large vehicles on the road. Nevertheless, the 6.77% of drivers classified in the low-tolerance group underscores a non-negligible minority who may experience significant discomfort or inconvenience in environments characterised by a high proportion of large vehicles. This heterogeneity underscores the importance of implementing tailored strategies in transportation planning to accommodate the varying tolerance levels of drivers. Examples of such measures include dedicating specific lanes for large vehicles, improving road infrastructure to mitigate congestion and enhancing public awareness of the benefits and challenges associated with large vehicles on the road.

2) The ordered logistic regression model identified several key factors influencing drivers' tolerance of large vehicle proportions. These factors include household annual income, household vehicle usage, weekly driving frequency and driving duration, all of which exert a significant positive impact on tolerance. Drivers with higher income, more frequent driving habits and greater driving experience are generally more tolerant of large vehicle proportions. Conversely, urban GDP, urban vehicle ownership and the urban peak congestion index negatively influence tolerance, with urban vehicle ownership emerging as the most influential factor, followed by driving duration. The findings suggest that in cities characterised by high vehicle ownership and severe congestion, drivers are likely to face greater challenges in navigating traffic, resulting in reduced tolerance for large vehicles. To address these challenges, cities with elevated vehicle ownership and congestion levels could adopt policies to alleviate traffic density, such as promoting public transportation, implementing congestion pricing mechanisms and encouraging the use of alternative transportation modes.
3) The study further investigated how the impact of various factors on drivers' tolerance varies across different tolerance levels. The findings indicate that as the urban peak congestion index, urban vehicle ownership, age and driving duration increase, the probabilities of medium and high tolerance exhibit stable linear upward and downward trends, respectively. This suggests that the likelihood of drivers tolerating large vehicle proportions changes predictably as these factors rise. In contrast, factors such as urban GDP, weekly driving frequency, household annual income and household vehicle usage demonstrate a diminishing effect on medium and high tolerance probabilities once certain thresholds are reached. Furthermore, the study revealed that the same factor may exert varying influences on different tolerance groups, highlighting that tolerance is shaped not only by individual factors but also by their interactions with other contextual conditions.

By analysing travellers' attitudes toward large vehicle proportions based on various personal attributes and travel characteristics, this study identifies key factors influencing drivers' tolerance of large vehicle proportions. These findings hold significant practical implications for road traffic management and contribute to enhancing drivers' travel safety. However, this study has certain limitations. The sample is primarily composed of young respondents from the eastern region of China, a demographic group that is more reliant on public transportation. Consequently, the proportion of households without a vehicle in this sample is relatively high. Furthermore, although this study is geographically focused on China, its findings are not necessarily applicable to other countries, due to differences in road conditions, traffic management practices and driving cultures. Nonetheless, the methodologies, approaches and conclusions of the study offer valuable insights for broader application, particularly in regions with comparable urbanisation trends. Additionally, the data used in this study are somewhat limited in terms of coverage, and future research could incorporate a more diverse range of data sources. For example, methods such as driving simulators and video-based surveys could be employed to further explore drivers' perceptions and assess the impact of various experimental scenarios. These approaches would enhance the reliability and generalisability of the findings.

ACKNOWLEDGEMENTS

We appreciate the constructive feedback and encouragement given to us throughout the study.

REFERENCES

[1] National Bureau of Statistics. National data. 2023. https://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A0G0J&sj=2023 [Accessed 20th Dec. 2024].

[2] Kong D, et al. Analyzing the impact of trucks on traffic flow based on an improved cellular automaton model. *Discrete dynamics in nature and society.* 2016;2016(1):1236846. DOI: 10.1155/2016/1236846.

[3] Haque MMS, Rilett LR, Zhao L. Impact of platooning connected and automated heavy vehicles on interstate freeway work zone operations. *Journal of Transportation Engineering, Part A: Systems.* 2023;149(3):04022160. DOI: 10.1061/jtepbs.teeng-7434.

[4] Wang Y, et al. An analysis of the interactions between adjustment factors of saturation flow rates at signalized intersections. *Sustainability.* 2020;12(2):665. DOI: 10.3390/su12020665.

[5] Dehman A, Farooq B. Capacity characteristics of long-term work zones on signalized intersection approaches. *Transportation Research Part A: Policy and Practice*. 2023;175:103791. DOI: 10.1016/j.tra.2023.103791.

[6] Wang P, Hao W, Jin Y. Fine-grained traffic flow prediction of various vehicle types via fusion of multisource data and deep learning approaches. *IEEE Transactions on Intelligent Transportation Systems.* 2020;22(11):6921-6930. DOI: 10.1109/TITS.2020.2997412.

[7] Zhao P, Lee C. Assessing rear-end collision risk of cars and heavy vehicles on freeways using a surrogate safety measure. *Accident Analysis & Prevention.* 2018;113:149-158. DOI: 10.1016/j.aap.2018.01.033.

[8] Wang Y, et al. A novel traffic conflict risk measure considering the effect of vehicle weight. *Journal of safety research.* 2022;80:1-13. DOI: 10.1016/j.jsr.2021.09.008.

[9] Manap N, et al. Identification of hotspot segments with a risk of heavy-vehicle accidents based on spatial analysis at controlled-access highway. *Sustainability.* 2021;13(3):1487. DOI: 10.3390/su13031487.

[10] Jagatheesvaran U, et al. A study on heavy vehicle visibility and crash analysis using ANSYS. *International Journal of Nanoelectronics & Materials.* 2022;15.

[11] Rahimi E, et al. Investigating the injury severity of single-vehicle truck crashes in a developing country. *Accident Analysis & Prevention.* 2020;137:105444. DOI: 10.1016/j.aap.2020.105444.

[12] Hu W, Monfort SS, Cicchino JB. The association between passenger-vehicle front-end profiles and pedestrian injury severity in motor vehicle crashes. *Journal of Safety Research.* 2024;90:115-127. DOI:10.1016/j.jsr.2024.06.007.

[13] Hasan AS, et al. Severity analysis of heavy vehicle crashes using machine learning models: A case study in New Jersey. *International Conference on Transportation and Development 2021.* 2021:285-296.

[14] Tan L, Ma C. Choice behavior of commuters' rail transit mode during the COVID-19 pandemic based on logistic model. *Journal of Traffic and Transportation Engineering (English Edition).* 2021;8(2):186-195. DOI: 10.1016/j.jtte.2020.07.002.

[15] Hosseinzadeh A, Moeinaddini A, Ghasemzadeh A. Investigating factors affecting severity of large truck-involved crashes: Comparison of the SVM and random parameter logit model. *Journal of safety research.* 2021;77:151-160. DOI: 10.1016/j.jsr.2021.02.012.

[16] Khalid B, et al. Regression approach to analyze the travel characteristics of university students. *Transportation Letters*. 2024;1-16. DOI:10.1080/19427867.2024.2366327.

[17] Azimi G, et al. Severity analysis for large truck rollover crashes using a random parameter ordered logit model. *Accident Analysis & Prevention.* 2020;135:105355. DOI: 10.1016/j.aap.2019.105355.

[18] Aaditya B, Rahul TM. Psychological impacts of COVID-19 pandemic on the mode choice behaviour: A hybrid choice modeling approach. *Transport policy.* 2021;108:47-58. DOI: 10.1016/j.tranpol.2021.05.003.

[19] Wang Z, et al. Risk factors affecting crash injury severity for different groups of e-bike riders: A classification tree-based logistic regression model. *Journal of safety research.* 2021;76:176-183. DOI: 10.1016/j.jsr.2020.12.009.

[20] Won M. Intelligent traffic monitoring systems for vehicle classification: A survey. *IEEE Access.* 2020;8:73340-73358. DOI: 10.1109/ACCESS.2020.

[21] China. *Guideline on the design of traffic operation of urban roads.* Beijing, China: State Administration for Market Regulation; 2018.

[22] China. *Technical standard of highway engineering.* Beijing, China: Ministry of transport of the People's Republic of China; 2014.

[23] Li H, et al. Uncertainty of available range in explaining the charging choice behavior of BEV users. *Transportation Research Part A: Policy and Practice.* 2023;170:103624. DOI: 10.1016/j.tra.2023.103624.

[24] Yan W, et al. The research of diesel engine assembly consistency influence NOx emission based on uniform design method. *Combustion Science and Technology.* 2024;1-17. DOI: 10.1080/00102202.2024.2409375.

[25] Zeng Q, et al. Measuring cyclists’ subjective perceptions of the street riding environment using K-means SMOTE-RF model and street view imagery. *International Journal of Applied Earth Observation and Geoinformation.* 2024;128:103739. DOI: 10.1016/j.jag.2024.103739.

[26] Brito B, Costa DG, Silva I. Geospatial risk assessment of cyclist accidents in urban areas: A K-means clustering approach. *IEEE 22nd Mediterranean Electrotechnical Conference (MELECON) 2024.* 2024. p. 744-749. DOI: 10.1109/MELECON56669.2024.10608791.

[27] Praus P. Photoreforming for microplastics recycling: A critical review. *Journal of Environmental Chemical Engineering*. 2024:112525. DOI: 10.1016/j.jece.2024.112525.

[28] Keum D, et al. Effects of developmental alcohol exposure on cortical multisensory integration. *European Journal of Neuroscience*. 2023;57(5):784-795. DOI: 10.1111/ejn.15907.