Computers and Artificial Intelligence；Vol.3, No.3; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n3.232238
Published by: Zeus Press

Big Data-Based Analysis of Urban Traffic Congestion: A Case Study of Beijing

Zihan Zhou*

School of Software, College of Intelligence and Computing, Tianjin University, Tianjin 300354, China

**Corresponding author: Zihan Zhou.*

Abstract

As urbanization continues to reshape large cities, traffic congestion remains a persistent challenge for urban transportation systems. Using Beijing as a case study, this paper examines the spatiotemporal evolution of urban traffic congestion from a deep learning perspective based on multi-source data. The results suggest that traffic congestion in Beijing displays a pronounced "dual-peak" pattern associated with daily commuting activities. Compared with the pre-pandemic period, weekday travel demand has generally recovered and in some cases exceeded previous levels, whereas holiday travel remains relatively subdued, accompanied by increasingly concentrated travel behavior. Among the models considered, Long Short-Term Memory (LSTM) performs particularly well in capturing nonlinear variations and temporal dependencies in traffic flow, leading to improved prediction accuracy. Between 2020 and 2025, the congestion index experienced a trajectory of decline, recovery, and subsequent stabilization, a pattern that appears to be associated with the gradual implementation of intelligent traffic management measures. These findings contribute to a better understanding of recent changes in urban traffic dynamics and may offer useful insights for future traffic planning and governance.

Keywords

urban traffic congestion, big data analysis, deep learning, LSTM

1. Introduction

Rapid urbanization has brought profound changes to urban transportation systems in China. The continuous growth of population size and vehicle ownership has placed increasing pressure on existing road networks, making traffic congestion a persistent issue in many large cities. Beyond longer travel times, congestion is also associated with higher energy consumption, increased pollutant emissions, and reduced efficiency of urban operations. Understanding the evolution of congestion patterns and improving prediction accuracy have therefore become important topics in urban traffic management.

Traffic flow is influenced by a wide range of factors, including travel demand, road network conditions, weather, and public events. As a result, traffic systems often exhibit strong nonlinearity and complex spatiotemporal dependencies. Traditional approaches, such as statistical forecasting methods and conventional machine learning models, can capture certain traffic characteristics but often struggle to represent these complex interactions. Recent advances in deep learning have provided new possibilities for traffic analysis by enabling the extraction of hidden patterns from large-scale data. For instance, Cui et al. improved network-level traffic state prediction through a stacked bidirectional Long Short-Term Memory (LSTM) framework [1], while Zhang et al. employed a generative diffusion model to enhance the reconstruction of traffic states under complex conditions [2]. Existing studies generally indicate that deep learning methods can offer advantages in modelling dynamic traffic processes and handling large volumes of heterogeneous data.

Beijing provides a particularly relevant case for examining urban traffic congestion. As one of China's largest metropolitan areas, the city accommodates more than 22 million residents and over 8.1 million registered motor vehicles. Despite sustained investment in transportation infrastructure and the implementation of various traffic management policies, congestion remains a recurring challenge. At the same time, changes in travel behavior and the growing application of intelligent transportation technologies have introduced new dynamics into the city's traffic system.

Against this background, this study draws on data from Gaode Map traffic reports, the Beijing Annual Traffic Development Report, and other relevant sources to investigate the spatiotemporal evolution of traffic congestion in Beijing. Particular attention is given to congestion characteristics observed in recent years and to the potential of deep learning approaches for understanding and predicting traffic dynamics. The analysis aims to provide an empirical perspective on urban congestion patterns and offer references for future traffic management and planning.

2. Big Data Foundation for Traffic Congestion Analysis

2.1 Main Data Types

Urban traffic data are characterized by strong temporal and spatial correlations. Traffic conditions often follow regular daily cycles, particularly during morning and evening commuting periods, while current traffic states are also influenced by conditions observed in preceding time intervals. Spatially, congestion in one area may affect surrounding roads, producing spillover effects across the road network [3]. The development of big data technologies has greatly expanded the range of information available for traffic analysis. Data collected from traffic monitoring systems, vehicle-mounted devices, and communication networks can now be processed on a large scale and with relatively high frequency, providing a more detailed picture of urban traffic operations.

Among the various data sources, traffic condition data, mobile phone signaling data, and Global Positioning System (GPS) data are the most widely used. Together, they provide information on vehicle movements, traffic flow, road conditions, traffic incidents, and travel behavior. Traffic condition data are particularly important because they directly reflect the operational status of the road network, including congestion levels, travel speeds, travel times, and traffic distribution. Mobile phone signaling data record interactions between mobile devices and communication base stations, making it possible to infer population mobility patterns and travel demand. GPS data offer detailed trajectory information and are widely used to examine vehicle movements and route choices. In studies of Beijing's transportation system, these data sources are often used in combination, although traffic condition data remain the primary basis for congestion assessment and analysis [4].

2.2 Core Analysis Methods for Traffic Congestion

Traffic congestion has become an increasingly prominent issue alongside rapid urbanization. As traffic systems grow more complex, conventional planning approaches are often unable to fully capture dynamic traffic conditions. The emergence of big data technologies has therefore opened up new possibilities for traffic analysis and management. By utilizing large-scale traffic data, researchers can identify operational patterns, evaluate network performance, and support decisions aimed at improving traffic efficiency and public transportation services. Before analysis can be conducted, raw traffic data generally require preprocessing. This step is important because traffic datasets often contain missing values, noise, and inconsistencies originating from different data sources. Common preprocessing procedures include data cleaning, normalization, standardization, and dimensionality reduction. These operations help improve data quality and provide a more reliable basis for subsequent modeling.

A range of analytical approaches has been applied to traffic congestion studies. Traditional time-series models remain widely used because of their relatively clear structure, low computational requirements, and ease of interpretation. They can provide reasonable short-term forecasts of traffic flow or speed, particularly when data availability is limited. However, their performance tends to decline when dealing with highly complex traffic systems, and they often struggle to capture long-range dependencies in large-scale urban networks [3]. Machine learning methods have expanded the scope of traffic analysis by allowing larger numbers of variables to be incorporated into prediction models. They are frequently used for tasks such as traffic state prediction and anomaly detection. Nevertheless, model performance often depends on feature selection and data quality, and some algorithms may have difficulty representing complex temporal relationships within traffic systems.

More recently, deep learning techniques have attracted considerable attention in transportation research. Their ability to learn nonlinear patterns directly from large datasets makes them particularly suitable for traffic applications. For example, CNN-based models can extract information from road surveillance images, while RNN-based architectures are designed to process sequential traffic data. Among these approaches, Long Short-Term Memory (LSTM) networks have been widely adopted because they are better able to capture temporal dependencies and nonlinear variations that commonly characterize traffic flow data.

Overall, the growing availability of multi-source traffic data has encouraged the use of deep learning approaches in congestion analysis. In particular, LSTM-based models provide an effective framework for describing and predicting the evolution of urban traffic congestion, often achieving higher predictive accuracy than conventional methods.

3. Application of Deep Learning in the Transportation Field: A Beijing Case Study

3.1 Overview of Beijing's Traffic Situation

3.1.1 Causes of Traffic Congestion in Beijing

Traffic congestion has become a persistent challenge in many Chinese cities as urbanization has accelerated and travel demand has continued to expand. Beijing provides a typical example of this trend. Despite ongoing investments in transport infrastructure, the growth of vehicle ownership and commuting demand has placed considerable pressure on the urban road network. One important factor is the rapid increase in the number of private vehicles. By the end of 2025, Beijing had approximately 8.101 million registered motor vehicles, representing an increase of 219,000 vehicles compared with the previous year [5]. Relative to 2000, vehicle ownership has expanded by more than 6.5 million over the past twenty-five years. At the same time, residential development on the urban periphery has generated growing commuting flows toward employment centers, increasing pressure on major corridors during peak periods.

Travel behavior and road space allocation also contribute to congestion. Although walking and cycling have regained some share in recent years, motor vehicles remain the dominant mode for many short-distance trips [6]. Parking shortages around schools, hospitals, and other high-demand destinations frequently lead to vehicles queuing on adjacent roads, reducing roadway efficiency. Data from Beijing Traffic Broadcasting suggest that congestion near tertiary hospitals lasts considerably longer than on ordinary road sections. In addition, the spatial separation between residential areas and employment centers produces large-scale commuting flows that are concentrated within relatively limited time periods.

Infrastructure constraints further intensify congestion. Key sections of the road network, particularly major ring roads, interchanges, and entrance-exit ramps, often operate close to or beyond their designed capacity. Seasonal fluctuations in travel demand, including holiday travel and tourism peaks, can place additional pressure on already congested corridors.

Figure 1: Trends in Daily Travel Frequency of Sample Vehicles in Beijing on Holidays and Non-Holidays, 2019–2023

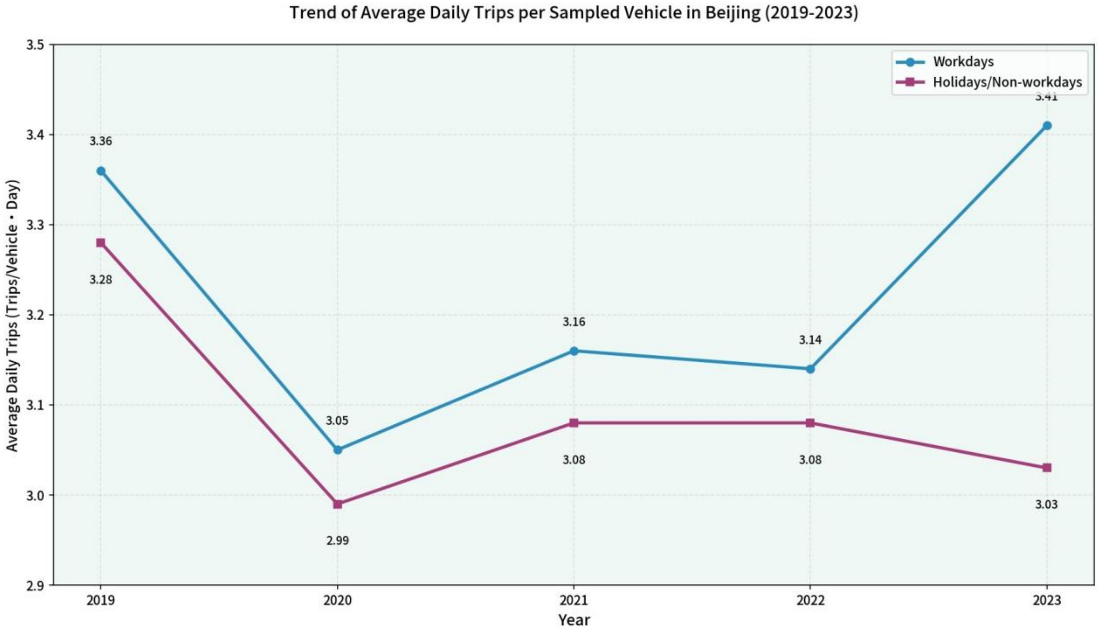

Figure 1 presents changes in the average daily travel frequency of sample vehicles in Beijing between 2019 and 2023 [5]. Travel activity declined sharply during the pandemic period, particularly in 2020. From 2020 to 2022, weekday travel gradually recovered, whereas holiday travel remained comparatively subdued. By 2023, average weekday travel had risen to 3.41 trips per day, exceeding the pre-pandemic level, while holiday travel remained below that recorded in 2019 [7]. These patterns suggest that commuting demand recovered more quickly than discretionary travel, reflecting differences in the pace of recovery across travel purposes.

3.1.2 Characteristics of Traffic Congestion in Beijing

Traffic congestion in Beijing exhibits clear temporal and spatial patterns. From a temporal perspective, congestion is concentrated during morning and evening commuting periods. Although traffic conditions fluctuate from day to day, overall patterns remain relatively stable, and congestion levels on weekdays tend to display a high degree of similarity. Previous studies have identified a pronounced dual-peak pattern, with severe congestion occurring during both the morning (7:30-9:00) and evening rush hours (17:30-19:30).

Spatially, congestion is unevenly distributed across the city and tends to cluster around major corridors, intersections, and transport nodes. Traffic flows often display tidal characteristics, with large volumes moving toward central urban areas during the morning peak and away from them during the evening peak. Congested sections are particularly common on major arterial roads and ring-road corridors, although the distribution varies across different parts of the city. In many cases, recurrent congestion occurs at similar locations and times each day, while nearby alternative routes frequently experience spillover effects due to network-wide traffic concentration.

3.2 Research on Traffic Congestion Prediction Based on Deep Learning

In recent years, deep learning has become an important approach in traffic congestion prediction. Compared with conventional statistical and machine learning methods, deep learning models are generally better suited to capturing complex temporal patterns and nonlinear relationships within large-scale traffic datasets [1–2]. These advantages have attracted growing attention in studies of Beijing's urban traffic system.

One area where deep learning has shown particular promise is the modeling of temporal dependencies in traffic flow. Traffic conditions in Beijing are influenced by recurring patterns such as morning and evening commuting peaks, weekend effects, and holiday travel. Capturing these patterns is often challenging because traffic states are affected not only by current conditions but also by events occurring over different time horizons. Recurrent Neural Networks (RNNs) and especially Long Short-Term Memory (LSTM) networks were developed to address this issue. Through their memory structures, these models are able to retain information from previous observations and learn temporal relationships that may extend over longer periods.

Deep learning models are also widely used because of their ability to represent nonlinear traffic dynamics. Traffic conditions do not change in a strictly linear manner. Incidents, adverse weather, fluctuations in travel demand, and other external factors can all trigger sudden changes in network performance. Traditional linear models often struggle to represent these transitions, whereas deep learning approaches can learn more complex relationships directly from the data. As a result, they are often applied in situations where traffic states change rapidly or unpredictably.

Another advantage of deep learning lies in its capacity to integrate information from multiple data sources. Beijing’s traffic monitoring system generates large volumes of heterogeneous data, including surveillance images, GPS trajectories, and sensor records. Different deep learning architectures can process different forms of information simultaneously. For example, CNN-based models are commonly used to extract spatial features from images, while LSTM networks are designed to analyze sequential traffic data. Combining these sources can provide a more comprehensive representation of traffic conditions and may improve prediction performance.

Figure 2: Trends in Congestion Index During Commuting Peaks in Beijing, 2020–2025

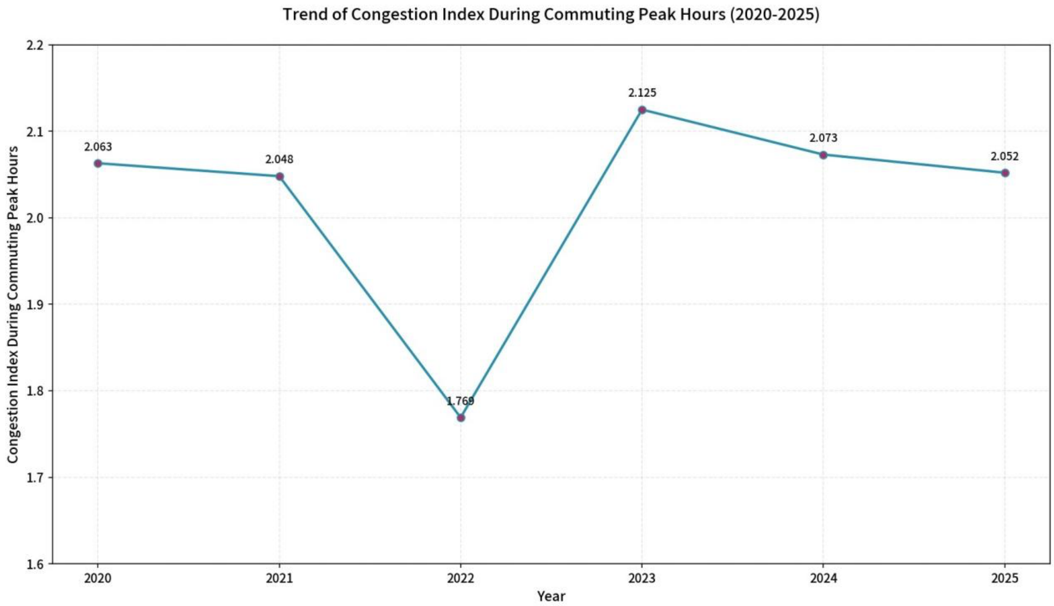

Figure 2 illustrates changes in the commuting peak congestion index in Beijing between 2020 and 2025 [5]. Overall, the congestion index followed a pattern of decline, recovery, and subsequent stabilization, which broadly corresponds to changes in travel demand during the same period. Between 2020 and 2022, the index fell from 2.063 to 1.769, representing a decrease of approximately 14.2%. This decline was largely associated with reduced travel activity during the pandemic period. In 2023, as economic and social activities resumed, commuting demand recovered rapidly and the congestion index increased to 2.125, a year-on-year rise of 20.1%. During 2024 and 2025, the congestion index remained relatively stable despite continued growth in weekday travel demand. The index fluctuated within a narrow range of 2.05–2.07 and reached 2.052 in 2025, around 3.4% lower than the 2023 peak. While holiday travel volumes remained below pre-pandemic levels, the relatively stable congestion index may also reflect the influence of traffic management measures and the continued role of public transportation in accommodating travel demand. Nevertheless, the interaction between these factors warrants further examination.

4. Research Gaps and Limitations

Despite the progress of deep learning in traffic prediction, most existing studies are still better suited to relatively stable and conventional traffic conditions. When applied to more complex urban systems such as Beijing, several limitations become more apparent.

One gap lies in the limited attention given to emerging mobility services and their impact on congestion patterns. Ride-hailing, shared bicycles, and delivery e-bikes have gradually become an important part of urban travel, but their influence is not yet fully reflected in many existing modeling frameworks. For instance, public data from Didi Chuxing indicate that daily ride-hailing orders in Beijing exceeded 2 million in 2024. Behaviors such as empty cruising and temporary roadside stopping may contribute to localized congestion, especially in central urban areas. At the same time, delivery e-bikes now operate at a very large scale, and issues such as non-standard parking and traffic violations are increasingly observed. However, these modes are still often treated as external disturbances rather than integrated variables in prediction models, which limits the explanatory power of current studies. More attention could therefore be given to how these mobility services shape congestion in both spatial and temporal dimensions, as well as how they might be incorporated into governance-oriented analysis [8].

Another limitation concerns traffic behavior under extreme or non-routine conditions. Existing models tend to perform well under regular commuting patterns but are less reliable during events such as severe weather, large public gatherings, or sudden disruptions. Some recent studies have explored the use of large-model-based systems or intelligent agents for traffic-related querying and prediction, but many of these applications remain at an early stage and are not yet widely implemented in operational traffic management. In practice, emergency response and capacity adjustment during periods such as the Spring Festival travel rush still rely heavily on rule-based scheduling and historical experience. This suggests that more robust modeling of rare or disruptive scenarios is still needed, particularly for urban road networks with high variability like Beijing.

From a longer-term perspective, future research may continue in several directions. One possible direction is the integration of multi-source mobility data, including ride-hailing trajectories and delivery vehicle movement data, into existing prediction frameworks. Another is the improvement of model robustness in low-frequency or extreme scenarios, where techniques such as data augmentation or adaptive learning could be explored. A further direction is the closer connection between data-driven models and traffic engineering practice, particularly in areas such as signal control and emergency traffic management. However, how these methods can be reliably transferred from experimental settings to real-world deployment remains an open question.

5. Conclusion

This paper examines traffic congestion in Beijing from both a spatiotemporal perspective and a modeling perspective, using multi-source traffic data combined with deep learning methods. The results suggest that congestion in Beijing is characterized by a relatively stable dual-peak pattern, with clear morning and evening commuting peaks. At the same time, travel behavior has shifted in recent years: weekday travel has returned to or slightly exceeded pre-pandemic levels, while holiday-related travel remains comparatively lower, indicating a more work-oriented travel structure.

From a modeling perspective, deep learning methods—particularly LSTM-based approaches—show better performance in handling the temporal structure and nonlinear variation of traffic flow compared with more traditional approaches. Their main contribution lies in improving the model’s ability to represent sequential dependencies in traffic data rather than relying on simplified linear assumptions. In terms of longer-term trends, congestion levels between 2020 and 2025 followed a general pattern of decline, recovery, and subsequent stabilization. This trajectory broadly corresponds to changes in travel demand during and after the pandemic period, and may also be related to ongoing traffic management and coordination measures, although the extent of their individual effects is not fully clear from the current analysis.

Future work could further incorporate data from emerging mobility services such as ride-hailing and delivery vehicles to improve the completeness of traffic representation. In addition, more attention may be given to model performance under non-routine or extreme conditions. Finally, the integration of data-driven prediction methods with practical traffic management applications remains an area that requires further exploration.

References

[1] Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*, 118, 102674. https://doi.org/10.1016/j.trc.2020.102674.

[2] Zhang, H., Dong, H., & Yang, Z. (2025). TSGDiff: Traffic state generative diffusion model using multi-source information fusion. *Transportation Research Part C: Emerging Technologies*, 174, 105081. https://doi.org/10.1016/j.trc.2025.105081.

[3] Wang, L. K., & Chen, C. (2024). Exploration of big data analysis and prediction models in transportation planning. *Theoretical Research on Urban Construction (Electronic Edition)*, (14), 151-153. https://doi.org/10.19569/j.cnki.cn119313/tu.202414051.

[4] Wang, B. L., Han, B. R., Dong, R., & Sun, Z. Y. (2024). Analysis of road congestion based on traffic status data. *Logistics Technology*, 47(15), 80-84. https://doi.org/10.13714/j.cnki.1002-3100.2024.15.020.

[5] Beijing Municipal Public Security Bureau Traffic Management Bureau. (2026, March). Business data release and government affairs disclosure (January 2026). https://jtgl.beijing.gov.cn/jgj/jgxx/ywsj/index.html.

[6] Amap. (2025). Amap traffic reports for Q1, Q2, and Q3 of 2025 (May, November, December 2025). Traffic Diagnosis Platform. https://report.amap.com/diagnosis/index.do.

[7] Beijing Transport Institute. (2026, April). Annual report on Beijing transport development (Over the years). Beijing Municipal Commission of Transport. https://jtw.beijing.gov.cn/xxgk/.

[8] Jiang, Y. S., Zhang, J., & Hu, L. (2022). Simulation study on interactive effects between ride-hailing and urban traffic congestion in large-scale scenarios. *Systems Engineering - Theory & Practice*, 42(11), 3079-3089.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.