electronics

MDPI

Article

Taxi Traffic Flow Prediction Based on Spatiotemporal-Fusion Graph Neural Networks

Nan Li [1,2], Guowei Jin [1,2,*], Pei Zhang [1,2,3], Wenlong Ma [1,2], Yuhang Tian [1,2], Shizheng Lu [1,2] and Guangtao Cao [1,2]

[1] School of Traffic and Transportation, Shijiazhuang Tiedao University, Shijiazhuang 050043, China; 1202306124@student.stdu.edu.cn (N.L.); 8202106013@student.stdu.edu.cn (P.Z.); 1202306015@student.stdu.edu.cn (W.M.); 1202306080@student.stdu.edu.cn (Y.T.); 1202306112@student.stdu.edu.cn (S.L.); 1202306040@student.stdu.edu.cn (G.C.)
[2] Hebei Key Laboratory of Traffic Safety and Control, Shijiazhuang 050043, China
[3] China Railway Jinan Group Co., Ltd., Jinan 250001, China
* Correspondence: jinguowei@stdu.edu.cn

Abstract

Accurate short-term traffic flow prediction in complex urban road networks is of great significance for capacity organization and dispatch optimization in intelligent transportation systems. Using publicly available historical taxi trip records released by the New York City Taxi and Limousine Commission from January to June 2016, this study develops a spatiotemporal fusion framework for short-term traffic flow prediction. To address the nonlinearity, sparsity, and complex spatiotemporal dependencies of traffic flow sequences, the raw trajectory data are first cleaned, spatially gridded, and temporally discretized. Based on the spatial adjacency relationships among grid nodes, a graph structure is then constructed, and a serially coupled graph convolutional network and long short-term memory model is developed to capture spatial dependency features and temporal dynamic features, respectively. Experimental results on the New York City taxi dataset show that, compared with baseline models including the historical average model, long short-term memory network, graph convolutional network, and Transformer, the proposed model achieves better performance in terms of mean absolute error, root mean square error, and coefficient of determination. Furthermore, the SHAP (SHapley Additive exPlanations) method is employed to ANALYZE the differences in feature contributions across nodes in different functional zones from both temporal and spatial perspectives. The results indicate that the model exhibits heterogeneous temporal dependency depths and spatial aggregation patterns across different types of regions within the study area. In addition, regions with high feature contributions show a certain degree of spatial correspondence with the major traffic corridors in Manhattan, suggesting that the model is able to capture part of the spatiotemporal correlation structure of traffic flow in this dataset. Finally, the limitations of the proposed method in terms of static graph structure, response to extreme events, and integration of external factors are discussed. It should be noted that these findings are derived from New York City taxi data from the first half of 2016, and their generalizability to other cities, time periods, or traffic scenarios remains to be further validated.

Keywords: taxi traffic flow; graph convolutional network (GCN); long short-term memory (LSTM); spatio-temporal fusion

Check for updates

Academic Editor: Maciej Ławryńczuk

Received: 7 March 2026
Revised: 6 April 2026
Accepted: 7 April 2026
Published: 13 April 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.

1. Introduction

As an essential component of urban transportation, taxis and ride-hailing services exhibit strong spatiotemporal variations in demand. Existing traffic flow prediction methods primarily rely on traditional statistical approaches or deep learning models, yet they often fail to adequately capture spatial dependencies or abrupt short-term fluctuations. With the rapid development of Intelligent Transportation Systems (ITS), accurate traffic flow prediction has become a core task in traffic management and planning. However, traffic flow data are characterized by high nonlinearity, randomness, and complex spatiotemporal dependencies, posing significant challenges to traditional forecasting methods. In recent years, researchers have proposed various deep learning-based models to more effectively capture the spatiotemporal features of traffic flow.

Ma and Liu [1] reviewed studies on short-term traffic flow prediction based on Long Short-Term Memory (LSTM) networks, analyzing their effectiveness in combined models. Mou et al. [2], addressing the daily periodic characteristics of traffic flow, proposed a temporal information-enhanced LSTM (T-LSTM), which improved prediction accuracy by capturing the intrinsic correlation between temporal information and traffic flow. Dai et al. [3] integrated spatiotemporal analysis with Gated Recurrent Units (GRU), utilizing a spatiotemporal feature selection algorithm to define optimal inputs, significantly enhancing the stability of road segment predictions. Ma et al. [4] further combined the advantages of unidirectional LSTM and bidirectional LSTM (BiLSTM), overcoming substantial prediction errors through time series smoothing and bidirectional dependency learning. Pure time-series models struggle to capture spatial correlations in traffic networks, prompting researchers to explore spatiotemporal fusion models. Qiao et al. [5] proposed a structure combining one-dimensional Convolutional Neural Networks (1DCNN) with LSTM (1DCNN-LSTM), using 1DCNN to extract spatial information and LSTM to capture temporal features, achieving regression prediction of spatiotemporal features. Zhuang and Cao [6] integrated the K-Nearest Neighbors (KNN) algorithm with BiLSTM, first filtering highly correlated spatial stations via KNN before feeding them into the model for prediction, resulting in a significant accuracy improvement compared to single models. Sun et al. [7] introduced a Multi-Component Attention (MCA) method, jointly modeling long-term trends, seasonality, and residual components in traffic flow through 1DCNN, bidirectional mechanisms, and attention mechanisms. To address the complex topological structure of road networks, Graph Convolutional Networks (GCN) have gradually become mainstream. Xing et al. [8] proposed the RL-GCN model, which integrates GCN for processing network features, LSTM for learning temporal information, and reinforcement learning algorithms to formulate optimal traffic control strategies. Chang et al. [9], addressing the limitations of fixed graph structures, proposed the Spatio-Temporal Self-Attention Graph Convolutional Network (STA-GCN), incorporating self-attention mechanisms into spatiotemporal convolutions to capture the dynamic characteristics of traffic flow. Zhong et al. [10] introduced a framework for analyzing discrete vehicle trajectory data, modeling spatiotemporal dependencies through Skip-Gram and GraphSAGE while integrating prediction patterns by considering intersection similarity. Ye Minghao [11] proposed a joint model combining improved LSTM with hierarchical dynamic graph convolution, leveraging dynamic graph convolution to break through the limitations of static topologies and capture multi-scale spatial relationships. Yang Guoliang et al. [12] proposed the STAR-DDRGC model, which adaptively generates node associations using dynamic graph convolutional recurrent networks and incorporates a decoupling module to handle multi-modal features, improving accuracy while reducing computational costs. With the evolution of attention mechanisms and pre-training techniques, traffic flow prediction has entered a higher-dimensional modeling stage. Lu Siyuan et al. [13], targeting temporal trends and spatial heterogeneity,

proposed a spatio-temporal-aware Transformer-based prediction model, using multi-scale diffusion convolution to capture local spatial features. Hao Chunlin and Zhang Jian [14] combined Temporal Convolutional Networks (TCN) with BiGRU and introduced a self-attention mechanism, addressing the challenge of feature extraction under sequence length constraints and enhancing model robustness. Lin et al. [15] employed an improved Particle Swarm Optimization algorithm (CAPSO) to optimize the hyperparameters of traditional Backpropagation Neural Networks, effectively avoiding local optima and improving sensitivity to initial weights. Liu et al. [16] introduced Timer-XL, a long-context Transformer for unified time series forecasting. This model achieved exceptional zero-shot prediction performance through large-scale pre-training, providing a new architecture for pre-trained time series models. In recent years, traffic flow prediction has gradually evolved from traditional single-city supervised learning toward more advanced frameworks involving cross-domain transfer, multi-frequency feature fusion, and complex scenario modeling. To address the issue of data scarcity, Wu et al. [17] proposed a cross-city deep transfer learning framework, which leverages a pretrained spatiotemporal model and a generative adversarial network (GAN) to achieve knowledge transfer and distribution alignment between source and target cities, thereby improving prediction performance in data-scarce urban environments. To model complex traffic disturbances, Sun et al. [18] developed the HST-GAN model, which dynamically adjusts graph structures through an event-aware mechanism and integrates global spatiotemporal feature extraction with a dynamic gating module. This approach enables the joint modeling of high-frequency local disturbances and low-frequency periodic patterns, while enhancing robustness against noise. In addition, Serdar et al. [19] investigated transportation network degradation under extreme conditions from a resilience perspective and proposed a multi-strategy percolation framework to evaluate road network robustness under sea-level rise scenarios. Their findings indicate that extreme disturbances not only affect directly impacted areas but also weaken the structural stability of the broader network. These studies suggest that traffic system analysis should not only focus on prediction accuracy, but also consider generalization ability, robustness, and adaptability under complex scenarios.

More recently, to overcome the limitations of inaccurate feature extraction and insufficient feature interaction in graph-based multi-source transfer learning, a spatial–channel collaborative multi-scale graph interaction deep transfer learning (SCMGIDTL) method has been proposed for unsupervised cross-domain machine fault diagnosis [20]. This method constructs high-quality graph prototypes from spatial and channel dimensions, establishes multi-scale graph interaction transfer networks, and designs category-level domain alignment constraints, achieving high diagnostic accuracy across multiple tasks. Although proposed in the fault diagnosis domain, its graph-based multi-source domain adaptation and multi-scale interaction mechanism provide valuable methodological references for spatiotemporal transfer learning in traffic flow prediction, especially under unsupervised target domain conditions.

In summary, although existing studies have made significant progress in spatiotemporal modeling and complex scenario analysis, limitations remain in the unified modeling of spatiotemporal coupling relationships and in the interpretability of regional-scale traffic flow prediction models. To address these issues, this study proposes a GCN-LSTM-based spatiotemporal fusion framework and further incorporates SHAP to interpret the contributions of temporal and spatial features in model predictions.

2. Data Description and Preprocessing

2.1. Data Source

The experimental data used in this study are derived from the publicly available historical trip records released by the New York City Taxi and Limousine Commission (TLC). This dataset is one of the commonly used public datasets in the fields of intelligent transportation systems and spatiotemporal data mining, characterized by its large volume, extensive spatial coverage, and strong temporal continuity. Specifically, taxi trip records from 1 January 2016, to 30 June 2016, are selected as the research object. Each record includes information such as pick-up and drop-off times, pick-up and drop-off coordinates, passenger count, and trip duration, enabling the construction of region-level short-term traffic flow sequences.

The selection of New York City data from the first half of 2016 is primarily motivated by the following considerations. First, the core urban area of New York City exhibits dense traffic activity and significant spatial heterogeneity, making it suitable for spatiotemporal prediction research. Second, the data from this period are characterized by large volume, temporal continuity, and complete fields, supporting stable partitioning into training, validation, and test sets. Third, this dataset has been widely used in urban transportation modeling studies, facilitating comparisons with existing methods. It should be noted that this study focuses on taxi ride-hailing demand; accordingly, references to traffic flow prediction throughout this paper refer to the prediction of taxi order volumes aggregated by spatial grids.

To ensure data quality and privacy, the raw data have been anonymized, containing no personally identifiable information (PII) of passengers or drivers. Furthermore, the dataset covers various temporal patterns, including weekdays, weekends, and holidays, as well as diverse functional areas such as the Manhattan Central Business District (CBD) and surrounding residential zones. This diversity in data composition is instrumental in comprehensively evaluating the generalization capability and robustness of prediction models. The raw data features selected for the experiment are detailed in Table 1.

Table 1. Description of Raw Data Fields in the NYC Taxi Dataset.

Field Name	Data Type	Description
id	String	A unique identifier for each trip record.
vendor_id	Categorical	The ID (1 or 2) of the taxi service provider.
pickup_datetime	Timestamp	The precise date and time of passenger pickup.
dropoff_datetime	Timestamp	The precise date and time of passenger drop-off.
passenger_count	Integer	The number of passengers.
pickup_longitude	Float	The longitude of the passenger pickup location.
pickup_latitude	Float	The latitude of the passenger pickup location.
dropoff_longitude	Float	The longitude of the passenger drop-off location.
dropoff_latitude	Float	The latitude of the passenger drop-off location.
store_and_fwd_flag	Categorical	A flag indicating whether the trip record was stored locally and forwarded later due to lack of network connectivity (Y/N).
trip_duration	Integer	The total duration of the trip (in seconds).

During the subsequent feature engineering phase, certain fields in the raw data (e.g.,store_and_fwd_flag), which exhibit weak correlation with the spatiotemporal distribution of traffic flow, will be either removed or retained only for reference purposes. It is worth noting that during the study period from January to June 2016, the overall traffic environment remained relatively stable, with no significant events such as large-scale, long-term lockdowns that would substantially alter normal travel patterns. Additionally, weather conditions were relatively mild. This ensures that the extracted spatiotemporal characteristics of traffic flow reflect genuine and generalizable patterns of regular traffic evolution, eliminating the interference of extreme external noise on baseline testing.

2.2. Data Cleaning

Due to factors such as sensor accuracy, signal transmission, and human operation, the original GPS data inevitably contain noise, missing values, and anomalous records. Directly modeling with such low-quality data would lead to the "Garbage In, Garbage Out" (GIGO) outcome, severely compromising the prediction accuracy of the model. Therefore, this section elaborates on the specific steps for data cleaning and preprocessing. To construct a high-quality traffic flow dataset, this study established stringent data screening rules, primarily encompassing two dimensions: spatial boundary constraints and logical rule filtering.

(1) Spatial Boundary Constraint

This research focuses on predicting traffic flow within the core area of New York City (primarily Manhattan Island). The raw dataset contains some trips destined for regions such as New Jersey, Connecticut, or the distant suburbs of Long Island. The sparse road networks in these peripheral areas, coupled with their distance from the region of interest, would result in an excessively sparse adjacency matrix, thereby interfering with feature extraction in the graph neural network. Consequently, a geo-fence was defined to retain only those records where both the pickup and drop-off locations fall within the following longitude and latitude ranges.

The specific screening boundaries are as follows:

$$40.60 \leq Latitude \leq 40.90 \tag{1}$$

$$-74.05 \leq Longitude \leq -73.75 \tag{2}$$

All anomalous records falling outside the defined geo-fence were filtered out. The spatial distribution of the raw data and the identification of noise are illustrated in Figure 1.

(2) Logical Rule Filtering

In addition to spatial location, certain trip records exhibit apparent errors in their physical attributes or internal logic (e.g., excessively short or long trip durations, or average speeds exceeding physical limits). This study established the following heuristic rules for secondary cleaning:

Trip Duration Outliers: Trips with durations of less than 60 s or greater than three hours were excluded from the dataset. Given the dense distribution of traffic signals and short block lengths in the core area of New York City, trips shorter than 60 s are generally insufficient to constitute a meaningful physical displacement. Such records typically originate from false orders caused by instantaneous cancellations, inadvertent meter activation by drivers, or transient GPS drift. The geofence of this study is strictly defined within the core commercial area of Manhattan and its surroundings, where the maximum straight-line distance does not exceed 20 km. Under normal urban traffic conditions, a single ride-hailing trip in this area rarely exceeds three hours. Records with durations beyond this threshold are highly likely to result from equipment malfunctions (e.g., failure to terminate billing in

a timely manner) or extreme atypical traffic events. Including such data would severely interfere with the extraction of regular spatiotemporal patterns by the model.

Trip Distance and Speed Anomalies: Records showing a travel distance of zero but a non-zero fare were eliminated. Records with a calculated average speed exceeding 150 km/h were also discarded, as this clearly violates urban speed limits and physical plausibility.

Invalid Field Filtering: Records missing critical spatiotemporal information (e.g., longitude or latitude values being NaN or 0) were filtered out.

The changes in the number of records before and after data cleaning, along with a summary of the filtering criteria, are presented in Table 2.

Table 2. Summary of data filtering criteria and record count changes.

Category	Filter Condition	Rationale	Records Removed	Proportion (%)
Spatial constraint	$Latitude \notin [40.60, 40.90]$ $Longitude \notin [-74.05, -73.75]$	Geofence limitation	5420	0.37%
Temporal logic	Duration < 60 s	Instantaneous cancellation or operational error	3150	0.22%
Temporal logic	Duration > 3 h	Equipment malfunction or atypical trip	1245	0.09%
Physical logic	Speed > 150 km/h	GPS signal jump	812	0.06%

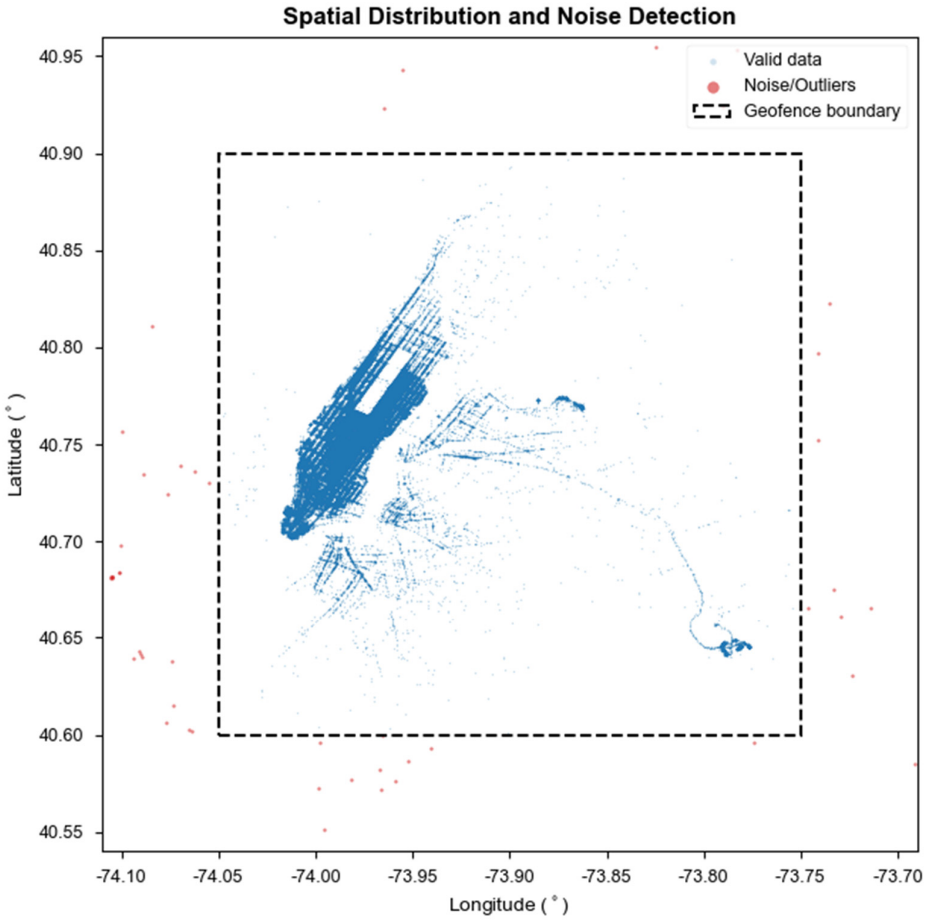


Figure 1. Spatial Distribution of Raw Data and Outlier Identification.

Following the aforementioned cleaning steps, the noise within the dataset was significantly reduced.

Although the logical rule-based filtering (Table 2) eliminated the vast majority of significant physical anomalies, a small number of residual outliers inevitably remain in the actual data (e.g., occasional minor congestion events or transient sensor drift). Z-Score standardization effectively limits the absolute magnitude of these residual outliers in the tensor by compressing the global data distribution.

More importantly, the LSTM module in the proposed model leverages its internal gating mechanism—particularly the forget gate-to adaptively identify and discard transient noise that lacks periodic patterns or temporal correlations with historical sequences. Meanwhile, the local Laplacian smoothing property of the GCN helps distribute and dilute anomalous pulses from individual grids to their neighboring regions in the spatial dimension. Consequently, even in the presence of a small number of residual outliers, the spatiotemporally coupled GCN-LSTM architecture maintains high predictive stability and noise robustness.

2.3. Spatiotemporal Grid Partitioning

Regarding the choice of spatial discretization method, this study adopts a uniform rectangular grid rather than irregular traffic analysis zones (TAZs) or actual road network partitions. This decision is primarily motivated by the following reasons. First, the actual road network topology in the core area of Manhattan exhibits a highly regular grid-like pattern, which renders rectangular grids a natural and physically consistent choice for spatial mapping. Second, uniform grids enable the mapping of discrete trajectories onto a standardized, unbiased spatial pixel array. This standardized underlying feature matrix facilitates efficient convolution operations in graph neural networks (GCNs).

To transform discrete GPS trajectory points into structured tensors suitable for neural network processing, a grid-based method is employed to discretize the study area.

(1) Spatial Dimension

In determining the specific grid resolution, this study fully considers the influence of the classical Modifiable Areal Unit Problem (MAUP) on feature extraction in spatial data mining. The selection of grid resolution requires a careful trade-off between the actual dispatch granularity and data sparsity. A grid that is too large (coarse granularity) may lead to excessive spatial smoothing, obscuring micro-scale traffic bottlenecks. Conversely, a grid that is too small (fine granularity) introduces a large number of water bodies or non-traffic areas, resulting in an extremely sparse spatiotemporal tensor.

To scientifically determine the optimal partitioning granularity, this study evaluates the sparsity characteristics under three resolutions—10 × 10, 20 × 20, and 30 × 30—and conducts preliminary experiments using the GCN-LSTM model constructed in subsequent sections for validation. The comparative results are presented in Table 3.

Table 3. Comparison of sparsity characteristics and model performance under different grid resolutions.

Grid Resolution	Number of Grid Cells	Data Sparsity (%)	RMSE	R^2
10 × 10	100	97.75%	0.2543	0.7250
20 × 20	400	96.56%	0.2644	0.6391
30 × 30	900	99.56%	0.0791	0.3694

As shown in Table 3, there is a clear trade-off among model performance, spatial representation capability, and data sparsity under different grid resolutions. Although the 10 × 10 grid achieves favorable performance on certain metrics, its relatively coarse spatial

granularity limits its ability to capture neighborhood-level traffic variations. In contrast, the 30 × 30 grid provides finer spatial resolution but significantly increases the proportion of empty grid cells, resulting in a sparser spatiotemporal tensor that is less conducive to stable modeling.

Considering the balance among prediction performance, spatial resolution, and data sparsity, this study ultimately adopts a 20 × 20 grid as the standard spatial discretization scheme for the study area. It should be noted that the selection of the 20 × 20 grid represents a compromised outcome based on empirical comparison, rather than a universally optimal partitioning. Alternative grid scales may be more suitable depending on specific task objectives.

For any given trip record k, its coordinate (lon_k, lat_k) is mapped to the corresponding grid coordinate (r, c). The calculation formulas are as follows:

$$r = \left[\frac{lat_k - lat_{\min}}{\Delta lat} \right] \tag{3}$$

$$c = \left[\frac{lon_k - lon_{\min}}{\Delta lon} \right] \tag{4}$$

(2) Temporal Dimension

The time axis is partitioned into fixed-length time slots. In this study, a time granularity of 5 min is adopted. Specifically, the number of trips within each grid cell is aggregated every 5 min. Through the aforementioned processing, the raw data are transformed into a three-dimensional tensor $X \in R^{T \times N \times C}$, where T denotes the total number of time steps, $N = 400$ represents the number of grid nodes, and C indicates the feature dimension (i.e., the traffic flow value).

It is noteworthy that, as observed in Figure 2b, the spatial distribution of traffic flow exhibits pronounced sparsity and imbalance. High traffic flow is predominantly concentrated on the left side of the grid matrix, corresponding to the central business district (CBD) of Manhattan Island, whereas surrounding areas show significantly lower volumes. This 'long-tail distribution' characteristic in space further underscores the necessity of employing the GCN-LSTM model for targeted spatiotemporal modeling, as conventional fully connected networks are inadequate for effectively capturing such localized, high-density spatial dependencies.

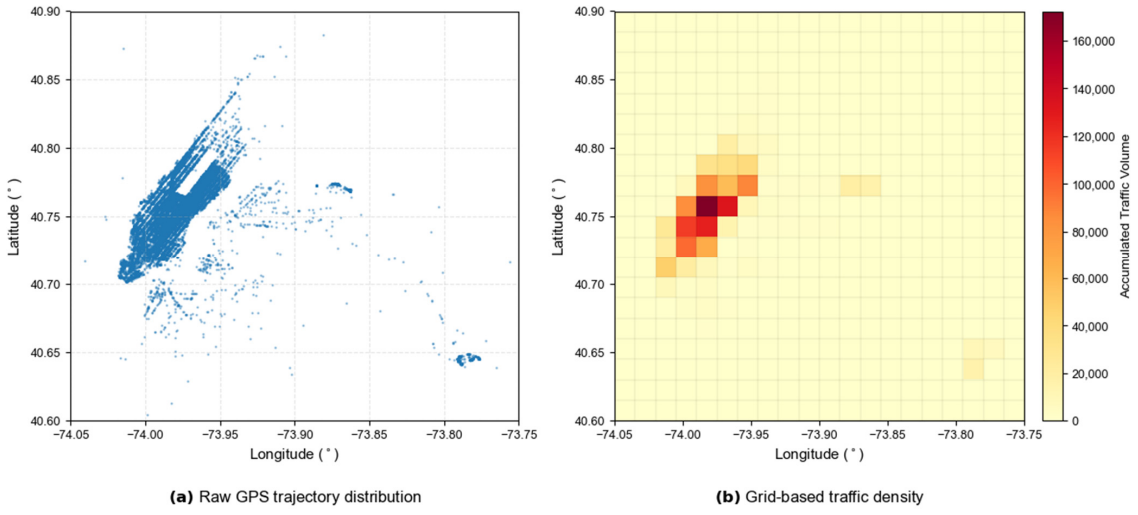


Figure 2. (**a**) Distribution of Raw GPS Trajectories; (**b**) Spatiotemporal Grid-based Traffic Flow Density.

2.4. Graph Structure Construction for Spatial Representation

(1) Definition of the Graph

The spatial distribution of traffic flow data exhibits pronounced non-Euclidean characteristics. To effectively leverage Graph Convolutional Networks (GCNs) for extracting the spatial topological features of the road network, this study first maps the partitioned grid areas into a graph structure. An undirected graph is defined to capture the spatial dependency among the grids. The undirected graph is constructed as $G = (V, E, \mathrm{A})$. Here, V denotes the set of nodes, where each partitioned $H \times W$ grid cell is treated as a node, resulting in a total number of nodes $|V| = N$. E represents the set of edges, signifying the connectivity or mutual influence relationships between the grid cells. A is the adjacency matrix, which mathematically describes the spatial relationships between nodes.

(2) Construction of the Adjacency Matrix

The construction of the adjacency matrix A directly determines the scope and strength of information aggregation within the GCN. Considering the diffusion effect of traffic flow in space—where the closer the physical distance between two grid cells, the stronger the correlation of their traffic states—this study employs a thresholded Gaussian kernel function based on Euclidean distance to construct the adjacency matrix.

Let the center coordinates of grid cell i and grid cell j be (x_i, y_i) and (x_j, y_j), respectively. The element A_{ij} of the adjacency matrix is calculated using the following formula:

$$A_{ij} = \begin{cases} \exp\left(-\frac{dist(v_i, v_j)^2}{\sigma^2}\right), & if\ dist(v_i, v_j) \le \varepsilon \\ 0 \qquad\qquad\qquad\quad , & otherwise \end{cases} \tag{5}$$

$$dist(v_i, v_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{6}$$

where $dist(v_i, v_j)$ denotes the Euclidean distance between the centroids of the two grid cells; σ^2 is the variance of the Gaussian kernel, controlling the rate at which the weight decays with distance; and ε is a distance threshold. To reduce computational complexity and mitigate interference from weakly correlated nodes, a connection is established only when the distance between two nodes is less than ε; otherwise, they are considered disconnected, i.e., $(A_{ij} = 0)$.

In the specific experiments, the parameters of the adjacency matrix were determined through preliminary experiments. The distance threshold ϵ was set to 2.0, and the Gaussian kernel variance σ was set to 1.5. This parameter configuration is primarily intended to achieve a balance between the utilization of local spatial information and the control of graph structural sparsity. If the threshold is too small, the neighborhood information available for aggregation becomes insufficient. Conversely, if the threshold is too large, distant nodes with weak correlations are excessively introduced, potentially increasing the risk of noise propagation. The variance parameter controls the rate at which edge weights decay with distance, thereby avoiding excessively strong or weak spatial information during the aggregation process.

It should be noted that the adjacency matrix constructed in this study primarily reflects spatial proximity relationships among grids and is not equivalent to the actual road connectivity graph. This limitation also represents an important direction for future work, in which road topology priors can be further incorporated. To facilitate an intuitive understanding of the above adjacency matrix construction process, a schematic diagram of the mapping from grid cells to the spatial graph structure is shown in Figure 3.

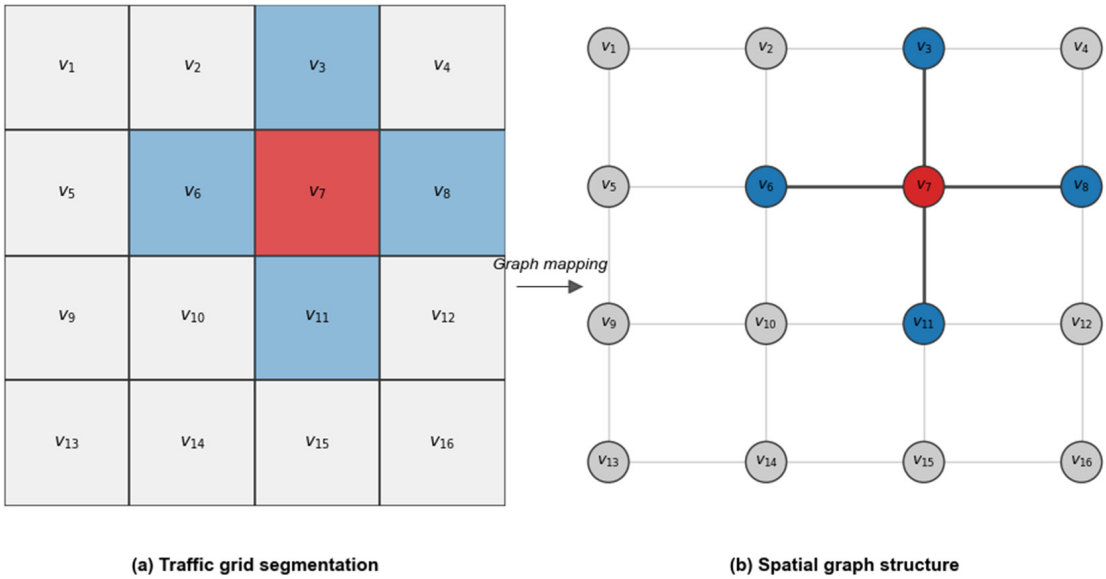


Figure 3. Schematic of Mapping from Traffic Grids to Spatial Graph Structure.

The physical adjacency relationships between grids are mapped to the edges ε in the graph. The connection weights between the red node and its cyan neighboring nodes are defined by the adjacency matrix A, enabling the GCN to aggregate localized spatial features.

(3) Matrix Normalization

Two grids are considered to have a connectivity relationship when they are spatially adjacent vertically or horizontally. Additionally, to avoid numerical instability in the feature aggregation process and to improve training stability, a symmetric normalization is applied to the adjacency matrix. Let the original adjacency matrix be denoted as A. The adjacency matrix with self-loops is defined as $\tilde{A} = A + I$, and the corresponding degree matrix is denoted as $\tilde{D}$. The normalized adjacency matrix is then expressed as:

$$\tilde{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \tag{7}$$

where $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. The normalized matrix $\tilde{A}$ serves as the input to the spatial convolution layer of the GCN, enabling stable aggregation of neighborhood features.

2.5. Data Partitioning and Standardization

To evaluate the model's generalization capability, the processed dataset was partitioned chronologically into training, validation, and test sets, with a ratio of 7:1:2.

Given the substantial numerical differences in traffic flow data across various time periods (e.g., morning peak hours versus late night), direct input to the model would lead to convergence difficulties. Therefore, this study applies Z-Score Standardization, transforming the data to a distribution with a mean of 0 and a standard deviation of 1. The calculation formula is as follows:

$$x' = \frac{x - \mu}{\sigma} \tag{8}$$

where x represents the original traffic flow data, x' denotes the standardized input data, μ is the mean, and σ is the standard deviation. This method ensures that the data conforms to a distribution with a mean of 0 and a standard deviation of 1, which facilitates the convergence of gradient descent optimizers.

3. Experimental Setup

3.1. Evaluation Metrics

To comprehensively evaluate the performance of the GCN-LSTM model in time series forecasting, this study employs three key evaluation metrics that quantify prediction accuracy, error distribution, and model interpretability from different perspectives. Widely adopted in the fields of machine learning and deep learning, these metrics provide an effective measure of the model's predictive capability and its reliability in practical applications.

$$MAE = \frac{1}{N}\sum_{i=0}^{N-1}|y_i - \hat{y}_i| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y}_i - y_i)^2} \tag{11}$$

Mean Absolute Error (MAE) is one of the most commonly used evaluation metrics in regression problems, measuring the average magnitude of absolute deviations between the model's predicted values and the true values.

Root Mean Squared Error (RMSE) is a widely adopted accuracy metric for regression model evaluation. It eliminates scale dependency by taking the square root of the Mean Squared Error (MSE). In practical applications, RMSE is often used in conjunction with MAE: RMSE reflects the stability of predictions (being more sensitive to large errors), while MAE indicates the average level of deviation. Their combined use allows for a comprehensive assessment of model performance.

Coefficient of Determination (R^2) measures the model's ability to explain the variance in the data. Its value ranges from 0 to 1, with an R^2 closer to 1 indicating a better fit of the model to the data.

3.2. GCN-LSTM Model Architecture

In urban transportation systems, the evolution of taxi traffic flow does not occur in isolation; rather, it is jointly constrained by the physical structure of the road network (spatial dimension) and historical traffic states (temporal dimension). In this study, GCN and LSTM are selected as the fundamental feature extractors because of their strong theoretical compatibility with the characteristics of traffic flow. Previous studies [21] have shown that representing a traffic network as a graph and jointly modeling it with graph convolution and recurrent neural networks can effectively capture both the spatial dependencies and temporal dynamics of traffic flow.

(1) Spatial Topological Feature Extraction Theory of GCN

Traditional convolutional neural networks (CNNs) are primarily designed to process Euclidean data with regular grid structures, such as images. However, when confronted with urban road networks, which represent typical high-dimensional non-Euclidean graph-structured data, CNNs are incapable of effectively capturing the irregular connectivity among nodes. The theoretical breakthrough of graph convolutional networks (GCNs) lies in introducing Fourier transform in the spectral domain to the graph Laplacian matrix. Through first-order Chebyshev polynomial approximation, GCNs are able to aggregate local spatial features with low computational complexity while preserving the global topological structure of the road network. This enables the model to transcend physical distance constraints and faithfully capture spatial spillover effects, such as the propagation of congestion from core commercial areas to surrounding neighborhoods.

(2) Temporal Dynamic Memory Theory of LSTM

Traffic flow sequences contain complex long- and short-term dependencies: in the short term, they exhibit high randomness due to sudden road conditions, while in the long term, they demonstrate strong periodicity, such as morning and evening rush hour patterns. Traditional recurrent neural networks (RNNs) are prone to gradient vanishing or gradient explosion when processing such long-sequence data. Long short-term memory (LSTM) networks introduce a sophisticated gating mechanism—input gate, forget gate, and output gate—along with a cell state that persists across time steps. Among these, the forget gate adaptively filters out redundant traffic noise, while the input gate writes critical instantaneous surges in traffic volume into long-term memory. This theoretical mechanism aligns perfectly with the dual requirements of traffic flow modeling: smoothing periodic fluctuations while sensitively capturing transient mutations.

(3) Cascaded Coupling Mechanism of GCN and LSTM

To achieve deep integration of spatial and temporal features, this study does not adopt a simple weighted combination of GCN and LSTM outputs. Instead, a deep spatial–temporal feature cascade mechanism is designed. The physical connection and tensor flow process of the two layers are as follows.

First, at any given time step t, the initial traffic flow features of all nodes in the network X_t are input together with the predefined normalized spatial adjacency matrix $\widetilde{A}$ into the GCN layer for spatial aggregation. The mathematical expression is as follows:

$$H_{gcn}^{(t)} = \sigma\left(\widetilde{A} X_t W_{gcn} + b_{gcn}\right) \tag{12}$$

where ReLU denotes the activation function, and W_{gcn} and b_{gcn} are the learnable weight matrix and bias term of the GCN layer, respectively. Through this step, the original isolated flow data are mapped into spatial feature representations $H_{gcn}^{(t)}$ that incorporate global road network topological awareness.

Subsequently, the model establishes a temporal serialization feature flow channel at the connection point. The GCN outputs from T consecutive historical time steps are concatenated along the temporal dimension to form a new spatiotemporal hybrid input sequence $[H_{gcn}^{(t-T+1)}, \ldots, H_{gcn}^{(t)}]$. This sequence is fed into the LSTM network, where the hidden state from the previous time step h_{t-1} is combined with the spatial features at the current time step $H_{gcn}^{(t)}$ to update the current cell state:

$$f_t = \sigma_g(W_f \cdot [h_{t-1}, H_{gcn}^{(t)}] + b_f) \tag{13}$$

$$i_t = \sigma_g(W_i \cdot [h_{t-1}, H_{gcn}^{(t)}] + b_i) \tag{14}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, H_{gcn}^{(t)}] + b_c) \tag{15}$$

$$h_t = o_t \odot \tanh(c_t) \tag{16}$$

This study employs a serially coupled GCN-LSTM architecture for traffic flow modeling. Specifically, the GCN module extracts spatial dependencies among different grid nodes, while the LSTM module captures temporal dynamic relationships in the historical sequence. At each time step, the current traffic flow features of all nodes are first input together with the normalized adjacency matrix into the GCN module to obtain node representations that incorporate neighborhood spatial information. Subsequently, the spatial representations from multiple consecutive time steps are organized into a sequence along the temporal order and fed into the LSTM module to learn the evolutionary patterns of traffic flow over time. Finally, the prediction results for each node at the next time step are

output through a fully connected layer. The core idea of this architecture is to first perform spatial encoding for each time step using GCN, followed by temporal modeling of the encoded sequential features using LSTM, thereby achieving sequential fusion of spatial and temporal features. Compared with models that model only temporal dependencies (e.g., LSTM) or only spatial dependencies (e.g., GCN), this architecture provides a more complete representation of the spatiotemporal coupling characteristics of urban traffic flow.

3.3. Experimental Environment and Parameter Configuration

To ensure the reproducibility of the experimental results, all model training and testing were conducted under a unified hardware and software environment. The hardware configuration consisted of an Intel® Core™ i5-10210U CPU, (Intel Corporation, Santa Clara, CA, USA) an NVIDIA Tesla T4 GPU (NVIDIA Corporation, Santa Clara, CA, USA), and 32 GB of RAM. The software environment included Python 3.8.10, PyTorch 1.12.1 (with CUDA 11.3), along with commonly used data processing and visualization libraries such as NumPy, Pandas, and Matplotlib. The detailed specifications are listed in Table 4.

Table 4. Experimental Environment Configuration.

Category	Item	Configuration/Details
Hardware	CPU	Intel® Core™ i5-10210U CPU @ 1.60 GHz
	GPU	NVIDIA Tesla T4
	RAM	16 GB
Software	Operating System (OS)	Windows 10
	Programming Language	Python 3.8.10
	Deep Learning Framework	PyTorch 1.12.1 (CUDA 11.3)
	Key Libraries	NumPy 1.21.5, Pandas 1.4.2, Matplotlib 3.5.1

To validate the effectiveness of the proposed GCN-LSTM model in traffic flow prediction and to ensure a fair comparison with baseline models (e.g., Transformer, GCN, LSTM), this experiment standardized the configurations for data input format, network training strategies, and core hyperparameters. The specific parameter settings are detailed in Table 5.

The time step duration (Δt) for the traffic flow data was set to 5 min. A sliding window mechanism was employed to construct temporal samples, with the input sequence length (T) set to 12, meaning historical traffic data from the past 60 min (12×5 min) was used as input features. The prediction horizon (H) was set to 1, corresponding to forecasting traffic flow for the next 5 min. The dataset was strictly partitioned into training, validation, and test sets in a 7:1:2 ratio to ensure an objective model evaluation.

During model training, the batch size was uniformly set to 64, and the maximum number of epochs was set to 200. The Adam optimizer was selected for parameter updates due to its ability to adaptively adjust the learning rate, thereby accelerating convergence. The initial learning rate (l_r) was set to 0.001, and a StepLR learning rate scheduler was introduced with a step size of 10 and a gamma value of 0.5. This configuration halves the learning rate every 10 epochs to prevent the model from oscillating around a local optimum during the later stages of training. The Mean Squared Error (MSE) loss function was adopted to quantify the difference between predicted and true values.

To ensure the fairness of comparative experiments, all baseline models were independently optimized using grid search to identify their optimal hyperparameter configurations,

ensuring that performance comparisons were conducted under the best possible conditions for each model. For the proposed GCN-LSTM model, the hidden layer dimension was set to 64. For the Transformer baseline model, additional parameters were configured: the number of attention heads was set to 4, the number of encoder layers to 2, and a dropout rate of 0.2 was applied to prevent overfitting.

The model training process is illustrated in Figure 4.

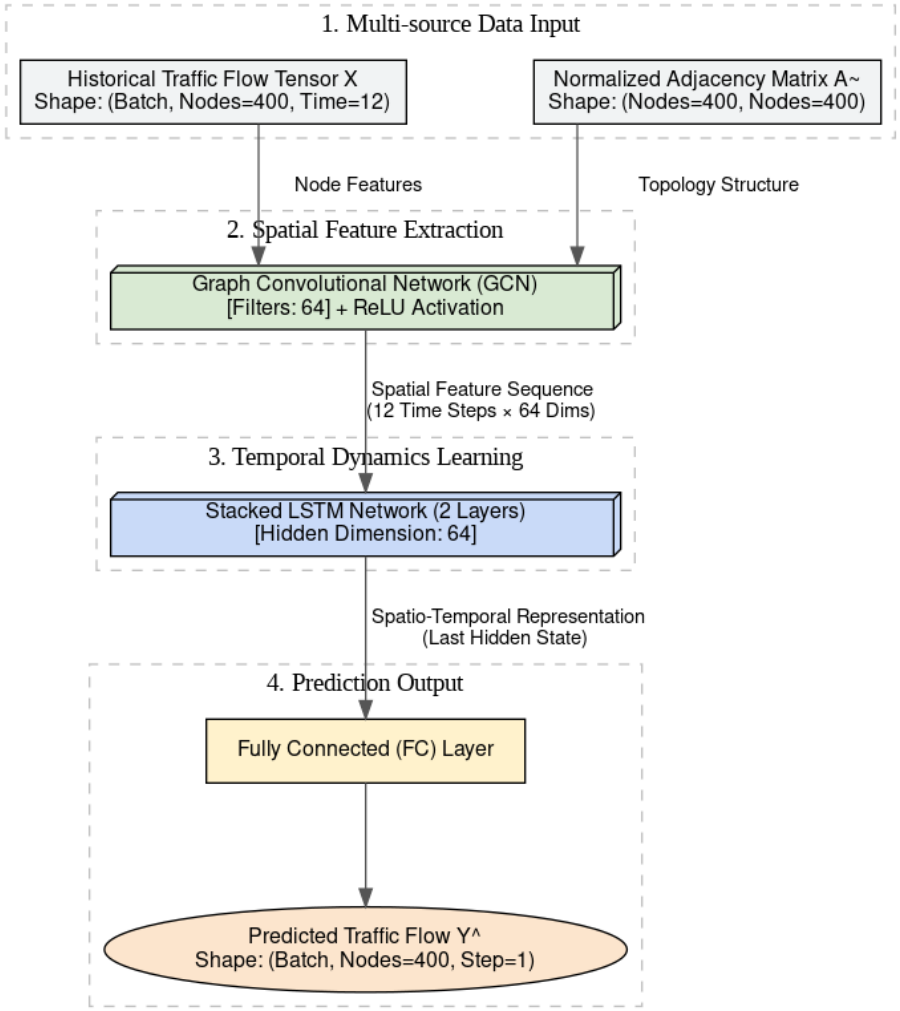


Figure 4. Training Process of the GCN-LSTM Model.

To further demonstrate the convergence behavior and training stability of the proposed GCN-LSTM model, the training and validation loss curves are presented in Figure 5. As shown, both losses decrease steadily during the early and middle stages of training and gradually stabilize after approximately 120 epochs, indicating good convergence and no obvious overfitting.

Table 5. Model Hyperparameter Configuration.

Parameter Category	Parameter Name	Symbol	Value/Setting
Data Configuration	Time Step Duration	Δt	5 min
	Input Sequence Length	T	12
	Prediction Horizon	H	1
	Train/Val/Test Split	-	7:1:2

Table 5. *Cont.*

Parameter Category	Parameter Name	Symbol	Value/Setting
Training Configuration	Batch Size	*Batch Size*	64
	Optimizer		Adam
	Initial Learning Rate	l_r	0.001
	Learning Rate Scheduler		StepLR(Step = 10, Gamma = 0.5)
	Maximum Epochs	*Epochs*	200
	Loss Function		MSE
Transformer	Hidden Dimension	d_{model}	64
	Number of Attention Heads	h	4
	Number of Encoder Layers	N	2
	Dropout Rate	p_{drop}	0.2

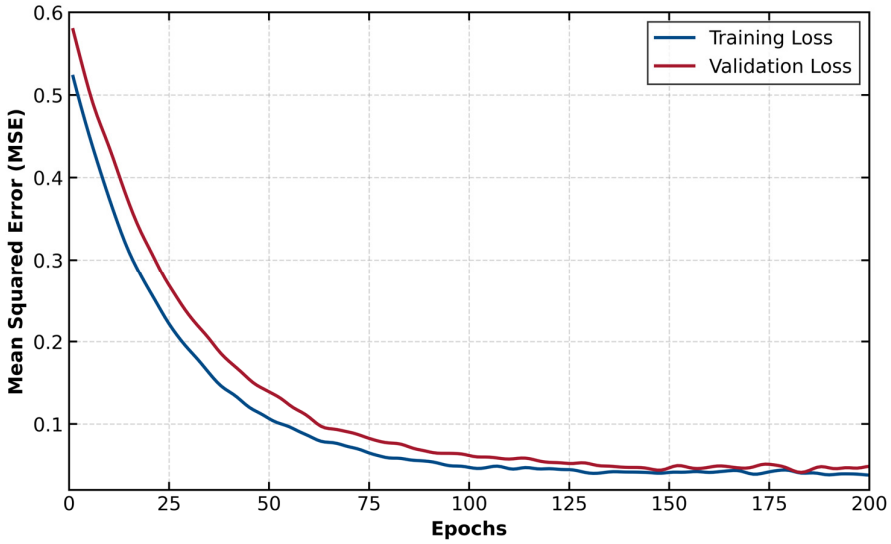


Figure 5. Training and validation loss curves of the proposed GCN-LSTM model.

4. Analysis of Prediction Results

4.1. Overall Performance Comparison

To validate the effectiveness of the proposed model, four baseline models—HA, LSTM, GCN, and Transformer—were selected for comparison. The results are presented in Table 6. Overall, the proposed GCN-LSTM model achieved the best or near-best performance across all three evaluation metrics—MAE, RMSE, and R^2—indicating that jointly modeling spatial dependencies and temporal dependencies enhances short-term traffic flow prediction performance.

Table 6. Comparative Analysis of Evaluation Metrics Across Models.

Models	MAE	RMSE	R^2
HA	0.0559	0.2896	0.5661
LSTM	0.0545	0.2716	0.6190
GCN	0.0550	0.2962	0.6059
Transformer	0.0536	0.2684	0.6280
GCN-LSTM	0.0522	0.2644	0.6389

Specifically, the HA model exhibited the weakest performance due to its inability to capture complex nonlinear spatiotemporal relationships. LSTM captured temporal dependencies but did not explicitly leverage spatial topological information. GCN emphasized spatial aggregation but lacked sufficient temporal dynamics modeling capability. Transformer demonstrated strong global modeling capacity and outperformed most baseline models overall. In contrast, the proposed GCN-LSTM model leverages the spatiotemporal coupling characteristics of traffic flow more effectively through its serial architecture of "spatial encoding followed by temporal modeling," thereby achieving superior comprehensive prediction performance.

Considering that deep learning model training results may be affected by random initialization and differences in optimization paths, this study further conducted five independent repeated experiments using different random seeds for both the proposed GCN-LSTM model and the best-performing baseline model, Transformer. The mean and standard deviation of the evaluation metrics were recorded for each model. The results indicate that the GCN-LSTM model exhibits good stability across repeated experiments.

Furthermore, based on the results of the five repeated experiments, a paired statistical test was conducted on the R^2 metric between GCN-LSTM and Transformer. The test results show that the performance improvement of GCN-LSTM over Transformer is statistically significant ($\rho < 0.05$). This demonstrates that the performance advantage of the proposed model is not attributable to random fluctuations but represents a stable and reliable outcome.

In addition to prediction accuracy, computational efficiency is also a critical indicator for determining whether a traffic flow prediction model can be practically deployed in real-world intelligent transportation systems (ITS). To address this, this study provides a computational overhead analysis for each model. Under a unified hardware environment (equipped with an NVIDIA GPU), the proposed GCN-LSTM model achieved a training time of approximately 42 s per epoch, with a total training duration of about 2.3 h for 200 epochs. During the inference stage, the model achieved an average inference latency of only 24 ms for generating a single-step (5 min) traffic flow prediction across all 400 nodes in the study area.

Given that the inference latency of several tens of milliseconds is far below the actual operational scheduling interval of 5 min, and that GPU memory usage during inference is only approximately 1.2 GB, these results fully demonstrate that the GCN-LSTM model not only offers significant advantages in prediction accuracy but also possesses the deployability and efficiency required for real-time online prediction in real-world urban traffic management systems.

To further evaluate the stability of the model under highly fluctuating traffic conditions, Table 7 presents a detailed breakdown of the prediction performance of the proposed GCN-LSTM model across different typical time periods: morning peak (07:00–09:00), evening peak (17:00–19:00), and full weekends. The results indicate that although the absolute errors (MAE, RMSE) exhibit a modest increase during peak hours due to the surge in traffic volume—which is within a reasonable range—the coefficient of determination (R^2) remains above 0.63. This demonstrates that the model maintains strong feature extraction and tracking capabilities even under extreme tidal traffic flow conditions.

Table 7. Performance breakdown of the GCN-LSTM model across different time periods.

Time Period	MAE	RMSE	R^2
Morning Peak (Weekdays 07:00–09:00)	0.0561	0.2812	0.6375
Evening Peak (Weekdays 17:00–19:00)	0.0583	0.2895	0.6358
Weekends (Full day)	0.0485	0.2466	0.6212
Overall Average	0.0522	0.2644	0.6280

4.2. Visual Analysis of Prediction Results

To comprehensively evaluate the spatial generalization capability of the model in complex urban road networks, this study moves beyond the limitation of evaluating only single high-traffic nodes. Based on the historical average traffic flow distribution across all 400 nodes in the study area, non-traffic areas (e.g., water bodies) were excluded, and three representative types of urban functional zones were intelligently selected for detailed comparison: a central business district (CBD, Node 204), a typical residential area (Node 182), and an urban fringe area (Node 206).

As shown in Figure 6, the horizontal axis represents the time step (each unit corresponds to 5 min), and the vertical axis represents the normalized traffic flow value. By comparing the predicted curves with the ground truth, the following detailed analysis can be performed:

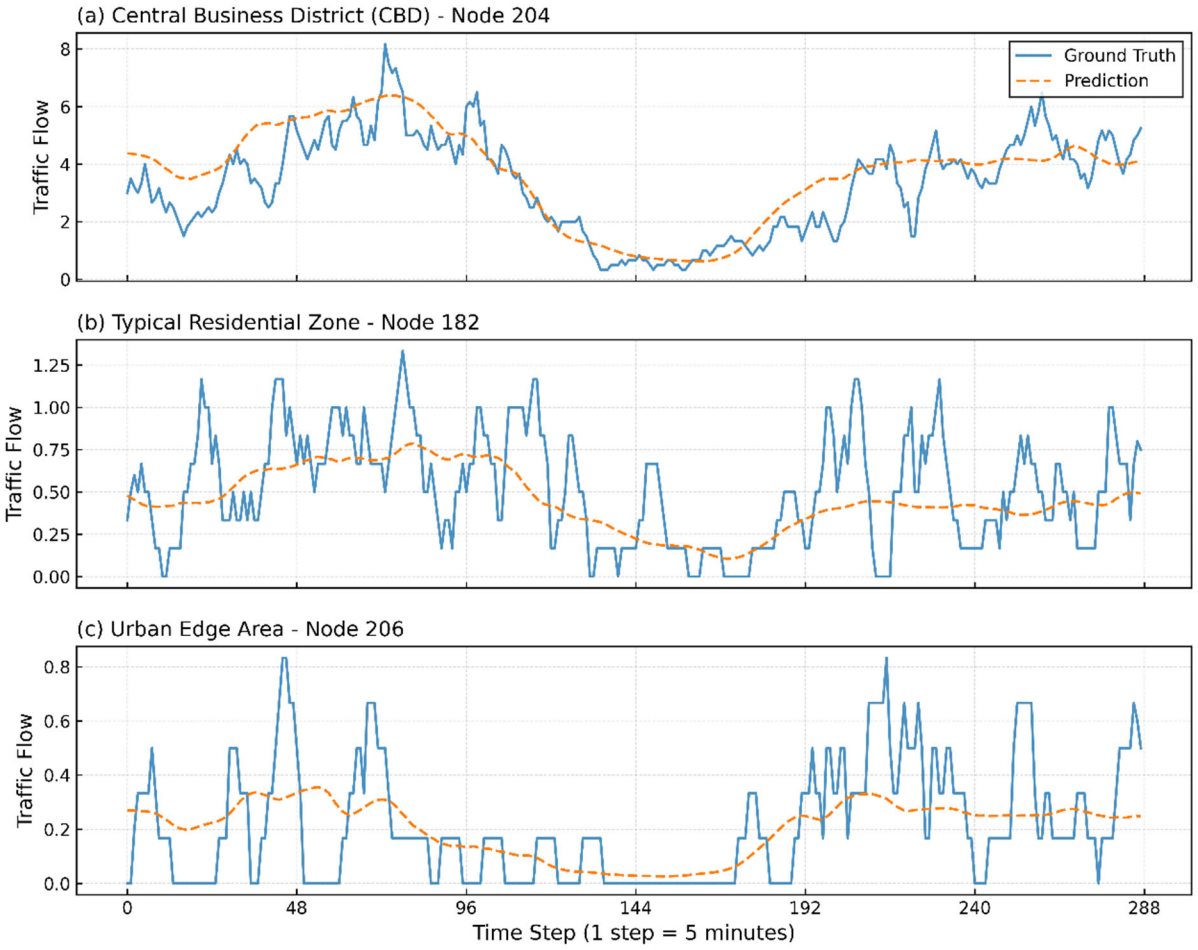


Figure 6. Comparison of predicted and ground truth traffic flow at representative urban functional zones.

To investigate the predictive performance of the model across different spatial functional zones, three representative nodes were selected from all valid nodes in the study area for visualization analysis: a central business district (CBD, Node 204), a typical residential area (Node 182), and an urban fringe area (Node 206). The results indicate that the proposed GCN-LSTM model is capable of effectively tracking traffic flow trends across different regions, though certain variations in performance are observed across node types.

In the CBD, the model successfully captures fluctuation patterns under high traffic volume conditions. In the residential area, it demonstrates strong fitting capability for the bimodal commuting pattern. In the urban fringe area, where the baseline traffic volume is lower and data sparsity is higher, the prediction curves are generally smoother but still maintain consistency with the ground truth trend during most time periods. Overall, the GCN-LSTM model exhibits robust adaptability across different scenarios characterized by spatial heterogeneity.

4.3. Analysis of Model Interpretability

Although the evaluation metrics presented above have demonstrated that the GCN-LSTM model achieves favorable performance in the traffic flow prediction task, deep learning models are often regarded as "black boxes", and their internal decision-making mechanisms lack intuitive interpretability. To examine whether the model has genuinely learned meaningful spatiotemporal dependencies rather than merely fitting numerical patterns, this section introduces SHAP (SHapley Additive exPlanations), a post hoc interpretation method grounded in game theory, to explain the model predictions [22]. SHAP values are based on the idea of feature attribution and can decompose an individual prediction into the marginal contributions of each input feature, including historical time steps and spatial neighboring nodes.

SHAP values are grounded in feature attribution theory and decompose the model's predictions into the marginal contributions of individual input features (including historical time steps and neighboring spatial nodes). In this section, three representative urban functional zones—the central business district (CBD, Node 204), a typical residential area (Node 182), and an urban fringe area (Node 206)—are selected as the analysis targets from the test set. A comparative investigation is conducted from two perspectives: temporal dependency (lag correlation) and spatial topological correlation. This analysis aims to comprehensively reveal the decision-making differences of the model across different physical scenarios.

(1) Feature Attribution Analysis in the Temporal Dimension

To investigate the capability of the LSTM module in capturing long- and short-term temporal dependencies, Figure 7 presents the average SHAP importance distribution of the input sequence (the past 12 time steps, corresponding to 60 min) for predicting the next-step traffic flow at three target nodes.

A detailed examination of the temporal feature contributions across different regions reveals that the model exhibits a highly adaptive temporal attention allocation strategy:

(1) Strong Short-Term Inertia and Deep Memory in the Central Business District (Node 204)

As shown in Figure 7a, the overall SHAP contributions for the CBD region are the highest among the three node types. The model demonstrates not only a strong reliance on recent physical inertia but also a smooth linear decay pattern in attention to historical information, maintaining relatively high attention weights even at distant time steps (e.g., $t - 12$ and earlier). This indicates that the formation and dissipation of congestion in the CBD region involve a prolonged and complex cumulative process. The model automatically learns and activates a deep temporal memory to accurately capture these complex evolutionary patterns.

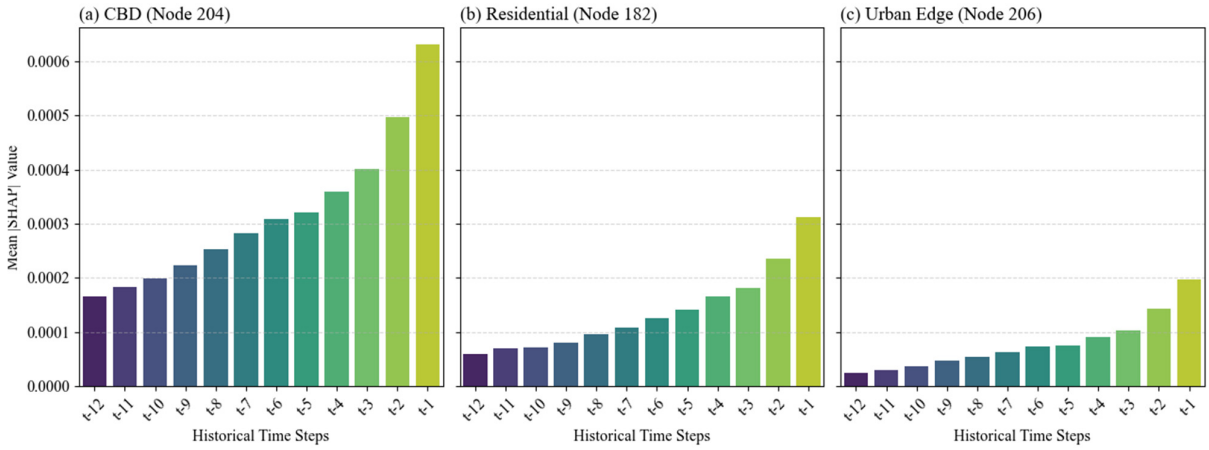


Figure 7. Temporal SHAP Value Distribution of Three Typical Nodes.

(2) Smooth Monotonic Dependency in the Typical Residential Area (Node 182)

As shown in Figure 7b, the overall magnitude of feature importance in this region decreases to a moderate level and exhibits a highly smooth, monotonic stepwise decay pattern. This reflects the fact that traffic flow in residential areas is predominantly governed by regular commuting tidal patterns, lacking the complex congestion spillover effects commonly observed in road networks. Consequently, when making short-term predictions at the micro level, the model primarily relies on recent baseline traffic fluctuations (e.g., $t-1$ to $t-3$) without assigning excessive importance to distant historical data.

(3) Adaptive Noise Filtering in the Urban Fringe Area (Node 206)

As shown in Figure 7c, the SHAP values for the urban fringe area are overall very low, with a pronounced sharp decay pattern. Aside from certain contributions from $t-1$ and $t-2$, the feature weights for earlier historical time steps (e.g., $t-3$ and earlier) are nearly negligible. This finding strongly demonstrates that when confronted with extremely sparse and random fringe area traffic flow, the model proactively truncates distant historical information that lacks physical relevance, adaptively filtering out system noise and effectively avoiding overfitting.

In summary, the feature extraction process of the GCN-LSTM model in the temporal dimension is not a rigid globally uniform pattern. Rather, the model intelligently and adaptively adjusts the depth of the temporal receptive field and the allocation of attention weights according to the physical characteristics and real-time traffic dynamics of different urban functional zones. This fully validates the scientific soundness and physical rationality of the model's underlying decision-making logic.

(2) Feature Attribution Analysis in the Spatial Dimension

In the field of spatiotemporal graph neural networks, to interpret whether the model has genuinely learned the underlying physical road network topology, this study maps the spatial SHAP attention weight matrices of three representative nodes back onto the actual street map of Manhattan, New York, and overlays them for visualization. As shown in Figure 8.

To further analyze the decision-making basis of the model in the spatial dimension, the spatial SHAP contribution results for the three representative nodes were mapped onto the street map of Manhattan for comparative visualization. The results reveal distinct spatial distribution patterns of high-contribution regions across the different node types.

For the central business district (CBD), the high-contribution grids are primarily concentrated in the vicinity of the target node and the surrounding major traffic corridors, exhibiting a certain directional distribution pattern. For the typical residential area, the high-contribution region is relatively more localized, mainly distributed in the neighborhoods

adjacent to the target node. For the urban fringe area, where traffic volume is relatively low and spatial activity is limited, the high-contribution region is further contracted, showing weak diffusion patterns near natural boundaries.

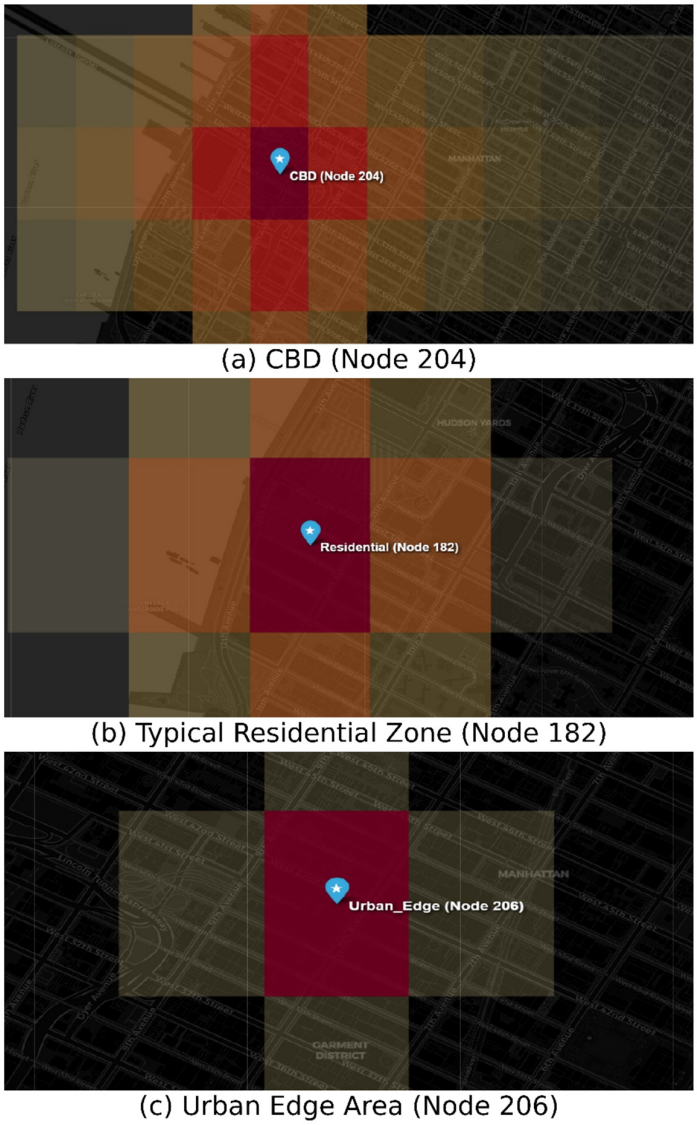


Figure 8. Spatial SHAP Attention Weight Mapping of Three Representative Functional Zones on the Real-World Road Network.

These findings indicate that the model learns differentiated spatial aggregation patterns across different spatial contexts. It should be noted that since the adjacency matrix in this study is primarily constructed based on spatial distances between grids, the SHAP mapping results are more appropriately interpreted as exhibiting a certain spatial correspondence with real-world traffic corridors, rather than serving as direct proof that the model has fully recovered the propagation mechanisms of the actual road network. This part of the analysis is intended primarily to enhance model interpretability rather than to provide rigorous validation of physical topology recovery.

5. Conclusions

This study constructs a GCN-LSTM spatiotemporal fusion framework for short-term traffic flow prediction based on taxi trip records from New York City between January and June 2016. Through data cleaning, spatiotemporal gridding, spatial graph construction, and

comparative experiments with multiple baseline models, the effectiveness of the proposed model is validated in terms of MAE, RMSE, and R^2. The results show that, on the New York City taxi dataset from the first half of 2016, the GCN-LSTM model is able to more effectively integrate spatial dependencies and temporal dynamics than baseline models such as HA, LSTM, GCN, and Transformer, thereby achieving better overall predictive performance.

In terms of interpretability analysis, this study further compares the SHAP contribution distributions of different functional zones from both temporal and spatial dimensions. The analytical results reveal that the model exhibits differentiated temporal dependency depths and spatial aggregation patterns across different regions, with high-contribution regions showing a certain spatial correspondence with major traffic corridors in Manhattan. This indicates that the GCN-LSTM model not only improves prediction accuracy but also captures, to some extent, the spatiotemporal evolution characteristics of traffic flow across different urban areas. It should be noted, however, that the adjacency matrix used in this study is constructed based on spatial distances and remains static; therefore, the current results are not sufficient to rigorously demonstrate that the model has fully recovered the propagation mechanisms of the actual road network.

Furthermore, this study still has several limitations. First, the current model is constructed on a static graph structure and does not explicitly characterize dynamic topological changes caused by factors such as traffic accidents, extreme weather, or road closures. Second, this study mainly focuses on taxi demand prediction under routine conditions, and the generalization capability of the model for low-traffic and highly sparse nodes, as well as under extreme perturbation scenarios, still requires further validation. Third, external influencing factors, such as weather conditions, holidays, and special events, are not incorporated in the present study; therefore, the mechanisms underlying traffic flow fluctuations under complex external conditions are not yet fully captured. It should also be noted that, although the parameters involved in adjacency matrix construction and the thresholds adopted in the data cleaning process were determined empirically, no systematic sensitivity analysis of these key settings was conducted. As a result, the potential effects of parameter selection on predictive performance, model stability, and experimental reproducibility have not been fully evaluated. This represents not only a limitation of the present study, but also an area that deserves further investigation in future work. Future research will therefore focus on dynamic graph structure modeling, the integration of multi-source external information, and systematic sensitivity and robustness analyses of key parameters and preprocessing thresholds, so as to improve the applicability, reliability, and reproducibility of the proposed model in real-world traffic management scenarios.

Author Contributions: Conceptualization, W.M. and S.L.; Methodology, N.L., W.M. and S.L.; Software, N.L.; Validation, N.L. and G.C.; Formal analysis, G.J. and Y.T.; Investigation, Y.T.; Resources, P.Z.; Data curation, P.Z., W.M. and G.C.; Writing—review & editing, N.L.; Visualization, Y.T. and S.L.; Supervision, G.C.; Project administration, G.J.; Funding acquisition, G.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: Author Pei Zhang was employed by the company China Railway Jinan Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ma, C.; Liu, T. Survey of short-term traffic flow prediction based on LSTM. *Int. J. Mod. Phys. C* **2025**, *36*, 2450177. [CrossRef]

2. Mou, L.; Zhao, P.; Xie, H.; Chen, Y. T-LSTM: A Long Short-Term Memory Neural Network Enhanced by Temporal Information for Traffic Flow Prediction. *IEEE Access* **2019**, *7*, 98053–98060. [CrossRef]
3. Dai, G.; Ma, C.; Xu, X. Short-term Traffic Flow Prediction Method for Urban Road Sections Based on Space-Time Analysis and GRU. *IEEE Access* **2019**, *7*, 143025–143035. [CrossRef]
4. Ma, C.; Dai, G.; Zhou, J. Short-Term Traffic Flow Prediction for Urban Road Sections Based on Time Series Analysis and LSTM_BILSTM Method. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 5615–5624. [CrossRef]
5. Qiao, Y.; Wang, Y.; Ma, C.; Yang, J. Short-term traffic flow prediction based on 1DCNN-LSTM neural network structure. *Mod. Phys. Lett. B* **2021**, *25*, 2150042. [CrossRef]
6. Zhuang, W.; Cao, Y. Short-Term Traffic Flow Prediction Based on a K-Nearest Neighbor and Bidirectional Long Short-Term Memory Model. *Appl. Sci.* **2023**, *13*, 2681. [CrossRef]
7. Sun, B.; Sun, T.; Zhang, Y.; Jiao, P. Urban traffic flow online prediction based on multi-component attention mechanism. *IET Intell. Transp. Syst.* **2020**, *14*, 1249–1258. [CrossRef]
8. Xing, H.; Chen, A.; Zhang, X. RL-GCN: Traffic flow prediction based on graph convolution and reinforcement learning for smart cities. *Displays* **2023**, *80*, 102513. [CrossRef]
9. Chang, Z.; Liu, C.; Jia, J. STA-GCN: Spatial-Temporal Self-Attention Graph Convolutional Networks for Traffic-Flow Prediction. *Appl. Sci.* **2023**, *13*, 6796. [CrossRef]
10. Zhong, R.; Hu, B.; Wang, F.; Feng, Y.; Li, Z.; Song, X.; Wang, Y.; Lou, S.; Tan, J. Multi-factor embedding GNN-based traffic flow prediction considering intersection similarity. *Neurocomputing* **2025**, *620*, 129193. [CrossRef]
11. Ye, M. Traffic Flow Prediction Based on Improved LSTM and Hierarchical Dynamic Graph Convolution. *Intell. Comput. Appl.* **2025**, *15*, 1–9. [CrossRef]
12. Yang, G.; Li, W.; Xiong, Z.; Zhu, X.; Ma, X. Traffic Prediction Based on Spatiotemporal Attention and Decoupled Dynamic Residual Graph Convolution. *J. Comput. Appl.* **2025**, 1–9. Available online: https://link.cnki.net/urlid/51.1307.TP.20251210.1129.004 (accessed on 6 April 2026).
13. Lu, S.; Shen, Q.; Bao, Y.; Gao, R.F.; Shi, Q. A Traffic Flow Prediction Model Based on Spatiotemporal-Aware Transformer. *Electron. Meas. Technol.* **2024**, *47*, 85–92. [CrossRef]
14. Hao, C.; Zhang, J. Traffic Flow Prediction Based on Self-Attention Mechanism TCN-BiGRU. *Electron. Meas. Technol.* **2024**, *47*, 61–68. [CrossRef]
15. Lin, Z.; Qiu, C.; Tang, H.; Yue, Y. Optimization of BP neural network traffic flow prediction model based on improved particle swarm algorithm. *Adv. Mech. Eng.* **2025**, *17*, 16878132251395508. [CrossRef]
16. Liu, Y.; Qin, G.; Huang, X.; Wang, J.; Long, M. Timer-XL: Long-Context Transformers for Unified Time Series Forecasting. In Proceedings of the 41st International Conference on Machine Learning (ICML), Vienna, Austria, 21–27 July 2024.
17. Wu, Q.; He, K.W.; Chen, X.; Yu, S.; Zhang, J.S. Deep Transfer Learning Across Cities for Mobile Traffic Prediction. *IEEE/ACM Trans. Netw.* **2022**, *30*, 1255–1267. [CrossRef]
18. Sun, W.; Deng, K.Q.; Gao, J.B. Traffic Flow Prediction Based on Generative Adversarial Network with Hybrid Spatio-Temporal Features Learning. *Clust. Comput.* **2025**, *28*, 621. [CrossRef]
19. Serdar, M.Z.; Marian, A.R.; Masad, E. A multi-strategy percolation framework for road network robustness under projected sea-level rise. *J. Environ. Manag.* **2026**, *403*, 129196. [CrossRef]
20. Wang, X.; Jiang, H.; Dong, Y.; Mu, M. Spatial-channel collaborative multi-scale graph interaction deep transfer learning for unsupervised rotating machinery fault diagnosis. *Eng. Appl. Artif. Intell.* **2026**, *176*, 114691. [CrossRef]
21. Cui, Z.; Henrickson, K.; Ke, R.; Wang, Y. Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting. *arXiv* **2018**, arXiv:1802.07007. [CrossRef]
22. Lundberg, S.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874. [CrossRef]

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.