Data-Driven Spatial-Temporal Modeling for Bicycle Traffic Prediction

Xiamei Wen

Propositions

accompanying the dissertation

Data-Driven Spatial-Temporal Modeling for Bicycle Traffic Prediction

by

Xiamei Wen

1. Accurate bicycle traffic prediction depends on the ability of models to capture nonlinear spatial-temporal dependencies while representing uncertainty arising from environmental variability. [Chapter 2]

2. LLM-based traffic prediction models offer significant potential for real-world deployment, particularly in contexts with limited computational resources. [Chapter 3]

3. Transferring learned knowledge from data-rich regions to data-scarce cities is an effective and economical way to improve prediction accuracy. [Chapters 3&4]

4. The availability of computational power can restrict or dictate the choice of models in practice.

5. In the field of traffic prediction, data-efficient methods offer practical and valuable solutions, especially in scenarios with limited data availability.

6. Scientific collaboration across disciplines is more a matter of mindset than methodology.

7. Effective supervision supports autonomy: a PhD candidate's judgment should take precedence in matters related to their research.

8. Research progress and inspiration are more often driven by the pressure of an approaching deadline than by free time or curiosity alone.

9. The less data you have, the more creative your approaches or maybe your excuses.

These propositions are regarded as opposable and defendable, and have been approved as such by the promotor prof. dr. ir. S.P. Hoogendoorn and copromotor dr. ir. D.C. Duives.

Data-Driven Spatial-Temporal Modeling for Bicycle Traffic Prediction

Front & Back Cover: designed by the author and ChatGPT 5.2

Data-Driven Spatial-Temporal Modeling for Bicycle Traffic Prediction

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus, Prof. dr. ir. H. Bijl,

chair of the Board for Doctorates

to be defended publicly on

Thursday 19, March 2026 at 12:30

by

Xiamei WEN

This dissertation has been approved by the promotor and the copromotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. S. P. Hoogendoorn	Delft University of Technology, promotor
Dr. ir. D. C. Duives	Delft University of Technology, copromotor

Independent members:

Prof. dr. ir. J. W. C. van Lint	Delft University of Technology
Prof. M. Saberi Kalaee	University of New South Wales, Australia
Prof. S.V. Ukkusuri	Purdue University, United States
Prof. dr. F. Viti	University of Luxembourg, Luxembourg
Dr. ir. W. Daamen	Delft University of Technology
Prof. dr. ir. B. van Arem	Delft University of Technology, reserve member

TRAIL Thesis Series no. T2026/7, The Netherlands Research School TRAIL

TRAIL
P.O. BOX 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-383-1

Printed in the Netherlands

The most incomprehensible thing about the universe is that it is comprehensible.

Albert Einstein

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Why Bicycle Traffic Prediction Matters?

With growing awareness of sustainable transportation development, low-carbon traffic modes have gained significant popularity as both individuals and governments increasingly recognize the environmental harm caused by conventional motorized transport [1]. To create more human-centered and resilient urban environments, many cities are actively encouraging cycling as a means of reducing greenhouse gas emissions, improving air quality, and decreasing dependence on private vehicles. These efforts are supported through an expanded cycling infrastructure and improved public services related to bicycle use. However, current efforts in building intelligent transportation systems (ITS) remain predominantly focused on motorized traffic [2], often overlooking bicycles as an important mode of transport. The support of ITS for bicycle traffic, which integrates road infrastructure and traffic management to reduce delays and improve safety, is still limited.

To build an inclusive and environmentally sustainable ITS, integrating bicycle traffic into monitoring and prediction frameworks is essential. Such integration enables city planners and traffic operators to optimize cycling infrastructure, allocate resources more efficiently, and implement timely safety interventions. However, bicycle traffic faces challenges similar to those of motorized transport, even in cycling-friendly countries such as the Netherlands, where bicycles are widely used for daily commute [3]. For example, peak commuting hours often lead to travel delays, while adverse weather conditions increase safety risks. Therefore, monitoring and predicting bicycle traffic is vital not only to improve operational control but also to improve the overall travel experience. By making cycling more attractive and reliable, such efforts encourage wider adoption and help to ensure that future transportation systems remain equitable, efficient, and environmentally sustainable.

In recent years, substantial progress has been made in traffic prediction research that aims to mitigate delays and improve safety, often through the use of advanced models that improve predictive accuracy. However, most existing studies have concentrated on motor vehicles or shared bicycle systems, which typically benefit from abundant data and well-defined usage patterns. In contrast, private bicycle traffic presents a different set of challenges. Due to limited resources and historical underinvestment in bicycle monitoring, data collection for private bicycles remains sparse, especially in small or medium-sized cities. Sensor coverage is often inconsistent, leading to fragmented datasets that are insufficient for traditional predictive modeling approaches. Although bicycle traffic may share certain temporal trends with motorized traffic, private bicycles exhibit distinct characteristics that complicate prediction. Their use is highly sensitive to external conditions, such as weather, road conditions, and urban form. In addition, private cyclists show greater route diversity, irregular travel schedules, and lack of digital tracking data, unlike shared bikes, which are often equipped with GPS. These differences result in complex and less predictable traffic patterns. As a result, there is a pressing need for tailored, data-driven prediction methods that can effectively model the unique behaviors of private bicycle traffic based on sensor-level observational data.

The remainder of this introduction is organized as follows. Starting with an overview of existing traffic prediction techniques. Building on this, we discuss spatial-temporal traffic prediction methods. Next, we review advanced paradigms, and finally, this introduction summarizes the current state of research on bicycle traffic prediction. Based on this background, the key scientific gaps are identified in Section 1.3. Section 1.4 outlines the research objectives and research questions that guide this study. Section 1.5 summarizes the main contributions of this thesis. Finally, Section 1.6 provides an outline of the thesis structure.

1.2 Traffic Prediction Methods and Their Potential for Bicycle Traffic Prediction

1.2.1 Overview of traffic prediction techniques

Traffic dynamics have been widely studied over the years in the transportation domain. Since the 1950s, researchers have developed physics-based models to analyze and simulate vehicular traffic behavior. Among the earliest and most influential are the Fundamental Diagram (FD) [4], the Lighthill-Whitham-Richards (LWR) model [5], and Car-Following Models [6]. The FD is a cornerstone of classical traffic flow theory and characterizes the relationships among traffic flow, density, and speed using three key diagrams: flow-density, speed-density, and speed-flow. FDs provide a macroscopic understanding of traffic behavior and are essential in applications such as traffic state estimation [7], congestion analysis [8], and control strategy design [9]. Despite its broad adoption, the FD assumes homogeneous traffic conditions, excluding lane-changing behavior and random disturbances of traffic, such as accidents or driver heterogeneity assumptions that limit its effectiveness in capturing real-world traffic complexity [10]. The LWR model, also developed in the mid-20th century, extends this macroscopic perspective by modeling traffic as a continuous fluid flow. It enables the study of traffic waves, including shock waves and congestion in bottlenecks, and has been widely implemented in macroscopic simulation tools, such as the Cell Transmission Model (CTM) for real-time prediction [11] and METANET for dynamic traffic management on highway networks [12]. Car-following models adopt a microscopic perspective of traffic flow prediction by focusing on the behavior of individual vehicles, such as their speed and acceleration. These models simulate how drivers adjust these behavior based on the actions of the vehicle ahead, and play a critical role in detailed traffic simulation tools such as AIMSUN and VISSIM [13]. These foundational models have laid the foundation for modern traffic prediction.

While physics-based models effectively describe and simulate traffic dynamics, they often fall short in real-time traffic prediction, particularly under the influence of unpredictable environmental factors such as accidents, adverse weather, or large-scale events [14]. These limitations led to a shift toward statistical modeling approaches beginning in the 1980s, which leveraged historical traffic data to predict future states through time series analysis [15]. Notable among these approaches is the Historical Average (HA), which estimates future values by averaging past observations. The AutoRegressive Integrated Moving Average (ARIMA) model [16] is also widely adopted, which can capture linear trends and seasonal patterns in traffic data. The Kalman Filter algorithm [17] has been widely used for recursive estimation of traffic states, providing a probabilistic framework for updating predictions as new observations become available. However, traffic systems are inherently nonlinear and stochastic, often influenced by complex interactions among traffic users, infrastructure, and external conditions. As a result, these three types of traditional statistical models struggle to fully capture the nonlinear dynamics and spatial-temporal dependencies characteristic of real-world traffic patterns [18].

In the 2000s, the emergence of statistical machine learning (ML) greatly advanced traffic prediction by enabling models to capture nonlinear patterns and spatial dependencies [15]. These capabilities improved the accuracy of real-time traffic dynamics prediction. Representative ML models applied in this domain include, Support Vector Regression (SVR) [19], Random Forests (RF) [20], Gradient Boosting algorithms (e.g., XGBoost, LightGBM) [21], K-Nearest Neighbors (KNN) [22], and shallow neural networks. SVR is known for its robustness to outliers and its ability to handle nonlinear relationships by mapping inputs into a high-dimensional feature space using kernel functions. Yet, the computational complexity of the SVR, arising from the need to solve a quadratic optimization problem, makes it more suitable for smaller datasets [23]. Ensemble learning methods, such as RF, XGBoost, and LightGBM, are capable of handling large-scale datasets efficiently and have demonstrated strong performance in various traffic prediction tasks [24]. However, these ensemble learning models are not inherently designed for sequential or temporal data, requiring the manual construction of time-based features [25]. As a result, they are typically more effective for low-frequency predictions, such as hourly or daily traffic flow prediction. The KNN could be used to predict traffic states by identifying historical observations similar to the current context and averaging their corresponding outcomes [26]. Although KNN is conceptually simple and interpretable, it suffers from high computational costs in large-scale networks or real-time applications, due to the need for continuous distance calculations across historical data [27].

Rapid advancement in computational power and the growing availability of large-scale traffic data have facilitated substantial improvements in traffic prediction through the adaptation of sophisticated deep learning techniques, which employ multi-layered neural networks to automatically learn complex spatial and temporal patterns from data. Temporal modeling of traffic patterns has evolved significantly with the adoption of recurrent architectures, particularly Long Short-Term Memory (LSTM) [28] and Gated Recurrent Unit (GRU) [29] networks, which effectively capture sequential dependencies while addressing the vanishing gradient problem. The field has also progressed with Transformer [30], which offers superior parallel processing capabilities and the ability to model both immediate fluctuations and extended temporal patterns. For spatial modeling, the inherent graph structure of transportation networks has motivated the widespread use of Graph Neural Networks (GNNs) [31], which naturally represent road networks through nodes and edges. Alternative approaches employ Convolutional Neural Networks (CNNs) [32] to process gridded representations of urban spaces, where the city is divided into regular zones, capturing localized spatial-temporal patterns.

Together, these models illustrate the evolution of temporal and spatial traffic prediction techniques, reflecting a shift from physics-based or simple machine learning models to deep learning approaches tailored for modeling large-scale, real-world time-series data.

1.2.2 Spatial-temporal traffic prediction

The essence of traffic prediction lies in identifying and learning the underlying temporal and spatial patterns that govern how traffic evolves over time and across the network [33]. Specifically, traffic at a given location often follows predictable short-term patterns. For example, morning and evening traffic demand peaks occur on most weekdays [34]. In the long term, traffic patterns may become apparent, for example, traffic volumes on weekdays follow different trends compared to weekends [35], reduced volumes during rainstorms compared to sunny days [36], or increased congestion in tourist cities during peak travel seasons [37]. These recurring patterns indicate that, if properly modeled, future traffic trends can be effectively predicted. Beyond the temporal dimension, traffic also exhibits strong spatial dependencies. Traffic operates within a connected road network. Conditions at one location can directly influence those at neighboring locations [38]. For example, congestion at a particular intersection may quickly propagate upstream to adjacent roads, forming a larger traffic bottleneck [39]. In addition to local spatial dependencies, traffic patterns also exhibit strong relationships between locations that are geographically far away from each other. For instance, areas near schools or universities often experience peak flows aligned with class schedules, while commercial districts follow different temporal patterns, such as lunchtime or evening rush hours [40]. Accurately understanding the traffic states of spatially influence is therefore essential for traffic prediction.

Modeling spatial-temporal dependencies involves combining spatial encoders, such as Convolutional Neural Networks (CNNs) for grid-based representations and Graph Neural Networks (GNNs) for irregular traffic networks, with temporal models, including Recurrent Neural Networks (RNNs) [41], attention mechanisms, and Transformer architectures. These hybrid approaches enable models to learn both localized and global spatial dependencies, as well as short and long-range temporal dependencies. Several representative models exemplify this integrated traffic prediction strategy. ST-ResNet [42], for instance, adopts a convolutional residual network to model temporal closeness, periodicity, and long-term trends in crowd traffic. Each temporal property is captured through a dedicated branch of residual convolutional units that also encode spatial features. DMVST-Net [43] introduces a multi-view architecture that integrates CNNs for local spatial correlation, LSTMs for temporal modeling, and a semantic embedding module to capture functional similarities among regions with shared demand patterns. ConvLSTM [44] extends fully connected LSTMs by incorporating convolutional structures in both input-to-state and state-to-state transitions, enabling spatial-temporal learning particularly effective in precipitation and traffic data. For graph-structured road networks, DCRNN [45] employs diffusion convolution to model spatial dependencies via bidirectional random walks, coupled with an RNN-based sequence-to-sequence framework for temporal traffic speed prediction. STGCN [46] combines graph convolutions with gated CNNs to separately model spatial and temporal features. ASTGCN [47] further enhances this approach by incorporating spatial and temporal attention mechanisms to dynamically adapt to varying conditions and capture latent dependencies across both domains.

Although spatial-temporal dependencies form the core of most traffic prediction models, real-world traffic systems are inherently stochastic and are significantly influenced by various exogenous factors [48], especially bicycle traffic. These include, but are not limited to, weather conditions, public holidays, special events, road accidents, variations in user behavior, factors that introduce irregularities and hidden uncertainties that might be caused by infrastructure [49]. Consequently, traffic prediction models that rely solely on historical traffic patterns may exhibit limited generalizability, particularly when deployed across heterogeneous urban environments or in non-stationary scenarios [50].

To address this limitation, recent research has emphasized the integration of external contextual information to improve the robustness and adaptability of traffic prediction models. For instance, ST-ResNet [42] employs a multi-branch residual convolutional framework to model temporal closeness, periodicity, and trend, while dynamically aggregating the outputs with learnable weights across regions. The model further incorporates external factors, such as weather and day of the week to enhance prediction accuracy of regional crowd flows. Similarly, the Spatial-Temporal Graph-to-Sequence Model (STG2Seq) [51] integrates external inputs including time of day, day of week, and holiday indicators, and uses a temporal attention mechanism to assign importance scores to historical time steps based on their relevance to future demand. The Graph Multi-Attention Network (GMAN) [52] encodes temporal contextual information by incorporating day-of-week and time-of-day features using one-hot encoding to produce enriched temporal embeddings. These studies collectively underscore the importance of external factor, highlighting its role in enhancing the generalization and real-world applicability of spatial-temporal traffic prediction models. However, these studies test their approaches on motorized traffic or shared-bike systems, which are not fully applicable to private bicycle traffic due to its distinct characteristics.

At the same time, several limitations still exist in considering external factors in the prediction models. First, the interactions between external factors and traffic flow have not been thoroughly explored or properly aligned; instead, they are often simply concatenated with traffic flow data in the same time slot. Second, external factors are often incorporated in static form, for example, without adequately modeling their dynamic effects on traffic behavior. In practice, the impact of an external event, such as rainfall or a large crowing event, can exhibit varying lag effects on traffic flow depending on the time of day, urban infrastructure, or local traffic regulations. These challenges are especially pronounced for bicycle traffic, which is far more sensitive to environmental conditions and urban design than motorized transport. As a result, existing models struggle to capture the complex, time-varying nature of exogenous influences, thereby limiting both predictive accuracy and interpretability in heterogeneous urban environments.

1.2.3 Advanced paradigms in traffic prediction

The application of deep learning algorithms introduces new challenges to the traffic prediction field, especially for bicycle traffic. Unlike motorized traffic systems, which often benefit from dense sensing infrastructure and abundant data, bicycle traffic data are frequently sparse or incomplete due to limited sensor coverage and infrastructure constraints [53]. Moreover, deep learning-based spatial-temporal traffic prediction models typically require large volumes of high-quality, labeled data, often aggregated from different sources [54]. In addition, data collected from sensitive sources, such as surveillance cameras or location-tracking systems, raises privacy and security concerns [55]. The centralized training of deep learning models often incurs substantial computational and communication overhead, limiting scalability and applicability in resource-constrained or distributed environments [56].

To overcome these limitations, several advanced paradigms have emerged that can reduce the dependency on centralized, large-scale datasets while enhancing model generalizability, preserving data locally, and lowering computational costs, offering promising directions for more practical and robust traffic prediction systems.

Underneath, three advanced paradigms are briefly introduced, namely self-supervised learning, transfer learning and federated learning.

Self-Supervised Learning (SSL) is a machine learning paradigm that enables models to learn meaningful representations from unlabeled data by generating supervisory signals from the data itself [57]. SSL methods can be broadly categorized into generative, contrastive, and generative-contrastive (adversarial) approaches [58]. Generative approaches aim to reconstruct or predict missing data. For instance, autoencoding models (e.g., variational autoencoders) are employed to impute missing traffic measurements by learning compact latent representations [59]. Autoregressive generative models, such as GPT-2, utilize Transformer decoder architectures to predict sequential data [60]. Unlike its predecessor GPT, GPT-2 removes task-specific fine-tuning and instead focuses on learning generalized representations from raw input sequences, enabling flexible adaptation to various downstream tasks [61]. Contrastive learning approaches aim to learn representations by distinguishing between similar (positive) and dissimilar (negative) samples [62]. Generative-contrastive learning, also referred to as adversarial representation learning, combines the strengths of both paradigms [63].

SSL for bicycle traffic - In bicycle traffic prediction, SSL, particularly autoregressive generative models such as GPT-2, can serve as pre-trained LLMs, transferring sequential representation learning to this domain, thereby reducing computational costs and improving prediction accuracy even with small datasets.

Transfer learning improves model performance on a target task by leveraging knowledge from a related, data-rich source domain. It can be categorized into four main solution types: instance-based, feature-based, parameter-based, and relational-based approaches [64]. In transportation, the instance-based approach assumes that source and target domains share similar traffic patterns except for differences in marginal distributions. For example, CrossTReS [65] is a selective transfer learning framework that adaptively re-weights source regions to assist target fine-tuning. The feature-based approach maps traffic data from both domains into a common feature space where their distributions align better. Common techniques include domain adaptation [66], adversarial learning [67], and meta-learning [68]. For instance, Ada-STGCN [67] integrates adversarial domain adaptation with spatial-temporal graph convolutional networks to learn discriminative, domain-invariant features, facilitating knowledge transfer from data-rich to data-scarce road networks. The parameter-based approach involves pretraining a traffic model on a large city and fine-tuning it on a smaller city with limited labeled data. For example, the TT-DL model [69] predicts real-time traffic flow in data-scarce roads by transferring knowledge from data-rich roads. The relational-based approach exploits similarities in road hierarchies or sensor layouts across districts, enabling knowledge transfer despite differing traffic volumes. Recent methods often combine multiple transfer learning strategies to leverage their complementary strengths. For example, TrafficTL [70] utilizes big data from other cities to assist data-scarce cities, identifying data similarity to reduce negative transfer and employing graph reconstruction techniques to mitigate data defects in small-data cities.

Transfer learning for bicycle traffic - Building on the above studies, in bicycle traffic prediction, where sensor networks are often sparse and data fragmented compared to motorized traffic systems, transfer learning offers a promising approach. By leveraging knowledge learned from related datasets, transfer learning can support model training in other urban settings while carefully mitigating the risk of negative transfer.

With increasing awareness of data privacy and security, along with the enforcement of data protection regulations such as the General Data Protection Regulation (GDPR) [55] and the California Consumer Privacy Act (CCPA) [71], individuals, enterprises, and governmental agencies increasingly prefer to store raw data locally and avoid sharing it externally. To address this, **Federated Learning** (FL), a distributed learning framework introduced by McMahan et al. [72], enables collaborative model training across decentralized clients without transmitting raw data to a central server. This paradigm preserves data privacy and reduces communication costs, making it particularly appealing in domains like traffic prediction. Even when traffic data is not inherently sensitive, competitive concerns among stakeholders and the high cost of data transfer further justify the adoption of FL.

FL for bicycle traffic - In the context of bicycle traffic prediction, FL is especially promising because bicycle traffic datasets are often distributed across multiple cities and each includes a small size of sensors. By training models locally and sharing only model parameters, FL allows different city districts to collaboratively improve predictive performance without exposing their raw datasets. However, most existing federated traffic prediction models primarily focus on capturing temporal dynamics, often neglecting the modeling of spatial dependencies and the loss of globally meaningful spatial correlations due to data decentralization [73, 74]. This limitation is critical for bicycle traffic, where accurate modeling of spatial correlations is essential for reliable prediction.

In summary, Self-Supervised Learning, Transfer Learning, and Federated Learning offer promising frameworks for tackling the challenges of bicycle traffic prediction, particularly those arising from limited data quality, data sparsity, and data security concerns.

1.3 Scientific Gaps

This thesis addresses several key scientific gaps in bicycle traffic prediction, including the influence of external factors, data sparsity, and data security concerns. These gaps are organized into the following four core research topics.

- This thesis investigates how to model spatial correlations in bicycle networks with high accuracy. With particular attention to the evolving influence of external environmental factors. These conclusions help guide the exploration of active mode traffic. [Chapter 2]

- This thesis studies how to effectively develop a bicycle traffic prediction framework by leveraging existing pretrained models rather than building entirely from scratch. These results will give insights into the potential for resource-efficient modeling in the transportation domain. [Chapter 3]
- This thesis investigates how to accurately predict bicycle traffic by leveraging knowledge learned from related datasets while avoiding negative knowledge transfer. This part addresses the data scarcity challenge commonly faced by small cities in traffic prediction. [Chapter 4]
- This thesis studies how to develop a spatial-temporal bicycle traffic prediction model that can achieve accurate predictions and does not require centralizing all data sources, thereby safeguarding raw data from direct sharing. These findings provide insight for the potential to minimize data leakage while facilitating collaborative modeling efforts across multiple data sources in traffic prediction. [Chapter 5]

The following sections provide a detailed explanation of the four identified scientific gaps.

Bicycle traffic is inherently less predictable than motorized traffic due to its sensitivity to external environmental factors [75]. Unlike motorized modes, which provide a physical enclosure that insulates travelers from weather and other external conditions, cyclists are directly exposed. Consequently, cyclist behavior is more influenced by weather, road conditions, and individual physical states, introducing a higher degree of variability and unpredictability. Most existing traffic prediction models are developed and evaluated using motorized traffic data or structured shared-bike systems [76], which are typically more controlled and organized through designated stations and docking infrastructure. These vehicular traffic-based models, however, are difficult to generalize to private bicycle traffic, which is more dynamic and unstructured. Furthermore, while current spatial-temporal traffic prediction approaches attempt to model spatial correlations through methods such as adaptive or dynamic graphs based on trainable parameters and data-driven correlation graphs which constructed from traffic features, they often overlook the influence of external contextual factors, particularly environmental conditions like weather, which are critical in bicycle traffic. In addition, existing fusion techniques, such as weighted averaging [77] or simple concatenation [78], may be insufficient for capturing the heterogeneous and context-dependent spatial relationships in bicycle traffic networks. Therefore, There is a need for more effective bicycle traffic prediction frameworks that account for external factors, such as weather, and employ advanced fusion strategies to integrate heterogeneous spatial correlations in bicycle traffic. [Chapter 2]

With the rise of Large Language Models (LLMs), their application has expanded into various domains, including traffic prediction. LLMs, trained on extensive text datasets, are designed to capture both semantic meaning and sequential dependencies in unstructured data. This makes them well-suited to support deep learning models in traffic prediction that rely on substantial training data. Despite the growing adoption of Large Language Models (LLMs) in traffic prediction, current applications primarily focus on motorized or public share-bike traffic [79, 80]. External factors such as weather or land use are often incorporated in a simplified manner [81], typically as static inputs, without adequately modeling their dynamic effects on traffic patterns. Building on insights from our preliminary study, we observe that weather conditions may exert lagged effects on spatial correlations in bicycle traffic networks, which in turn influence traffic flow due to their strong interdependence. These lagged impacts can affect bicycle ridership demand, route choice, and operational behavior not only in real time but also minutes or hours later. Moreover, spatial heterogeneity stemming from land use patterns and the presence or absence of dedicated cycling infrastructure contributes further to the variability of bicycle traffic. Therefore, there is a critical need to develop an LLM-based traffic prediction framework specifically designed for bicycle traffic. Such a framework must account for the multimodal and context-sensitive nature of active mobility and dynamically model the temporal and spatial influences of external factors to improve prediction accuracy. [Chapter 3]

Bicycle traffic is especially underrepresented in data collection efforts, as urban planning has historically prioritized motorized transport, resulting in inadequate monitoring systems for active modes of travel. Data scarcity and quality limitations significantly reduce the accuracy of traffic prediction models. Transfer learning offers a promising solution by enabling knowledge transfer from data-rich source domains to improve predictions in data-scarce target domains. However, most existing approaches are limited to single-source and single-target transfer settings [70], with little attention given to multi-source ensemble transfer learning, especially where source domains themselves may contain sparse or heterogeneous data. This challenge is amplified in the context of bicycle traffic, which lacks standardized road hierarchies and exhibits highly variable spatial-temporal patterns across different cities. Such variability complicates the identification of consistent and transferable trends. Therefore, there is a critical need for a robust multi-source transfer learning framework that can effectively integrate diverse sources and adapt to the unique characteristics of bicycle traffic, enabling more accurate predictions in data-scarce environments. [Chapter 4]

Aggregating traffic data from multiple sources can alleviate data sparsity and enhance predictive accuracy. However, centralized data aggregation introduces several challenges: it demands high network bandwidth and storage capacity, raises significant security concerns, and may lead to competitive risks when sensitive traffic data are shared across organizations. Federated learning provides a promising alternative by enabling model training across decentralized devices while keeping raw data local and secure [72]. Recent research has shown that federated learning can be effectively integrated with deep learning-based traffic prediction models, achieving a balance between performance and data security [54, 82]. However, a major challenge remains: capturing global spatial correlations, which are critical in traffic prediction but not directly accessible through local data alone. These spatial dependencies extend beyond physical road network connectivity to include traffic pattern similarities derived from sensor data across locations. Some studies address this by assuming only physical connectivity-based correlations [83, 84], while others introduce random noise to approximate spatial relations in a privacy-preserving manner [85]. However, such methods often fail to learn interpretable and semantically meaningful global spatial correlations within the federated setting. Effective prediction thus requires a nuanced understanding of both spatial interactions and complex network structures [86]. [Chapter 5]

1.4 Research Objective and Questions

According to the discussion above on the research background and scientific gaps, the main research question of this thesis is articulated as follows:

What are the promising solutions for accurately predicting bicycle traffic in complex networks given the effects of weather, data sparsity, and data security concern?

This thesis aims to address the above research gaps and achieve the overall objective by answering the following four key research questions.

Question 1: To what extent can a dynamic heterogeneous spatial-temporal traffic prediction model accurately capture spatial correlations specific by considering external influence on bicycle traffic?

The spatial correlations and temporal patterns among nodes in the bicycle traffic network are neither fixed nor uniform, as they are influenced by a variety of internal and external factors, such as weather. In this context, developing a dynamic heterogeneous spatial-temporal traffic prediction model is essential for accurately capturing the complexity of bicycle traffic. Graph Convolutional Networks (GCNs) and multi-head self-attention mechanisms offer powerful tools to model both spatial and temporal embeddings, while incorporating external factors to reflect the evolving relationships within the network. Furthermore, multi-head self-attention can effectively fuse heterogeneous spatial correlations, enabling the model to generate an optimal representation of graph-based spatial dependencies tailored to bicycle traffic. These capabilities not only improve prediction accuracy but also provide valuable insights into how external conditions impact mobility patterns, supporting more adaptive and resilient traffic management strategies in the face of changing environments. [Chapter 2]

Question 2: How can a pre-trained LLM-based framework be applied for bicycle traffic prediction that performs effectively across full-sample and limited-sample conditions?

Pre-trained LLMs, having been trained on large-scale textual data, possess strong generalization capabilities that can be leveraged for sequential data modeling, such as time-series traffic patterns. By utilizing this pre-trained knowledge, the model can support accurate predictions even when domain-specific data is scarce. To fully capture the spatial-temporal dynamics of bicycle traffic, the proposed framework integrates 1D convolutional neural networks (1D-CNNs) for temporal feature extraction and graph convolutional networks (GCNs) for modeling spatial dependencies among road segments. Additionally, an external processor module is incorporated to account for external influences, such as geographic semantic context and weather conditions. The findings aim to enables the framework to generalize effectively across cities, allowing for rapid adaptation to new urban environments through fine-tuning, thereby eliminating the need for costly model training from scratch. [Chapter 3]

Question 3: What are the opportunities in developing an effective spatial-temporal traffic prediction model for bicycle traffic in data-scarce scenarios by leveraging knowledge from multi-source traffic data collected across diverse urban environments?

Transfer learning is a well-known framework that enables the use of knowledge learned from data-rich environments to support modeling in data-scarce scenarios. However, the critical challenge lies in how to transfer this knowledge effectively. In the context of bicycle traffic, which can be heavily influenced by geographic factors such as infrastructure, land use, and the density of bicycle lane, it is important to take into account spatial variability. To reduce the risk of negative transfer, a clustering approach can be applied to first group sensors with similar traffic patterns, and then reconstruct spatial correlations based on traffic data of each sensors. Additionally, a learnable adaptive module should be designed to transfer knowledge from the multi-source domain to the target domain while adjusting for contextual differences, thereby minimizing negative transfer. This framework aims to provide an effective solution for accurate bicycle traffic prediction in cities with limited data availability. [Chapter 4]

Question 4: What challenges arise in managing the heterogeneity of bicycle traffic patterns to develop accurate spatial-temporal traffic prediction models while ensuring that raw data remain stored locally?

Federated learning is a promising machine learning paradigm that can be integrated with traffic prediction models to enable decentralized training. In this framework, traffic models are trained locally on devices using large volumes of traffic data, and only model parameters or updates are transmitted to a central server, thereby preserving data security. In the context of transportation systems, spatial-temporal traffic prediction must account not only for parameter updates but also for spatial correlations across different subnetworks. To address this, a spatial correlation aggregation mechanism must be designed specifically for the federated learning framework. Furthermore, bicycle traffic data are inherently heterogeneous across both time and space. For example, sensors located on main roads typically collect more data than those on secondary roads. To capture such heterogeneity, a multi-head self-attention mechanism is applied to model temporal traffic patterns. Additionally, Graph Convolutional Networks (GCNs) combined with a 2D convolution-based global multi-head self-attention mechanism are employed to capture complex and heterogeneous spatial correlations among sensors within each subnetwork at each client. The proposed framework is expected to enable distributed model training with comparable performance to centralized approaches, while enhancing security by ensuring that raw traffic data remain local. This supports effective and secure traffic data management in intelligent transportation systems. [Chapter 5]

1.5 Contributions

By addressing the four key research questions and achieving the overall research objective, this thesis contributes to the field of bicycle traffic prediction by developing accurate prediction approaches that effectively account for diverse conditions. In this subsection, we outline the major contributions of this thesis and the specific contributions of each paper in bridging scientific gaps and tackling key challenges.

- This thesis proposes advanced spatial-temporal methods for predicting bicycle traffic, incorporating the influence of weather and addressing the challenges of data sparsity.
- This thesis investigates distributed learning approaches for spatial-temporal traffic prediction, ensuring that raw data are retained locally.

1.5.1 Dynamic spatial-temporal model for active mode prediction

[Chapter 2]

- This study examines the influence of weather conditions on dynamic attention-based graph spatial correlations in bicycle traffic, by evaluating the prediction performance of the proposed dynamic attention-based spatial-temporal graph convolution network model (DyASTGCN).
- We propose a heterogeneous spatial correlation fusion approach that integrates distance-based, adaptive, and dynamic attention data-based graph correlations within the bicycle network, to enhance spatial representation and improve traffic prediction accuracy.
- We propose the DyASTGCN model for bicycle traffic prediction, combining multi-head self-attention for temporal dependencies and GCNs to capture spatial correlations, including weather influences and heterogeneous spatial

structures within the bicycle network. The proposed framework provides a foundation for developing robust and context-aware prediction models for active mobility systems.

1.5.2 LLM-based framework for bicycle traffic prediction

[Chapter 3]

- We propose a bicycle traffic prediction framework based on pre-trained LLMs, which incorporates spatial-temporal traffic patterns and external influences on bicycle traffic to enhance representation learning during fine-tuning.
- To account for the delayed impact of weather on bicycle traffic, we develop a lag-aware module that dynamically integrates weather embeddings with spatial-temporal traffic pattern representations.
- We evaluate the proposed framework using multi-city datasets under both full-sample and limited-sample conditions. The results demonstrate its effectiveness and adaptability for bicycle traffic prediction across varying data availability scenarios.

1.5.3 Transfer learning for bicycle traffic prediction

[Chapter 4]

- We propose a robust multi-source transfer learning spatial-temporal graph neural network (MultiTLSTGCN) specifically designed to address the challenges of data scarcity in spatial-temporal bicycle traffic prediction, enabling reliable predictions in data-constrained environments.
- Our approach begins with a traffic pattern clustering approach that groups similar patterns across source cities and reconstructs virtual spatial correlations for each cluster, enabling effective knowledge transfer while reducing the risk of negative transfer to target cities.
- Recognizing that only specific source clusters align with target city traffic patterns, we propose an adaptive transfer learning approach to enable more accurate and targeted knowledge transfer.

1.5.4 Federated spatial-temporal learning for active traffic prediction

[Chapter 5]

- We propose a heterogeneous spatial-temporal graph neural network based on federated learning (FedHSTGCN), which enables accurate traffic prediction across distributed clients while ensuring raw data remain stored locally. The results provide insights into the ability of the model in balancing prediction accuracy with data security.

- The heterogeneous spatial-temporal graph neural network (HSTGCN) is designed to investigate how spatially statistical heterogeneity among sensors influences the prediction of active mode traffic flow within a traffic network.

- We design the federated learning global spatial aggregation mechanism (FedGSAM) for the FedHSTGCN model to aggregate spatial correlations across client subnetworks, enabling the model to capture spatial influences from distributed traffic networks.

1.6 Thesis Outline

The structure of this paper-based thesis is illustrated in **Fig. 1.1**. Following the introductory and concluding chapters, the core of the thesis consists of four main chapters, each corresponding to a standalone research paper and addressing a specific research question. At the beginning of each chapter, the publication status of the associated paper is clearly indicated. If the paper has been published, the name of the journal or conference is provided; if not, the chapter notes whether the paper is under review and the venue of submission. To ensure consistency and transparency, each chapter presents the paper in its original form, identical to the version published or submitted to the respective outlet.

The four core chapters of this thesis present the development of several spatial-temporal traffic prediction approaches tailored for bicycle transportation. These chapters are grouped into two main categories, as visually distinguished by colorful rectangular in **Fig. 1.1**. Chapters 2 through 4 focus on centralized learning approaches, while Chapter 5 explores a distributed learning approach. Specifically, Chapter 2 introduces a dynamic spatial-temporal traffic prediction model that integrates weather factors to capture their lagged effects on spatial correlations in bicycle traffic. This study identifies two key challenges in the prediction of bicycle traffic: limited availability of bicycle data and high variability in the flow of bicycle traffic. These insights shape the research gap in Chapters 3, 4, and 5, with a variant of the base Transformer model from Chapter 2 being applied in Chapter 5. Chapter 3 presents a novel bicycle traffic prediction framework that leverages pre-trained large language models to enhance the prediction capability by integrating external information. Chapter 4 addresses the problem of data scarcity by proposing a transfer learning-based approach to improve prediction performance in data-limited settings. Chapter 5 transitions from centralized to distributed learning by introducing a federated learning framework. This approach allows for collaborative model

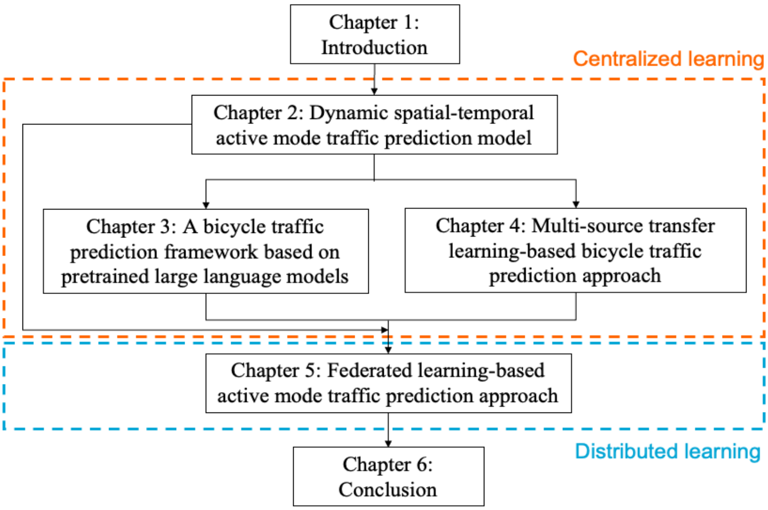


Figure 1.1 Outline of this thesis

training across multiple clients without sharing raw data, using a subnetwork spatial correlation fusion mechanism on the central server. Finally, Chapter 6 synthesizes the key findings from the previous chapters and offers a comprehensive conclusion. This last chapter also reflects on the broader implications of this work for practical applications and outlines promising directions for future investigation.

References

[1] U. Sotvoldiev. "THE PROBLEM OF ENVIRONMENTAL PROTECTION AND THE HARMFUL EFFECTS OF MOTOR TRANSPORT ON THE ENVIRONMENT". In: *Western European Journal of Modern Experiments and Scientific Methods* 2.6 (2024), pp. 274–279.

[2] B. Singh and A. Gupta. "Recent trends in intelligent transportation systems: a review". In: *Journal of transport literature* 9 (2015), pp. 30–34.

[3] E. Heinen, B. Van Wee, and K. Maat. "Commuting by bicycle: an overview of the literature". In: *Transport reviews* 30.1 (2010), pp. 59–96.

[4] C. F. Daganzo and N. Geroliminis. "An analytical approximation for the macroscopic fundamental diagram of urban traffic". In: *Transportation Research Part B: Methodological* 42.9 (2008), pp. 771–781.

[5] L. Leclercq. "Hybrid approaches to the solutions of the "Lighthill–Whitham–Richards" model". In: *Transportation Research Part B: Methodological* 41.7 (2007), pp. 701–709.

[6] P. G. Gipps. “A behavioural car-following model for computer simulation”. In: *Transportation research part B: methodological* 15.2 (1981), pp. 105–111.

[7] R. Shi, Z. Mo, K. Huang, X. Di, and Q. Du. “A physics-informed deep learning paradigm for traffic state and fundamental diagram estimation”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.8 (2021), pp. 11688–11698.

[8] P.-A. Cuniasse, C. Buisson, J. Rodriguez, E. Teboul, and D. De Almeida. “Analyzing railroad congestion in a dense urban network through the use of a road traffic network fundamental diagram concept”. In: *Public Transport* 7.3 (2015), pp. 355–367.

[9] M. D. Simoni, A. J. Pel, R. A. Waraich, and S. P. Hoogendoorn. “Marginal cost congestion pricing based on the network fundamental diagram”. In: *Transportation Research Part C: Emerging Technologies* 56 (2015), pp. 221–238.

[10] L. Zhang, Z. Yuan, L. Yang, and Z. Liu. “Recent developments in traffic flow modelling using macroscopic fundamental diagram”. In: *Transport Reviews* 40.6 (2020), pp. 689–710.

[11] A. Roy, M. Hossain, and Y. Muromachi. “Enhancing the prediction Performance of real-time crash prediction models: A cell transmission-dynamic Bayesian network approach”. In: *Transportation research record* 2672.38 (2018), pp. 58–68.

[12] Y. Wang, A. Messmer, and M. Papageorgiou. “Freeway network simulation and dynamic traffic assignment with METANET tools”. In: *Transportation Research Record* 1776.1 (2001), pp. 178–188.

[13] S. Panwai and H. Dia. “Comparative evaluation of microscopic car-following behavior”. In: *IEEE Transactions on intelligent transportation systems* 6.3 (2005), pp. 314–325.

[14] Z. Zhang, X. T. Yang, and H. Yang. “A review of hybrid physics-based machine learning approaches in traffic state estimation”. In: *Intelligent Transportation Infrastructure* 2 (2023), liad002.

[15] A. A. Kashyap, S. Raviraj, A. Devarakonda, S. R. Nayak K, S. KV, and S. J. Bhat. “Traffic flow prediction models–A review of deep learning techniques”. In: *Cogent Engineering* 9.1 (2022), p. 2010510.

[16] B. K. Nelson. “Time series analysis using autoregressive integrated moving average (ARIMA) models”. In: *Academic emergency medicine* 5.7 (1998), pp. 739–744.

[17] G. Welch, G. Bishop, *et al.* “An introduction to the Kalman filter”. In: (1995).

[18] D. A. Tedjopurnomo, Z. Bao, B. Zheng, F. M. Choudhury, and A. K. Qin. “A survey on modern deep neural network for traffic prediction: Trends, methods and challenges”. In: *IEEE Transactions on Knowledge and Data Engineering* 34.4 (2020), pp. 1544–1561.

[19] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. “Support vector regression machines”. In: *Advances in neural information processing systems* 9 (1996).

[20] L. Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.

[21] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz. “A comparative analysis of gradient boosting algorithms”. In: *Artificial Intelligence Review* 54.3 (2021), pp. 1937–1967.

[22] L. E. Peterson. “K-nearest neighbor”. In: *Scholarpedia* 4.2 (2009), p. 1883.

[23] L. Liu. “A short-term traffic flow prediction method based on svr”. In: *2021 2nd International Conference on Urban Engineering and Management Science (ICUEMS)*. IEEE. 2021, pp. 1–4.

[24] Z. Mei, F. Xiang, and L. Zhen-hui. “Short-term traffic flow prediction based on combination model of xgboost-lightgbm”. In: *2018 International Conference on Sensor Networks and Signal Processing (SNSP)*. IEEE. 2018, pp. 322–327.

[25] H. Xia, X. Wei, Y. Gao, and H. Lv. “Traffic prediction based on ensemble machine learning strategies with bagging and lightgbm”. In: *2019 IEEE international conference on communications workshops (ICC workshops)*. IEEE. 2019, pp. 1–6.

[26] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong. “An improved k-nearest neighbor model for short-term traffic flow prediction”. In: *Procedia-Social and Behavioral Sciences* 96 (2013), pp. 653–662.

[27] H. Yu, N. Ji, Y. Ren, and C. Yang. “A special event-based K-nearest neighbor model for short-term traffic state prediction”. In: *Ieee Access* 7 (2019), pp. 81717–81729.

[28] A. Graves and A. Graves. “Long short-term memory”. In: *Supervised sequence labelling with recurrent neural networks* (2012), pp. 37–45.

[29] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. “Empirical evaluation of gated recurrent neural networks on sequence modeling”. In: *arXiv preprint arXiv:1412.3555* (2014).

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[31] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. “A comprehensive survey on graph neural networks”. In: *IEEE transactions on neural networks and learning systems* 32.1 (2020), pp. 4–24.

[32] K. O’shea and R. Nash. “An introduction to convolutional neural networks”. In: *arXiv preprint arXiv:1511.08458* (2015).

[33] T. Anwar, C. Liu, H. L. Vu, M. S. Islam, and T. Sellis. “Capturing the spatiotemporal evolution in road traffic networks”. In: *IEEE Transactions on Knowledge and Data Engineering* 30.8 (2018), pp. 1426–1439.

[34] S. Yang and S. Qian. "Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents". In: *IEEE Intelligent Transportation Systems Magazine* 11.3 (2019), pp. 12–28.

[35] T. Zhou, L. Gao, and D. Ni. "Road traffic prediction by incorporating online information". In: *Proceedings of the 23rd International Conference on World Wide Web*. 2014, pp. 1235–1240.

[36] S. Dunne and B. Ghosh. "Weather adaptive traffic prediction using neurowavelet models". In: *IEEE Transactions on Intelligent Transportation Systems* 14.1 (2013), pp. 370–379.

[37] Z. Wang, Y. Chen, J. Su, Y. Guo, Y. Zhao, W. Tang, C. Zeng, and J. Chen. "Measurement and prediction of regional traffic volume in holidays". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE. 2019, pp. 486–491.

[38] Y. Kamarianakis and P. Prastacos. "Space–time modeling of traffic flow". In: *Computers & Geosciences* 31.2 (2005), pp. 119–133.

[39] B. S. Kerner. "Traffic congestion, modeling approaches to". In: *Encyclopedia of complexity and systems science*. Springer, 2009, pp. 9302–9355.

[40] Z. Bao, Y. Ou, S. Chen, and T. Wang. "Land use impacts on traffic congestion patterns: A tale of a Northwestern Chinese City". In: *Land* 11.12 (2022), p. 2295.

[41] L. R. Medsker, L. Jain, *et al.* "Recurrent neural networks". In: *Design and applications* 5.64-67 (2001), p. 2.

[42] J. Zhang, Y. Zheng, and D. Qi. "Deep spatio-temporal residual networks for citywide crowd flows prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017.

[43] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li. "Deep multi-view spatial-temporal network for taxi demand prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. 1. 2018.

[44] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting". In: *Advances in neural information processing systems* 28 (2015).

[45] Y. Li, R. Yu, C. Shahabi, and Y. Liu. "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting". In: *arXiv preprint arXiv:1707.01926* (2017).

[46] B. Yu, H. Yin, and Z. Zhu. "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting". In: *arXiv preprint arXiv:1709.04875* (2017).

[47] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan. "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 922–929.

[48] S. Berkani, B. Guermah, M. Zakroum, and M. Ghogho. "Spatio-temporal forecasting: A survey of data-driven models using exogenous data". In: *IEEE Access* 11 (2023), pp. 75191–75214.

[49] J. Ye, S. Xue, and A. Jiang. "Attention-based spatio-temporal graph convolutional network considering external factors for multi-step traffic flow prediction". In: *Digital Communications and Networks* 8.3 (2022), pp. 343–350.

[50] L. Wang, D. Chai, X. Liu, L. Chen, and K. Chen. "Exploring the generalizability of spatio-temporal traffic prediction: Meta-modeling and an analytic framework". In: *IEEE Transactions on Knowledge and Data Engineering* 35.4 (2021), pp. 3870–3884.

[51] L. Bai, L. Yao, S. Kanhere, X. Wang, Q. Sheng, *et al.* "Stg2seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting". In: *arXiv preprint arXiv:1905.10069* (2019).

[52] C. Zheng, X. Fan, C. Wang, and J. Qi. "Gman: A graph multi-attention network for traffic prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. 01. 2020, pp. 1234–1241.

[53] B. Wang, P. Wang, Y. Zhang, X. Wang, Z. Zhou, and Y. Wang. "Condition-guided urban traffic co-prediction with multiple sparse surveillance data". In: *IEEE Transactions on Vehicular Technology* 74.1 (2024), pp. 166–178.

[54] Y. Liu, J. James, J. Kang, D. Niyato, and S. Zhang. "Privacy-preserving traffic flow prediction: A federated learning approach". In: *IEEE Internet of Things Journal* 7.8 (2020), pp. 7751–7763.

[55] P. Regulation. "General data protection regulation". In: *Intouch* 25 (2018), pp. 1–5.

[56] T. Liu, Y. Wang, H. Zhou, J. Luo, and F. Deng. "Distributed short-term traffic flow prediction based on integrating federated learning and TCN". In: *IEEE Access* (2024).

[57] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon. "A survey on contrastive self-supervised learning". In: *Technologies* 9.1 (2020), p. 2.

[58] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang. "Self-supervised learning: Generative or contrastive". In: *IEEE transactions on knowledge and data engineering* 35.1 (2021), pp. 857–876.

[59] S. Zhang, X. Hu, J. Chen, W. Zhang, and H. Huang. "An effective variational auto-encoder-based model for traffic flow imputation". In: *Neural Computing and Applications* 36.5 (2024), pp. 2617–2631.

[60] Y. Xu and M. Liu. "GPT4TFP: Spatio-temporal fusion large language model for traffic flow prediction". In: *Neurocomputing* 625 (2025), p. 129562.

[61] J. Ye, W. Zhang, K. Yi, Y. Yu, Z. Li, J. Li, and F. Tsung. "A survey of time series foundation models: Generalizing time series representation with large language model". In: *arXiv preprint arXiv:2405.02358* (2024).

[62] H. Hu, X. Wang, Y. Zhang, Q. Chen, and Q. Guan. "A comprehensive survey on contrastive learning". In: *Neurocomputing* 610 (2024), p. 128645.

[63] Y. Wang, L. Sang, Y. Zhang, Y. Zhang, and X. Wu. "Generative-contrastive heterogeneous graph neural network". In: *IEEE Transactions on Big Data* (2025).

[64] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. "A comprehensive survey on transfer learning". In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.

[65] Y. Jin, K. Chen, and Q. Yang. "Selective cross-city transfer learning for traffic prediction via source city region re-weighting". In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2022, pp. 731–741.

[66] J. Shao, S. Li, K. Zhang, A. Wang, and M. Li. "Cross-City traffic prediction based on deep domain adaptive transfer learning". In: *Transportation Research Part C: Emerging Technologies* 176 (2025), p. 105152.

[67] Z. Yao, S. Xia, Y. Li, G. Wu, and L. Zuo. "Transfer learning with spatial–temporal graph convolutional network for traffic prediction". In: *IEEE Transactions on Intelligent Transportation Systems* 24.8 (2023), pp. 8592–8605.

[68] J. Mo and Z. Gong. "Cross-city multi-granular adaptive transfer learning for traffic flow prediction". In: *IEEE Transactions on Knowledge and Data Engineering* 35.11 (2022), pp. 11246–11258.

[69] B. Wang, Z. Yan, J. Lu, G. Zhang, and T. Li. "Road traffic flow prediction using deep transfer learning". In: *Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference (FLINS 2018)*. World Scientific. 2018, pp. 331–338.

[70] Y. Huang, X. Song, Y. Zhu, S. Zhang, and J. J. Yu. "Traffic prediction with transfer learning: A mutual information-based approach". In: *IEEE Transactions on Intelligent Transportation Systems* 24.8 (2023), pp. 8236–8252.

[71] E. Goldman. "An introduction to the california consumer privacy act (ccpa)". In: *Santa Clara Univ. Legal Studies Research Paper* (2020).

[72] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. "Communication-efficient learning of deep networks from decentralized data". In: *Artificial intelligence and statistics*. PMLR. 2017, pp. 1273–1282.

[73] Y. Liu, S. Zhang, C. Zhang, and J. James. "Fedgru: Privacy-preserving traffic flow prediction via federated learning". In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2020, pp. 1–6.

[74] S. S. Sepasgozar and S. Pierre. "Fed-ntp: A federated learning algorithm for network traffic prediction in vanet". In: *IEEE Access* 10 (2022), pp. 119607–119616.

[75] P. Spencer, R. Watts, L. Vivanco, and B. Flynn. “The effect of environmental factors on bicycle commuters in Vermont: influences of a northern climate”. In: *Journal of Transport Geography* 31 (2013), pp. 11–17.

[76] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li. “Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 5668–5675.

[77] W. Zhang, F. Zhu, Y. Lv, C. Tan, W. Liu, X. Zhang, and F.-Y. Wang. “AdapGL: An adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks”. In: *Transportation Research Part C: Emerging Technologies* 139 (2022), p. 103659.

[78] X. Ta, Z. Liu, X. Hu, L. Yu, L. Sun, and B. Du. “Adaptive spatio-temporal graph neural network for traffic forecasting”. In: *Knowledge-based systems* 242 (2022), p. 108199.

[79] C. Liu, S. Yang, Q. Xu, Z. Li, C. Long, Z. Li, and R. Zhao. “Spatial-temporal large language model for traffic prediction”. In: *2024 25th IEEE International Conference on Mobile Data Management (MDM)*. IEEE. 2024, pp. 31–40.

[80] Y. Ren, Y. Chen, S. Liu, B. Wang, H. Yu, and Z. Cui. “TPLLM: A traffic prediction framework based on pretrained large language models”. In: *arXiv preprint arXiv:2403.02221* (2024).

[81] Z. Li, L. Xia, J. Tang, Y. Xu, L. Shi, L. Xia, D. Yin, and C. Huang. “Urbangpt: Spatio-temporal large language models”. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, pp. 5351–5362.

[82] Y. Qi, M. S. Hossain, J. Nie, and X. Li. “Privacy-preserving blockchain-based federated learning for traffic flow prediction”. In: *Future Generation Computer Systems* 117 (2021), pp. 328–337.

[83] C. Zhang, S. Zhang, S. Yu, and J. James. “Graph-based traffic forecasting via communication-efficient federated learning”. In: *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2022, pp. 2041–2046.

[84] X. Yuan, J. Chen, J. Yang, N. Zhang, T. Yang, T. Han, and A. Taherkordi. “Fedstn: Graph representation driven federated learning for edge computing enabled urban traffic flow prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* 24.8 (2022), pp. 8738–8748.

[85] C. Zhang, S. Zhang, J. James, and S. Yu. “FASTGNN: A topological information protected federated learning approach for traffic speed forecasting”. In: *IEEE Transactions on Industrial Informatics* 17.12 (2021), pp. 8464–8474.

[86] J. Lin and Y. Ban. “Complex network topology of transportation systems”. In: *Transport reviews* 33.6 (2013), pp. 658–685.

Chapter 2

Dynamic Spatial-temporal Model for Active Mode Prediction

This chapter introduces the Dynamic Attention-based Spatial-Temporal Graph Convolution Network (DyASTGCN) model, which incorporates the influence of weather on spatial correlations within the active mode traffic network. The experimental results reveal that weather changes exhibit a lagging effect on these spatial correlations, highlighting the dynamic interplay between environmental factors and traffic patterns. Additionally, we introduce a fusion approach to integrate various heterogeneous spatial correlations, aiming to represent the optimal spatial correlations within the active mode network. Given the uncertain traffic states and the highly sparse nature of active mode data, this fusion approach proves adept at identifying and capturing critical spatial correlations, which are essential for precise traffic flow prediction. By leveraging this enhanced understanding of complex graph correlations and traffic patterns, our model achieves improved prediction accuracy and offers deeper insights into the dynamics of active mode networks. These findings underscore the significance of addressing the sparsity and complexity inherent in active mode data, which lays the foundation for designing robust prediction models explored in the subsequent chapters.

This chapter is published as a journal article: Wen, X., Krishnakumari, P. and Hoogendoorn, S. Dynamic Spatial-Temporal Graph Convolutional Neural Networks Approach for Active Mode Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2025. The content is the same as the journal paper.

2.1 Introduction

Recently, low-carbon transportation has gained popularity as people become more aware of the negative impacts of motorized traffic modes on the environment [1]. Active mode transportation such as walking and cycling as a sustainable, low-carbon and healthy alternative to driving is promoted in more countries [2]. However, the growing inclination towards active mode traffic introduces familiar traffic issues, such as the travel time delay, increased number of stops at (controlled) intersections, increased travel time unreliaiblity, along with new challenges, such as insufficient and narrow bike lanes and sidewalks, increasing number of accidents, etc. These factors can make users of active mode feel unsafe and encounter difficulties while cycling or walking. Furthermore, cycling or walking in bad or even extreme weather, such as high winds and thunderstorms, could increase the risk of accidents [3]. To alleviate the delays and risks faced by active mode travelers, several measures can be implemented. These include adopting traffic control and management strategies that prioritize active modes during adverse weather conditions, utilizing advanced traffic monitoring systems to optimize flow and enhance safety, and developing mobile applications that provide real-time traffic information and personalized recommendations for those using active modes. By implementing these measures, cities can create a more responsive and safer environment for pedestrians and cyclists. In this context, accurate prediction of active mode traffic is gaining significant attention from researchers owing to its crucial role in optimizing travel routes and facilitating informed decisions for transportation planning and infrastructure investment.

However, predicting active mode traffic presents unique challenges owing to its inherent uncertainty, sparse data availability, and significant noise in the datasets. Unlike traditional traffic predictions, which primarily involve motorized vehicles, prediction for active modes like walking or cycling is influenced by a multitude of factors that shape travel patterns. These factors include individual preferences dictating route choices, fluctuations in travel speeds influenced by physical conditions, and sensitivity to external elements such as weather variations, infrastructure quality, and prevailing traffic conditions. These complexities render active mode traffic more variable and less predictable compared to motorized traffic [4]. Therefore, achieving accurate predictions for active mode traffic requires a comprehensive understanding of both the temporal relationship and the uncertain spatial relationship. Existing research on active mode traffic prediction has mainly focused on predicting bike sharing usage [5]. In bike-sharing prediction, two main approaches have been utilized: traditional statistical models like regression analysis [6], and machine learning models such as neural networks [7]. These studies focus mainly on determining routes and estimating rental demand for docking stations or parking areas [8–10]. Bike-sharing systems are usually meticulously organized, featuring predetermined stations and docking systems. The supply and arrangement of bikes are centrally controlled, providing a high degree of system control and predictability. In contrast, private bike owners enjoy the flexibility to use their bikes according to personal schedules, destinations, and travel habits. This variability poses a challenge in collecting traffic data for private bikes and predicting bicycle traffic flow, resulting in increased uncertainty compared to the more structured and predictable nature of bike-sharing systems.

Although active mode traffic prediction might be more complex than motorized traffic, the methods developed for motorized traffic modes could still prove useful for active mode traffic prediction. In recent years, with the availability of large amounts of traffic data and the help of advanced machine learning algorithms, traffic prediction for motorized traffic modes has seen great success [11–14]. Most studies focus on developing spatial-temporal traffic prediction methods to capture the change in traffic patterns over time and across different locations. For example, Yao et al. [15] proposed a Spatial-Temporal Dynamic Network (STDN) for traffic prediction, which captures the spatial and temporal information using local Convolutional Neural Network (CNN) and Long Short-term Memory (LSTM), respectively. The evaluation of the taxi data of New York City (NYC), bike sharing data from NYC, and Jinan Road camera data proved the model's effectiveness. To capture the spatial, short-term and long-term periodical dependencies of traffic patterns, Shi et al. [16] developed an Attention-based Periodic-Temporal Neural Network (APTN) model. Its prediction results outperform state-of-the-art methods. However, these methods were predominantly crafted and validated based on the traffic characteristics of motorized traffic. As a consequence, the distinctive attributes of active modes, such as sensitivity to weather conditions and the uncertainty in route selection, which can significantly influence the spatial dynamics of active modes, are not adequately captured in existing traffic models. For example, conventional graph neural networks frequently depend on fixed or adaptive adjacency matrices to extract spatial information [17]. The fixed adjacency matrix is typically constructed based on physical connections and distances between nodes, providing a static representation of network relationships. In contrast, adaptive and dynamic adjacency matrices are generated using trainable parameters, allowing them to evolve based on data-driven insights and changing conditions. However, these matrices may not fully capture the intricate details of active modes, which are highly responsive to external features and exhibit dynamic behaviors. Furthermore, the diverse types of adjacency matrices symbolize distinct spatial relationships among traffic modes. Employing a weighted averaging method [18] or a concatenation approach [17] might not be adequate to seamlessly merge this varied information into a unified adjacency matrix that captures the relationships within the graph structure. Consequently, a fused approach is necessary to integrate these heterogeneous spatial relationships effectively.

To address the challenges outlined above, this study proposes an approach called the DyASTGCN for accurate active mode traffic prediction. Furthermore, this method enhances prediction accuracy by integrating the impact of weather conditions on active mode dynamic attention data-based graph spatial correlations and effectively fusing heterogeneous spatial correlations within the active mode network.

- We investigate the impact of weather conditions on dynamic attention data-based graph spatial correlations within active mode traffic by analyzing the prediction accuracy of the proposed DyASTGCN traffic prediction model. Specifically, we examine three scenarios with different combinations of traffic flow and weather factors to derive dynamic attention data-based graph spatial correlations for active mode traffic, each scenario incorporating various lag sizes between traffic flow and weather conditions. Experimental results reveal that weather significantly impacts active mode traffic, with weather effects exhibiting a lagged influence on changes in spatial correlations.

- To capture the optimal representation of graph spatial correlations in active mode traffic, we proposed a heterogeneous spatial correlation fusion approach. This method aggregates multiple heterogeneous spatial graph correlations including predefined distance-based graph correlations, parameter-based adaptive graph correlations, and dynamic attention data-based graph correlations within the active mode network to enhance traffic prediction accuracy. Experimental results reveal that the fused graph spatial correlations provide the best representation of the spatial dynamics of active mode traffic, when compared to other individual graph correlations. Consequently, the prediction results of the DyASTGCN model, which utilizes the fused graph spatial correlations, outperform those based on other individual graph correlation methods.

- Given the temporal and spatial characteristics of active mode traffic, we introduce the DyASTGCN model to predict traffic flow accurately. This model leverages multi-head self-attention to capture intricate temporal relationships and utilizes Graph Convolution Networks (GCN) to incorporate spatial information, including weather influences and heterogeneous spatial correlations within the active mode network. Our prediction results demonstrate that DyASTGCN outperforms baseline models in forecasting active mode traffic.

The rest of this paper is structured as follows. Section 2 provides an overview of related work in the field of active mode prediction. Section 3 details the methodology of our proposed model. Section 4 covers our data, including data collection and filling approaches, the dataset used for assessment, experimental configurations, as well as the findings and analysis from our experiments. Section 5 discusses the implications and contributions of our work.

2.2 Related Work

In this section, we first review the existing traffic prediction studies. We then explore research on traffic prediction using graph neural networks, particularly focusing on how these studies capture the spatial correlations of the graph network for prediction. Finally, we review studies that address heterogeneity in traffic networks.

2.2.1 Traffic prediction

Predicting traffic is a vital element of the intelligent transportation system [19–21] and plays a key role in contemporary traffic control and management. With the growing volume and variety of available traffic data, data-driven methods, including statistical methods, traditional machine learning methods, and deep learning methods, have gained significant attention in traffic prediction research [22]. Especially deep learning methods that can be highly effective for predicting traffic patterns in real time and can adapt to changes in traffic flow patterns over time [23–25]. Duan et al. [26] proposed a deep hybrid neural network by integrating a Convolutional Neural Network (CNN) and Long Short-term Memory (LSTM) to predict urban traffic flow using real GPS taxi trajectory data from Xi'an city. A greedy policy is used in training to reduce computation time; experimental results show that the proposed method outperforms existing methods. Similarly, Wu et al. [27] integrated deep Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) as a novel traffic flow prediction method that can capture the spatial and temporal features to predict traffic flow; the experimental results show that the proposed method can learn specific knowledge from large traffic flow data. However, these traffic prediction methods are developed primarily for motorized traffic modes.

For active mode traffic prediction, because of its complexity, lack of data and sensitivity to external factors (e.g., weather), it is more difficult to capture traffic patterns accurately on the road network. Existing studies on active mode mainly focus on bike-sharing systems [28–30]. For example, to predict the number of bicycles rented and returned to each parking station, Li et al. [31] developed a hierarchical prediction model and evaluated its performance in two bike-sharing systems in New York City and Washington, DC, respectively, which shows the advantage of the model. However, these studies primarily focus on predicting demand at fixed stations for bikes, thereby ignoring the traffic state of shared bikes on the road network. Additionally, predicting traffic patterns for active modes such as private or shared bicycles on road network, which lack fixed stations or parking zones, remains more unpredictable due to the uncertainty surrounding trip origins and destinations.

2.2.2 Graph neural networks on traffic prediction

The advent of Graph Neural Networks (GNNs) has significantly enhanced the capacity to investigate spatial relationships in traffic prediction. GNNs are specifically engineered to capture non-Euclidean spatial structural data, which aligns more closely with the intricacies of the structure of the traffic network [32]. In this context, capturing the complexity and nonlinear traffic patterns seems possible when combining GNNs with existing temporal dependencies learning models [33].

For example, Zhao et al. [34] proposed a novel network-based Temporal Graph Convolutional Network model (T-GCN), which could simultaneously capture spatial and temporal dependences to predict road traffic based on the SZ-taxi dataset and the Los-loop dataset. To achieve an accurate prediction of traffic flow, Li et al. [35] explored the spatial-temporal features in traffic flow using a Graph and Attention-based Long Short-Term Memory Network (GLA); The results show that this method performs better than most previous methods based on PeMS dataset. However, these studies mainly predefined the graph structure relationship based on the Euclidean distance or learned the graph structure relationship according to some attributes of the road network, such as POI distribution and regional function, which is insufficient to contain all the valuable information for the complicated active mode traffic prediction scenario.

Some graph-based traffic prediction approaches aim to learn the underlying dependencies of graph relationships adaptively. Wu et al. [36] proposed Graph WaveNet for spatial-temporal graph modeling, which retains the dependency matrix by capturing hidden spatial dependencies in the data in an adaptive way. Zhang et al. [18] designed four different types of relationships between nodes—origin-destination (OD) relationship, transfer relationship, distance relationship, and correlation coefficient relationship to help adaptively exploit hidden correlations between nodes. These studies could capture the underlying spatial relationships between nodes adaptively to some extent. However, the spatial relationships might vary over time depending on traffic conditions and the traffic network environment. To address the dynamic nature of spatial relationships, Hu et al. [37] designed a graph learning module to learn spatial dependencies in the traffic network based on input data, complemented by a dilated causal convolution network with a gating mechanism to capture long-term temporal correlations in the traffic data. Additionally, Ta et al. [17] developed an Adaptive Spatial-Temporal Graph Neural Network (ASTGNN) for multi-step traffic condition forecasting, which captured the optimal graph structure considering node attributes and complex spatial-temporal correlations using a spatial-temporal convolution architecture. However, active modes traffic are influenced by weather conditions and the physical well-being of travelers, introducing a higher level of complexity in their spatial relationships compared to traditional traffic modes. Given this context, there is an increasing demand for novel traffic prediction methods capable of capturing these uncertain spatial relationships inherent in active modes.

2.2.3 Heterogeneous spatial correlations handling in traffic networks

Existing graph architectures are often developed for homogeneous graphs with identical types of nodes and edges. For example, some researchers have focused on graph embedding for undirected and unweighted homogeneous graphs, considering only the structural information of the graph [38]. To extract more information from the graph, weighted graphs with identical types of nodes and edges have also been explored [39]. Some studies have considered directed graphs, which can provide a more precise graph representation [40]. However, these approaches are not well-suited for graphs that contain different types of nodes and edges, as they cannot properly capture the intricate representation and interaction of diverse types of nodes and edges within a graph. To address this limitation, researchers have developed methods for heterogeneous networks. For example, Chang et al. [41] designed a deep embedding algorithm to capture correlations between heterogeneous data in a network, demonstrating the effectiveness and scalability of their approach. In traffic systems, the represented graphs of traffic networks also contain various types of nodes and edges, including weighted edges based on distance, similarity functions, and traffic patterns, as well as nodes from different traffic modes. Several studies have explored traffic patterns and characteristics in heterogeneous traffic networks. For instance, Liang et al. [42] proposed a Multi-Relational Spatial-Temporal Graph Neural Network (ST-MRGNN) for multimodal demand prediction, accounting for diverse spatial units and heterogeneous spatial-temporal correlations across subway and ride-hailing modes. This model outperforms existing methods. Similarly, Guo et al. [43] developed an Attention-based Spatial-Temporal Graph Neural Network (ASTGNN) for highway traffic prediction by considering periodicity and spatial heterogeneity through embedding modules, surpassing state-of-the-art baselines. Given this context, exploring heterogeneous dynamic graph relationships in active mode is crucial for accurately capturing and understanding traffic patterns.

2.3 Proposed Methodology for Active Mode Traffic Prediction

In this section, we begin with a brief overview of the definition and problem statement of this study. We then delve into the methodology employed to capture the various spatial correlations within the active mode network traffic and the approach used to fuse these spatial correlations to achieve optimal spatial correlations for active mode traffic. Finally, we systematically present the structure of the proposed DyASTGCN model for active mode traffic prediction.

2.3.1 Definition and problem statement

Definition (Graph traffic network): To streamline computations while effectively capturing significant traffic patterns, we define the traffic network as an undirected graph $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_t)$, where $\mathcal{G}_t = (V, E, \mathbf{A}_t)$ denotes the active mode graph at time t, V is the set of nodes representing sensors that record the traffic flow of active modes on the road network; E is a set of edges; and $\mathbf{A}_t \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ is the adjacency matrix of $\mathcal{G}_t$ with N nodes at time t recorded as a weighted adjacency matrix.

Problem (Multi-step active mode traffic prediction): The historical active mode traffic data for each input includes data from the past hour, represented as: $\mathbf{X} = (\mathbf{X}_{t-k+1}, \mathbf{X}_{t-k+2}, ..., \mathbf{X}_t) \in \mathbb{R}^{N \times F \times T_k}$, where T_k are the time steps of each input, $\mathbf{X}_t = (\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, ..., \mathbf{x}_{t,N}) \in \mathbb{R}^{N \times F}$ indicates the vector of characteristics of each node v at time t, F is the dimension of the characteristic of each node. Our goal is to find a function f to predict the following T time steps data $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_{t+1}, \hat{\mathbf{x}}_{t+2}, ..., \hat{\mathbf{x}}_{t+T}) \in \mathbb{R}^{N \times F \times T}$, that is:

$$(\hat{\mathbf{x}}_{t+1}, \hat{\mathbf{x}}_{t+2}, ..., \hat{\mathbf{x}}_{t+T}) = f_\theta((\mathbf{X}_{t-k+1}, ..., \mathbf{X}_t), \mathbf{A}_t) \tag{2.1}$$

here, θ represents the learnable parameters of the function. Considering the sensitivity of active mode traffic to external factors, our approach utilizes the most recent hour of historical data as input to predict the traffic for the subsequent hour.

2.3.2 Graph spatial dependencies learning

Active mode traffic graph spatial correlations encompass various types of relationships. Within a traffic network, the flow at a road section is influenced by its connections, suggesting that distance-based predefined graph correlations can reflect the geographic network correlations for active modes. Additionally, different road structures and facilities can impact these correlations, which could be explored through a parameter-based adaptive graph [36]. Furthermore, surrounding traffic patterns and weather conditions contribute to changing spatial correlations over time in the active mode network. To capture these dynamics, we introduce a dynamic attention data-based graph correlation matrix to model active mode spatial correlations.

However, these individual spatial correlations only capture partial aspects of the spatial relationships in active mode traffic. To comprehensively capture the overall graph correlations in active mode, we introduce a fusion approach. This method integrates these heterogeneous graph correlations to derive optimal spatial relationships for the analysis of active mode traffic.

(1) Distance-based predefined graph spatial correlations learning

Spatial proximity between sensors often implies similarity in the traffic patterns they detect. Therefore, we model the sensors as nodes in a graph, where the connectivity between two nodes is determined by the shortest path obtained using Dijkstra's algorithm [44] based on the active mode road network. To capture the spatial dependency, we define the edge weights based on the distance of the shortest path between two nodes, employing a threshold Gaussian kernel weighting function [45] as follows,

$$A_d^{i,j} = \begin{cases} \exp(-\frac{|\mathrm{dist}(i,j)|^2}{2\theta^2}) & \text{if } dist(i,j) < K \\ 0 & \text{else} \end{cases} \tag{2.2}$$

where $A_d^{i,j}$ denotes the distance weight of the graph, with $|\mathrm{dist}(i,j)|$ representing the shortest path distance between node i and node j. Here, θ signifies the standard deviation of distances, and K serves as the threshold.

(2) Parameter-based adaptive graph spatial correlations learning

Active mode traffic flow is highly dynamic, influenced by various factors such as events, accidents, traffic lights, road connectivity, and surrounding environmental conditions. These factors introduce complexities in capturing active mode spatial correlations. Parameter-based adaptive methods offer flexibility in adjusting the correlations between nodes. Therefore, a parameter-based adaptive graph structure learning method [36] is utilized to capture the hidden spatial dependencies within the active mode graph. This method utilizes two embedding dictionaries with learnable parameters to derive the spatial dependency weight among nodes $\mathbf{E}_1$,$\mathbf{E}_2 \in \mathbb{R}^{\mathbb{N}\times\mathbb{P}}$, as follows,

$$\mathbf{A}_a = SoftMax(ReLU(\mathbf{E}_1\mathbf{E}_2^T) \tag{2.3}$$

where P is the hidden dimensions of each node. $\mathbf{E}_1$ is the source node embedding dictionary, $\mathbf{E}_2$ is the target node embedding dictionary. ReLU is an activation function used to introduce nonlinearity into the matrix and ensure the values in matrix $\mathbf{A}_a$ are non-negative. The SoftMax function is then applied to normalize the matrix, converting it into a probability distribution where the sum of the values is equal to 1.

(3) Dynamic attention data-based graph spatial correlations learning

Geographical proximity alone may not adequately capture spatial correlations between nodes, as unconnected nodes can exhibit stronger correlations than those in close physical proximity. For instance, nodes with similar functions or roads at the same hierarchy level may demonstrate similar traffic patterns, even if they are geographically distant. Therefore, capturing spatial correlations based on traffic flow feature involves exploring pattern similarities that do not rely on geographical closeness. Moreover, active mode traffic is notably susceptible to various external factors, particularly bad weather conditions. These factors can significantly impact cyclists' route choices or prompt shifts to other traffic modes, thereby increasing the uncertainty, complexity, and difficulty of capturing the spatial correlations within the active mode traffic network based on traffic flow feature. Weather changes do not immediately impact traffic spatial correlations; their effects manifest gradually. For example, consider a sudden onset of rain during peak commuting hours. Initially, there might be a minor reduction in the number of cyclists on the roads, as some individuals might continue their commute despite the rain. However, as the rain persists and intensifies, more cyclists may opt to seek shelter or switch to public transportation. This gradual shift in behavior alters the spatial correlations within the traffic network over time, as illustrated in **Fig. 2.1**. This lag in the manifestation of weather effects highlights the importance of capturing dynamic, evolving spatial correlations for accurate traffic predictions.

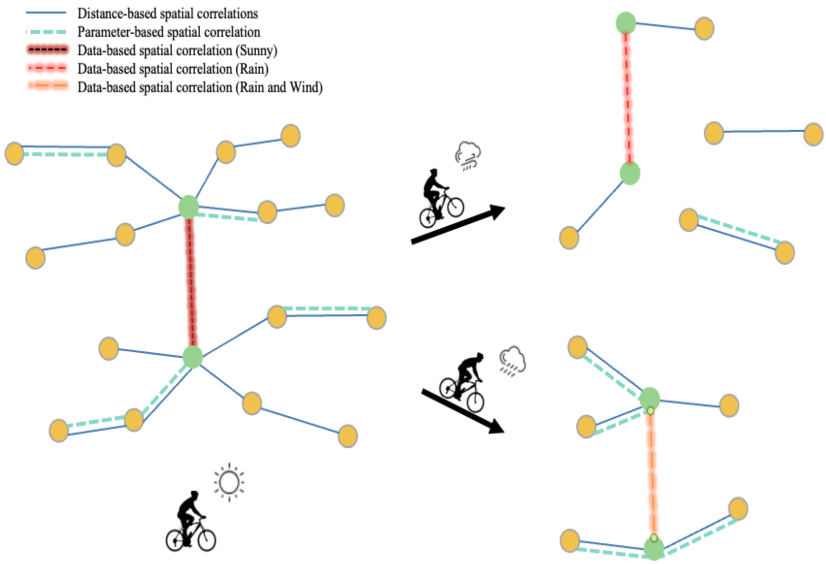


Figure 2.1 Framework of the overall model

To investigate the spatial correlations of active mode traffic based on traffic flow features and how variations in weather information alter these correlations, we first propose employing an attention mechanism, as shown in Equation (2.4), to capture the dependencies of traffic flow features and weather information for each sensor.

$$\mathbf{M}_{at,i} = attention(\mathbf{M}_{tq,i}, \mathbf{M}_{tk,i}, \mathbf{M}_{tv,i}) \tag{2.4}$$

where $\mathbf{M}_{t,i}$ is the traffic flow input sequence and the corresponding weather data for prediction at time t, $\mathbf{M}_{t,i}$ is projected onto separate learned linear subspaces to obtain the queries, keys, and values $\mathbf{M}_{tq,i}$, $\mathbf{M}_{tk,i}$, and $\mathbf{M}_{tv,i}$ of attention mechanism, respectively. $\mathbf{M}_{at,i}$ represents information dependencies of the traffic flow feature and weather information that are most relevant for capturing spatial correlations. In this paper, $attention(\cdot)$ function is a Scaled Dot-Product Attention [46] as follow,

$$attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_m}})\mathbf{V} \tag{2.5}$$

where $\mathbf{Q}$, $\mathbf{K}$ and $\mathbf{V}$ are the query, key and value, respectively; d_m is the scaling factor, which is used to balance the complexity and capacity of the model.

After obtaining the dependency representations of each sensor, we derive the spatial pattern similarity between the sensors by calculating the dot product [17] of the dependency representations between sensors, as follows:

$$\mathbf{A}_{t,attri} = ReLU(\mathbf{M}_{at}\mathbf{M}_{at}^{T}) \tag{2.6}$$

where $\mathbf{M}_{at}$ is the matrix of the representations of all sensors with utmost pertinent details of traffic data for the spatial correlations extraction at time t; $\mathbf{M}_{at}^{T}$ is the transpose of $\mathbf{M}_{at}$; $\mathbf{A}_{t,attri}$ is the dynamic attention data-based adjacency matrix at time t.

(4) Heterogeneous graph spatial correlations fusion approach

The representation of spatial correlations in a graph can be complex and multifaceted. Not only does a physical relationship, such as a distance-based predefined graph, play a role in describing the graph, but also a semantic or contextual relationship, such as a self-adaptive graph or a dynamic attention data-based graph, is essential in capturing the graph's relationships between nodes. These diverse types of graph representations can provide complementary information to the relationships between nodes. However, balancing the utilization of these graphs with different types of edge directly in a graph architecture can be challenging. To address this challenge, we have designed a fusion approach to derive an optimal graph for active mode spatial correlations representations.

Specifically, we first concatenate all matrices along the feature dimension as shown in Equation (2.7).

$$\mathbf{A}_{t,all} = concat(\mathbf{A}_{d}, \mathbf{A}_{a}, \mathbf{A}_{t,attri}) \tag{2.7}$$

To get a fully represented graph structure based on the heterogeneous graph spatial relationship, we apply two-layer 1D convolutions to capture optimal graph correlations of active mode.

2.3.3 Short-term active mode traffic prediction

Active mode traffic is flexible and subject to change based on personal choices, motives and trip purposes, and preferences. To make an accurate traffic prediction for active mode, we develop a spatial-temporal traffic prediction module based on the Transformer encoder-decoder structure as shown in **Fig. 2.2**, which allows for the capture of complex relationships of spatial and temporal characteristics, enabling the model to effectively learn and predict dynamic changes in active mode traffic. Specifically, we adopt a "sandwich structure" [47] in both the encoder and decoder. This structure comprises two multi-head self-attention layers to capture temporal information of the traffic flow, with a spatial information capturing module in between to capture spatial information. To facilitate deep training, we integrate residual connections and layer normalization between each layer. The comprehensive design incorporates one encoder layers and one decoder layers, providing the model with the capability to proficiently acquire and anticipate spatial-temporal traffic patterns.

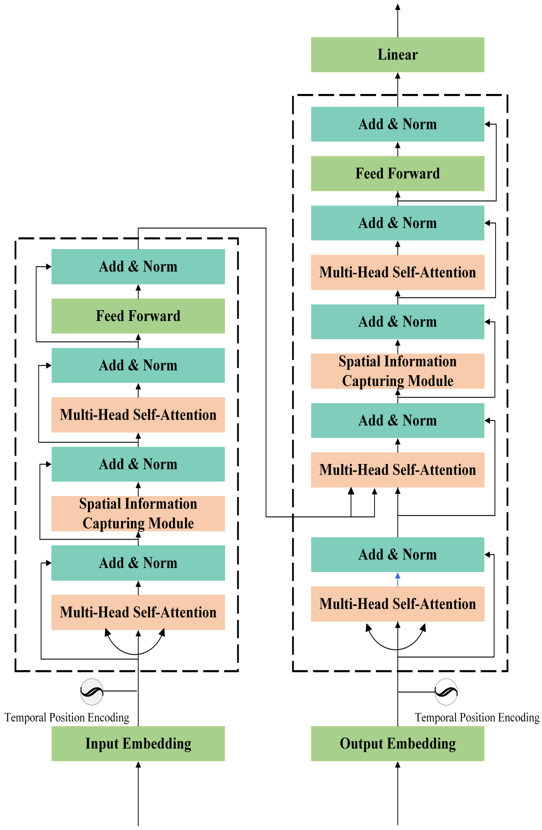


Figure 2.2 Active mode traffic prediction module

(1) Temporal dependency learning

Given the unpredictable nature of traffic patterns in active modes, extracting crucial dependencies from historical traffic flow data becomes essential to improve prediction accuracy. However, handling long sequences of data can potentially lead to the loss of vital information during the learning process. To address this challenge, we propose integrating the multi-head self-attention module into our model. This module, derived from the Transformer architecture [46], allows the model to selectively focus on different segments of the input sequence. This capability ensures that temporal dependencies are effectively captured, enabling the model to retain and utilize significant information for more accurate traffic predictions.

The multi-head self-attention module, while highly effective in capturing global dependencies between input elements, is not explicitly designed to model the order of the input sequence in the same way as recurrent or convolutional neural networks. In applications such as active mode traffic prediction, the temporal order of the input sequence plays a critical role in capturing important information for accurate prediction. For instance, weather and traffic conditions at a particular time are likely to be more similar to those at subsequent time steps. To account for this, it is necessary to place more weight on information that is closer to the prediction time, especially when dealing with active mode traffic with uncertain patterns. By emphasizing the importance of recent data, models can more effectively capture the temporal dependencies between input elements and improve their predictive performance. Therefore, we combine the fixed positional encodings [46] with the input embeddings so that multi-head self-attention module could still make use of temporal order information of traffic data. The equations are shown below,

$$P_{(pt,2dim)} = \sin(pt/10000^{2dim/d_m}) \tag{2.8}$$

$$P_{(pt,2dim+1)} = \cos(pt/10000^{2dim/d_m}) \tag{2.9}$$

where pt is the position index of the traffic flow input sequence; dim is dim-th dimension of the positional encoding vector; d_m is the dimension of the positional encoding, which is the same as the dimension of input embeddings, so that they could be added element-wise.

(2) Spatial dependency learning

To explore the geographic and hidden spatial correlations of the active mode graph network, the spatial Graph Convolutional Network (GCN) [48] can be used. The spatial-based GCN is a type of neural network that can operate on non-Euclidean graph-structured data, allowing it to learn feature representations by aggregating the features of a node's neighbors and incorporating both local and global information in the graph.

In the context of active mode traffic flow prediction, the graph represents the spatial correlations between different sensors, and each node represents a specific sensor. The spatial-based GCN operates by conducting successive graph convolutions on the input traffic flow features using a dynamic adjacency matrix $\mathbf{A}_{t,f}$. This matrix updates the feature representations of each node by incorporating the traffic flow features of its neighbors. In our model, the input adjacency matrix comprises the output of a heterogeneous graph spatial correlations fusion approach. This approach considers the spatial correlations influenced by weather and integrates heterogeneous features from various types of spatial correlations. The output of the Spatial-based GCN is a set of new traffic flow feature representations for each node. Specifically, this spatial-based GCN is able to handle large graphs by expanding the adjacency matrix to incorporate additional connections, without fundamentally changing the architecture.

$$GCN(\mathbf{X}_t) = ReLU(\hat{\mathbf{D}}^{-1}\hat{\mathbf{A}}_t\mathbf{X}_t\mathbf{W}) \tag{2.10}$$

where $\hat{\mathbf{A}}_t = \mathbf{A}_{t,f}$; $\hat{\mathbf{D}}_{ii} = \sum_j \hat{\mathbf{A}}_{ij}$, $\mathbf{W}$ is the trainable weight matrix. $ReLU(\cdot)$ is an activation function.

(3) Residual connection and feedforward networks

To mitigate the vanishing gradient problem and enhance model performance, we incorporate residual connections between each layer, as shown in the Equation (2.11). Additionally, a fully connected feedforward network [46] is introduced to each encoder and decoder layer, enabling the model to learn complex input-output relationships and introducing nonlinearity as depicted in the Equation (2.12).

$$\mathbf{X}^{l+1} = layer(\mathbf{X}^l) + \mathbf{X}^l \tag{2.11}$$

$$FeedForward(\mathbf{X}) = ReLU(\mathbf{X}\mathbf{W}_0 + b_0)\mathbf{W}_1 + b_1 \tag{2.12}$$

where $layer(\mathbf{X}^l)$ is the output of layer l, $\mathbf{X}^l$ is the input of layer l, $\mathbf{X}^{l+1}$ is the output after residual connection. $layer(\cdot)$ is the temporal or spatial information capturing operation function.

(4) Multi-step traffic prediction

In the reference section of this model, we will employ an autoregressive inference to forecast traffic conditions over multiple future time steps. This involves using the previously generated prediction result as the input to predict the next time step. This approach provides a more comprehensive and extensive view of anticipated traffic conditions, offering valuable insights for various applications in transportation planning, management, and decision making.

2.4 Experimental Results and Benchmarking

In this section, we present the result of applying the proposed approach. Before presenting the results, we will first provide an overview of the experimental set-up, including a description of the data, the metrics used for performance assessment, and the baseline models used in this paper.

2.4.1 Dataset description

To evaluate the effectiveness of our proposed model, we conducted experiments using real-world bicycle traffic flow data from the National Road Traffic Data Portal of the Netherlands. This dataset, collected by loop detector sensors in Rotterdam,

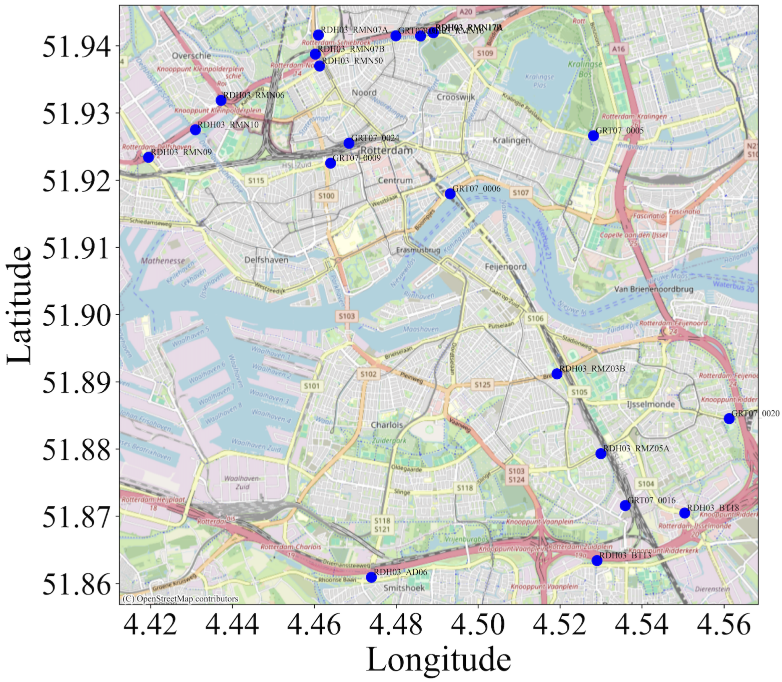


Figure 2.3 Location of sensors in Rotterdam (2022).

spans from January 1, 2022, to December 31, 2022, and includes data from 21 sensors. As shown in **Fig. 2.3**, the sensors are sparsely distributed. Missing data were imputed using the average data from the corresponding time spots over the previous two days. The dataset was aggregated into 5-minute intervals, resulting in 12 data points per hour. Standard normalization was applied to the input dataset to enhance model convergence and stability during training. Our objective was to perform multi-step active mode traffic predictions, using one-hour historical data (12 data points) to predict traffic for the next 60, 45, and 15 minutes. To maintain evaluation integrity, the dataset was divided chronologically into training, validation, and test sets with a split ratio of 6:2:2.

The weather feature data from the Royal Netherlands Meteorological Institute (KNMI) at the Rotterdam Station was utilized. This dataset includes hourly observations of precipitation amounts and mean wind speeds. Precipitation data reflects the millimeters of rain recorded during the preceding hour. The mean wind speed (in 0.1 m/s) is recorded for the hour preceding the observation timestamp. Since the weather data is provided at hourly intervals, each data point was divided into 12 intervals. Specifically, the mean wind speed data was duplicated for each interval, while the precipitation data was evenly distributed across the intervals to represent the average millimeters of rain over 5 minutes. The timestamps of these weather data intervals were then aligned with or lagged by 5 to 30 minutes from the traffic flow data timestamps to be incorporated into the model.

2.4.2 Evaluation metrics

In this paper, the prediction results of active mode traffic flow are evaluated by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Weighted Absolute Percentage Error (WAPE). The formulations to calculate these metrics are shown below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2.13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2.14}$$

$$WAPE = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\sum_{i=1}^{n}|y_i|} \tag{2.15}$$

where $\hat{y_i}$ and y_i represent the predicted value and ground truth data of the value i. n is the number of sample values.

2.4.3 Baseline models

To evaluate the performance of our proposed DyASTGCN model, we compare it with several traditional traffic prediction models and state-of-the-art spatial-temporal traffic prediction models. The baseline models used for comparison are as follows.

- HA: The Historical Average (HA) model predicts traffic flow by taking the average value of historical data.
- SVR: Support Vector Regression (SVR) [49] is a regression algorithm that aims to minimize the discrepancy between the predicted value and a predefined margin, which is an extension of Support Vector Machines (SVM) for regression tasks.
- LSTM: Long Short-Term Memory (LSTM) [50] is a specialized variant of Recurrent Neural Networks (RNNs) specifically designed to effectively capture and model long-term dependencies in sequential data.
- STGCN: Spatio-Temporal Graph Convolutional Network (STGCN) [47] is proposed to tackle the time series traffic prediction problem by harnessing

comprehensive spatial-temporal correlations.

- ASTGCN: Attention-based Spatial-Temporal Graph Convolutional Networks(ASTGCN) [43] is designed for high nonlinearity and complex traffic flow prediction by integrating the spatial-temporal attention mechanism with spatial-temporal convolution.

- Ada-STNet: Adaptive Spatio-Temporal Graph Neural Network(Ada-STNet) [17] developed a graph structure learning component and a dedicated spatial-temporal convolution architecture to capture spatial relationships and temporal dependencies of traffic data.

- STMFGNN: Spatial-Temporal Multifactor Fusion Graph Neural Network (STMFGNN) [51] leverages dynamic similarity and static adjacency graphs for parallel graph convolution, integrating global hidden and local prior knowledge. A gated fusion module adaptively learns dynamic influence weights to capture multiscale spatial dependencies. The model employs gated tanh unit convolution, multireceptive fields, and gated recurrent units for temporal feature extraction, enabling comprehensive traffic flow prediction by considering multiscale factors.

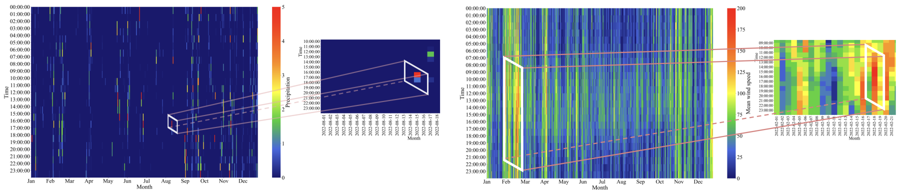
2.4.4 Experimental settings

The experiments are conducted using Google Colab, a cloud-based Python environment. The computing environment included a Tesla L4 GPU with a CUDA version of 12.0. The CPU used was an Intel(R) Core(TM) i9-9900KS clocked at 4 GHz. We implemented all the deep learning models using the PyTorch framework in Python. The models were optimized using the Adam optimizer, with nn.L1Loss() employed as the loss function in PyTorch. This loss function calculates the Mean Absolute Error (MAE) by measuring the average absolute difference between the predicted and actual target values. The hyperparameters for all deep learning models were carefully tuned through a validation set. Specifically, for the DyASTGCN model, we set the model dimension d_{model} to 64, the number of attention heads h to 8, the convolution kernel size to 3, the learning rate is gradually decreased from 0.001 to 0.00001, and batch size is 64.

2.4.5 Experimental results and discussion

(1) Active mode spatial correlations with weather influence

Weather conditions, such as precipitation and wind speed, significantly impact cyclists' route choices, leading to changes in traffic patterns and altering the spatial flow dynamics within the network. During extreme weather conditions, cyclists may avoid certain paths in favor of routes that offer more shelter or better drainage.



(a) (b)

Figure 2.4 Overview of weather in Rotterdam (2022). (a) Precipitation. (b) Mean wind speed.

Others might opt for alternative modes of transportation, resulting in notable changes in traffic flow compared to the same time periods under normal weather conditions.

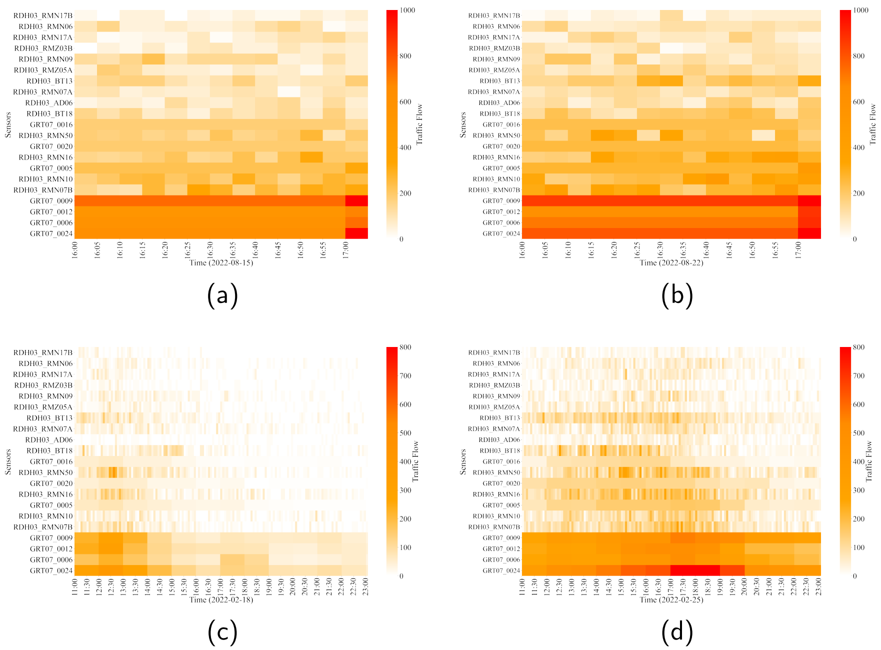


Figure 2.5 Traffic flow. (a) Heavy precipitation. (b) No precipitation. (c) Strong wind. (d) Calm wind.

As illustrated in **Fig. 2.4**, there was heavy precipitation from 16:00 to 17:00 on August 15, 2022, and strong wind from 11:00 to 23:00 on February 18, 2022. The traffic flow patterns during these periods, shown in **Fig. 2.5** (a) and (c) respectively, demonstrate significant differences when compared to traffic flow patterns during the same time periods on similar weekdays without extreme weather, as depicted in **Fig. 2.5** (b) and (d). The comparison clearly indicates a decrease in traffic flow during extreme weather conditions.

With the change in traffic flow influenced by weather, the spatial correlations within the traffic network might also shift. However, cyclists' responses to weather changes might not be instantaneous. For instance, if it starts raining, cyclists already on the road might continue their journey until they find suitable shelter, while others might switch to alternative modes of transportation in extreme weather. This lag in response causes temporal shifts in traffic flow, thereby affecting the spatial correlations within the network.

To investigate how weather influences spatial correlations in traffic flow data, we introduce various lags in the weather data to capture the dynamic data-based graph spatial relationships. This approach aligns earlier weather conditions with later traffic flow observations, allowing us to assess if this improves the accuracy of the proposed DyASTGCN model. We conduct three types of experiments to illustrate the lagged effect of different weather factors:

- Precipitation data only: This experiment includes only precipitation data along with traffic flow data to capture dynamic data-based graph spatial correlations for traffic prediction.
- Wind speed data only: This experiment includes only wind speed data along with traffic flow data to capture dynamic data-based graph spatial correlations for traffic prediction.
- Combined precipitation and wind speed data: This experiment includes both precipitation and wind speed data along with traffic flow data to capture dynamic data-based graph spatial correlations for traffic prediction.

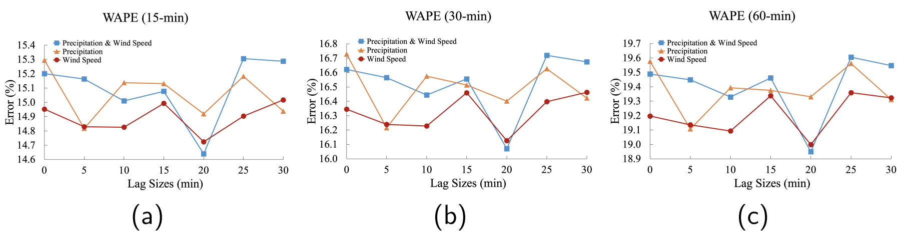


Figure 2.6 Performance of DyASTGNN with weather data. (a) WAPE (15-min). (b) WAPE (30-min). (c) WAPE (60-min).

As illustrated in **Fig. 2.6**, our study demonstrates that the DyASTGCN model achieves optimal performance with distinct lag times for precipitation and wind speed data relative to the active mode traffic flow data. Specifically, the model shows the highest accuracy with a 5-minute lag for precipitation data, a 20-minute lag for wind speed data, and a 20-minute lag for the combined precipitation and wind speed data. In particular, the 20-minute lag for the combined precipitation and wind speed data perform the best in all the scenarios.

In the Netherlands, particularly in cities like Rotterdam, cycling is a popular mode of transportation, which is sensible to the weather change. Rain prompts immediate adjustments in cycling routes and behaviors. This rapid response is reflected in the model's optimal performance with a short 5-minute lag for precipitation data. Cyclists often alter their routes or temporarily wait for rain to subside, leading to an immediate but brief impact on traffic flow patterns. In contrast, wind speed feature has a more gradual impact on cycling behavior. Cyclists may take longer to adjust to windy conditions, resulting in slower speeds or route changes that develop over time. The model's optimal performance with a 20-minute lag for wind speed data reflects this extended adjustment period. Wind affects the ease and safety of cycling, causing sustained changes in cycling behavior, which are captured by the longer lag period. When combining precipitation and wind speed data, the model shows the best performance with a 20-minute lag. This finding indicates that while rain causes immediate changes, the dominant and prolonged impact of wind requires a longer lag period to accurately capture its effect on traffic flow. The interaction between rain and wind creates complex conditions for cyclists, who might wait until conditions improve or adjust more slowly to persistent windy conditions. The 20-minute lag effectively captures the overall impact, particularly the dominant influence of wind, on traffic patterns.

Table 2.1 PERFORMANCE COMPARISON OF MULTI-STEP TRAFFIC PREDICTION

Baselines	**60 min**			**30 min**			**15 min**		
	MAE	RMSE	WAPE	MAE	RMSE	WAPE	MAE	RMSE	WAPE
HA	84.80	143.51	71.02%	84.80	143.51	71.02%	84.80	143.50	71.02%
SVR	40.23	82.36	38.03%	31.47	66.82	29.75%	26.62	58.32	25.16%
LSTM	29.08	65.81	27.49%	22.91	53.72	21.65%	19.34	45.07	18.29%
STGCN	24.82	52.63	23.47%	22.12	46.83	20.91%	19.99	42.24	18.90%
ASTGCN	25.03	53.44	23.67%	21.38	47.46	20.21%	18.79	46.02	17.78%
Ada-STNet	20.12	40.57	19.02%	17.64	37.74	16.68%	16.35	37.05	15.45%
STMFGNN	20.97	38.31	19.83%	19.04	34.82	18.00%	17.96	32.96	16.97%
DyASTGCN	**19.67**	**39.37**	**18.59%**	**16.82**	**35.01**	**15.90%**	**15.42**	**32.75**	**14.58%**

(2) Overall active mode traffic prediction comparison

Table 2.1 displays the performance of baseline methods and proposed model, presenting the overall average errors of 15 minutes, 30 minutes, and 60 minutes prediction horizons. Our DyASTGCN model demonstrates superior performance compared to all baseline methods on the dataset of active mode in terms of MAE, RMSE and WAPE, respectively.

Specifically, among traditional statistical methods in the context of time series forecasting, The Historical Average (HA) exhibits the poorest performance when compared to alternative baseline methods. This can be attributed to its reliance only on past data. SVR and LSTM possess the capability to encompass both linear and nonlinear patterns, making them more adept at capturing intricate details within active mode data. However, only considering temporal features of active mode data results in their performance deterioration compared to methods based on graph neural networks. To gain a deeper understanding of traffic patterns, STGCN and ASTGCN leverage graph convolution neural networks to consider spatial interconnections between traffic components, thus amplifying the precision of predictions. In particular, ASTGCN introduces an attention mechanism into the STGCN model, fostering the capture of dynamic spatial and temporal patterns through an encoder and decoder structure. This augmentation yields superior outcomes on the 60 minutes prediction horizon of active mode traffic prediction. Ada-STNet not only considers temporal dependencies but also incorporates the spatial characteristics of road networks through graph convolution and causal convolution. Additionally, Ada-STNet takes node attributes into account to create dynamic and self-adaptive graph structures, STMFGNN parallelly utilizes dynamic similarity graphs and static adjacency graphs to capture the multiscale spatial dependencies between nodes, surpassing the performance of ASTGCN.

Ada-STNet extracts the node attributes based on a convolution operation, which is well suited to capture relevant features of traffic data. However, in the context of active mode traffic prediction, active mode data might exhibit varying patterns over an hour due to factors such as rush hours, events, or weather changes. Attention mechanism is more adaptable to capturing the attributes of the active mode data and changing relationships between nodes. Therefore, we utilize the attention mechanism to capture the data attribute for graph structure generation over time and capturing temporal and spatial dependencies with graph convolutional neural networks and multi-head self-attention. Furthermore, Ada-STNet fused graph structure with different spatial relationships by summing up all the adjacency matrices, which did not consider the heterogeneity of different adjacency matrices. By contrast, we design a fusion approach to aggregate the matrices properly for active mode traffic prediction. The performance of the prediction model DyASTGCN surpasses Ada-STNet. This outcome highlights the capability of DyASTGCN to effectively capture the variations by considering the influence of weather factors on graph spatial correlations and produce an appropriate adjacency matrix that accurately represents the intricate relationships within the active mode network.

(3) The influence of spatial correlations

Bicycle traffic flow often exhibit greater fluctuations compared to car traffic flows due to bicycles being highly sensitive to immediate environmental and situational changes. For instance, cyclists are more affected by weather conditions such as strong winds or rain, which can drastically alter their riding behavior and route choices. Local events or community activities often attract large numbers of cyclists, creating sudden spikes in traffic flow that are less predictable than the more stable

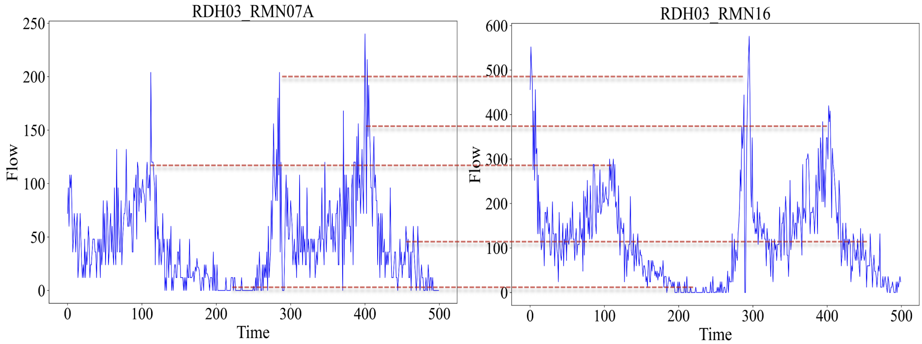


Figure 2.7 Bicycle traffic flow distribution

patterns observed in car traffic. Furthermore, cyclists frequently make spontaneous decisions to deviate from their routes, such as stopping at local shops or changing paths to avoid congested areas, contributing to the irregular and fluctuating nature of bicycle traffic. As shown in **Fig. 2.7**, the traffic flow at sensor RDH03_RMN16 fluctuates dramatically, with the traffic flow of the previous five minutes differing significantly from the next five minutes. In this case, the local historical traffic flow of this sensor might not be sufficient for accurately predicting future traffic flow, as sudden changes in traffic patterns can occur. However, we observe that at the same time slot, some sensors exhibit similar pattern changes as point out in **Fig. 2.7**, which could be explained by the interconnected nature of the network where patterns among nodes influence each other.

To explore and capture these spatial correlations for accurate traffic prediction, we introduced various types of spatial correlations: distance-based predefined graph spatial correlations, parameter-based adaptive graph spatial correlations, dynamic attention data-based graph spatial correlations considering weather influences, and optimal fusion-based graph spatial correlations that combine all the above using the proposed fusion approach. The prediction results are shown in **Fig. 2.8**. **Fig. 2.8**(a) presents the prediction results of the DyASTGCN model with optimal fusion-based graph spatial correlations. Although this model does not perfectly predict the ground-truth values, it successfully captures spikes in traffic flow to some extent. In comparison, the other models with distance-based predefined graph spatial correlations shown in **Fig. 2.8** (b), parameter-based adaptive graph spatial correlations shown in **Fig. 2.8** (c), and dynamic attention data-based graph spatial correlations shown in **Fig. 2.8** (d), demonstrate that the DyASTGCN model with optimal fusion-based graph spatial correlations is more sensitive to fluctuations in bicycle traffic flow.

Distance-based predefined graph spatial correlations can represent geographical spatial correlations based on the road network; however, two sensors close to each other could exhibit dramatically different traffic flows at the same time slot due to

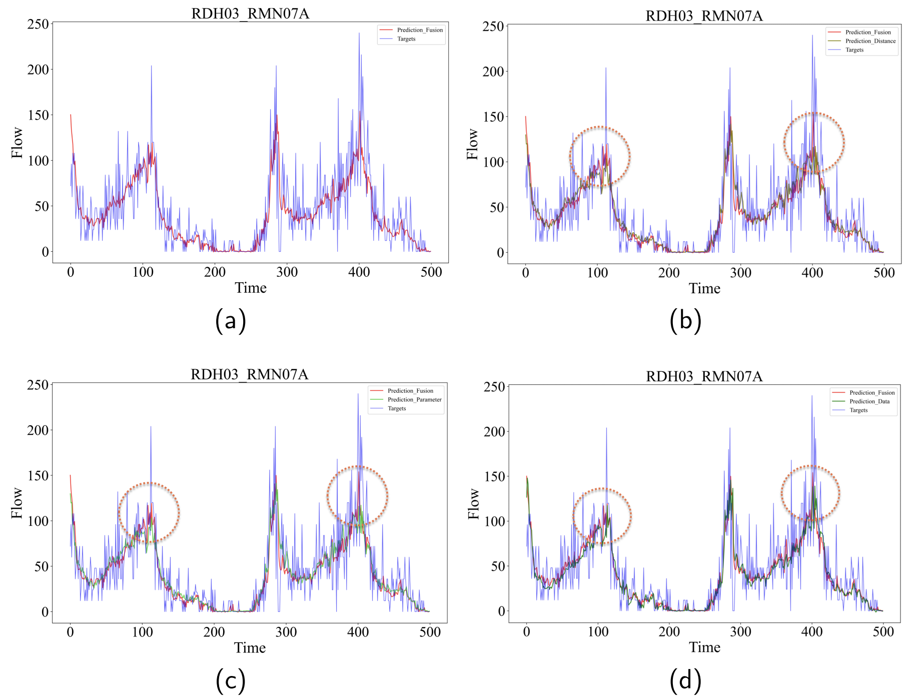


Figure 2.8 Bicycle traffic spatial correlations. (a) Optimal fusion-based graph spatial correlations. (b) Distance-based predefined graph spatial correlations. (c) Parameter-based adaptive graph spatial correlations. (d) Dynamic attention data-based graph spatial correlations.

differences in road levels or access destinations. Parameter-based adaptive graph spatial correlations capture some hidden spatial correlations among sensors, but these are quite limited. While dynamic attention data-based graph spatial correlations consider traffic flow and weather conditions, they overlook the sensor connections in the actual road network. Therefore, by considering all factors, optimal fusion-based graph spatial correlations more comprehensively represent the spatial correlations among sensors, enabling the capture of fluctuating traffic flow patterns of bicycles.

2.4.6 Ablation study

In order to gain deeper insights into the impact of various components in DyASTGCN, we performed ablation experiments and analyzed the resulting outcomes using the same active mode data as mentioned earlier.

(1) Ablation study of proposed DyASTGCN

To evaluate the impact of integrating weather data on capturing dynamic attention-based graph spatial correlations in active mode traffic, two experiments were conducted. Our approach involved training a model both with and without the incorporation of weather data while maintaining consistent experimental settings. The primary objective was to assess how effectively our model captured dynamic spatial correlations enhanced by weather information. **Fig. 2.9** demonstrate that the model incorporating dynamic attention-based graph spatial correlations with weather data outperforms the model that does not consider weather factors. This finding underscores the profound influence of weather conditions on the complex dynamics of active mode traffic networks.

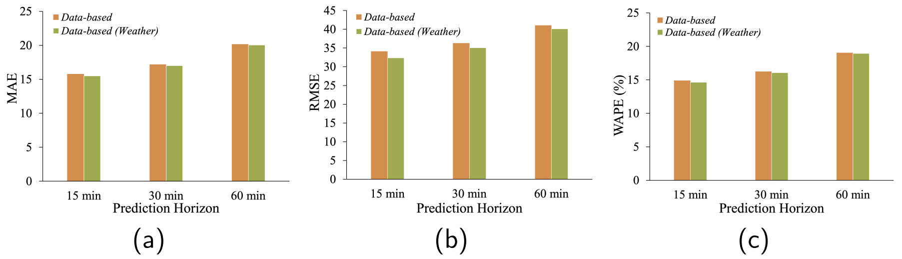


Figure 2.9 Graph spatial correlation ablation study. (a) MAE. (b) RMSE. (c) WAPE.

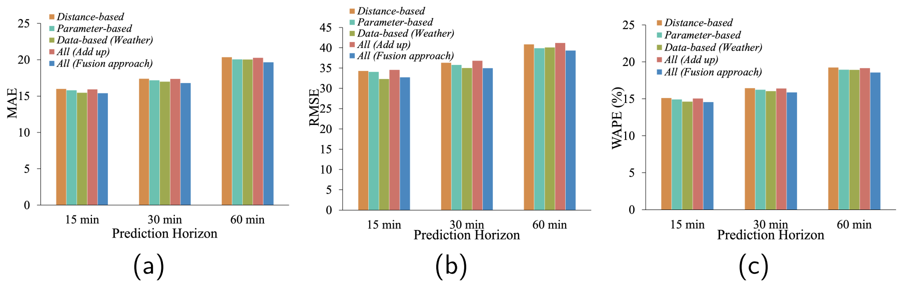


Figure 2.10 Fusion approach ablation study. (a) MAE. (b) RMSE. (c) WAPE.

Furthermore, we conducted experiments to evaluate the performance of our proposed heterogeneous graph spatial correlations fusion approach, aimed at combining various types of graph spatial correlations to optimize predictions for active mode traffic networks. As depicted in **Fig. 2.10**. we compared our fusion approach against several scenarios:

- Distance-based: This model exclusively incorporates predefined graph spatial correlations based on distances.
- Parameter-based: This model solely utilizes adaptive graph spatial correlations derived from parameters.

- Data-based: This model relies solely on dynamic attention data-based graph spatial correlations cooperate with 20-minute lag weather information.

- All (Add up): This model integrates all three types of graph spatial correlations by simply summing them.

- All (Fusion approach): Our proposed approach, where the heterogeneous graph spatial correlations are fused to derive optimal spatial correlations for active mode traffic.

The ablation study revealed that the model employing our heterogeneous graph spatial correlations fusion approach consistently outperformed models that do not integrate various spatial correlations or do not use a systematic approach to harness the utility of diverse spatial correlations. This underscores the critical importance of effectively integrating and leveraging information embedded within diverse spatial correlations for accurate predictions of active mode traffic flow.

(2) Hyperparameter analysis

Hyperparameter tuning is an essential step in optimizing the performance of DyASTGCN. As detailed in the Experimental Settings section, all hyperparameters for the deep learning models were carefully calibrated using a validation set. While many hyperparameters influence model performance, this section highlights batch size and model dimension as illustrative examples. The impact of these hyperparameters on the WAPE error of DyASTGCN is shown in **Fig. 2.11**, demonstrating that the model achieves optimal performance when the batch size and model dimension are both set to 64.

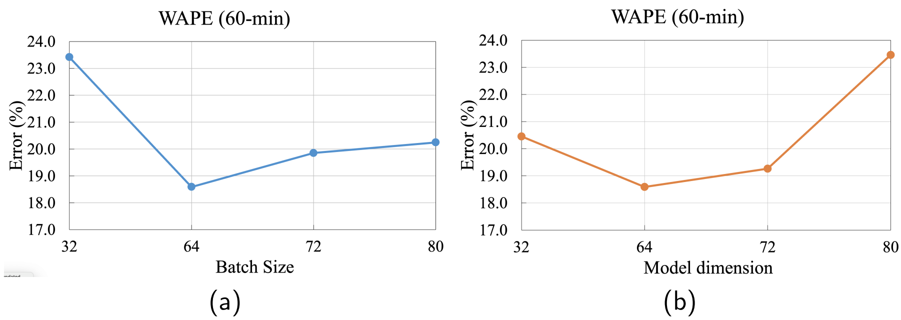


Figure 2.11 Hyperparameter analysis. (a) Batch size. (b) Model dimension.

2.5 Conclusion

In this study, given the sensitivity of active mode traffic to weather conditions and the complexity of heterogeneous spatial correlations within active mode graph networks, we introduce a Dynamic Attention-based Spatial-Temporal Graph Convolutional Network model (DyASTGCN) for predicting active mode traffic flow. Our model incorporates the influence of weather on traffic graph spatial correlations and proposes a fusion approach to derive optimal spatial correlations that accurately represent active mode traffic dynamics. Experimental results highlight a lag effect of weather on active mode traffic spatial correlations. Specifically, precipitation exhibits a 5-minute lag relative to active mode flow, while mean wind speed shows a 20-minute lag. Including both precipitation and mean wind speed with a 20-minute lag relative to active mode flow yields the best performance compared to above individual scenarios. Our proposed heterogeneous graph spatial correlations fusion approach demonstrates that effectively integrating diverse spatial correlations leads to optimal spatial representations for precise prediction of active mode traffic. This approach ensures that the model captures and utilizes the nuanced interactions between weather factors and traffic dynamics, thereby enhancing prediction accuracy.

Overall, this research underscores the importance of considering weather impacts and leveraging heterogeneous spatial correlations to advance the understanding and prediction of active mode traffic behavior in urban environments. In future research, our aim is to evaluate the model's performance across diverse datasets collected from expansive regions where active mode transportation is prevalent, offering a thorough evaluation of the robustness of the proposed DyASTGCN. Additionally, we plan to explore the influence of various external variables such as individual preferences, road conditions, and other factors on the temporal dynamics of spatial correlations within the active mode traffic. This investigation will broaden our understanding of how these factors interact with and influence the predictive capabilities of our models, thereby enhancing their applicability and robustness in real-world scenarios.

Acknowledgements

This research is supported by the China Scholarship Council. This support is gratefully acknowledged.

References

[1] N. Jia, L. Li, S. Ling, S. Ma, and W. Yao. “Influence of attitudinal and low-carbon factors on behavioral intention of commuting mode choice–A cross-city study in China”. In: *Transportation research part A: policy and practice* 111 (2018), pp. 108–118.

[2] D. Ton, D. C. Duives, O. Cats, S. Hoogendoorn-Lanser, and S. P. Hoogendoorn. "Cycling or walking? Determinants of mode choice in the Netherlands". In: *Transportation research part A: policy and practice* 123 (2019), pp. 7–23.

[3] M. Meng, J. Zhang, Y. D. Wong, and P. H. Au. "Effect of weather conditions and weather forecast on cycling travel behavior in Singapore". In: *International journal of sustainable transportation* 10.9 (2016), pp. 773–780.

[4] H. Xu, J. Ying, H. Wu, and F. Lin. "Public bicycle traffic flow prediction based on a hybrid model". In: *Applied Mathematics & Information Sciences* 7.2 (2013), pp. 667–674.

[5] S. Sohrabi and A. Ermagun. "Dynamic bike sharing traffic prediction using spatiotemporal pattern detection". In: *Transportation research part D: transport and environment* 90 (2021), p. 102647.

[6] A. Biehl, A. Ermagun, and A. Stathopoulos. "Community mobility MAUP-ing: A socio-spatial investigation of bikeshare demand in Chicago". In: *Journal of Transport Geography* 66 (2018), pp. 80–90.

[7] G. Xiao, R. Wang, C. Zhang, and A. Ni. "Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks". In: *Multimedia Tools and Applications* 80 (2021), pp. 22907–22925.

[8] P.-C. Chang, J.-L. Wu, Y. Xu, M. Zhang, and X.-Y. Lu. "Bike sharing demand prediction using artificial immune system and artificial neural network". In: *Soft Computing* 23 (2019), pp. 613–626.

[9] Y. Li and Y. Zheng. "Citywide bike usage prediction in a bike-sharing system". In: *IEEE Transactions on Knowledge and Data Engineering* 32.6 (2019), pp. 1079–1091.

[10] P. Hulot, D. Aloise, and S. D. Jena. "Towards station-level demand prediction for effective rebalancing in bike-sharing systems". In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2018, pp. 378–386.

[11] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang. "Traffic flow prediction with big data: A deep learning approach". In: *Ieee transactions on intelligent transportation systems* 16.2 (2014), pp. 865–873.

[12] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang. "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data". In: *Transportation Research Part C: Emerging Technologies* 54 (2015), pp. 187–197.

[13] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li. "Deep multi-view spatial-temporal network for taxi demand prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. 1. 2018.

[14] S. R. Chandra and H. Al-Deek. "Predictions of freeway traffic speeds and volumes using vector autoregressive models". In: *Journal of Intelligent Transportation Systems* 13.2 (2009), pp. 53–72.

[15] H. Yao, X. Tang, H. Wei, G. Zheng, Y. Yu, and Z. Li. "Modeling spatial-temporal dynamics for traffic prediction". In: *arXiv preprint arXiv:1803.01254* 1.9 (2018).

[16] X. Shi, H. Qi, Y. Shen, G. Wu, and B. Yin. "A spatial–temporal attention approach for traffic prediction". In: *IEEE Transactions on Intelligent Transportation Systems* 22.8 (2020), pp. 4909–4918.

[17] X. Ta, Z. Liu, X. Hu, L. Yu, L. Sun, and B. Du. "Adaptive spatio-temporal graph neural network for traffic forecasting". In: *Knowledge-based systems* 242 (2022), p. 108199.

[18] W. Zhang, F. Zhu, Y. Lv, C. Tan, W. Liu, X. Zhang, and F.-Y. Wang. "AdapGL: An adaptive graph learning algorithm for traffic prediction based on spatiotemporal neural networks". In: *Transportation Research Part C: Emerging Technologies* 139 (2022), p. 103659.

[19] W. Jiang and J. Luo. "Graph neural network for traffic forecasting: A survey". In: *Expert systems with applications* 207 (2022), p. 117921.

[20] X. Kong, K. Wang, M. Hou, F. Xia, G. Karmakar, and J. Li. "Exploring human mobility for multi-pattern passenger prediction: A graph learning framework". In: *IEEE Transactions on Intelligent Transportation Systems* 23.9 (2022), pp. 16148–16160.

[21] X. Kong, Z. Shen, K. Wang, G. Shen, and Y. Fu. "Exploring bus stop mobility pattern: a multi-pattern deep learning prediction framework". In: *IEEE Transactions on Intelligent Transportation Systems* 25.7 (2024), pp. 6604–6616.

[22] M. Shaygan, C. Meese, W. Li, X. G. Zhao, and M. Nejad. "Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities". In: *Transportation research part C: emerging technologies* 145 (2022), p. 103921.

[23] Y. Li, R. Yu, C. Shahabi, and Y. Liu. "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting". In: *arXiv preprint arXiv:1707.01926* (2017).

[24] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He. "Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies". In: *Transportation research part C: emerging technologies* 105 (2019), pp. 297–322.

[25] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang. "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning". In: *Transportmetrica A: Transport Science* 15.2 (2019), pp. 1688–1711.

[26] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain. "Improved deep hybrid networks for urban traffic flow prediction using trajectory data". In: *Ieee Access* 6 (2018), pp. 31820–31827.

[27] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang. “A hybrid deep learning based traffic flow prediction method and its understanding”. In: *Transportation Research Part C: Emerging Technologies* 90 (2018), pp. 166–180.

[28] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, and J. Jakubowicz. “Dynamic cluster-based over-demand prediction in bike sharing systems”. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2016, pp. 841–852.

[29] C. Xu, J. Ji, and P. Liu. “The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets”. In: *Transportation research part C: emerging technologies* 95 (2018), pp. 47–60.

[30] Z. Yang, J. Hu, Y. Shu, P. Cheng, J. Chen, and T. Moscibroda. “Mobility modeling and prediction in bike-sharing systems”. In: *Proceedings of the 14th annual international conference on mobile systems, applications, and services*. 2016, pp. 165–178.

[31] Y. Li, Y. Zheng, H. Zhang, and L. Chen. “Traffic prediction in a bike-sharing system”. In: *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*. 2015, pp. 1–10.

[32] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin. “Deep learning on traffic prediction: Methods, analysis, and future directions”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2021), pp. 4927–4943.

[33] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan. “Attention based spatial-temporal graph convolutional networks for traffic flow forecasting”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 922–929.

[34] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li. “T-GCN: A temporal graph convolutional network for traffic prediction”. In: *IEEE transactions on intelligent transportation systems* 21.9 (2019), pp. 3848–3858.

[35] Z. Li, G. Xiong, Y. Chen, Y. Lv, B. Hu, F. Zhu, and F.-Y. Wang. “A hybrid deep learning approach with GCN and LSTM for traffic flow prediction”. In: *2019 IEEE intelligent transportation systems conference (ITSC)*. IEEE. 2019, pp. 1929–1933.

[36] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang. “Graph wavenet for deep spatial-temporal graph modeling”. In: *arXiv preprint arXiv:1906.00121* (2019).

[37] N. Hu, D. Zhang, K. Xie, W. Liang, and M.-Y. Hsieh. “Graph learning-based spatial-temporal graph convolutional neural networks for traffic forecasting”. In: *Connection Science* 34.1 (2022), pp. 429–448.

[38] X. Zhao, A. Chang, A. D. Sarma, H. Zheng, and B. Y. Zhao. “On the embeddability of random walk distances”. In: *Proceedings of the VLDB Endowment* 6.14 (2013), pp. 1690–1701.

[39] S. Cao, W. Lu, and Q. Xu. “Deep neural networks for learning graph representations”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30. 1. 2016.

[40] L. Liu, W. K. Cheung, X. Li, and L. Liao. “Aligning users across social networks using network embedding.” In: *Ijcai*. Vol. 16. 2016, pp. 1774–80.

[41] S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. S. Huang. “Heterogeneous network embedding via deep architectures”. In: *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 2015, pp. 119–128.

[42] Y. Liang, G. Huang, and Z. Zhao. “Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach”. In: *Transportation research part C: emerging technologies* 140 (2022), p. 103731.

[43] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong. “Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting”. In: *IEEE Transactions on Knowledge and Data Engineering* 34.11 (2021), pp. 5415–5428.

[44] N. A. Ojekudo and N. P. Akpan. “Anapplication of Dijkstra’s Algorithm to shortest route problem”. In: *IOSR Journal of Mathematics (IOSR-JM)* 13.3 (2017), p. 14.

[45] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains”. In: *IEEE signal processing magazine* 30.3 (2013), pp. 83–98.

[46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[47] B. Yu, H. Yin, and Z. Zhu. “Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting”. In: *arXiv preprint arXiv:1709.04875* (2017).

[48] T. N. Kipf and M. Welling. “Semi-supervised classification with graph convolutional networks”. In: *arXiv preprint arXiv:1609.02907* (2016).

[49] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. “Support vector regression machines”. In: *Advances in neural information processing systems* 9 (1996).

[50] A. Graves and A. Graves. “Long short-term memory”. In: *Supervised sequence labelling with recurrent neural networks* (2012), pp. 37–45.

[51] H. Jia, Z. Yu, Y. Chen, and H. Xia. “Spatial-temporal multi-factor fusion graph neural network for traffic prediction”. In: *Applied Intelligence* 54.19 (2024), pp. 9464–9487.

Chapter 3

LLM-based Framework for Bicycle Traffic Prediction

In the previous chapter, we demonstrated that weather conditions exert a lagging effect on the spatial correlations within the active mode traffic network. Building on this insight, this chapter introduces a Bicycle Traffic Prediction Framework Based on Pretrained LLMs (BiSTLLM), designed to deliver accurate and robust bicycle traffic flow predictions. To effectively model the unique characteristics of bicycle traffic time series, we develop a spatial-temporal embedding module that integrates one-dimensional convolution (1DConv), Graph Convolutional Networks (GCNs), and a domain-specific external feature processing module. This module is tailored specifically for bicycle traffic scenarios. Notably, it incorporates geographic semantic embeddings to represent location-specific patterns and introduces a lagging processor to capture the delayed impact of weather conditions. The lagging processor dynamically fuses weather-related embeddings with spatial-temporal traffic flow representations. Comprehensive experiments conducted on real-world bicycle traffic datasets confirm that BiSTLLM significantly outperforms a variety of state-of-the-art baseline models, highlighting its effectiveness in capturing complex spatial-temporal dependencies and external contextual influences in bicycle traffic prediction.

This chapter is submitted to a journal (under review, R1 revision): Wen, X., Duives, D. and Hoogendoorn, S. A Bicycle Traffic Prediction Framework Using Pretrained Large Language Models. The content is the same as the paper.

3.1 Introduction

The precise prediction of bicycle traffic flow is crucial for efficient traffic management and potentially contributes significantly to improving service quality in a wide range of traffic-related applications that impact daily life [1]. The traffic state within a road network is influenced not only by recent historical traffic flow at a specific location but also by the upstream and adjacent segments due to spatial connectivity within the network. Accurate traffic prediction, therefore, requires models capable of capturing these complex spatial and temporal dependencies. In recent years, deep learning models, which are a class of neural networks composed of more than one hidden layer, organized in deep nested network architectures [2–6], have gained prominence in the domain of traffic prediction due to their ability to effectively model intricate traffic patterns across the network.

However, the accuracy of deep learning models is highly dependent on being trained with large volumes of high-quality traffic data [7]. In the case of motorized traffic, such data is abundantly collected by companies and governments [8]. However, for active mode traffic, the availability of large-scale datasets is very limited, while shared-bike systems collect substantial demand data due to their investment returns. One key reason is that private bicycle traffic, unlike motorized traffic, is not typically associated with severe congestion issues. Consequently, it is often deprioritized in terms of resource allocation for data collection. The data sparsity undermines the ability of prediction models to capture network-wide influences, reducing their utility for traffic management and related applications [9]. As a result, cyclists face greater uncertainty in their journeys, particularly during peak hours, as reliable information about traffic conditions and potential delays is unavailable. This limitation is especially pronounced in cycling-dominant countries like the Netherlands [10].

To achieve reliable traffic flow prediction, especially in essential monitoring points lacking sensor coverage or where sensors operate intermittently, pre-trained LLMs offer a transformative solution by leveraging their ability to infer patterns from incomplete data, integrate multimodal contextual signals [11], and adapt to unseen scenarios through few-shot learning [12]. Although LLMs were originally developed for natural language processing tasks, such as text generation [13], translation [14], and semantic reasoning [15], their core architecture (self-attention mechanisms) [16] and pre-training paradigm (transfer learning) [17] have proven equally powerful for time-series data modeling [18].

Recent studies have adapted LLMs for motorized traffic prediction, achieving good performance in tasks such as highway flow prediction [19], and urban demand prediction [20]. Some efforts have also explored the application of LLMs to shared bicycle traffic prediction by incorporating features such as business Points of Interest [21], special event days, and weather conditions [22]. However, these approaches remain underdeveloped for private bicycle traffic. Private bicycle traffic exhibits unique behavioral and environmental sensitivities that existing models often fail to capture. In particular, prior studies tend to overlook critical multimodal factors, such as the lagged effects of weather on bicycle flow [23], where meteorological conditions influence bicycle demand, ridership patterns, route choices, and even operational needs minutes or even hours later, as well as geographic land use properties and the accessibility of nearby bicycle lanes.

To fully support effective traffic prediction and management strategies, this study proposes a Bicycle Traffic Prediction Framework Based on Pre-trained LLMs (BiSTLLM) for accurate bicycle traffic flow prediction, particularly in sensor networks with essential monitoring points that lack coverage or experience intermittent sensor activity. The proposed framework integrates spatial-temporal embeddings with pre-trained LLMs by leveraging its strong representation learning and transfer capabilities to enhance predictive performance in complex and data-sparse environments. In addition to the spatial-temporal representation of bicycle flow, BiSTLLM incorporates weather embeddings that capture the lagged effects of meteorological conditions on cycling behavior, geographic semantic embeddings derived from Points of Interest (POIs) to reflect location-specific activity patterns, and bicycle lane accessibility embeddings that model the semantics of cycling infrastructure and connectivity. These multi-modal contextual features enable the model to generalize effectively across diverse urban settings, even in the presence of limited sensor data. The main contributions of this paper are summarized as follows:

- We propose a Bicycle Traffic Prediction Framework Based on Pre-trained LLMs (BiSTLLM) for accurate bicycle traffic flow prediction. To enable LLMs to effectively interpret and model the dynamics of bicycle traffic, we develop a dedicated spatial-temporal embedding module that integrates one-dimensional convolution (1Dconv), Graph Convolutional Networks (GCNs), and a domain-specific external contextual feature processing module tailored to bicycle traffic. This design allows the model to capture and fuse rich spatial and temporal dependencies, as well as contextual semantics, that are intrinsic to bicycle traffic patterns.

- To address the delayed effects of weather on bicycle traffic, we design a lagging processor that dynamically fuses weather embedding with spatial-temporal bicycle traffic flow embedding. Specifically, we employ a flow-gated mechanism to adaptively scale the influence of weather conditions based on the current traffic flow. Subsequently, a Causal Temporal Convolution (CausalConv) is applied to model the time-lagged impact of weather on traffic dynamics effectively.

- To evaluate the practical applicability of the proposed approach in realistic bicycle traffic scenarios, we conduct experiments using data from multiple cities under both high-data and low-data conditions. Extensive experimental results demonstrate that BiSTLLM consistently outperforms a range of state-of-the-art baseline models.

The remainder of this paper is structured as follows. Section 2 reviews related work on spatial-temporal traffic prediction and the application of pre-trained LLMs to time series prediction. Section 3 formulates the problem addressed in this study. Section 4 details the architecture and components of the proposed BiSTLLM framework. Section 5 describes the experimental setup and presents the empirical results. Finally, Section 6 discusses the broader implications and key contributions of our approach.

3.2 Related Work

3.2.1 Data-driven spatial-temporal traffic prediction

Traditional traffic prediction approaches, such as historical average (HA) and time-series statistical prediction model like AutoRegressive Integrated Moving Average (ARIMA) [24] have shown limited effectiveness in capturing the complex, nonlinear spatial-temporal dependencies inherent in traffic data. To address some of these limitations, machine learning models, including Support Vector Regression (SVR) [25] and Random Forests [26], were introduced. However, these approaches primarily focused on modeling nonlinear temporal patterns while neglecting spatial dependencies. With the advent of deep learning, models such as Long Short-Term Memory (LSTM) networks [27], Gated Recurrent Unit (GRU) [28] have been widely adopted to learn temporal dynamics. Convolutional Neural Networks (CNNs) [29] have been used to model local spatial and temporal correlations through convolutional operations, while Graph Neural Networks (GNNs) [30] have been leveraged to capture spatial relationships. These deep learning models align closely with the intrinsic characteristics of traffic flow, offering significant improvements in the accuracy and robustness of spatial-temporal traffic prediction.

More recent hybrid models integrate deep learning architectures to jointly capture spatial and temporal features, further enhancing the performance of traffic prediction systems. For example, a novel Attention-based Periodic-Temporal neural Network (APTN) is proposed by Shi et al.[31] as an end-to-end solution for traffic prediction, designed to capture spatial, short-term, and long-term periodic dependencies. Several studies have further incorporated external factors such as weather, events, and road conditions to enrich spatial-temporal representations. For instance, Qi et al. [32] proposed a spatial-temporal fusion graph convolutional network (STFGCN) that leverages multi-feature fusion to explicitly account for weather conditions, demonstrating improved prediction performance over representative baselines. Similarly, Ye et al. [33] proposed an Attention-Based Spatio-Temporal Graph Convolutional Network considering External Factors (ABSTGCN-EF), which models contextual features such as time of day, day of the week, and traffic accident events. Their results show that incorporating external information can significantly enhance predictive accuracy. While these approaches improve representation learning by integrating contextual features, they generally rely on dense and continuous sensor data to achieve strong performance. However, in practice many bicycle traffic monitoring locations suffer from insufficient sensor coverage or intermittent operation, resulting in sparse and incomplete historical records. Under such data-limited conditions, conventional deep learning models often struggle to generalize effectively. This limitation highlights the need for models with stronger transferability and robustness in low-data scenarios.

3.2.2 Pre-trained LLMs for time-series prediction

Recent advancements in pre-trained LLMs have significantly expanded their applications beyond natural language processing into time-series prediction tasks [34]. This cross-domain transferability stems from their architecture of modeling sequential dependencies through self-attention mechanisms [16]. Recently, LLMs have demonstrated remarkable potential in capturing both short-term fluctuations and long-term trends in complex time-series data in domains such as finance, healthcare, and energy systems [35]. For example, Yu et al. [36] present a novel study that harnesses the knowledge and reasoning capabilities of Large Language Models (LLMs) for explainable financial time-series forecasting. Li et al. [37] propose Multimodal ECG-Text Self-supervised pre-training (METS) to improve electrocardiogram (ECG) data classification, enabling zero-shot prediction and improving model generalization without relying on annotated data.

In the domain of traffic prediction, pre-trained LLMs have recently been explored as a promising alternative to conventional spatial-temporal traffic prediction models, due to their ability to exploit structural similarities between traffic data and language, such as sequences and context, and to adapt through fine-tuning to capture complex temporal dependencies. For example, Liu et al. [38] proposed a Spatial-Temporal Large Language Model (ST-LLM) for traffic prediction based on a partially frozen attention strategy to adapt the LLM to capture global spatial-temporal dependencies for traffic prediction. Experiments on the NYCTaxi and CHBike datasets offer evidence that ST-LLM is a powerful spatial-temporal learner that outperforms state-of-the-art models. Ren et al. [39] introduced a novel traffic prediction framework leveraging LLMs (TPLLM). A sequence embedding layer based on Convolutional Neural Networks (CNNs) and a graph embedding layer based on Graph Convolutional Networks (GCNs) is used to extract sequence features and spatial features. Experiments on PeMS04 and PeMS08 datasets collected by California Department of Transportation demonstrate that TPLLM exhibits commendable performance in both full-sample and few-shot prediction scenarios. Xu and Liu [40] propose a Spatio-temporal Fusion Large Language model (GPT4TFP) for traffic flow prediction, which is divided into four components: the spatio-temporal embedding layer, the spatio-temporal fusion layer, the frozen pre-trained LLM layer, and the output linear layer. The experimental results on NYCTaxi, CitiBike, PEMS04 and PEMS08 traffic flow datasets show that the proposed model outperforms a set of state-of-the-art baseline models. Current applications of LLMs in traffic prediction have predominantly focused on motorized vehicles and shared-bike systems. However, private bicycle traffic prediction remains relatively underexplored despite its unique challenges. Private bicycle flows exhibit greater behavioral variability [41]. Moreover, the influence of environmental factors and localized infrastructure on private bicycle usage has yet to be thoroughly investigated.

3.3 Problem Formulation

The ability to reliably predict traffic conditions at locations with limited data collection is particularly crucial for infrastructure planning, real-time mobility management, and safety enhancement in bicycle mobility and traffic flow, as accurate prediction in these locations could support better-informed planning and resource allocation, reduce uncertainty in infrastructure investment decisions, and provides a more complete understanding of traffic dynamics for the users and decision-makers. To address these challenges, this study develops a novel prediction framework that is capable of providing an accurate and reliable prediction for bicycle traffic, even in cases where historical data from some sensors is limited.

Definition (Active-mode traffic network): Given the challenges associated with collecting bidirectional bicycle traffic data, we define the bicycle traffic network as an undirected graph $\mathcal{G} = (V, E, \mathbf{A})$, where V represents the set of nodes corresponding to sensors that monitor bicycle traffic flow within the road network; E denotes the set of edges capturing the connectivity between these sensors; and $\mathbf{A} \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ is the weighted adjacency matrix of $\mathcal{G}$ encoding the structural relationships among the N nodes.

Problem (Multi-step short-term traffic prediction): The input for traffic prediction consists of sensor traffic flow data from the past hour. For each sensor, the traffic flow data is represented as $\mathbf{X} = (\mathbf{X}_{t-T_k+1}, \mathbf{X}_{t-T_k+2}, \ldots, \mathbf{X}_t) \in \mathbb{R}^{N\times F\times T_k}$, where T_k denotes the number of time steps in each input sequence. In addition, the data at each location and each time step such as $\mathbf{X}_{t-T_k+2}$ is treated as a token. At each time step t, the feature matrix is given by $\mathbf{X}_t = (\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, \ldots, \mathbf{x}_{t,N}) \in \mathbb{R}^{N\times F}$, where N is the number of sensors, and F represents the feature dimension of each node.

Our goal is to find a function f to predict the following T time steps data $\hat{\mathbf{y}} = (\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_{t+2}, \ldots, \hat{\mathbf{y}}_{t+T}) \in \mathbb{R}^{N\times F\times T}$, that is:

$$(\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_{t+2}, \ldots, \hat{\mathbf{y}}_{t+T}) = f_\theta((\mathbf{X}_{t-T_k+1}, \ldots, \mathbf{X}_t), \mathbf{A}_t) \tag{3.1}$$

here, θ represents the learnable parameters of the function.

3.4 The BiSTLLM Framework

This section presents the details of the proposed BiSTLLM. As illustrated in **Fig. 3.1**, the framework employs a pre-trained LLM as the core prediction engine, augmented with a customized architecture designed to capture domain-specific spatial-temporal embeddings for bicycle traffic flow prediction. BiSTLLM comprises three primary components: (1) extraction of spatial-temporal bicycle traffic flow embeddings, (2) modeling of bicycle traffic-specific external contextual embeddings, and (3) LLM-based prediction utilizing enriched spatial-temporal embedding.

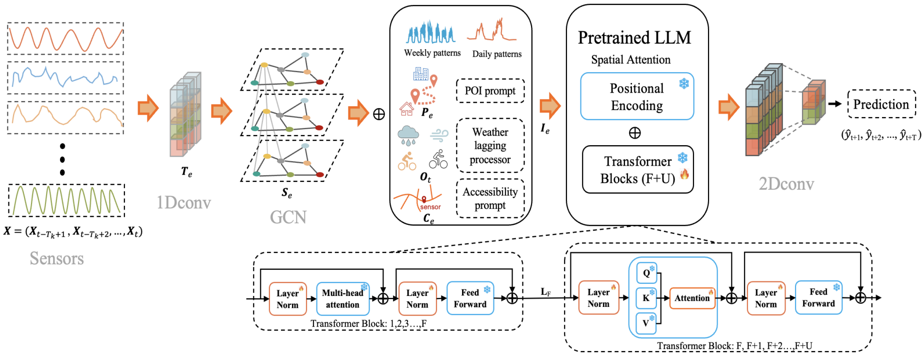


Figure 3.1 The Framework of BiSTLLM. Here, Q, K, and V refer to the Query, Key, and Value representations in multi-head self-attention, respectively. In our case, they are traffic embeddings, obtained by applying learned linear transformations to the input traffic data.

The framework begins by processing the bicycle traffic data through parallel modules for spatial-temporal embeddings capturing:

- **Temporal embedding capturing:** 1DConv [42] is used to encode sequential patterns in historical bicycle flow data, effectively capturing local temporal dependencies.

- **Spatial embedding capturing:** The GCN generates node embeddings that encode structural dependencies within the bicycle traffic network.

To enhance the representational capacity of the model and capture the unique characteristics of bicycle traffic, the following bicycle traffic-specific external contextual feature embedding are incorporated:

- **Weather condition embedding capturing:** Meteorological data undergoes temporal alignment through a weather lagging processor that accounts for delayed weather impacts on bicycle flows.

- **Geographic context embedding capturing:** Contextual features derived from POIs and bicycle lane accessibility are encoded through prompt-based semantic extraction using the pre-trained LLM, capturing geographic and infrastructural factors that influence cycling behavior.

These embeddings are then fused and organized into a unified input stream for the LLM:

- **Input embeddings construction:** The spatial-temporal traffic flow embeddings are fused with temporal indicators, processed weather embedding, geographic semantic context, and node-specific embeddings to form a rich, structured representation.

The pre-trained LLM processes these inputs based on the fine-tuning strategy of section 3.4.4, enabling the model to jointly capture complex spatial and temporal dependencies. This design facilitates accurate, context-aware predictions of bicycle traffic flow across diverse urban environments.

3.4.1 Spatial-temporal bicycle traffic flow embeddings

(1) Temporal dependency capture module

While LLMs are powerful sequence learners, they are primarily optimized for processing natural language and may lack the inductive biases necessary to effectively capture short-term, domain-specific temporal patterns critical for traffic flow prediction. To address this limitation, we introduce a series of 1DConv layers as a preprocessing module prior to inputting the data into the pre-trained LLM. This design choice is particularly relevant for traffic sequences, where temporal fluctuations are often influenced by dynamic external factors such as weather conditions, special events, and variations in travel demand behavior. The 1DConv layers explicitly encode these fine-grained temporal features, enhancing the quality and informativeness of the input representations. By capturing temporal dependencies of bicycle traffic before feeding data into the pre-trained LLM, the model may be better equipped to learn complex temporal patterns, potentially improving accuracy and robustness in traffic flow predictions.

$$\mathbf{T_e} = \text{1DConv}(\mathbf{X}; \mathbf{W}, \mathbf{b}) \tag{3.2}$$

where $\mathbf{W} \in \mathbb{R}^{k \times C \times C'}$ denotes the convolutional kernel weights with kernel size k, mapping from C input channels to C' output channels; $\mathbf{b} \in \mathbb{R}^{C'}$ is the bias term; $\mathbf{T_e} \in \mathbb{R}^{T' \times C'}$ is the output temporal embedding; T' is the output sequence length; $\mathbf{X}$ is the input traffic flow data, which includes only feature flow.

(2) Spatial dependency capture module

The bicycle traffic network is modelled as a graph, where nodes represent sensors, and edges capture their spatial correlations. Each node is associated with traffic features derived from historical traffic flow data. Spatial-based GCN [43] performs successive graph convolutions using an adjacency matrix $\mathbf{A}$, allowing the model to aggregate information from neighboring nodes. This enables the network to learn feature representations that encapsulate local and global spatial dependencies within the traffic network. Through this process, spatial-based GCN effectively refines the feature representations of each node, facilitating improved downstream pre-trained LLM prediction performance. The corresponding operational equation 3.3 is formulated as follows:

$$\mathbf{S_e} = ReLU(\mathbf{A}\mathbf{X}_t\mathbf{W}) \tag{3.3}$$

where the trainable weight matrix, denoted by $\mathbf{W}$ is used to transform the node features during the graph convolution operation. The Rectified Linear Unit $ReLU(\cdot)$ is employed as the activation function to introduce non-linearity into the model, ensuring its capacity to capture complex patterns in the data. $\mathbf{X}_t$ is the output from temporal dependency capture module.

The sensors distributed across the road network are modeled as nodes within a graph, where the Dijkstra algorithm [44] is employed to compute the shortest paths between nodes. These shortest paths define the topological connectivity of the network, capturing the structural relationships between sensors. Recognizing that traffic patterns at a given sensor are often influenced by the behavior of neighboring sensors, where closer sensors tend to exhibit more similar traffic patterns, a distance-based adjacency matrix is introduced to establish weighted connections between sensors. The weights in this matrix are derived using a Gaussian kernel weighting function [45], as formulated in Equation (3.4), which assigns higher weights to pairs of sensors that are geographically closer and lower weights to those that are farther apart. To further capture latent spatial correlations and enhance model adaptability, we treat this adjacency matrix as a learnable parameter. This allows the model to dynamically adjust spatial dependencies during training, thereby more accurately modeling the influence of neighboring sensors on local traffic flow patterns.

$$Dis_d^{i,j} = \begin{cases} \exp(-\frac{|\mathrm{dist}(i,j)|^2}{2\hat{\theta}^2}) & \text{if } dist(i,j) < H \\ 0 & \text{else} \end{cases} \tag{3.4}$$

where the weight of the graph's edges is denoted by $Dis_d^{i,j}$, $|\mathrm{dist}(i,j)|$ indicates the Euclidean distance between node i and node j. In this context, $\hat{\theta}$ represents the standard deviation of distances, and H is utilized as the threshold parameter.

3.4.2 Bicycle traffic-specific external contextual feature processing module

(1) Weather lagging processor

Bicycle traffic is highly sensitive to external environmental factors, particularly weather conditions [41, 46, 47]. For instance, bicycle flow typically decreases during periods of rainfall or strong winds. However, this reduction in traffic flow does not necessarily occur immediately with the onset of adverse weather; instead, a lagged effect is often observed between the change in weather and its impact on bicycle traffic [23], as shown in **Fig. 3.2**. Moreover, the influence of weather is nonuniform across different traffic volumes. For example, weather effect on low traffic volumes may differ from its impact during peak periods. To accurately model the influence of weather on bicycle

Figure 3.2 The delayed effect of weather conditions on bicycle traffic patterns. The black and blue bicycle icons indicate different cycling directions. The orange and red bicycle icons represent cyclists exhibiting a lagged response to adverse weather conditions across different cycling directions: some cyclists stop and wait immediately, while others continue their trips despite unfavorable weather.

traffic, The weather impacts is first dynamically scaled based on real-time traffic conditions using a sigmoid-based adjustment mechanism. Subsequently, we employ a causal temporal convolution [48] to capture the delayed effects of weather on traffic flow, as shown in Equation (3.5).

$$\mathbf{o}_t = \mathrm{DepthwiseConv1D}\big((\mathbf{W}_r\mathbf{p}_t + \mathbf{W}_w\mathbf{s}_t) \otimes \sigma(\mathbf{W}_f\mathbf{X}_t)\big) \tag{3.5}$$

where $\mathbf{o}_t \in \mathbb{R}^d$ is the output lagging weather representation at time t, $\mathbf{W}_r, \mathbf{W}_w, \mathbf{W}_f \in \mathbb{R}^{d\times 1}$ are learnable projection weights, $\mathbf{p}_t, \mathbf{s}_t, \mathbf{X}_t \in \mathbb{R}$ denote rainfall, wind speed, and bicycle traffic flow produced by the spatial dependency capture module respectively, $\sigma(\cdot)$ is the sigmoid activation function, $\otimes$ represents the Hadamard (element-wise) product, and $\mathrm{DepthwiseConv1D}(\cdot)$ applies a causal depthwise convolution so that the convolutional filter never sees the future weather information.

(2) Geographic feature embedding capture

POI-based semantic embedding generation: Traffic patterns within a road network are strongly influenced by the surrounding land use characteristics [49]. For example, sensors deployed in areas with a high density of restaurants often exhibit distinct flow patterns due to increased human activity, while zones dominated by educational institutions, such as schools, reflect different temporal behaviors. Regions that share similar land use characteristics, such as areas dense with restaurants or commercial establishments, may exhibit analogous traffic patterns, even if geographically distant. To capture these latent spatial correlations, we extract textual descriptions of POIs surrounding each sensor location using OpenStreetMap (OSM) data. These textual POI metadata include categories such as food services, retail, entertainment, public services, lodging, and cultural institutions. To ensure consistency and interpretability, we construct structured natural language descriptions using a predefined template. Each description summarizes the distribution of nearby POI categories for a given node. We then leverage a pre-trained LLM, GPT-2, as a semantic feature extractor to convert the POI textual metadata into dense vector embeddings. Specifically, the POI text associated with each sensor location is tokenized using GPT-2's tokenizer and fed through GPT-2's embedding layers to generate semantic-rich feature vectors $\mathbf{P}_e$, as defined in Equation (3.6). These embeddings serve as auxiliary input embeddings, which are concatenated with traffic flow spatial-temporal embedding in the proposed BiSTLLM framework to enhance the model's understanding of spatial context.

$$\mathbf{P}_e = \mathcal{E}_{\text{GPT-2}} \left(\mathbf{T}_{\text{POI}} \left(P_{\text{id}}(\text{Lat}, \text{Lon}) \right) \right) \tag{3.6}$$

where $P_{\text{id}}(\text{Lat}, \text{Lon})$ denotes the geographic coordinates of the sensor. $\mathbf{T}_{\text{POI}}(\cdot)$ denotes POI-related textual metadata; $\mathcal{E}_{\text{GPT-2}}(\cdot)$ indicates the embedding extraction operation using GPT-2.

Bicycle lane accessibility semantic embedding generation: The availability and connectivity of bicycle lanes significantly influence urban traffic dynamics, especially for regions with a high density of non-motorized mobility, and areas with extensive and well-connected bicycle lane infrastructure tend to become active cycling corridors, which in turn shape local traffic flow patterns [50]. Furthermore, geographically distant regions that share similar levels of bicycle infrastructure accessibility may exhibit analogous traffic behaviors. To capture this form of structural spatial context, A bicycle lane accessibility prompt is designed, which encodes the degree of bicycle lane integration at each sensor location. Specifically, for each traffic sensor, we extract the sub-network of bicycle lanes based on OSM located within a 100-meter radius of its geographic position. Within this localized network, we identify the number of distinct bicycle lane segments directly connected to the topological lane on which the sensor resides. This connectivity count serves as a proxy for local infrastructure integration and accessibility. The resulting numerical accessibility value is then transformed into a descriptive textual prompt. This textual metadata is tokenized using GPT-2's tokenizer and subsequently passed through the embedding layer to obtain the semantic representation $\mathbf{C}_e$, as shown in Equation (3.7):

$$\mathbf{C}_e = \mathcal{E}_{\text{GPT-2}} \left(\mathbf{T}_{\text{bike}} \left(\text{Net}_{\text{sensor}}(\text{Lat}, \text{Lon}) \right) \right) \tag{3.7}$$

here, $\text{Net}_{\text{sensor}}(\text{Lat}, \text{Lon})$ denotes the local bicycle lane sub-network centered at the sensor's geographic location. $\mathbf{T}_{\text{bike}}(\cdot)$ generates the lane-related textual metadata, and $\mathcal{E}_{\text{GPT-2}}(\cdot)$ represents the GPT-2-based embedding function applied to the tokenized prompt.

3.4.3 Input embedding fusion

We concatenate the previously described spatial-temporal embeddings into a unified representation, which serves as the input to the pre-trained LLM. To align these embeddings with the input structure expected by the LLM, we employ a two-dimensional convolutional neural network (2DConv) for formatting and transformation, which serves purely as a linear projection for feature fusion without applying any activation functions. This step ensures that spatial-temporal embeddings are not only effectively fused but also organized in a tokenized structure compatible with the LLM architecture. The 2DConv module captures local spatial and temporal dependencies across both dimensions, thereby enhancing the contextual coherence of the input sequence. This transformation allows the LLM to better interpret domain-specific traffic dynamics and facilitates end-to-end learning for downstream prediction tasks. The equation is shown below:

$$\mathbf{I}_e = 2Dconv(Concate(\mathbf{T_e}, \mathbf{S}_e, \mathbf{T}_d w, \mathbf{N}_e, \mathbf{o}_t, \mathbf{C}_e, \mathbf{P}_e), \mathbf{W}_{2d}) \tag{3.8}$$

where $\mathbf{T}e$ denotes the temporal input embedding to the pre-trained LLM, and $\mathbf{S}e$ represents the spatial input embedding. $\mathbf{o}_t$ corresponds to the lagged weather embedding, while $\mathbf{T}_d w$ denotes the time-of-day and day-of-week embeddings, $\mathbf{P}_e$ represents the POI embedding, $\mathbf{C}_e$ is the bicycle lane accessibility embedding, and $\mathbf{N}_e$ corresponds to the node-specific adaptive embedding used to encode the traffic network structure [38]. $\mathbf{W}_{2d}$ represents the learnable parameters of 2Dconv.

3.4.4 Pre-trained LLM fine-tuning strategy

Fine-tuning the pre-trained LLM is critical to facilitate downstream tasks based on domain-specific datasets, enabling the models to better capture task-relevant semantics and contextual nuances [51]. Prior approaches such as the Frozen Pretrained Transformer (FPT) [52] and the Partially Frozen Attention (PFA) LLM [38] have demonstrated strong performance across a variety of downstream tasks. However, short-term bicycle flow prediction presents unique challenges that may limit their effectiveness. To enhance the performance of pretrain-LLM for short-term bicycle traffic prediction, we introduce a fine-tuning strategy for the proposed BiSTLLM framework. Specifically, we freeze the multi-head attention (MHA) and feed-forward network (FFN) sublayers in the first F transformer blocks while allowing updates only to the layer normalization (LN) components. This design preserves the general linguistic and structural priors learned during pretraining while enabling modest task-specific adaptation. In the subsequent U transformer layers, we further allow updates to the attention mechanisms by unfreezing the attention computations and LN sublayers, while keeping the FFN and the query, key, and value projection matrices frozen. Additionally, we retain the positional encoding in a fixed state, as we focus on spatial dimensions in pre-trained LLM and positional encoding along this dimension is not semantically meaningful. This fine-tuning strategy balances adaptability and efficiency, allowing the model to capture domain-specific patterns in bicycle traffic flow data while minimizing the risk of catastrophic forgetting and reducing computational overhead. In contrast, randomly freezing ignores the functional roles of different modules and can lead to unstable training dynamics or degraded performance.

The fine-tuning approach adopted in this study is based on a pre-trained LLM constructed using the GPT-2 Transformer architecture with small size. GPT-2, as an autoregressive model built upon the Transformer decoder, offers a relatively compact parameter structure, making it well-suited for traffic flow prediction tasks due to its balance between expressiveness and computational efficiency [53]. The architectural modifications applied to different components of the model during fine-tuning are illustrated in the lower section of **Fig. 3.1**. The formulas are shown in Equations (3.9) and (3.10)

$$\begin{aligned} \bar{\mathbf{E}}_i &= \mathrm{MHA}(\mathrm{LN}(\mathbf{E}_i)) + \mathbf{E}_i \\ \mathbf{E}_{i+1} &= \mathrm{FFN}(\mathrm{LN}(\bar{\mathbf{E}}_i)) + \bar{\mathbf{E}}_i \end{aligned} \tag{3.9}$$

where $i \in 1, \ldots, F-1$; $\mathbf{E}_1 = \mathbf{I}e + \mathbf{P}_E$, with $\mathbf{P}_E$ denoting the frozen positional encoding; $\mathbf{E}_i$ is the input to the i-th Transformer block; $\bar{\mathbf{E}}_i$ represents the intermediate output after applying the unfrozen layer normalization $\mathrm{LN}(\cdot)$ followed by the frozen multi-head attention $\mathrm{MHA}(\cdot)$; and $\mathbf{E}_{i+1}$ is the final output after applying the unfrozen $\mathrm{LN}(\cdot)$ and the frozen feed-forward network (FFN) to $\bar{\mathbf{E}}_i$.

$$\begin{aligned} \bar{\mathbf{E}}_{F+U-1} &= \mathrm{MHA}(\mathrm{LN}(\mathbf{E}_{F+U-1})) + \mathbf{E}_{F+U-1} \\ \mathbf{E}_{F+U} &= \mathrm{FFN}(\mathrm{LN}(\bar{\mathbf{E}}_{F+U-1})) + \bar{\mathbf{E}}_{F+U-1} \end{aligned} \tag{3.10}$$

where $\bar{\mathbf{E}}_{F+U-1}$ denotes the intermediate representation at the $(F+U-1)$-th layer, obtained by applying the unfrozen layer normalization $\mathrm{LN}(\cdot)$ and the unfrozen multi-head attention $\mathrm{MHA}(\cdot)$ with frozen query, key, and value projections; and $\mathbf{E}_{F+U}$ represents the final output after applying the unfrozen $\mathrm{LN}(\cdot)$ and the frozen feed-forward network (FFN) to $\bar{\mathbf{E}}_{F+U-1}$.

3.4.5 Loss function

Due to the presence of noise and outliers in bicycle traffic flow data caused by sensor failures or environmental interference, we adopt the robust L1 loss function in BiSTLLM. L1 loss is less sensitive to extreme deviations, making it more suitable for handling anomalies in real-world traffic datasets.

$$\mathcal{L}_{\text{loss}} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{3.11}$$

here, N is the total number of samples; y_i denotes the ground truth value of the i-th sample; and $\hat{y}_i$ is the corresponding predicted value. The overall workflow of BiSTLLM is outlined in Algorithm 1.

Algorithm 1: The BiSTLLM Framework

Input: Bicycle traffic features $\mathbf{X}_t$ over T_k historical timesteps; training epochs E_r; learning rate η; batch size B_z; optimizer: Ranger21.
Output: Trained BiSTLLM model.
foreach *epoch* $r \in \{1, 2, \ldots, E_r\}$ **do**
    Shuffle the training data;
    foreach *batch* $\mathbf{X}_T$ *in training data* **do**
        $\mathbf{T_e}, \mathbf{S}_e, \mathbf{T}_dw, \mathbf{N}_e, \mathbf{o}_t, \mathbf{C}_e, \mathbf{P}_e \leftarrow$ Spatial and Temporal Embedding using Equations (2), (3), (5), (6), and (7);
        $\mathbf{I}_e \leftarrow$ Fuse embeddings using Equation (8);
        for $i = 1$ **to** $F + U$ **do**
            if $i \leq F$ **then**
                Compute $\mathbf{E}^{i+1}$ using Equation (9) with $\mathbf{E}^i$;
            else
                Compute final output $\mathbf{E}^{F+U}$ using Equation (10) with $\mathbf{E}^{F+U-1}$;
            end
        end
        $\hat{y} \leftarrow$ The prediction output;
        Update all learnable parameters by minimizing the loss in Equation (11) via Ranger21 optimizer.
    end
end

3.5 Experimental Results and Analysis

In this section, we first describe the datasets and evaluation metrics employed in our experiments. We then outline the baseline models and detail the experimental settings used for performance comparison. Finally, we present and analyze the prediction results, followed by an ablation study to evaluate the contribution of each component in the proposed framework.

3.5.1 Dataset analysis

To assess the effectiveness of the proposed model, we conducted extensive experiments on real-world bicycle traffic flow data obtained from the National Road Traffic Data Portal of the Netherlands. The dataset comprises measurements recorded by loop detector sensors strategically deployed across two Dutch cities: Rotterdam and Leiden. For Rotterdam, the dataset spans from September 1, 2024, to November 30, 2024, covering 18 sensor locations. Similarly, for Leiden, data were collected from 18 sensors over the period from January 1, 2025, to April 30, 2025. The sensors distribution are shown in **Fig. 3.3**. The raw sensor readings were preprocessed by aggregating the traffic flow data into 5-minute intervals, resulting in 12 observations per hour. This temporal resolution offers a practical balance between granularity and computational efficiency. To facilitate efficient model training and improve convergence, the aggregated data were standardized using z-score normalization, ensuring zero mean and unit variance across features. For a rigorous and unbiased evaluation, the dataset was chronologically partitioned into training, validation, and test sets using a 6:2:2 split. This forward-looking split methodology mitigates data leakage and ensures that the model is evaluated on future, unseen observations, thereby providing a realistic assessment of its predictive performance and generalization capability.

Table 3.1 DATASETS INFORMATION

Datasets	Time Span	Number of Sensors	Aggregation Granularity
Rotterdam	2024/09/01-2024/11/30	18	5-min
Leiden	2025/01/01-2025/04/30	18	5-min

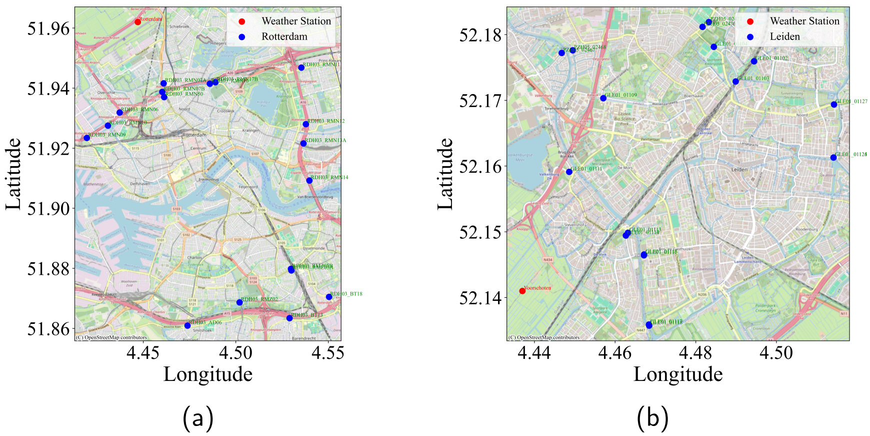


Figure 3.3 Sensors and weather station distribution. (a) Rotterdam and Rotterdam weather station. (b) Leiden and Voorschoten weather station.

Weather information was obtained from the Royal Netherlands Meteorological Institute (KNMI), specifically from the Rotterdam and Voorschoten weather stations. The dataset includes hourly observations of key meteorological variables, namely precipitation (measured in millimeters) and mean wind speed (measured in 0.1 m/s). Precipitation represents the total rainfall recorded in the preceding hour, while wind speed reflects the average wind conditions over the same period. To align the weather data temporally with the 5-minute bicycle traffic flow intervals, each hourly weather observation was divided into 12 equal sub-intervals. The mean wind speed was replicated across all 12 sub-intervals, assuming consistent wind conditions within each hour. For precipitation, the hourly value was uniformly distributed across the 12 intervals to approximate the average rainfall per 5-minute period. This interpolation strategy allows for fine-grained temporal alignment between meteorological data and traffic flow observations, thereby enhancing the model's ability to capture lagged or cumulative weather effects on cycling behavior.

In this paper, we use traffic flow data from the previous hour (12 observations per hour) to predict traffic flow for the next hour at 5-minute intervals.

3.5.2 Evaluation metrics

In this paper, the prediction results of bicycle traffic flow are evaluated by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Weighted Absolute Percentage Error (WAPE). The formulations to calculate these metrics are shown below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{3.12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.13}$$

$$WAPE = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\sum_{i=1}^{n}|y_i|}, \tag{3.14}$$

where $\hat{y_i}$ and y_i represent the predicted value and ground truth data of the sample i, respectively. n is the number of sample values.

3.5.3 Baseline methods

To assess the performance of our proposed BiSTLLM model in predicting bicycle traffic flow, we compare it with several previously used traffic prediction models. Through this comparison, our aim is to highlight BiSTLLM's capability in predicting bicycle traffic. The baseline models used for comparison are briefly described below:

- HA: The Historical Average (HA) model predicts traffic flow by taking the average value of historical data. Specifically, the HA baseline is computed as the per-sensor, per historical observations average.
- LSTM: Long Short-Term Memory (LSTM) [27] is a specialized variant of Recurrent Neural Networks (RNNs) specifically designed to effectively capture and model long-term relations in sequential data.

- STGCN: Spatial-Temporal Graph Convolutional Network (STGCN) [54] is proposed to address the problem of time series traffic prediction by harnessing comprehensive spatial-temporal correlations.
- ASTGNN: The Attention-based Spatial-temporal Graph Neural Network (ASTGNN) [55] is designed to capture the dynamics of traffic data across both temporal and spatial dimensions.
- STMFGNN: Spatial-Temporal Multifactor Fusion Graph Neural Network (STMFGNN)[56] leverages dynamic similarity and static adjacency graphs for parallel graph convolution, integrating global hidden and local prior knowledge. A gated fusion module adaptively learns dynamic influence weights to capture multiscale spatial dependencies. The model employs gated tanh unit convolution, multireceptive fields, and gated recurrent units for temporal feature extraction, enabling comprehensive traffic flow prediction by considering multiscale factors.
- OFA: One Fits All (OFA) [53] demonstrates a unified framework for time series prediction by leveraging pretrained language models, specifically GPT-2, without modifying its core architectural components such as self-attention mechanisms or feed-forward residual blocks. We adopt a complementary perspective on traffic data within the OFA framework, optimizing its representation and processing pipeline to enhance predictive performance.
- ST-LLM: Spatial-Temporal Large Language Model (ST-LLM) [38] incorporates spatial-temporal embeddings and a fusion convolution layer for unified representation, along with a partially frozen attention mechanism to adapt pretrained LLMs for traffic prediction.
- LLAMA2: LLAMA2 is a suite of pretrained and fine-tuned large language models developed by Meta, designed for a wide range of natural language understanding and generation tasks. In our baseline setting, we adopt a variant of ST-LLM [38] in which the original GPT-2 backbone is replaced with LLAMA2-7B.

3.5.4 Experimental setting

All experiments were conducted using Google Colab, a cloud-based Python environment equipped with an NVIDIA L4 GPU (CUDA 12.4, 23 GB memory) and an Intel® Core™ i9-9900KS CPU running at 4.0 GHz. Deep learning models were implemented using the PyTorch framework. For training the LLM-based models, we employed the Ranger21 optimizer, while the Adam optimizer was used for all other deep learning baselines. Hyperparameters, including batch size and learning rate, were carefully tuned through a validation set, with batch sizes ranging from 64 to 256 and learning rates varying between 0.001 and 0.0001. Specifically, the proposed BiSTLLM model utilized a GPT-2 architecture with five transformer layers and a learning rate of 0.001, while the other LLM-based baselines were configured with six GPT-2 layers using the same learning rate. LLAMA2-7B is used in the baseline. We implemented the baselines as described in their original paper.

3.5.5 Full-sample prediction

To assess the effectiveness of the proposed BiSTLLM framework in standard traffic flow prediction tasks, we conduct comprehensive experiments using the full-sample bicycle traffic flow prediction based on the Rotterdam and Leiden datasets. The comparative results against baseline models are presented in **Table 3.2** with average 15-minute, 30-minute and 60-minute prediction errors. Overall, the proposed BiSTLLM exhibits the highest accuracy.

Table 3.2 PERFORMANCE COMPARISON OF MULTI-STEP TRAFFIC PREDICTION (THE UNIT IS THE COUNT OF BICYCLE TRAFFIC PER 5 MINUTES). THE BOLD RESULTS ARE THE BEST

Datasets	**Baselines**	**15 min**			**30 min**			**60 min**		
		MAE	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**
Rotterdam	HA	58.77	81.47	85.95%	58.76	81.46	85.97%	58.76	81.45	86.01%
	LSTM	25.83	42.57	40.75%	27.32	44.95	43.12%	30.24	50.42	47.74%
	STGCN	24.94	39.70	39.35%	25.40	40.58	40.08%	26.59	42.91	41.99%
	ASTGCN	24.20	38.14	38.18%	24.65	38.94	38.90%	26.00	41.63	41.05%
	STMFGNN	23.24	37.04	36.67%	23.59	37.66	37.22%	24.54	39.16	38.73%
	OFA	26.64	43.18	42.04%	26.98	43.83	42.57%	27.56	45.08	43.51%
	ST-LLM	22.48	35.83	35.47%	22.68	36.24	35.79%	23.01	37.13	36.33%
	LLAMA2	23.55	36.84	37.16%	23.75	37.22	37.47%	24.01	37.86	37.90%
	BiSTLLM	**22.26**	**35.31**	**35.12%**	**22.34**	**35.47**	**35.26%**	**22.61**	**36.07**	**35.70%**
Leiden	HA	82.26	123.25	73.42%	82.27	123.25	73.42%	82.27	123.26	73.42%
	LSTM	29.23	51.07	26.69%	31.86	56.05	28.76%	36.51	65.61	32.97%
	STGCN	27.13	44.48	24.78%	29.66	50.00	26.78%	33.10	56.96	29.89%
	ASTGCN	24.78	40.84	22.63%	26.60	44.58	24.02%	29.27	50.09	26.43%
	STMFGNN	23.55	38.64	21.51%	25.15	41.50	22.71%	26.71	45.73	24.11%
	OFA	28.02	46.36	25.59%	29.13	49.37	26.30%	30.68	53.09	27.70%
	ST-LLM	22.68	38.42	20.71%	23.38	40.07	21.11%	24.23	43.04	21.88%
	LLAMA2	22.69	37.27	20.72%	23.46	39.08	21.19%	24.56	42.20	22.18%
	BiSTLLM	**21.59**	**35.89**	**19.72%**	**22.40**	**37.72**	**20.22%**	**23.36**	**41.18**	**21.09%**

In particular, conventional spatial-temporal deep learning models such as STMFGNN, ASTGCN, and STGCN show comparatively lower performance compared to BiSTLLM. This may be attributed to their limited ability to capture the complex and highly variable spatial-temporal patterns of bicycle traffic. Furthermore, simpler models like HA and LSTM, which either lack spatial modeling or struggle with non-linear temporal dependencies, also exhibit notably lower predictive accuracy.

Among the LLM-based baselines, BiSTLLM outperforms LLAMA2, ST-LLM, and OFA, achieving reductions in average MAE of 5.83%, 1.74%, and 17.96% respectively, on the Rotterdam dataset for the 60-minute prediction horizon. Similarly, on the Leiden dataset, BiSTLLM achieves MAE reductions of 4.89%, 3.59%, and 23.86%, respectively. These improvements highlight BiSTLLM's enhanced capacity to model complex bicycle traffic dynamics by accurately capturing representative traffic embeddings and effectively fine-tuning the pre-trained LLM. The difference in prediction errors across datasets such as BiSTLLM has a WAPE of 35.70% in Rotterdam compared to 21.09% in Leiden is primarily due to differences in the magnitude and distribution of bicycle traffic volumes between the two cities.

In summary, BiSTLLM demonstrates strong generalization across locations and time scales, offering significant improvements over both traditional deep learning models and existing LLM-based baselines in the bicycle traffic flow prediction task.

3.5.6 Ablation study

The results of different ablation studies show the impact of including different components in the BiSTLLM modeling framework, as well as the impact of the proposed fine-tuning strategy.

(1) BiSTLLM components ablation study

To assess the contribution of each component within the proposed BiSTLLM framework, we conduct a series of ablation experiments. These experiments systematically include or exclude specific modules to evaluate their individual impact on the model's overall performance in bicycle traffic flow prediction. Specifically, we examine several variants of BiSTLLM, including: (1) w/o Temporal: a variant with the temporal dependency modeling module removed; (2) w/o Spatial: a variant with the spatial dependency modeling module excluded; (3) w/o LLM: a version without the pre-trained LLM component; (4) w/o Lag: a variant omitting the weather lag processing module; (5) w/o Weather: a version in which weather embeddings are removed; (6) w/o POI: a variant without the POI embedding; and (7) w/o Accessibility: a version excluding the bicycle lane accessibility embedding.

Fig. 3.4 presents the results of the ablation study, highlighting the impact of removing individual components from the BiSTLLM framework. The removal of the pre-trained LLM component (w/o LLM) results in a substantial increase in prediction errors across MAE, RMSE, and WAPE metrics. This suggests that the LLM is a critical element in capturing complex dependencies inherent in bicycle traffic flow data. Furthermore, the exclusion of the temporal (w/o Temporal) and spatial (w/o Spatial) modules also leads to significant performance degradation, underscoring the importance of spatial-temporal representations for accurately modeling traffic dynamics. In addition, removing external contextual embeddings, specifically weather embeddings (w/o Weather), POI embedding (w/o POI), and bicycle lane accessibility embedding (w/o Accessibility), consistently worsens prediction performance. This demonstrates the utility of integrating external factors that influence bicycle traffic patterns. Notably, the w/o Lag

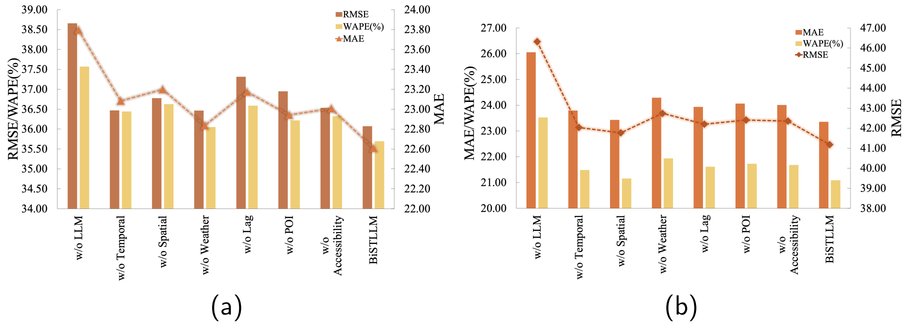


Figure 3.4 BiSTLLM components ablation study. (a) Rotterdam. (b) Leiden. Each component contributes positively to the overall performance of the model.

variant, which omits the weather lagging mechanism, exhibits a marked increase in error. This may be attributed to a misalignment between weather conditions and their delayed impact on traffic flow; without accounting for this temporal offset, the weather input may act as noise rather than informative context.

(2) Pre-trained LLM fine-tuning strategy ablation study

In this section, we conduct an ablation study to evaluate the effectiveness of our fine-tuning strategy for the pre-trained LLM to do bicycle traffic flow prediction. Our approach builds upon the Partially Frozen Attention (PFA) mechanism proposed by Liu et al. [38], with key distinctions in whether the positional encoding module and components of the final U layers are frozen or unfrozen.

To systematically assess the impact of these variations, we designed a set of model configurations by selectively freezing or unfreezing different components. The experimental variants are as follows: (1) Frozen-Full: A fully frozen model where no layers or modules are fine-tuned. (2) Frozen-Pe & Frozen-Att-Proj(ours): The positional encoding module is frozen. Additionally, within the final U layers, the query, key, and value projection matrices of the attention mechanism are kept frozen. (3) Unfrozen-Pe & Frozen-Att-Proj: The positional encoding module is unfrozen, while the attention projections in the last U layers remain frozen. (4) Unfrozen-Pe & Unfrozen-Att-Proj: Both the positional encoding module and the attention projection matrices in the last U layers are unfrozen, allowing maximum adaptation. (5) Frozen-Pe & Unfrozen-Att-Proj: The positional encoding module remains frozen, while the attention projection matrices in the final U layers are fine-tuned.

The ablation results, presented in **Table 3.3**, demonstrate that freezing both the positional encoding and the attention projection layers yields the best performance among all fine-tuning strategies. In contrast, variants that involve unfreezing either the positional encoding or the attention projection layers result in increased prediction errors. One possible explanation is that the pre-trained LLM in our framework primarily leverages spatial attention mechanisms, which are not inherently dependent on any sequential order among nodes. Since traffic nodes lack a natural ordering, unfreezing the positional encoding may introduce noise or confusion into the model's spatial reasoning. Moreover, while the pre-trained projections were originally optimized for linguistic tasks, their generic ability to model pairwise interactions (via query/key/value mappings) may transfer to traffic nodes. Freezing these layers preserves this inductive bias, preventing overfitting to limited fine-tuning data.

Table 3.3 ABLATION STUDY OF FINE-TUNING STRATEGY (AVERAGE 60-MINUTE HORIZON)

Datasets	Baselines	MAE	RMSE	WAPE
Rotterdam	Unfrozen-Pe & Frozen-Att-Proj	22.86	36.69	36.08%
	Unfrozen-Pe & Unfrozen-Att-Proj	22.81	36.25	36.00%
	Frozen-Full	22.79	36.31	35.97%
	Frozen-Pe & Unfrozen-Att-Proj	22.77	36.53	35.94%
	Frozen-Pe & Frozen-Att-Proj (Ours)	**22.61**	**36.07**	**35.70%**
Leiden	Unfrozen-Pe & Frozen-Att-Proj	23.54	41.74	21.26%
	Unfrozen-Pe & Unfrozen-Att-Proj	23.83	42.18	21.51%
	Frozen-Full	23.42	41.34	21.15%
	Frozen-Pe & Unfrozen-Att-Proj	23.83	42.53	21.52%
	Frozen-Pe & Frozen-Att-Proj (Ours)	**23.36**	**41.18**	**21.09%**

3.5.7 Parameter analysis

The performance of the BiSTLLM framework is notably sensitive to its hyperparameter settings. To identify the optimal configuration for bicycle traffic flow prediction, we conducted a comprehensive grid search across multiple hyperparameter combinations. As illustrated in **Fig. 3.5**, we investigated various configurations of the pre-trained LLM in terms of total number of layers and the number of unfrozen layers U used during fine-tuning. Motivated by previous findings [53], which highlight a favorable trade-off between performance and computational cost with a 6-layer GPT-2 variant, we examined models with 5, 6, and 7 layers. We further evaluated the impact of varying the number of unfrozen layers U on predictive performance. The results, presented in **Fig. 3.5c** and **Fig. 3.5f**, reveal that the best performance was achieved using a 5-layer LLM with 4 unfrozen layers on the Rotterdam dataset, and a 5-layer LLM with 5 unfrozen layers on the Leiden dataset. These findings suggest that deeper fine-tuning of shallower LLMs can effectively enhance domain adaptation for traffic prediction tasks. Additionally, we performed a grid search on the batch size and learning rate. As shown in **Fig.3.5a**, **Fig.3.5d**, **Fig.3.5b**, and **Fig.3.5e**, we observed that a batch size of 64 and a learning rate of 0.001 consistently yielded the highest prediction accuracy on both datasets. These hyperparameter settings were used in all subsequent experiments.

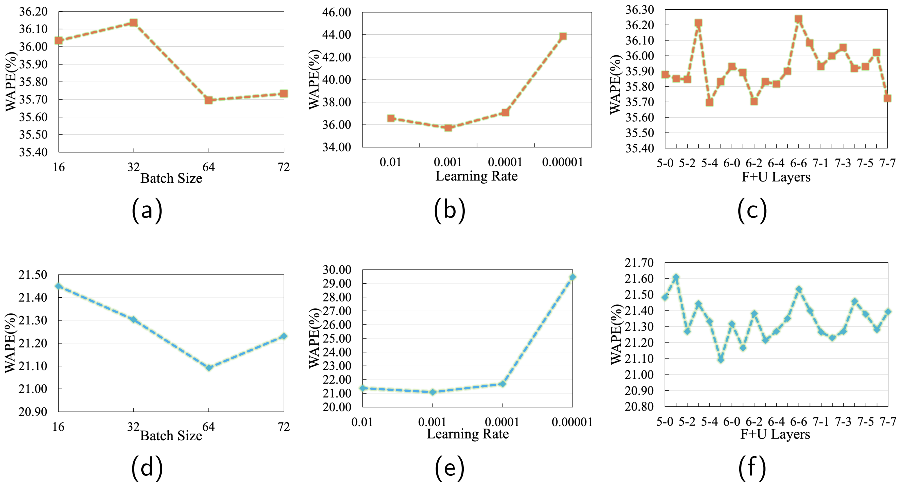


Figure 3.5 Parameter analysis. (a) Batch Size(Rotterdam). (b) Learning Rate(Rotterdam). (c) $F+U$ layers(Rotterdam). (d) Batch Size(Leiden). (e) Learning Rate(Leiden). (f) $F+U$ layers(Leiden). Here $F+U$ is the total number of layers and U is the number of unfrozen layers by allowing updates to the attention computations in attention mechanisms and to the LN sublayers during fine-tuning.

3.5.8 Few-shot prediction

Collecting high-quality bicycle traffic data presents significant challenges due to both technical and resource limitations. Unlike car traffic, which often benefits from extensive sensing infrastructure and continuous monitoring systems, bicycle traffic monitoring typically relies on fewer and often less accurate sensors deployed in limited locations. Often, these sensors operate only during specific time periods to reduce operational costs. As a result, the data collected is often sparse, inconsistent, or incomplete over time. In such scenarios, traditional data-hungry prediction models may underperform or fail to generalize. To address this limitation, we adopt a few-shot learning framework, which enables effective traffic flow prediction from limited training samples, thereby improving model adaptability in real-world, low-resource settings.

Consistent with the full-sample experiments, the datasets are chronologically divided into training, validation, and testing sets. However, for the few-shot prediction setting, only last 10% of the original training set is used, while the validation and test sets remain identical to those used in the full-sample experiments. Consequently, the final data split corresponds to 6% for training, 20% for validation, and 20% for testing. As reported in **Table 3.4**, the average prediction errors across 15-minute, 30-minute, and 60-minute horizons demonstrate that BiSTLLM consistantly outperforms other baselines, specifically, BiSTLLM outperforms LLAMA2, ST-LLM, and OFA, achieving reductions in average MAE of 17.68%, 3.13%, and 18.22%, respectively, on the Rotterdam dataset for the 60-minute prediction horizon. Similarly, on the Leiden dataset, BiSTLLM achieves MAE reductions of 14.40%, 10.17%, and 23.63%, respectively. These results underscore the robustness ability of BiSTLLM, particularly in data-scarce scenarios.

Table 3.4 PERFORMANCE COMPARISON OF FEW-SHOT TRAFFIC PREDICTION. THE BOLD RESULTS ARE THE BEST

Datasets	**Baselines**	**15 min**			**30 min**			**60 min**		
		MAE	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**
Rotterdam	LSTM	33.92	57.94	53.52%	34.85	58.46	54.99%	37.21	62.09	58.74%
	STGCN	37.35	60.73	58.93%	38.51	61.41	60.77%	43.88	67.69	69.28%
	ASTGCN	32.94	55.09	51.97%	36.52	56.76	57.63%	41.13	61.50	64.94%
	STMFGNN	26.20	42.79	41.34%	26.92	43.86	42.49%	28.77	47.23	45.42%
	OFA	28.88	46.67	45.56%	30.01	48.94	47.35%	32.09	53.55	50.66%
	ST-LLM	25.75	41.89	40.63%	25.97	42.02	40.98%	27.09	44.04	42.77%
	LLAMA2	29.13	45.00	45.96%	30.02	46.53	47.37%	31.88	49.68	50.33%
	BiSTLLM	**25.23**	**41.79**	**39.80%**	**25.37**	**41.76**	**40.04%**	**26.24**	**43.28**	**41.43%**
Leiden	LSTM	84.40	138.31	77.07%	85.74	140.23	77.42%	85.83	140.49	77.50%
	STGCN	82.69	131.77	75.51%	84.62	135.45	76.41%	86.68	139.52	78.27%
	ASTGCN	66.10	105.35	60.36%	73.79	109.52	66.63%	80.91	114.10	73.05%
	STMFGNN	27.04	48.16	24.70%	29.22	51.72	26.38%	32.19	57.05	29.06%
	OFA	30.32	51.31	27.69%	32.40	55.59	29.26%	35.85	63.72	32.37%
	ST-LLM	29.22	49.87	26.68%	29.71	50.94	26.82%	30.48	52.32	27.52%
	LLAMA2	30.55	45.92	27.90%	30.59	47.76	27.62%	31.98	51.00	28.88%
	BiSTLLM	**24.89**	**41.59**	**22.73%**	**25.83**	**43.50**	**23.32%**	**27.37**	**47.21**	**24.72%**

3.6 Conclusions

Bicycle traffic exhibits significant fluctuations due to its strong sensitivity to external environmental factors. This inherent variability makes it considerably less predictable than vehicular traffic. To address these challenges and effectively capture the fluctuations in bicycle traffic flow, we have proposed BiSTLLM, a novel framework for bicycle traffic flow prediction based on pre-trained LLM.

In the BiSTLLM framework, we introduce a weather-lagging processor and incorporate semantic embeddings of POIs and bicycle lane accessibility, enabling the model to learn infrastructure-dependent traffic behaviors and localized activity patterns. These enhancements facilitate the model's ability to capture both local and global spatial-temporal interactions. To adapt the general knowledge encoded in a pre-trained LLM to the domain of bicycle traffic, we implement a domain-specific fine-tuning strategy. Experiments demonstrate that BiSTLLM consistently outperforms existing state-of-the-art models across various prediction horizons. Given that bicycle traffic data is often sparse, seasonal, and unevenly distributed over time, we also evaluate BiSTLLM under few-shot prediction scenarios, where only 10% of the training data is used. Notably, BiSTLLM maintains strong performance in these low-data conditions, highlighting its robustness and adaptability.

In conclusion, this study illustrates the potential of leveraging pre-trained LLM for accurate bicycle traffic prediction, even in data-sparse environments. However, several limitations remain. One key omission is the assumption that limited historical data can be supplemented effectively through fine-tuning based on pre-trained LLM without fully evaluating the differences of urban contexts. Additionally, due to constraints in data availability and scope, we were unable to evaluate the model's performance across a wider network of cities. While BiSTLLM is designed as a general spatiotemporal framework and does not rely on region-specific features, the experimental validation in this study is limited to bicycle traffic datasets from the Netherlands. Differences in cycling behavior, infrastructure, and data collection practices across regions may affect model performance.

For future work, exploring the computational complexity and time cost of the proposed model in real world practice and further developing a generalized bicycle traffic prediction model capable of predicting flow at sensor locations with no historical data is necessary. Additionally, examining the interpretability of the proposed model with directional neworks through transportation theory is necessary to better understand its functioning. In addition, including the evaluation to more diverse cycling datasets from other cities or countries is an important direction to assess the generalizability and scalability of the proposed approach. Addressing these challenges will further enhance the real-world applicability of the bicycle traffic prediction model, contributing to the broader goal of providing effective solutions in data-scarce scenarios and supporting more sustainable transportation planning.

Acknowledgements

This research is supported by the China Scholarship Council. This support is gratefully acknowledged.

References

[1] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin. "Deep learning on traffic prediction: Methods, analysis, and future directions". In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2021), pp. 4927–4943.

[2] C. Janiesch, P. Zschech, and K. Heinrich. "Machine learning and deep learning". In: *Electronic markets* 31.3 (2021), pp. 685–695.

[3] X. Yao, S. C. Calvert, and S. P. Hoogendoorn. "Driving heterogeneity identification using machine learning: A review and framework for analysis". In: *Transportation Research Interdisciplinary Perspectives* 32 (2025), p. 101511.

[4] R. Zhao, Z. Sun, and A. Ji. "A deep reinforcement learning approach for automated on-ramp merging". In: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2022, pp. 3800–3806.

[5] A. Ji, D. Li, Z. Dai, M. Cui, L. Yu, and Z. Duan. "A hybrid graph memory network approach with multi-level feature representation for traffic flow forecast". In: *Expert Systems with Applications* (2025), p. 130316.

[6] A. Ji, Z. Liu, L. Su, and Z. Dai. "A hybrid framework for spatio-temporal traffic flow prediction with multi-scale feature extraction". In: *Information Sciences* (2025), p. 122259.

[7] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li. "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 5668–5675.

[8] A. Naghib, N. Jafari Navimipour, M. Hosseinzadeh, and A. Sharifi. "A comprehensive and systematic literature review on the big data management techniques in the internet of things". In: *Wireless Networks* 29.3 (2023), pp. 1085–1144.

[9] J. Liu, G. P. Ong, and X. Chen. "GraphSAGE-based traffic speed forecasting for segment network with sparse data". In: *IEEE Transactions on Intelligent Transportation Systems* 23.3 (2020), pp. 1755–1766.

[10] D. Ton, D. C. Duives, O. Cats, S. Hoogendoorn-Lanser, and S. P. Hoogendoorn. "Cycling or walking? Determinants of mode choice in the Netherlands". In: *Transportation research part A: policy and practice* 123 (2019), pp. 7–23.

[11] D. Zollner, R. Vasiliev, B. Castellano, G. Molnar, and P. Janssen. "A technical examination to explore conditional multimodal contextual synthesis in large language models". In: (2024).

[12] C. H. Song, J. Wu, C. Washington, B. M. Sadler, W.-L. Chao, and Y. Su. "Llm-planner: Few-shot grounded planning for embodied agents with large language models". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, pp. 2998–3009.

[13] A. Ramezani and Y. Xu. "Knowledge of cultural moral norms in large language models". In: *arXiv preprint arXiv:2306.01857* (2023).

[14] H. Xu, A. Sharaf, Y. Chen, W. Tan, L. Shen, B. Van Durme, K. Murray, and Y. J. Kim. "Contrastive preference optimization: Pushing the boundaries of llm performance in machine translation". In: *arXiv preprint arXiv:2401.08417* (2024).

[15] Z. Li, Y. Cao, X. Xu, J. Jiang, X. Liu, Y. S. Teo, S.-W. Lin, and Y. Liu. "Llms for relational reasoning: How far are we?" In: *Proceedings of the 1st International Workshop on Large Language Models for Code*. 2024, pp. 119–126.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[17] L. Torrey and J. Shavlik. “Transfer learning”. In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

[18] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, *et al.* “Time-llm: Time series forecasting by reprogramming large language models”. In: *arXiv preprint arXiv:2310.01728* (2023).

[19] X. Lee and Y. Gu. “STEP-LLM: Spatial-Temporal-Enriched Prompting with Large Language Models for Traffic Prediction”. In: *International Conference of Electrical, Electronic and Networked Energy Systems*. Springer. 2024, pp. 248–255.

[20] H. Qu, H. Li, L. You, R. Zhu, J. Yan, P. Santi, C. Ratti, and C. Yuen. “ChatEV: Predicting electric vehicle charging demand as natural language processing”. In: *Transportation Research Part D: Transport and Environment* 136 (2024), p. 104470.

[21] P. Li, Y. Pang, and J. Ren. “Spatio-Temporal Graph Convolutional Network Combined Large Language Model: A Deep Learning Framework for Bike Demand Forecasting”. In: *arXiv preprint arXiv:2403.15733* (2024).

[22] X. Huang. “Enhancing traffic prediction with textual data using large language models”. In: *arXiv preprint arXiv:2405.06719* (2024).

[23] X. Wen, P. Krishnakumari, and S. Hoogendoorn. “Dynamic Spatial-Temporal Graph Convolutional Neural Networks Approach for Active Mode Traffic Prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* (2025).

[24] G. E. Box and D. A. Pierce. “Distribution of residual autocorrelations in autoregressive-integrated moving average time series models”. In: *Journal of the American statistical Association* 65.332 (1970), pp. 1509–1526.

[25] M. Awad, R. Khanna, M. Awad, and R. Khanna. “Support vector regression”. In: *Efficient learning machines: Theories, concepts, and applications for engineers and system designers* (2015), pp. 67–80.

[26] L. Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.

[27] A. Graves and A. Graves. “Long short-term memory”. In: *Supervised sequence labelling with recurrent neural networks* (2012), pp. 37–45.

[28] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. “Learning phrase representations using RNN encoder-decoder for statistical machine translation”. In: *arXiv preprint arXiv:1406.1078* (2014).

[29] K. O’shea and R. Nash. “An introduction to convolutional neural networks”. In: *arXiv preprint arXiv:1511.08458* (2015).

[30] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. “A comprehensive survey on graph neural networks”. In: *IEEE transactions on neural networks and learning systems* 32.1 (2020), pp. 4–24.

[31] X. Shi, H. Qi, Y. Shen, G. Wu, and B. Yin. “A spatial–temporal attention approach for traffic prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* 22.8 (2020), pp. 4909–4918.

[32] X. Qi, J. Yao, P. Wang, T. Shi, Y. Zhang, and X. Zhao. “Combining weather factors to predict traffic flow: A spatial-temporal fusion graph convolutional network-based deep learning approach”. In: *IET Intelligent Transport Systems* 18.3 (2024), pp. 528–539.

[33] J. Ye, S. Xue, and A. Jiang. “Attention-based spatio-temporal graph convolutional network considering external factors for multi-step traffic flow prediction”. In: *Digital Communications and Networks* 8.3 (2022), pp. 343–350.

[34] Y. Jiang, Z. Pan, X. Zhang, S. Garg, A. Schneider, Y. Nevmyvaka, and D. Song. “Empowering time series analysis with large language models: A survey”. In: *arXiv preprint arXiv:2402.03182* (2024).

[35] J. Ye, W. Zhang, K. Yi, Y. Yu, Z. Li, J. Li, and F. Tsung. “A survey of time series foundation models: Generalizing time series representation with large language model”. In: *arXiv preprint arXiv:2405.02358* (2024).

[36] X. Yu, Z. Chen, Y. Ling, S. Dong, Z. Liu, and Y. Lu. “Temporal data meets LLM–explainable financial time series forecasting”. In: *arXiv preprint arXiv:2306.11025* (2023).

[37] J. Li, C. Liu, S. Cheng, R. Arcucci, and S. Hong. “Frozen language model helps ecg zero-shot learning”. In: *Medical Imaging with Deep Learning*. PMLR. 2024, pp. 402–415.

[38] C. Liu, S. Yang, Q. Xu, Z. Li, C. Long, Z. Li, and R. Zhao. “Spatial-temporal large language model for traffic prediction”. In: *2024 25th IEEE International Conference on Mobile Data Management (MDM)*. IEEE. 2024, pp. 31–40.

[39] Y. Ren, Y. Chen, S. Liu, B. Wang, H. Yu, and Z. Cui. “TPLLM: A traffic prediction framework based on pretrained large language models”. In: *arXiv preprint arXiv:2403.02221* (2024).

[40] Y. Xu and M. Liu. “GPT4TFP: Spatio-temporal fusion large language model for traffic flow prediction”. In: *Neurocomputing* 625 (2025), p. 129562.

[41] X. Wen, M. Khosla, and S. Hoogendoorn. “Multi-Source Transfer Learning With Spatial-Temporal Graph Neural Network for Short-Term Bicycle Traffic Prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* (2025).

[42] S. Bai, J. Z. Kolter, and V. Koltun. “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling”. In: *arXiv preprint arXiv:1803.01271* (2018).

[43] T. N. Kipf and M. Welling. “Semi-supervised classification with graph convolutional networks”. In: *arXiv preprint arXiv:1609.02907* (2016).

[44] N. A. Ojekudo and N. P. Akpan. “Anapplication of Dijkstra’s Algorithm to shortest route problem”. In: *IOSR Journal of Mathematics (IOSR-JM)* 13.3 (2017), p. 14.

[45] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains". In: *IEEE signal processing magazine* 30.3 (2013), pp. 83–98.

[46] S. Hanson and P. Hanson. "Evaluating the impact of weather on bicycle use". In: *Transportation Research Record* 629 (1977), pp. 43–48.

[47] L. Böcker, M. Dijst, and J. Prillwitz. "Impact of everyday weather on individual daily travel behaviours in perspective: a literature review". In: *Transport reviews* 33.1 (2013), pp. 71–91.

[48] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. "Mobilenets: Efficient convolutional neural networks for mobile vision applications". In: *arXiv preprint arXiv:1704.04861* (2017).

[49] A. Izanloo, A. K. Rafsanjani, and S. P. Ebrahimi. "Effect of commercial land use and accessibility factor on traffic flow in Bojnourd". In: *Journal of Urban Planning and Development* 143.2 (2017), p. 05016016.

[50] J. Dill and T. Carr. "Bicycle commuting and facilities in major US cities: if you build them, commuters will use them". In: *Transportation research record* 1828.1 (2003), pp. 116–123.

[51] K. VM, H. Warrier, Y. Gupta, *et al.* "Fine tuning llm for enterprise: Practical guidelines and recommendations". In: *arXiv preprint arXiv:2404.10779* (2024).

[52] K. Lu, A. Grover, P. Abbeel, and I. Mordatch. "Frozen pretrained transformers as universal computation engines". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 36. 7. 2022, pp. 7628–7636.

[53] T. Zhou, P. Niu, L. Sun, R. Jin, *et al.* "One fits all: Power general time series analysis by pretrained lm". In: *Advances in neural information processing systems* 36 (2023), pp. 43322–43355.

[54] B. Yu, H. Yin, and Z. Zhu. "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting". In: *arXiv preprint arXiv:1709.04875* (2017).

[55] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong. "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting". In: *IEEE Transactions on Knowledge and Data Engineering* 34.11 (2021), pp. 5415–5428.

[56] H. Jia, Z. Yu, Y. Chen, and H. Xia. "Spatial-temporal multi-factor fusion graph neural network for traffic prediction". In: *Applied Intelligence* 54.19 (2024), pp. 9464–9487.

Chapter 4

Transfer Learning for Bicycle Traffic Prediction

In chapter 2, we explored the complexities of the active mode traffic system and recognized that data size could represent a potential limitation affecting the robustness of active mode traffic prediction. While the results do not explicitly confirm this, our understanding suggests that data availability and size warrant further consideration in designing effective prediction models. Building on these insights, this chapter focuses on enhancing prediction accuracy in data-scarce target cities by leveraging knowledge from multiple source domains with more abundant data. Our approach begins with a traffic pattern clustering technique, grouping traffic patterns from diverse source cities to facilitate more effective single-source knowledge transfer to target cities. To further maximize the utility of this knowledge, we introduce an adaptive transfer learning framework that integrates the most relevant insights from multiple source models, tailoring them specifically for active mode traffic prediction in the target city. Experimental results demonstrate that our proposed multi-source transfer learning spatial-temporal graph neural network (Multi-TLSTGCN) model not only surpasses baseline models but also outperforms single-source transfer approaches. Furthermore, as data availability in the target city decreases, our model exhibits superior robustness compared to non-transfer models, underscoring its effectiveness in delivering accurate predictions in data-scarce urban environments.

This chapter is published as a journal article: Wen, X., Khosla, M. and Hoogendoorn, S. Multi-Source Transfer Learning with Spatial-Temporal Graph Neural Network for Short-term Bicycle Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2025. The content is the same as the journal paper.

4.1 Introduction

Accurate short-term traffic predictions are crucial for efficient transportation operations in urban areas. These predictions enable proactive decision-making by traffic management authorities, helping to mitigate traffic delay, reduce travel time, and enhance overall road network efficiency [1, 2]. Traffic prediction has been approached through simulation, parametric, and machine learning methods. Simulation techniques like discrete event simulation (DES) [3] and agent-based modeling (ABM) [4] replicate traffic dynamics but require extensive system knowledge and may lack generalizability. Parametric methods, including regression models [5], autoregressive integrated moving average (ARIMA) [6], and Kalman filters [7], offer quantitative insights but struggle with complex, non-linear traffic patterns.

Machine learning has shifted traffic prediction towards data-driven models that capture intricate patterns and adapt to dynamic environments. Methods such as k-nearest neighbors (KNN) [8], support vector machines (SVM) [9], and artificial neural networks (ANN) [10] improve prediction but often fail to capture deep traffic dependencies. Deep learning further enhances accuracy by modeling spatial-temporal dependencies. Recurrent neural networks (RNNs) [11], long short-term memory networks (LSTMs) [12], and attention mechanisms [13] extract temporal trends, while graph neural networks (GNNs) [14] and convolutional neural networks (CNNs) [15] model spatial interactions. By integrating both, deep learning provides more precise and comprehensive traffic forecasts.

The effectiveness of machine learning models depends on data quality and quantity. Despite abundant traffic data from IoT [16], real-world datasets often suffer from missing values, sensor errors, and inconsistencies. Resource constraints further limit traffic monitoring, especially in smaller cities. Privacy concerns may also restrict data access. Bicycle traffic faces even greater challenges due to inadequate monitoring infrastructure, as urban planning has historically prioritized motorized transport. Data scarcity and quality issues undermine prediction accuracy, making reliable insights into bicycle traffic increasingly critical as cities shift toward reduced car dependency.

Transfer learning addresses data scarcity by leveraging knowledge from data-rich environments (source domain) to improve predictions in data-limited areas (target domain) [17]. In transportation, it enhances traffic prediction by transferring patterns from well-monitored cities to smaller ones. Some studies achieve this by freezing lower model layers while fine-tuning upper layers with target data [18], while others mitigate domain shifts using adversarial training [19], maximum mean discrepancies (MMD) [20], or Dynamic Time Warping (DTW) [21]. Despite progress, transfer learning for traffic prediction, especially bicycle traffic, remains an ongoing challenge.

4.1.1 Current research gaps and our contributions

One major limitation of existing transfer learning models is their narrow focus on single-source and single-target domains, with limited exploration of ensemble transfer learning. This is particularly problematic for bicycle traffic, where sensor distribution is generally sparse, and data from a single city may not provide sufficient coverage for accurate predictions. The lack of research into multi-source transfer learning, and specifically what information should be transferred, leaves a significant gap, especially for bicycle traffic. In motorized traffic, road structures are relatively consistent across cities, leading to similar distribution patterns that can be easily transferred between urban environments. In contrast, bicycle traffic is far more complex, with no clear road hierarchies, and traffic patterns vary widely from city to city. This variability makes it difficult to identify consistent trends that can be transferred effectively across different urban environments.

To overcome the above challenges, we develop a robust multi-source transfer learning framework which we refer to as Multi-TLSTGCN tailored to the requirements of data-scarce spatial-temporal bicycle traffic prediction. Such a framework would be instrumental in overcoming the limitations of current models and providing accurate predictions. Our approach enhances traffic prediction accuracy in data-scarce target cities by leveraging information learned from multiple source domains with more abundant data. Our approach begins with a traffic pattern clustering technique, which groups traffic patterns from various source cities, enabling more effective single-source knowledge transfer to the target cities. Subsequently, to maximize the utility of knowledge from all sources, we employ an adaptive transfer learning approach that integrates the most relevant aspects from the multi-source models, tailoring them for bicycle traffic prediction in the target city. We conducted extensive experiments using real-world bicycle traffic datasets from six cities to validate the proposed framework. Results indicate that our proposed Multi-TLSTGCN model outperforms both baseline models and single-source transfer approaches. Notably, as target city data availability decreases, our model's robustness surpasses that of models without transfer support, demonstrating its effectiveness for accurate bicycle traffic prediction in data-scarce urban settings.

The paper is organized as follows. Section II provides an overview of existing research on spatial-temporal traffic flow prediction, followed by an exploration of research on transfer learning in transportation. Section III offers detailed insights into the proposed Multi-TLSTGCN model. Section IV introduces the methodology of this study, including the spatial-temporal prediction model, clustering approach and the propose adaptive transfer learning approach. Section V presents the experimental setup, datasets used, experiment results, discussions, and an ablation study. Finally, Section VI concludes this work.

4.2 Related Work

4.2.1 Spatial-temporal traffic flow prediction

In recent years, the burgeoning availability of traffic data has sparked a growing interest in deep learning traffic prediction research. Deep learning models for traffic prediction, For example, Recurrent Neural Networks (RNNs) [11] can leverage vast amounts of historical traffic data to learn complex nonlinear temporal patterns and make accurate predictions. Graph Neural Networks (GNNs) [22] are tailored to handle non-Euclidean spatial structural data, making them well-suited for modeling the intricate structure of traffic networks [23]. By integrating GNNs with temporal dependencies, it becomes promising to capture the complexity and non-linear spatial-temporal traffic patterns effectively. Bao et al. [24] investigated spatial–temporal accurate traffic flow prediction based on Spatial–Temporal Complex Graph Convolution Network (ST-CGCN). It combines spatial and temporal features while considering node correlations and external interferences. By leveraging dynamic weights and complex correlation matrices, ST-CGCN outperforms existing models in real-world datasets. Shi et al. [25] designed a novel Attention-based Periodic-Temporal neural Network (APTN) to address the challenges of accurate traffic forecasting in smart cities. By leveraging attention mechanisms, APTN effectively captures spatial, short-term, and long-term periodical dependencies in traffic data. The experiments demonstrated significant improvements of APTN in traffic forecasting applications over existing state-of-the-art models based on real-world datasets PeMSD4 and PeMSD8 from California. However, the effectiveness of these advanced methods can decline when trained on limited data, potentially resulting in overfitting when the dataset is not sufficiently large.

In recent years, significant amounts of motorized traffic data have been gathered, particularly in large metropolitan areas with abundant resources. This is largely due to the variety of data collection sources available, such as traffic cameras, GPS devices integrated into vehicles, and advanced smart infrastructure sensors [26]. These data sources have enabled detailed analysis and prediction of traffic patterns for cars and other motorized vehicles in large metropolitan areas [24] [27]. However, when it comes to bicycle traffic, the situation is different. In countries like the Netherlands, where bicycle is a primary mode of urban transport, particularly for short-distance travel and eco-friendly commuting [28], the collection of bicycle traffic data has been problematic. The lack of dedicated sensors and tracking systems, along with frequent malfunctions of the existing ones, significantly hampers the ability to gather comprehensive bicycle traffic data [29]. The problem is even more pronounced in smaller municipalities, where limited resources make sustained data collection efforts difficult, further complicating the analysis and prediction of bicycle traffic patterns. While general bicycle traffic data is difficult to obtain, bike-sharing systems offer a notable exception. In large metropolitan areas, data from these systems is abundant, thanks to the robust digital infrastructure that supports them [30]. However, this ease of data collection is not applicable to general bicycle traffic, which remains largely unmonitored. Consequently, most existing studies on bicycle traffic flow prediction focus on bike-sharing systems [31], leaving general bicycle traffic, despite its significance as a major urban transportation mode, largely neglected.

4.2.2 Transfer learning in transportation

Transfer learning is a technology focuses on utilizing knowledge gained from one domain or task and applying it to a different but related domain or task [17]. In the context of traffic prediction, transfer learning has shown significant promise in addressing challenges related to traffic data scarcity, variability in traffic patterns, and the need for robust predictive models. Li et al. [18] applied transfer learning techniques with multiple transfer strategies to machine learning methods for short-term traffic prediction using data from highways England road networks in the UK. To address the challenge of data scarcity in small cities, Huang et al. [21] proposed a cross-city traffic prediction approach called TrafficTL. This method leverages big data from other cities to assist data-scarce cities by identifying similarities between data sets and mitigating negative transfer effects caused by differences in data distributions from distant locations. Yao et al. [32] introduces an Adversarial Domain Adaptation with Spatial-Temporal Graph Convolutional Network (Ada-STGCN) model designed to predict traffic indicators for data-scarce target road networks by transferring knowledge from data-rich source road networks. Experimental results on real-world traffic datasets demonstrate that Ada-STGCN outperforms state-of-the-art baseline methods in traffic flow prediction tasks, delivering superior prediction accuracy. These studies have explored various methods for transferring knowledge from data-rich cities to data-scarce ones. However, these methods are typically designed for motorized traffic, which benefits from abundant traffic data both temporally and spatially, with extensive sensor coverage across the road network. In contrast, sensor distribution for private bicycle traffic is often sparse, resulting in limited source data that may cover only a small portion of the road network. This lack of diverse traffic patterns in the source data makes it challenging to transfer knowledge effectively. Consequently, existing methods may not be suitable for private bicycle traffic prediction.

Single-source transfer may not perfectly align with the features of the target domain. In contrast, multi-source transfer can encompass a broader range of scenarios and subdomains within the target domain, leveraging multiple related domains to improve the effectiveness of transfer learning. Li et al. [33] proposed a physics-guided multi-source transfer learning method for multi-region traffic flow, utilizing adversarial training and MFD-based weighting. This approach achieves domain adaptation and assigns weights based on traffic network properties. To mitigate the impact of noise or potentially negative knowledge from source cities, Jin et al. [34] proposed a selective transfer learning framework called CrossTReS for traffic prediction. This framework adaptively re-weights source regions to enhance fine-tuning in the target domain. Mo and Gong [35] proposed a transfer learning method for traffic prediction called Cross-city Multi-Granular Adaptive Transfer Learning (MGAT). The model is trained on multiple source cities using meta-learning algorithm to achieve a strong initialization. It then captures multi-granular regional characteristics and employs an Adaptive Transfer module with Spatial-Attention and Multi-head Attention mechanisms to automatically select the most relevant features, ensuring optimal knowledge transfer. The significant success of these studies highlights the potential of multi-source transfer learning. However, existing approaches primarily focus on multi-region transfer, with a balanced distribution of source data across different regions. For bicycle traffic, where source data may be sparse within each city, this assumption may not hold, making traditional multi-region transfer less effective.

4.3 Preliminaries and Problem Definition

This section provides an overview of the preliminaries relevant to transfer learning traffic prediction. First, we define the structure of the traffic network, followed by an explanation of traffic networks in both the source and target domains. Finally, we outline the specific research problem addressed in this work.

4.3.1 Physical traffic network description

To simplify bicycle spatial correlations and account for the absence of bidirectional traffic flow data, we define the bicycle traffic network as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where $\mathcal{V}$ is the set of nodes representing sensors that record the traffic flow of bicycle on the road network; $\mathcal{E}$ is a set of edges; and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix of $\mathcal{G}$ with N nodes recorded as a weighted adjacency matrix. Note that the traffic flow feature collected by each sensor changes over time. Specifically, $\mathbf{X}_t = (\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, ..., \mathbf{x}_{t,i}, ..., \mathbf{x}_{t,N}) \in \mathbb{R}^{F \times N}$ denotes the traffic feature of the graph $\mathcal{G}$ at time t, where $\mathbf{x}_{t,i}$ indicates the feature value of the $i - th$ node, and F represents the total number of traffic features for each node.

In this paper, sensors are depicted as nodes within a graph, and the Dijkstra algorithm [36] is used to determine the shortest path between nodes within the road network. The connectivity of the topological structure is defined by these shortest paths between nodes. Considering that the traffic patterns observed at a particular sensor can be influenced by the behavior of neighboring sensors, we establish weighted connections between sensors using a distance-based adjacency matrix $\mathbf{A}$. This entails applying a Gaussian kernel weighting function [37], with weights determined by sensor distances, as elaborated in the following.

$$\mathbf{A}(i,j) = \begin{cases} \exp(-\frac{|\mathrm{dist}(i,j)|^2}{2\hat{\theta}^2}) & \text{if } dist(i,j) < H \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

where $|\mathrm{dist}(i, j)|$ indicates the shortest path between node i and node j. In this context, $\hat{\theta}$ represents the standard deviation of distances, and H is utilized as the threshold parameter.

(1) Traffic networks of source domain

We define the multi-source domains graph as $\mathcal{G}^S = \mathcal{G}^{S_1}, ..., \mathcal{G}^{S_i}, ..., \mathcal{G}^{S_M}$ to describe the source road network of the cities, where M represent the number of source cities. $\mathcal{G}^{S_i} = (\mathcal{V}^{S_i}, \mathcal{E}^{S_i}, \mathbf{A}^{\mathbf{S_i}})$ denotes the graph of $i - th$ source city, where $\mathcal{V}^{S_i}$ is the set of nodes representing sensors that record the traffic flow on the road network of $i - th$ source city ; $\mathcal{E}^{S_i}$ is a set of edges; and $\mathbf{A}^{\mathbf{S_i}} \in \mathbb{R}^{N^{S_i} \times N^{S_i}}$ is the adjacency matrix of $\mathcal{G}^{S_i}$ in city i with N^{S_i} nodes recorded as a weighted adjacency matrix. $\mathbf{X}_{\mathbf{t}}^{\mathbf{S_i}} \in \mathbb{R}^{N^{S_i}}$ is the feature matrix of $i - th$ source domain graph at time t.

(2) Traffic networks of target domain

We define the target domain graph as $\mathcal{G}^T = (\mathcal{V}^T, \mathcal{E}^T, \mathbf{A}^T)$, where $\mathcal{V}^T$ is the set of nodes representing sensors that record the traffic flow of the road network; $\mathcal{E}^T$ is a set of edges; and $\mathbf{A}^T \in \mathbb{R}^{N^T \times N^T}$ is the adjacency matrix of $\mathcal{G}^T$ with N^T nodes recorded as a weighted adjacency matrix of target city. The traffic flow feature on the target domain graph changes over time. $\mathbf{X}_t^T \in \mathbb{R}^{N^T}$ is the feature matrix of target domain graph at time t.

4.3.2 Problem definition

Our objective is to develop a precise bicycle traffic prediction model for a data-scarce target road network within a city by leveraging insights gained from other cities. In this paper, both the source and target domains share a common task, which involves predicting future traffic state based on historical observations.

For both the source and target domains, the historical bicycle traffic flow feature for each input of each node includes data from one hour earlier, which can be represented as $\mathbf{X} = (\mathbf{X}_{t-k+1}, \mathbf{X}_{t-k+2}, ..., \mathbf{X}_t) \in \mathbb{R}^{F \times T_k}$, where T_k are the time steps of each slice source, $\mathbf{X}_t = (\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, ..., \mathbf{x}_{t,N}) \in \mathbb{R}^{\mathbb{F}}$. F indicates the vector features of each node at time t. Our goal is to find a function $f(\cdot)$ to predict the following T time steps data $\hat{\mathbf{X}} = (\hat{\mathbf{X}}_{t+1}, \hat{\mathbf{X}}_{t+2}, ..., \hat{\mathbf{X}}_{t+T})$, that is:

$$(\hat{\mathbf{X}}_{t+1}, \hat{\mathbf{X}}_{t+2}, ..., \hat{\mathbf{X}}_{t+T}) = f_\theta((\mathbf{X}_{t-k+1}, ..., \mathbf{X}_t), \mathbf{A}_t) \tag{4.2}$$

To align with the need for immediate traffic management and planning, such as adjusting traffic signals or deploying resources to manage traffic, our approach involves utilizing the most recent one-hour historical data as input for our model to predict the traffic for the subsequent hour, where θ represents the learnable parameters of the function.

4.4 Methodology

To address the challenge of data scarcity in traffic flow prediction, we introduce the Multi-TLSTGCN model, which leverages abundant data from other cities to inform learning in the target city. The core idea is to first build a pre-trained model using sensor data from multiple cities, capturing a wide variety of traffic patterns and dynamics. By leveraging this diverse datasets, the model gains a richer understanding of traffic behaviors, leading to better generalization. Once pre-trained, the model is fine-tuned on the limited data from the target city, resulting in significantly improved predictive performance, even with scarce target data. The overall framework is shown in **Fig. 4.1**.

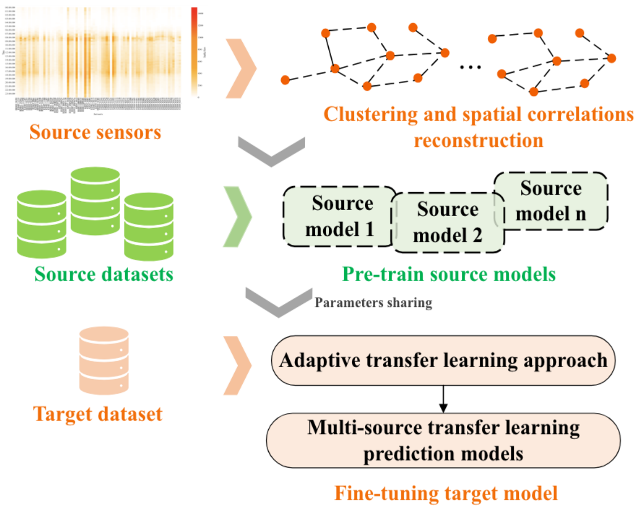


Figure 4.1 Overall framework of Multi-TLSTGCN model. Cluster-based multi-source transfer with reconstructive spatial network.

The *first* challenge in realizing our objective is that, unlike motorized traffic, which follows organized road hierarchies with dense sensor coverage, bicycle traffic lacks such structure, and the irregular distribution of sensors often fails to cover all regions. Additionally, road usage patterns vary significantly between cities, making it difficult to effectively transfer regional models to different bicycle networks. To overcome this, we first cluster sensors from multiple cities into groups based on similar bicycle traffic patterns. Clustering sensor data ensures that those with comparable behaviors are grouped together, enabling us to train a separate source model for each cluster. This process (detailed in Section 4.4.1) improves the likelihood that models will identify familiar patterns when transferred to the target city, thereby reducing the risk of negative knowledge transfer, where mismatched knowledge could degrade performance. Furthermore, clustering helps manage the complexity and variability of bicycle traffic, making models more adaptable and focused within each cluster.

However, even within a cluster of sensors with similar patterns, local differences in traffic behavior still exist at the individual sensor level, resulting in the *second* challenge. To address these variations, we incorporate data-driven and adaptive spatial correlations within each cluster (detailed in Section 4.4.2), allowing the model to adjust for local discrepancies. This ensures that predictions remain accurate and context-sensitive, as the model considers how traffic at one location influences other locations with similar traffic behaviors. (refer to Section 4.4.3)

Relying solely on single-source knowledge of bicycle traffic may not provide sufficient predictive performance in the target domain. This is primarily because bicycle traffic flow can vary significantly across different regions due to a multitude of factors, including population density, urban layout, and the availability of transportation infrastructure for bicycle. Moreover, the distribution of sensors used to collect bicycle traffic data is often sparse, particularly in lower-priority or less densely populated areas. To overcome this *third* challenge, we introduce an adaptive multi-source transfer learning approach that aggregates contributions from multiple source domains to improve prediction accuracy in the target domain (details in Section 4.4.4).

4.4.1 Bicycle traffic pattern clustering

To capture the overall traffic patterns for each sensor, we calculated the average traffic flow at 5-minute intervals throughout the day, creating a general representation of daily traffic trends for each sensor. To categorize sensors with similar daily traffic patterns into distinct groups, we implemented a two-step approach that combined deep learning with clustering techniques. This approach enabled us to effectively group sensors based on their unique traffic flow characteristics.

First, we used a temporal convolutional model to extract key temporal features from the general daily traffic patterns of each sensor. Specifically, a 1D convolutional neural network (1D-CNN) was designed, where the input channel size corresponded to each sensor's half-day time-series traffic patterns. The model processed the average traffic flow daily patterns of each sensor through two 1D-CNN layers with ReLU activation in between, capturing essential temporal characteristics. This process yields a reduced-dimensional matrix encapsulating the critical temporal dynamics of each sensor's traffic flow. Second, we applied Z-score normalization [38] to the transformed data, ensuring that all features contributed equally to the clustering process. Last, we employed the K-means algorithm [39] to cluster sensors with similar traffic patterns.

To determine the optimal number of clusters, we used the Elbow Method and Silhouette Score analysis [40], as detailed in Equation (4.3) and Equation (4.4). The Elbow Method helped identify the point where adding more clusters no longer significantly reduced the sum of squared distances, while the Silhouette Score assessed the cohesion and separation of the clusters. We determine the optimal number of clusters by identifying the point where the sum of squared errors begins to decrease more gradually, while also considering the Silhouette Score, aiming for a higher score to indicate better clustering quality.

$$SSE = \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_j - \mu_i\|^2, \tag{4.3}$$

where SSE is the within-cluster sum of squared errors; k is the number of clusters; C_i represents the i-th cluster; x_j is a data point within cluster C_i; μ_i is the centroid of cluster C_i.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \tag{4.4}$$

where $s(i)$ is Silhouette Score of data point i; $a(i)$ is the average distance from data point i to all other points in the same cluster; $b(i)$ is the minimum average distance from the data point i to all the points in the nearest different cluster.

This combined approach allowed us to effectively cluster sensors based on their daily traffic flow patterns, ensuring that the clusters were both meaningful and representative of the underlying traffic behaviors.

4.4.2 Virtual spatial correlations reconstruction within each cluster

Since road network structures vary significantly across cities, directly transferring spatial correlations based on geographical proximity from source to target domains can lead to ineffective or even negative knowledge transfer. To address this challenge, we introduce virtual spatial correlations, which are defined by traffic pattern similarity rather than physical proximity in source domains. By focusing on similarity, we more effectively identify relevant spatial correlations in the target domain, improving the overall performance of transfer learning. Specifically, we *first* construct a data-driven graph that captures spatial correlations within each cluster based on the traffic attributes of sensors. This graph reflects the underlying similarities in traffic patterns, allowing for more meaningful connections between sensors. *Second*, we implement a parameter-based adaptive graph mechanism that dynamically adjusts spatial correlations over time to obtain overall graph spatial correlations. This adaptive graph continuously refines the model's understanding of spatial dynamics within each cluster, enabling it to account for temporal changes and local variations. This approach allows for more accurate and context-aware predictions. Below we explain the key two steps in detail.

(1) Data-based graph spatial correlations learning

To capture the data-based spatial correlations in bicycle, we begin by employing a 2D convolutional neural network (2D-CNN) to analyze the traffic flow information from each sensor based on the input data. Once the feature representations for each node are obtained, we calculate spatial correlations between nodes using the dot product, as shown in [41]. Instead of using a dynamic adjacency matrix for each input, which can complicate computations and introduce noise, we simplify the process by averaging the values across each batch to generate a stable data-based graph spatial correlations. We compute the the data-based adjacency matrix corresponding to the batch, $\mathbf{A}_{attri}$ as follows:

$$\mathbf{A}_{attri} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{ReLU}(M_i M_i^\top), \tag{4.5}$$

where M_i is the traffic flow representations of all nodes in i-th batch; $M_i^\top$ is the transpose of M_i; n denotes the batch size.

(2) Parameter-based adaptive graph spatial correlations learning

To refine the dynamics of data-based graph spatial correlations over time and uncover hidden spatial relationships, we employ a parameter-based adjacency matrix derived from two learnable embedding dictionaries E_1, $E_2 \in \mathbb{R}^{N\times P}$. These dictionaries are used to capture and model the evolving spatial dependencies among nodes, enabling a more adaptive and nuanced representation of spatial correlations.

$$\mathbf{A}_a = E_1 E_2^\top \tag{4.6}$$

where P is the hidden dimensions of each node. Both E_1, E_2 are the source node embedding dictionaries.

(3) Overall graph spatial correlations learning

The overall graph spatial correlations, as defined in Equation (4.7), are obtained by multiplying the data-based graph spatial correlations with the parameter-based adaptive graph spatial correlations. Rectified Linear Unit activation function (ReLU) is then applied to ensure that the resulting spatial correlation values remain non-negative.

$$\mathbf{A}_{all} = ReLU(\mathbf{A}_{attri}\mathbf{A}_a) \tag{4.7}$$

4.4.3 Spatial-temporal dependency learning

Considering the fact that bicycle traffic network can encompass temporal attributes and diverse spatial features to portray different facets of spatial and temporal dynamics within the transportation system. In this study, we utilize an integrated attention-based spatial-temporal graph convolutional neural network known as ASTGCN, combining both the multi-head self-attention mechanism and graph convolutional neural network, to effectively capture the spatial-temporal dependencies inherent in traffic data of source and target domains.

(1) Temporal dependency capture module

Traffic flow patterns exhibit intricate temporal dependencies, where the flow at any given time is intricately linked to preceding time intervals. The multi-head attention mechanism with position encoding as shown in Equation (4.11) and Equation (4.12) adeptly captures these traffic flow dependencies, this mechanism permits the model to discern various facets of the input sequence through multiple attention heads, facilitating the comprehensive capture of both short-term fluctuations and long-range trends within traffic flow data. In this scenario, we employ the multi-head self-attention mechanism [42], as depicted in Equation (4.8) and Equation (4.9), to capture the temporal feature representation of the traffic flow data.

$$Multi - head(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Concat(head_1, ..., head_h) \tag{4.8}$$

where $head_i$ represents a self-attention mechanism, particularly a Scaled Dot-Product Attention [42], as outlined in Equation (4.10). The traffic flow input sequence $\mathbf{X}$ intended for prediction undergoes projection onto distinct learned linear subspaces, yielding the query $\mathbf{Q}$, key $\mathbf{K}$, and value $\mathbf{V}$ for the attention mechanism, respectively.

$$head_i = attention(\mathbf{X}_i\mathbf{W}_i^Q, \mathbf{X}_i\mathbf{W}_i^k, \mathbf{X}_i\mathbf{W}_i^V) \tag{4.9}$$

let $\mathbf{Q}_i = \mathbf{X}_i\mathbf{W}_i^Q$, $\mathbf{K}_i = \mathbf{X}_i\mathbf{W}_i^K$, and $\mathbf{V}_i = \mathbf{X}_i\mathbf{W}_i^V$, where $\mathbf{W}_i^Q$, $\mathbf{W}_i^K$, and $\mathbf{W}_i^V$ represent the learnable weights of the linear projection layers.

$$attention(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = softmax(\frac{\mathbf{Q}_i\mathbf{K}_i^T}{\sqrt{d_m}})\mathbf{V}_i \tag{4.10}$$

where d_m represents the scaling factor utilized to balance the complexity and capacity of the model.

$$P_{(pt,2dim)} = sin(pt/10000^{2dim/d_m}) \tag{4.11}$$

$$P_{(pt,2dim+1)} = cos(pt/10000^{2dim/d_m}) \tag{4.12}$$

the traffic flow input sequence $\mathbf{X}$ intended for prediction undergoes projection onto distinct learned linear subspaces, where pt represents the position index of the traffic flow input sequence, while dim corresponds to the dim_{th} dimension of the positional encoding vector. Additionally, d_m denotes the dimension of the positional encoding, aligning with the dimension of input embeddings to enable element-wise addition.

(2) Spatial dependency capture module

Bicycle traffic network frequently display intricate spatial connections and irregular geometries, rendering them non-Euclidean in nature. This complexity poses challenges for conventional Euclidean-based methods when applied to graph-based traffic prediction tasks. Spatial-based GCN [43] is a neural network architecture tailored to address this issue. GCN employ convolutional operations within graph structures to aggregate information from neighboring nodes, enabling them to learn representations that encapsulate both local and global spatial characteristics.

In this study, we model the traffic network as a graph, where nodes represent individual sensors and edges represent the connections or correlations between them. Each node is associated with traffic features extracted from its historical traffic flow data. To capture the spatial dependencies, a GCN is employed, which applies a series of graph convolutions to the traffic features of each node. These convolutions use the adjacency matrix $\hat{\mathbf{A}}_t$ to update a node's feature representation by aggregating information from its neighboring nodes. As a result, the GCN produces updated feature representations for each node, reflecting both local and global traffic patterns. The operation of the spatial-based GCN is described by Equation (4.13).

$$GCN(\mathbf{X}_t) = ReLU(\hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}_t\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}_t\mathbf{W}) \tag{4.13}$$

where $\hat{\mathbf{A}}_t$ represents the graph adjacency matrix. $\hat{\mathbf{D}}$ is a diagonal matrix where the i_{th} element on the diagonal corresponds to the degree of the i_{th} node, defined as $\hat{\mathbf{D}}_{ii} = \sum_j \mathbf{A}_{ij}$, here j loops through all nodes in the graph. $\mathbf{W}$ represents the trainable weight matrix, while $ReLU(\cdot)$ denotes the activation function.

(3) Residual connection and feedforward networks

During the training of deep learning neural networks, particularly those with numerous layers, the gradients of the loss function concerning the parameters may diminish significantly as they propagate backward through the layers. Consequently, the earlier layers' weights may cease to learn effectively due to their gradients approaching zero. To mitigate the vanishing gradient problem and enhance model performance, we introduce residual connections between each layer. This mechanism enables the input of a specific layer to bypass one or more layers, directly adding it to the subsequent layer's output, as demonstrated in Equation (4.14). Additionally, we integrate a fully connected feedforward network [42] into each encoder layer. This incorporation empowers the model to capture intricate input-output relationships and introduce nonlinearity, as depicted in Equation (4.15).

$$\mathbf{X}^{l+1} = layer(\mathbf{X}^l) + \mathbf{X}^l \tag{4.14}$$

$$FeedForward(\mathbf{X}) = ReLU(\mathbf{X}\mathbf{W}_0 + b_0)\mathbf{W}_1 + b_1 \tag{4.15}$$

where $layer(\mathbf{X}^l)$ represents the traffic flow embedding output of layer l, $\mathbf{X}^l$ denotes the traffic flow embedding input of layer l, and $\mathbf{X}^{l+1}$ signifies the traffic flow output after the residual connection. The function $layer(\cdot)$ denotes the operation responsible for capturing temporal or spatial information.

(4) Multi-step traffic prediction

In this study, we adopt a multi-step traffic prediction approach, which offers a holistic perspective on future traffic conditions. Additionally, within our model architecture, we include a fully connected layer in the reference section. This layer is designed to forecast traffic flow across multiple future time steps simultaneously, preventing the accumulation of prediction errors from prior time steps.

4.4.4 Bicycle multi-source adaptive transfer learning approach

Multi-source data can help overcome the scarcity of bicycle data by providing a broader and more comprehensive dataset. Carefully selecting data from multiple sources can help mitigate the risk of negative transfer, ensuring that the model benefits from relevant and diverse information.

(1) Adaptive transfer learning approach

In this context, we propose an adaptive transfer learning approach. First, each cluster source model is trained separately on the target domain dataset. In the loss function as shown in Equation (4.16), we adaptively aggregate the contributions from all cluster-based transfer models using learnable weighted parameters to fine-tune the models. This method allows the model to adapt to the unique characteristics of the target domain while effectively utilizing source domain knowledge, minimizing the risk of negative transfer. The formula for combining the prediction results is provided in Equation (4.17).

$$\mathcal{L}_{\text{loss}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{4.16}$$

where N is the number of samples; y_i represents the traffic flow true value of the i-th sample; $\hat{y}_i$ represents the traffic flow predicted value of the i-th sample.

$$\hat{y}_i = \sum_{i=j}^{n} \alpha_j \cdot \hat{y}_{i,j} \tag{4.17}$$

where $\hat{y}_{i,j}$ represents the fine-tuned traffic flow prediction result of the j-th source model for the i-th sample on the target dataset; α_j denotes the learnable weight assigned to the j-th source model, which is initialized randomly using a standard normal distribution; n is the total number of cluster-based transfer models.

Algorithm 2: The Multi-TLSTGCN Model

Input: Bicycle traffic features $\mathbf{X}$ over T_k historical timesteps; Number of training epochs E_r; learning rate η; batch size B_z; Number of prediction steps T; prediction output $\hat{\mathbf{X}}$; Learnable weights α_j; Optimizer: Adam.
Output: Trained Multi-TLSTGCN model.
Source Domain Training:;
Cluster the source bicycle traffic patterns into k_c clusters;
Reconstruct virtual spatial correlations within each cluster;
foreach *cluster in source domain* **do**
    foreach *epoch* $r \in \{1, 2, \ldots, E_r\}$ **do**
        Shuffle the training data;
        foreach *batch in training data* **do**
            $\hat{\mathbf{X}}_s \leftarrow f_\theta(\mathbf{X}_s)$;
        end
    end
end
Target Domain Adaptation:;
Load pretrained source models;
foreach *epoch* $r \in \{1, 2, \ldots, E_r\}$ **do**
    Shuffle the training data;
    foreach *batch in training data* **do**
        foreach *pretrained source model* **do**
            $\hat{\mathbf{X}}_r \leftarrow f_\theta(\mathbf{X}_r)$;
        end
        Compute $\hat{\mathbf{X}}$ using Equation (17);
        Update learnable parameters by minimizing the loss in Equation (16) using Adam;
    end
end

(2) Transfer learning process

The model training process comprises two primary phases. Initially, the first phase occurs predominantly within the source domains, focusing on acquiring knowledge from the source domains. Subsequently, the second phase involves transferring the acquired knowledge from the source domains to the target domains for accurate traffic prediction. The overall workflow of Multi-TLSTGCN is outlined in Algorithm 2.

Source domain training In the pre-training phase of multi-source transfer learning, sensors from various source cities are grouped into distinct clusters, each representing a different source domain, denoted as $\mathcal{S}_i$. Separate models are trained on the data within each of these source domains. Once trained, these source models are transferred to the target domain for subsequent fine-tuning.

Target domain training After acquiring pre-trained models from multiple source domains, we fine-tune them using the target dataset. An adaptive transfer learning approach is then applied to combine the outputs from these fine-tuned models, producing the final prediction results, as described in Equation (4.17).

4.5 Experiments

4.5.1 Dataset description

In this section, we first introduce the datasets used in both the source and target domains for our analysis. We then detail the cluster patterns identified within the source datasets using the K-means algorithm, which formed the basis for our subsequent modeling.

(1) Overall dataset description

To evaluate the effectiveness of our proposed model in predicting bicycle traffic with scarce datasets, we conducted experiments using bicycle datasets from six cities in the Netherlands: Delft, The Hague, Rotterdam, Leiden, Dordrecht, and Gouda. Notably, the Delft dataset was collected by smart cameras located on the TU Delft campus, while the datasets for the other five cities were sourced from publicly accessible bicycle traffic flow data provided by the National Road Traffic Data Portal of the Netherlands, which were collected using loop detector sensors. We utilized the datasets from The Hague, Rotterdam, Leiden, and Dordrecht as source datasets. These datasets include data from 18, 21, 21, and 17 sensors respectively, recorded between January 1, 2022, and December 31, 2022. To fine-tuning the target model, we conducted experiments separately using two datasets from Delft and Gouda. Both target datasets include data from 10 sensors, recorded between April 1, 2022, and May 31, 2022. Additionally, all datasets contain geographic information regarding the sensor locations.

The bicycle traffic flow datasets are aggregated into 5-minute intervals, providing a balance that smooths out fluctuations while still capturing significant traffic patterns. This results in 12 data points per hour, allowing for detailed analysis. To improve model convergence and training stability, we apply Z-score standardization to normalize the input datasets. Our goal is to perform multi-step bicycle traffic prediction by using one hour of historical data to forecast traffic flow for the next 60, 30, and 15 minutes. All datasets are sequentially divided into training, validation, and test sets, following a 6:2:2 split ratio, respectively.

(2) Clustering source datasets

To divide the source bicycle traffic flow data into distinct clusters, we begin by analyzing the overall average daily traffic patterns for each source sensor. As shown in **Fig. 4.2**, the morning peak traffic patterns are clearly pronounced across these sensors, while the evening peak appears more dispersed. Additionally, distinct traffic patterns emerge throughout the day. For example, some sensors consistently record high traffic volumes throughout the day without displaying significant peak hours, whereas other sensors show little to no traffic throughout the entire day. Next, we design a 1D convolutional neural network (CNN) to capture the unique traffic patterns of each sensor, which are then used as the initial features for K-means clustering. To determine the optimal number of clusters, we apply K-means clustering across a range of 2 to 9 clusters, calculating the sum of squared errors (SSE) for each scenario and performing a silhouette score analysis. As depicted in **Fig. 4.3**, the rate of SSE decline significantly slows at 4 clusters, and the silhouette score is also high at this point. This suggests that clustering the traffic patterns into 4 distinct groups is the most effective choice.

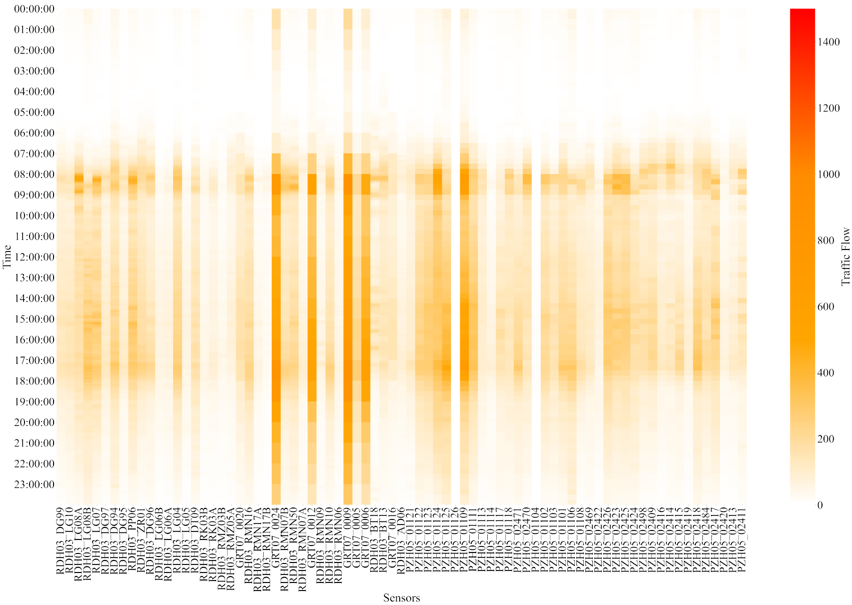


Figure 4.2 Aggregated daily average bicycle traffic patterns in five-minute intervals.

Based on the above analysis, we grouped sensors from all source cities into four distinct clusters using the K-means algorithm, as illustrated in **Fig. 4.4**. Specifically, Cluster0 and Cluster3 exhibit similar peak traffic in the morning but differ in their afternoon peak patterns. In contrast, Cluster1 shows more pronounced morning peak hours compared to Cluster0 and Cluster3. Cluster2, however, stands out with significantly higher peak traffic, and unlike the other clusters, its morning peak begins slightly later.

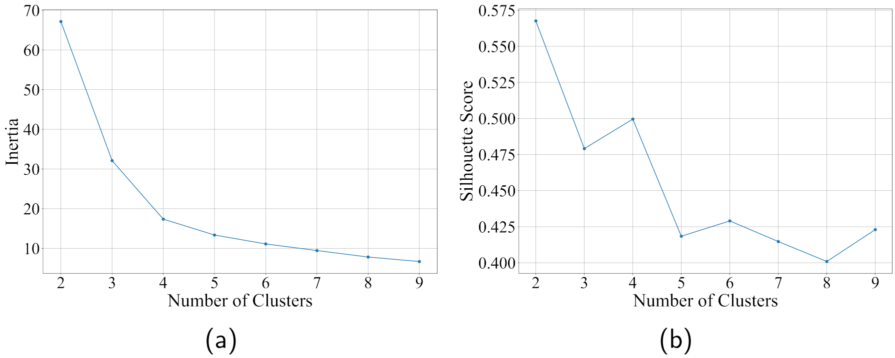


Figure 4.3 The optimal number of clusters. (a)The sum of squared errors (SSE), the rate of SSE decline significantly slows at 4 clusters. (b)Silhouette score, the 4-cluster configuration achieved the second-highest score.

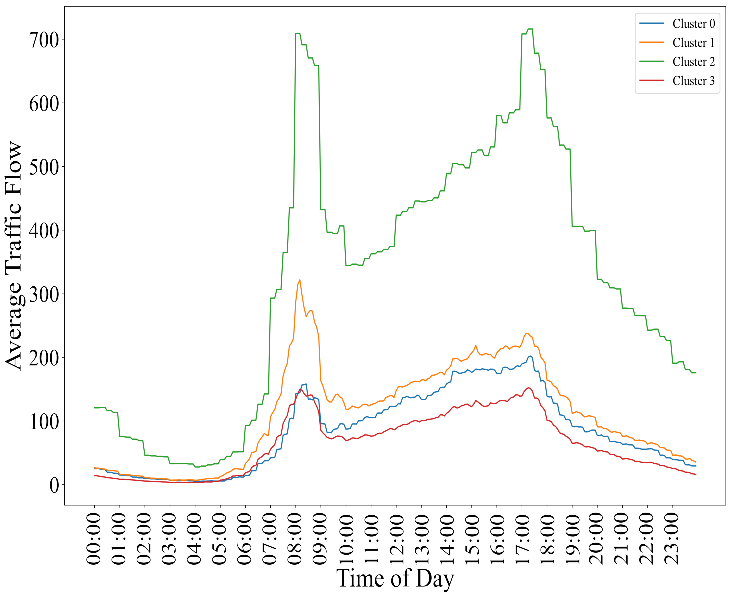


Figure 4.4 Cluster result. Clustering based on aggregated daily average bicycle traffic patterns.

4.5.2 Evaluation metrics

In this paper, the prediction results of bicycle traffic flow are evaluated by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Weighted Absolute Percentage Error (WAPE). The formulations to calculate these metrics are shown below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y_i} - y_i| \tag{4.18}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2} \tag{4.19}$$

$$WAPE = \frac{\sum_{i=1}^{n}|\hat{y_i} - y_i|}{\sum_{i=1}^{n}|y_i|}, \tag{4.20}$$

$$RP_{WAPE} = \frac{WAPE_{source-target} - WAPE_{target-only}}{WAPE_{target-only}}, \tag{4.21}$$

where $\hat{y_i}$ and y_i represent the predicted value and ground truth data of the sample i, respectively. n is the number of sample values.

4.5.3 Baseline models

To assess the performance of our proposed Multi-TLSTGCN model in predicting bicycle traffic flow, we compare it with several widely recognized traffic prediction models, from simple historical and statistical models (HA, ARIMA) to advanced machine learning and deep learning techniques commonly applied in traffic prediction (SVR, LSTM, STGCN and ASTGCN). Through this comparison, we aim to highlight Multi-TLSTGCN's capability in forecasting bicycle traffic, particularly in situations where the data is scarce or noisy. The baseline models used for comparison are as follows.

- HA: The Historical Average (HA) model predicts traffic flow by taking the average value of historical data.
- SVR: Support Vector Regression (SVR) [44] is a regression algorithm that aims to minimize the discrepancy between the predicted value and a predefined margin, which is an extension of Support Vector Machines (SVM) for regression tasks.
- LSTM: Long Short-Term Memory (LSTM) [45] is a specialized variant of Recurrent Neural Networks (RNNs) specifically designed to effectively capture and model long-term dependencies in sequential data.

- STGCN: Spatial-Temporal Graph Convolutional Network (STGCN) [46] is proposed to tackle the time series traffic prediction problem by harnessing comprehensive spatial-temporal correlations.
- ASTGNN: The attention-based spatial-temporal graph neural network (ASTGNN) [47] is designed to capture the dynamics of traffic data across both temporal and spatial dimensions.
- TransGTR: TransGTR [48] is a transferable traffic forecasting framework that jointly learns and transfers graph structures and forecasting models across cities. It comprises a node feature network trained with knowledge distillation for city-agnostic feature extraction, a structure generator with temporal decoupled regularization to ensure spatial features share similar distributions across cities, enabling effective knowledge transfer, and a forecasting model.

4.5.4 Experimental settings

The experiments were conducted using Google Colab, a cloud-based Python environment. The computing environment included a Tesla L4 GPU with a CUDA version of 12.0. The CPU used was an Intel(R) Core(TM) i9-9900KS clocked at 4 GHz. We implemented all the deep learning models using the PyTorch framework in Python. The models were optimized using the Adam optimizer. The hyperparameters for all deep learning models were carefully tuned through a validation set. Specifically, for the Multi-TLSTGCN model, we set the model dimension d_{model} to 64, the number of attention heads h to 8, the convolution kernel size to 3, The learning rate is reduced by a factor of 0.1 from 0.0001 to 0.00001, and batch size is 64. To prevent overfitting and ensure optimal model performance, we implemented an early stopping mechanism during the training process. Specifically, we used an early stopping criterion with a patience value of 20. This means that if the performance of the model on the validation set does not improve by at least 0.000001 for consecutive 20 epochs, the training process is halted.

Table 4.1 HYPERPARAMETERS OF THE MULTI-TLSTGCN

Hyperparameters	**Values**
batch size	64
d_{model}	64
kernel size	3
initial learning rate	0.0001
patience value	20
attention heads	8

4.5.5 Experimental results and discussion

In this section, we first evaluate the performance of single-city and cluster-based source transfer learning for predicting bicycle traffic flow in a target city. To assess the effectiveness of multi-source transfer learning in enhancing prediction accuracy, we conduct an experiment using the proposed Multi-TLSTGCN model. The performance of this model is compared against widely-used traffic prediction models to demonstrate that multi-source transfer learning is a promising approach for bicycle traffic prediction, particularly when dealing with scarce bicycle traffic data.

(1) Single-source transfer results

As all datasets are sourced from cities across the Netherlands, traffic patterns such as peak and off-peak hours are generally consistent across these locations, with the exception of Delft. This consistency supports a balanced assessment of the model's effectiveness. Delft, however, presents a unique case due to its educational land use and campus-centered sensor distribution, resulting in multiple traffic peaks that correspond to the academic calendar. These characteristics make Delft's traffic patterns distinct from those of other cities in the dataset. To thoroughly evaluate the robustness of our approach, we include Delft as a target city in our analysis.

Table 4.2 PERFORMANCE OF TRAFFIC PREDICTION USING SINGLE-CITY SOURCE TRANSFER

Cities	15 min			30 min			60 min		
	MAE	RMSE	WAPE	MAE	RMSE	WAPE	MAE	RMSE	WAPE
Delft (Target)	**0.90**	**2.10**	**34.68%**	**0.98**	**2.28**	**37.82%**	**1.09**	**2.56**	**42.03%**
The Hague–Delft	0.87	2.04	33.56%	0.95	2.23	36.35%	1.04	2.44	39.91%
Rotterdam–Delft	0.87	2.00	33.56%	0.97	2.24	37.18%	1.10	2.58	42.39%
Leiden–Delft	0.87	2.10	33.36%	0.94	2.27	36.15%	1.03	2.48	39.56%
Dordrecht–Delft	0.86	1.93	33.04%	0.94	2.13	36.01%	1.04	2.40	40.14%
Gouda(Target)	**17.37**	**39.23**	**21.87%**	**21.16**	**44.99**	**26.65%**	**24.41**	**50.19**	**30.76%**
The Hague–Gouda	17.12	35.46	21.54%	20.22	40.33	25.45%	23.08	45.69	29.09%
Rotterdam–Gouda	17.76	38.59	22.36%	21.60	44.80	27.20%	25.05	50.51	31.56%
Leiden–Gouda	15.41	33.17	19.39%	19.19	38.60	24.16%	22.39	44.73	28.22%
Dordrecht–Gouda	16.53	36.49	20.81%	20.63	43.06	25.97%	24.16	49.15	30.45%

Single-city transfer To evaluate the effectiveness of knowledge transfer from individual source cities to the target city and to investigate the influence of source-specific features on transfer performance, we first implement single-source transfer learning for each target city independently. This approach involves training models on datasets from individual source cities to capture their respective traffic patterns and subsequently fine-tuning these models using the target city's dataset. The goal is to improve the accuracy of bicycle traffic flow predictions for the target city. As shown in **Table 4.2**, using The Hague, Leiden, and Dordrecht as source cities improves the prediction accuracy for Delft, with Leiden achieving the highest performance 5.88% reduction in WAPE compared to the baseline. In contrast, employing Rotterdam as a source city adversely impacts the accuracy of predictions over a 60-minute horizon with 0.86% degradation on WAPE, indicating a negative transfer effect. This suggests that Rotterdam's traffic patterns diverge significantly from those in Delft, particularly over longer time frames. This discrepancy may be due to Rotterdam's larger and more complex bicycle traffic network, which complicates long-term forecasting.

For the target city Gouda, as indicated in **Table 4.2**, The Hague, Leiden, and Dordrecht contribute to improved prediction accuracy for bicycle traffic flow. Notably, Leiden achieves the highest performance among the source cities, with a 8.26% reduction in WAPE compared to using Gouda's own dataset alone. This improvement may be attributed to the similarity in sensor distribution between Leiden and Gouda, where sensors are predominantly located on ring roads around the city, resulting in comparable traffic patterns. In contrast, using Rotterdam as a source leads to a 2.60% increase in WAPE, which is worse than using Gouda's own dataset. This negative transfer effect suggests that while general traffic patterns may appear similar, variations in traffic distribution and sensorspecific patterns across cities can hinder effective knowledge transfer. These findings highlight the importance of selecting source cities with compatible sensor layouts and traffic characteristics to ensure successful transfer learning.

Table 4.3 PERFORMANCE OF TRAFFIC PREDICTION USING SINGLE-CLUSTER SOURCE TRANSFER

Cities	**15 min**			**30 min**			**60 min**		
	MAE	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**
Delft (Target)	**0.90**	**2.10**	**34.68%**	**0.98**	**2.28**	**37.82%**	**1.09**	**2.56**	**42.03%**
Cluster1–Delft	0.85	1.96	32.57%	0.92	2.12	35.18%	1.01	2.33	38.76%
Cluster2–Delft	0.85	1.99	32.71%	0.92	2.15	35.34%	1.02	2.37	39.11%
Cluster3–Delft	0.85	1.93	32.61%	0.92	2.12	35.45%	1.01	2.34	38.99%
Cluster4–Delft	0.86	1.92	32.84%	0.93	2.11	35.76%	1.03	2.36	39.83%
Gouda(Target)	**17.37**	**39.23**	**21.87%**	**21.16**	**44.99**	**26.65%**	**24.41**	**50.19**	**30.76%**
Cluster1–Gouda	15.83	34.30	19.92%	19.54	39.81	24.60%	22.70	45.62	28.61%
Cluster2–Gouda	15.91	33.81	20.02%	19.71	39.37	24.81%	22.93	45.44	28.89%
Cluster3–Gouda	15.64	34.37	19.69%	19.52	39.69	24.58%	22.79	45.71	28.71%
Cluster4–Gouda	16.17	34.77	20.35%	20.23	41.20	25.47%	23.61	47.31	29.75%

Single-cluster transfer To mitigate the risk of negative transfer in bicycle traffic prediction, it is crucial to account for the sparse distribution of sensors and the complexity of bicycle road networks across different cities. Unlike motorized traffic, which follows a more structured and hierarchical road system, bicycle networks tend to be more intricate and irregular. Directly transferring spatial-temporal traffic data to the target city may lead to negative transfer due to these structural differences. In this study, we cluster sensors from all source cities to identify traffic patterns in the source domains that closely resemble those in the target city. This approach makes knowledge transfer more relevant and effective, reducing the likelihood of negative transfer and ensuring that the transferred information is beneficial. We then transfer information from each single-source cluster individually to improve traffic predictions in the target city. The results in **Table 4.3** show that each single-source cluster positively contributes to bicycle traffic predictions in the target cities.

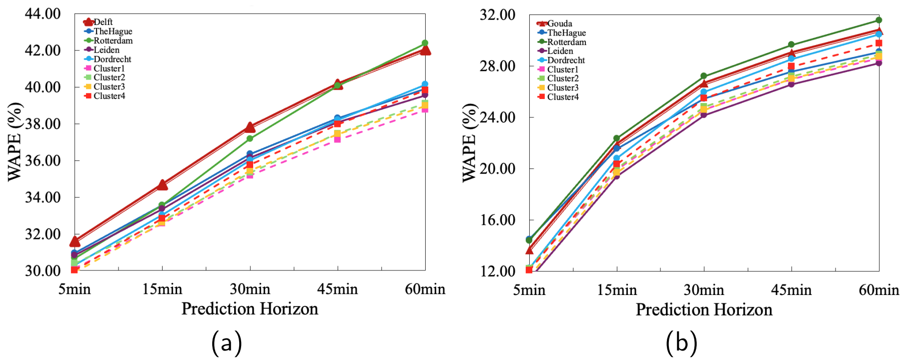


Figure 4.5 Performance comparison between single-city source and single-cluster source. (a) WAPE(Delft). (b) WAPE(Gouda).

Comparison of single-city and single-cluster source transfer Fig. 4.5 demonstrates that single-city source transfer learning generally performs worse compared to single-cluster source transfer learning. For example, in the Delft prediction task, the WAPE difference between the best single-city transfer (Dordrecht→Delft, 39.91%) and the best cluster-based transfer (Cluster-based, 38.76%) is 1.15 percentage points, this corresponds to a 2.88% relative improvement, which is meaningful given the inherent noise and variability in traffic data. While certain individual source-target city pairs (e.g., Leiden→Gouda) may exceed cluster-based performance in specific cases, cluster-based transfer demonstrates more consistent average performance and offers reduced risk of negative transfer. This makes it especially advantageous in scenarios where prior knowledge of structural or behavioral similarities between cities is unavailable.

Single-city transfer relies heavily on the alignment between the source and target cities' road networks and sensor distributions. When well-matched, it can yield high accuracy. However, mismatches, such as complex or dissimilar traffic patterns, can introduce noise, leading to negative transfer and reduced predictive performance.

In contrast, single-cluster transfer draws from more focused and homogeneous subsets of traffic patterns. This helps maintain alignment with the target city's dominant characteristics, mitigates the introduction of irrelevant or conflicting information, and supports more stable performance overall. Given the sparse and complex nature of bicycle traffic data, cluster-based transfer learning emerges as a more robust and scalable approach for active mode traffic prediction.

(2) Multi-source transfer learning results

Table 4.4 PERFORMANCE COMPARISON OF MULTI-STEP BICYCLE TRAFFIC PREDICTION

Target	**Baselines**	**15 min**			**30 min**			**60 min**		
		MAE	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**	**MAE**	**RMSE**	**WAPE**
Delft	HA	3.29	5.22	104.96%	3.29	5.22	104.96%	3.29	5.22	104.96%
	SVR	1.07	2.41	41.03%	1.22	2.65	47.03%	1.56	3.19	60.22%
	LSTM	1.01	2.31	38.92%	1.10	2.54	42.34%	1.23	2.86	47.31%
	STGCN	0.93	2.08	35.66%	1.00	2.31	38.51%	1.10	2.61	42.37%
	ASTGNN	0.86	2.04	32.94%	0.93	2.25	35.86%	1.03	2.49	39.55%
	Multi-TLSTGCN	**0.82**	**1.85**	**31.56%**	**0.89**	**2.03**	**34.20%**	**0.97**	**2.24**	**37.52%**
Gouda	HA	69.83	104.51	76.91%	69.82	104.50	76.94%	69.81	104.49	76.99%
	SVR	24.25	51.73	30.52%	26.48	54.51	33.34%	29.44	59.20	37.10%
	LSTM	18.15	39.75	22.84%	22.77	47.35	28.67%	27.29	55.43	34.39%
	STGCN	18.01	37.27	22.66%	20.98	42.56	26.42%	23.49	47.38	29.60%
	ASTGNN	16.38	35.53	20.61%	19.87	40.18	25.01%	22.99	45.73	28.97%
	Multi-TLSTGCN	**15.21**	**32.88**	**19.15%**	**19.17**	**38.60**	**24.14%**	**22.50**	**44.82**	**28.35%**

In this study, we clustered bicycle data from The Hague, Rotterdam, Leiden, and Dordrecht into different groups to serve as sources supporting predictions in Delft and Gouda. The results, presented in **Table 4.4**, demonstrate that the proposed Multi-TLSTGCN significantly outperforms other baselines and the single-cluster source model. This superior performance highlights the effectiveness of incorporating knowledge from multiple cities, making it a promising approach for improving bicycle traffic prediction. In comparison, the baseline HA model exhibits poor performance compared to other models. This is because HA relies on predicting traffic states by averaging past observations, which fails to capture the inherent fluctuations and variability of bicycle traffic. SVR and LSTM can effectively capture both linear and nonlinear patterns, making them well-suited for the complexities of bicycle traffic flow data. However, their focus on temporal features alone limits their performance, as spatial correlations among sensors in the bicycle road network are also crucial. STGCN and ASTGNN address this by using graph convolutional neural networks to capture spatial correlations of bicycle traffic, thereby improving prediction accuracy. Nevertheless, these deep learning models require large datasets to capture features accurately, posing a significant challenge for bicycle traffic prediction due to the inherent data scarcity. We also compared the performance of Multi-TLSTGCN with the state-of-the-art transfer learning method TransGTR. Although TransGTR incorporates a node feature extractor and a structure generator to capture city-agnostic features, it still struggles with bicycle traffic prediction—particularly when Delft is the target city, where bicycle flow is strongly influenced by academic schedules, which are difficult to generalize across cities.

(3) Multi-source transfer learning with different levels of target data scarcity

To assess the performance of the proposed Multi-TLSTGCN model in real-world situations with varying levels of bicycle flow data availability, we designed a series of diverse scenarios for traffic prediction of target cities, each with different levels of bicycle target dataset scarcity. By reducing the available target training dataset from 90% down to 0%, we aim to evaluate the model's robustness and generalization across different level of data-scarce environments.

For each target city, we conduct two scenarios: the first involves training the model under varying levels of data scarcity in the target city without leveraging knowledge from source models, while the second trains the model using Multi-TLSTGCN, benefiting from the transfer knowledge acquired from source models.

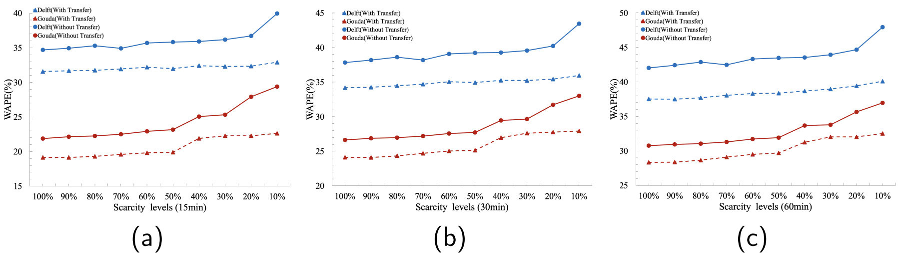


Figure 4.6 Performance across varying levels of target data sparsity.

As shown in **Fig. 4.6**, the scenarios without transfer knowledge exhibit increasing error rates as the amount of training data decreases. In Gouda, the error begins to rise dramatically when only 50% of the training dataset is included. For Delft, the error steadily increases from the full 100% training dataset down to 20%; however, once the training dataset falls below 10%, the error escalates sharply. In contrast, the scenarios utilizing Multi-TLSTGCN, which incorporates transfer knowledge, show a more gradual increase in error. For Gouda, a noticeable increase occurs when only 50% of the training dataset is utilized, but the error stabilizes when the dataset is reduced to 30%, in comparison to scenarios without transfer knowledge. For Delft, throughout the various levels of data scarcity, the error only slightly increases as the training dataset decreases. The WAPE values differ across cities Gouda and Delft, primarily due to variations in traffic flow stability, sensor coverage, and urban layout.

These findings suggest that Multi-TLSTGCN effectively transfers valuable information to the target domain, enabling accurate traffic predictions even when target data is limited. This underscores the model's robustness in supporting bicycle traffic prediction under data-scarce conditions.

Our assessment further indicates that the proposed model can be applied across a variety of real-world scenarios. As shown in Figure X, the model maintains strong performance, with only a slight degradation as the size of the training dataset decreases. This suggests that the approach is robust even under limited data conditions.

For practical deployment, the amount of training data required will depend on two key factors: (i) the availability of local data and (ii) the level of predictive accuracy desired. These two aspects must be balanced in real-world applications. In data-rich environments, larger datasets will naturally support higher accuracy, while in data-scarce contexts, the model can still be operationalized with smaller datasets, but with somewhat reduced precision.

4.5.6 Ablation study

To evaluate the impact of virtual spatial correlations from each cluster-based transfer model on traffic prediction in the target domain and to assess the effectiveness of the adaptive transfer learning approach in multi-source transfer learning, we conducted ablation experiments. These experiments were carried out using the same bicycle datasets previously described. Finally, we perform a comprehensive hyperparameter analysis to demonstrate the rationale behind our hyperparameter selection process.

(1) The impact of virtual spatial correlations

Given the complexity of bicycle road networks and their variability across different cities, we grouped sensors from all cities into distinct clusters based on traffic patterns, disregarding geographical spatial correlations. However, even within the same cluster, traffic patterns among sensors can vary. To account for these differences and capture potential spatial correlations within each cluster, we established a virtual spatial correlation among sensors. To evaluate whether this virtual correlation effectively captures useful spatial information from source data to enhance traffic prediction in target cities, we conducted ablation experiments.

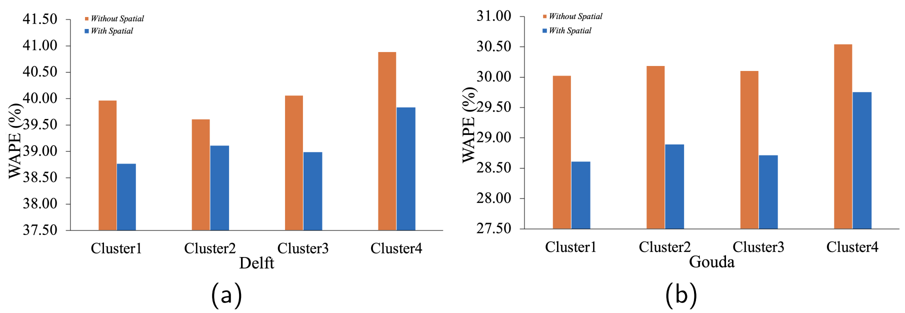


Figure 4.7 Virtual spatial correlations ablation study. (a) WAPE(Delft). (b) WAPE(Gouda).

As shown in **Fig. 4.7**, we evaluated the impact on target city. Generally, source models that incorporated virtual spatial correlations outperformed those without them in enhancing traffic predictions for the target city. However, in Delft, the inclusion of virtual spatial dependencies did not significantly improve performance. This can be attributed to Delft's limited sensor network coverage, where traffic patterns across different locations are relatively similar, leading the model to rely more on temporal correlations rather than spatial ones. In contrast, Gouda's sensors are spread across a larger, more complex network, and the results indicate that leveraging virtual spatial information significantly improves traffic prediction accuracy in Gouda.

(2) The effectiveness of the multi-source adaptive transfer learning approach

To effectively transfer knowledge from multiple data sources, we propose a multi-source adaptive transfer learning approach that combines prediction outputs from various source models using target domain data. As shown in **Fig. 4.8**, our method significantly outperforms a basic aggregation approach for all cluster-based models and city-based models, which merely averages the parameters of all source models. These results underscore the strength of our adaptive transfer learning strategy in harnessing diverse knowledge sources to enhance prediction accuracy for the target city.

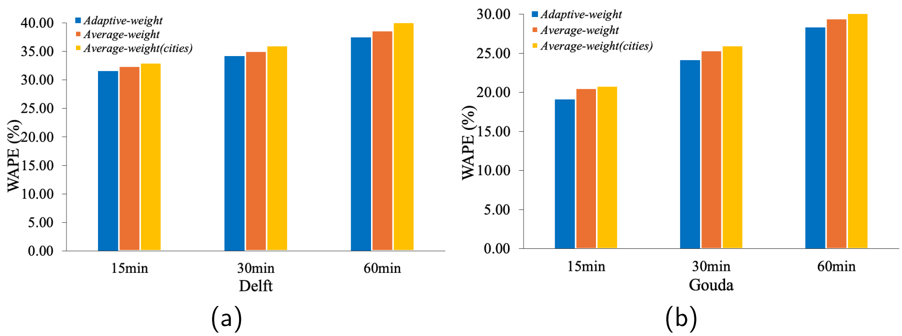


Figure 4.8 Multi-source transfer learning approach ablation study. (a) WAPE(Delft). (b) WAPE(Gouda).

(3) Hyperparameter analysis

Hyperparameters play a crucial role in controlling the training dynamics of deep learning models, including the proposed Multi-TLSTGCN. As detailed in the details in Section 4.5.4, all hyperparameters were carefully tuned using a validation set to ensure optimal model performance and generalization. We analyze the model's sensitivity using two key hyperparameters as examples: batch size and learning rate.

The impact of batch size and learning rate on the WAPE is visualized in **Fig. 4.9**.

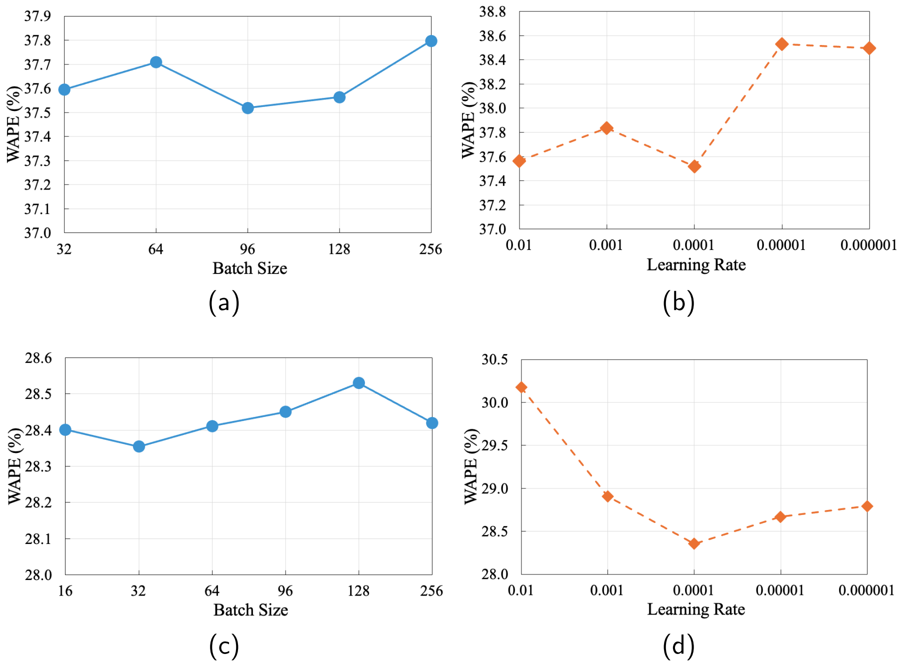


Figure 4.9 Hyperparameter analysis. (a) Batch size (Delft). (b) Learning rate (Delft). (c) Batch size (Gouda). (d) Learning rate (Gouda).

The results demonstrate that the optimal configuration for the source city of Delft is achieved with a batch size of 96 and a learning rate of 0.0001. In contrast, for the source city of Gouda, the model achieves its best performance with a batch size of 32 and the same learning rate of 0.0001.

4.6 Conclusion

Given the scarcity of bicycle data and the significant requirements for accurate traffic prediction, we propose the Multi-TLSTGCN model, which leverages datasets from other cities through transfer learning to enhance prediction accuracy in the target city. The proposed process begins by clustering sensors from different source cities into distinct groups based on their traffic patterns. Next, we reconstruct the spatial correlations within each group based on the similarity of traffic patterns among sensors. These pre-trained source group models are then fine-tuned using the target city's dataset. Finally, adaptive transfer learning approach is used to aggregate the predictions from these multiple source models, optimizing the traffic predictions for the target city.

In this study, we conducted three key experiments to evaluate the feasibility and robustness of our model in predicting bicycle traffic under data scarcity conditions. First, we perform single-city and single-cluster experiments to explore how traffic knowledge transfer can avoid negative transfer. Second, we implement a multi-source transfer learning experiment using Multi-TLSTGCN for traffic prediction, assessing the model's performance under data scarcity. The results demonstrate that our approach outperforms baselines trained solely on the target dataset. We evaluated the robustness of Multi-TLSTGCN by training the model under varying levels of data scarcity. Comparing this to the same model without transfer learning, we observe that Multi-TLSTGCN consistently outperforms its counterpart. As the amount of target training data decreases, the accuracy of the non-transfer learning model declines significantly, whereas Multi-TLSTGCN maintains its predictive performance, demonstrating its effectiveness in data-scarce environments.

Overall, this study demonstrates the potential for achieving accurate bicycle traffic prediction in cities with limited data collection devices or insufficient traffic data. While the Multi-TLSTGCN framework is designed to be data-driven and flexible, its generalizability to cities with less developed cycling infrastructure may be influenced by variations in sensor availability, traffic dynamics, and urban topology. Nevertheless, the clustering-based transfer learning strategy and the virtual spatial correlation module are model-agnostic, offering adaptability to datasets from cities outside the Netherlands. Moreover, the proposed approach is designed for typical operating scenarios and is not specifically tailored to conditions such as extreme weather, irregular events, unique urban topologies, or distinct policy interventions. Exploring these challenging scenarios could be an important work for future research. In addition, this paper does not incorporate time-related factors such as day of the week, seasonality, or weather conditions. These factors could be integrated in real-world deployments and have the potential to further improve prediction accuracy. For deployment purposes, the real-time computational cost is primarily associated with the target model, which is initialized using the pretrained source model. Since the pretrained model can be generated offline and reused across tasks, the runtime cost during deployment remains modest and practical for real-world applications. In future work, we plan to evaluate the model on data from cities with diverse cycling cultures and infrastructure characteristics to further validate its robustness and applicability. Additionally, we aim to enhance the framework by developing synthetic data generation techniques tailored for regions with extremely sparse historical records. These enhancements will expand the model's utility, offering effective solutions for traffic prediction in a wide range of urban settings challenged by data scarcity and variability.

Acknowledgements

This research is supported by the China Scholarship Council. This support is gratefully acknowledged.

References

[1] Y. Jia, J. Wu, and M. Xu. “Traffic flow prediction with rainfall impact using a deep learning method”. In: *Journal of advanced transportation* 2017.1 (2017), p. 6575947.

[2] M. C. Poelman. “Predictive Traffic Signal Control under Uncertainty: Analyzing and Reducing the Impact of Prediction Errors”. In: (2024).

[3] S. Yu, Y. Li, G. Sheng, and J. Lv. “Research on short-term traffic flow forecasting based on KNN and discrete event simulation”. In: *Advanced Data Mining and Applications: 15th International Conference, ADMA 2019, Dalian, China, November 21–23, 2019, Proceedings 15*. Springer. 2019, pp. 853–862.

[4] B. Zhao, K. Kumar, G. Casey, and K. Soga. “Agent-based model (ABM) for city-scale traffic simulation: A case study on San Francisco”. In: *International Conference on Smart Infrastructure and Construction 2019 (ICSIC) Driving data-informed decision-making*. ICE Publishing. 2019, pp. 203–212.

[5] I. Alam, D. M. Farid, and R. J. Rossetti. “The prediction of traffic flow with regression analysis”. In: *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 2*. Springer. 2019, pp. 661–671.

[6] C. Xu, Z. Li, and W. Wang. “Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming”. In: *Transport* 31.3 (2016), pp. 343–358.

[7] S. V. Kumar. “Traffic flow prediction using Kalman filtering technique”. In: *Procedia Engineering* 187 (2017), pp. 582–587.

[8] J. M. Keller, M. R. Gray, and J. A. Givens. “A fuzzy k-nearest neighbor algorithm”. In: *IEEE transactions on systems, man, and cybernetics* 4 (1985), pp. 580–585.

[9] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. “Support vector machines”. In: *IEEE Intelligent Systems and their applications* 13.4 (1998), pp. 18–28.

[10] J. Zou, Y. Han, and S.-S. So. “Overview of artificial neural networks”. In: *Artificial neural networks: methods and applications* (2009), pp. 14–22.

[11] L. R. Medsker, L. Jain, *et al.* “Recurrent neural networks”. In: *Design and applications* 5.64-67 (2001), p. 2.

[12] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, *et al.* “Long short term memory networks for anomaly detection in time series”. In: *Proceedings*. Vol. 89. 9. 2015, p. 94.

[13] S. Hao, D.-H. Lee, and D. Zhao. “Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system”. In: *Transportation Research Part C: Emerging Technologies* 107 (2019), pp. 287–300.

[14] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. “The graph neural network model”. In: *IEEE transactions on neural networks* 20.1 (2008), pp. 61–80.

[15] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.* “Recent advances in convolutional neural networks”. In: *Pattern recognition* 77 (2018), pp. 354–377.

[16] M. Ge, H. Bangui, and B. Buhnova. “Big data for internet of things: a survey”. In: *Future generation computer systems* 87 (2018), pp. 601–614.

[17] K. Weiss, T. M. Khoshgoftaar, and D. Wang. “A survey of transfer learning”. In: *Journal of Big data* 3 (2016), pp. 1–40.

[18] J. Li, F. Guo, A. Sivakumar, Y. Dong, and R. Krishnan. “Transferability improvement in short-term traffic prediction using stacked LSTM network”. In: *Transportation Research Part C: Emerging Technologies* 124 (2021), p. 102977.

[19] Z. Fang, D. Wu, L. Pan, L. Chen, and Y. Gao. “When Transfer Learning Meets Cross-City Urban Flow Prediction: Spatio-Temporal Adaptation Matters.” In: *IJCAI*. Vol. 22. 2022, pp. 2030–2036.

[20] S. Wang, H. Miao, J. Li, and J. Cao. “Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.5 (2021), pp. 4695–4705.

[21] Y. Huang, X. Song, Y. Zhu, S. Zhang, and J. J. Yu. “Traffic prediction with transfer learning: A mutual information-based approach”. In: *IEEE Transactions on Intelligent Transportation Systems* 24.8 (2023), pp. 8236–8252.

[22] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. “A comprehensive survey on graph neural networks”. In: *IEEE transactions on neural networks and learning systems* 32.1 (2020), pp. 4–24.

[23] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin. “Deep learning on traffic prediction: Methods, analysis, and future directions”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2021), pp. 4927–4943.

[24] Y. Bao, J. Huang, Q. Shen, Y. Cao, W. Ding, Z. Shi, and Q. Shi. “Spatial–temporal complex graph convolution network for traffic flow prediction”. In: *Engineering Applications of Artificial Intelligence* 121 (2023), p. 106044.

[25] X. Shi, H. Qi, Y. Shen, G. Wu, and B. Yin. “A spatial–temporal attention approach for traffic prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* 22.8 (2020), pp. 4909–4918.

[26] A. Naghib, N. Jafari Navimipour, M. Hosseinzadeh, and A. Sharifi. “A comprehensive and systematic literature review on the big data management techniques in the internet of things”. In: *Wireless Networks* 29.3 (2023), pp. 1085–1144.

[27] Y. Xu, X. Cai, E. Wang, W. Liu, Y. Yang, and F. Yang. “Dynamic traffic correlations based spatio-temporal graph convolutional network for urban traffic prediction”. In: *Information Sciences* 621 (2023), pp. 580–595.

[28] E. Heinen, K. Maat, and B. Van Wee. “The effect of work-related factors on the bicycle commute mode choice in the Netherlands”. In: *Transportation* 40 (2013), pp. 23–43.

[29] M. El Esawey, A. I. Mosa, and K. Nasr. “Estimation of daily bicycle traffic volumes using sparse data”. In: *Computers, Environment and Urban Systems* 54 (2015), pp. 195–203.

[30] K. Kim. “Investigation of modal integration of bike-sharing and public transit in Seoul for the holders of 365-day passes”. In: *Journal of Transport Geography* 106 (2023), p. 103518.

[31] P. Deng, Y. Zhao, J. Liu, X. Jia, and M. Wang. “Spatio-temporal neural structural causal models for bike flow prediction”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 37. 4. 2023, pp. 4242–4249.

[32] Z. Yao, S. Xia, Y. Li, G. Wu, and L. Zuo. “Transfer learning with spatial–temporal graph convolutional network for traffic prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* 24.8 (2023), pp. 8592–8605.

[33] J. Li, C. Liao, S. Hu, X. Chen, and D.-H. Lee. “Physics-Guided Multi-Source Transfer Learning for Network-Scale Traffic Flow Prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* (2024).

[34] Y. Jin, K. Chen, and Q. Yang. “Selective cross-city transfer learning for traffic prediction via source city region re-weighting”. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2022, pp. 731–741.

[35] J. Mo and Z. Gong. “Cross-city multi-granular adaptive transfer learning for traffic flow prediction”. In: *IEEE Transactions on Knowledge and Data Engineering* 35.11 (2022), pp. 11246–11258.

[36] N. A. Ojekudo and N. P. Akpan. “Anapplication of Dijkstra’s Algorithm to shortest route problem”. In: *IOSR Journal of Mathematics (IOSR-JM)* 13.3 (2017), p. 14.

[37] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains”. In: *IEEE signal processing magazine* 30.3 (2013), pp. 83–98.

[38] S. Patro and K. K. Sahu. “Normalization: A preprocessing stage”. In: *arXiv preprint arXiv:1503.06462* (2015).

[39] K. Krishna and M. N. Murty. “Genetic K-means algorithm”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 29.3 (1999), pp. 433–439.

[40] P. Bholowalia and A. Kumar. “EBK-means: A clustering technique based on elbow method and k-means in WSN”. In: *International Journal of Computer Applications* 105.9 (2014).

[41] X. Ta, Z. Liu, X. Hu, L. Yu, L. Sun, and B. Du. “Adaptive spatio-temporal graph neural network for traffic forecasting”. In: *Knowledge-based systems* 242 (2022), p. 108199.

[42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[43] T. N. Kipf and M. Welling. “Semi-supervised classification with graph convolutional networks”. In: *arXiv preprint arXiv:1609.02907* (2016).

[44] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. “Support vector regression machines”. In: *Advances in neural information processing systems* 9 (1996).

[45] A. Graves and A. Graves. “Long short-term memory”. In: *Supervised sequence labelling with recurrent neural networks* (2012), pp. 37–45.

[46] B. Yu, H. Yin, and Z. Zhu. “Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting”. In: *arXiv preprint arXiv:1709.04875* (2017).

[47] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong. “Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting”. In: *IEEE Transactions on Knowledge and Data Engineering* 34.11 (2021), pp. 5415–5428.

[48] Y. Jin, K. Chen, and Q. Yang. “Transferable graph structure learning for graph-based traffic forecasting across cities”. In: *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. 2023, pp. 1032–1043.

Chapter 5

Federated Spatial-temporal Learning for Active Traffic Prediction

The previous chapter addressed the data scarcity challenge in active mode traffic systems by utilizing data from multiple sources. However, this approach introduced potential issues related to data storage and security. To tackle these limitations, this chapter presents a novel heterogeneous spatial-temporal graph neural network based on federated learning (FedHSTGCN). This model combines federated learning with a Graph Convolutional Network (GCN) to effectively capture heterogeneous spatial patterns and employs a self-attention mechanism to model temporal patterns, all while ensuring that raw data remains secure and localized. Additionally, the proposed federated learning mechanism enables distributed data sources to share spatial correlations, thereby improving prediction accuracy. The experimental results demonstrate that FedHSTGCN not only addresses data heterogeneity but also enhances traffic prediction performance, outperforming baseline models in bicycle traffic flow forecasting. This makes it a robust, secure, and efficient solution for distributed traffic prediction.

This chapter is submitted to a journal (under review, R1 revision): Wen, X., Khosla, M. and Hoogendoorn, S. Federated Learning-based Heterogeneous Spatial-Temporal Graph Neural Network for Active Mode Traffic Prediction. The content is the same as the paper.

5.1 Introduction

Intelligent transportation systems (ITS), vast amounts of traffic data are gathered from different sources, including traffic cameras, GPS devices, and sensors, to monitor and analyze traffic conditions across various domains in real-time. Traffic prediction, as a crucial component of ITS, can benefit from these data by creating accurate models using deep learning techniques such as graph neural networks (GNNs) [1] and recurrent neural networks (RNNs) [2]. However, these models typically require extensive, centralized datasets, necessitating data sharing across multiple organizations. This reliance on centralized data introduces challenges, particularly for active mode traffic (e.g., pedestrian and bicycle traffic), where data collection is inherently sparse and decentralized.

Despite growing interest in sustainable mobility, ITS infrastructure has traditionally been designed around motorized traffic, with limited investment in monitoring systems for active modes of transportation. As a result, existing sensor networks often fail to capture the complex and dynamic nature of pedestrian and bicycle movements. Moreover, most existing graph-based architectures are tailored to homogeneous graphs, which fail to account for skewed data distributions and the inherent heterogeneity of active mode traffic networks [3]. While some studies have addressed road-level heterogeneity for motorized traffic [4], these methods fall short in capturing the statistical heterogeneity inherent in active mode traffic, which lacks clear road-level classifications.

Aggregating data from multiple sources can help mitigate data sparsity and improve predictive accuracy. However, centralized data aggregation poses challenges: it demands substantial network bandwidth and storage, increases privacy risks, and raises competitive concerns when sensitive traffic data are shared. Thus, addressing data sparsity, heterogeneity, and privacy concerns is crucial for effective active mode traffic prediction.

Federated learning offers a promising solution framework by training models on decentralized devices, keeping data local and private [5]. Recent studies have demonstrated that federated learning can be effectively integrated with deep learning traffic prediction models, striking a balance between model performance and data privacy [6, 7]. However, most of these studies focus on exploring the application of the federated learning framework with temporal traffic prediction models but do not consider the spatial dependencies features of traffic data in their federated learning-based traffic prediction models [8, 9]. Nevertheless, transportation systems are complex networks that require a deep understanding of their intricate spatial relationships to accurately predict traffic patterns [10].

To address the issue of traffic prediction using sparse active mode traffic data while ensuring data security, this paper introduces the heterogeneous spatial-temporal graph neural network based on federated learning (FEDHSTGCN) model. The primary contributions of our work are as follows:

- We design a heterogeneous spatial-temporal graph neural network (HSTGCN) model to explore the influence of spatially statistical heterogeneous traffic patterns among sensors in an active mode traffic network on the prediction of bicycle traffic flow. Our proposed model demonstrates superior performance in accurately predicting bicycle flow, highlighting its effectiveness in capturing both temporal and spatial intricacies within the heterogeneous bicycle traffic network.

- We propose the federated learning global spatial aggregation mechanism (FedGSAM) aggregation mechanism for the federated learning-based HSTGCN active mode traffic prediction model to capture the spatial correlations of active mode traffic network among different data source clients. Analysis of the captured adjacency matrix indicates that the proposed aggregation mechanism effectively captures the global spatial relationships within the network.

- The proposed FEDHSTGCN model enables accurate active mode traffic prediction by training models across different clients with sparsity data distribution while maintaining the security of raw data. The performance of our model showcases its excellent capacity in balancing accurate traffic prediction and data security.

The subsequent section offers an overview of existing research on federated learning in traffic prediction. Section 3 provides detailed insights into the proposed FEDHSTGCN model. Section 4 introduces the experimental setup, dataset used, experiment results, discussions, and an ablation study. Finally, Section 5 presents the conclusion of this work.

5.2 Related Work

5.2.1 Spatial-temporal traffic prediction model

Spatial-temporal traffic prediction approaches excel in capturing complexities of urban traffic systems. These comprehensive approaches yields highly accurate prediction results, significantly enhancing decision-making processes in the realm of transportation and urban planning [11–15]. In temporal information capturing, traffic prediction models, such as RNNs [2, 16], multi-head self-attention [17], and temporal convolutional networks (TCNs) [18], aim to predict future traffic state based on historical data, considering the temporal aspects of the transportation system. Simultaneously, in spatial information capturing, the emergence of convolutional neural networks (CNNs) [19] and GNNs has greatly bolstered our ability to explore spatial relationships in traffic prediction. The integration of spatial and temporal information provides a holistic understanding of traffic dynamics, leading to more accurate and informed predictions of traffic state. Li et al. [20] investigated spatial-temporal features in traffic flow utilizing a graph and attention-based long short-term memory network (GLA). The findings indicate that this approach outperforms many previous methods, as demonstrated on the PeMS dataset. Guo et al. [21] introduced an attention based spatial-temporal graph neural network (ASTGNN) model designed to capture dynamic traffic patterns in a flexible manner. The ASTGNN model demonstrates superior long-term prediction accuracy compared to state-of-the-art traffic prediction methods. By consolidating insights from diverse studies, it becomes evident that the incorporation of both temporal and spatial traffic information is imperative for achieving accurate traffic predictions.

While spatial-temporal traffic prediction methods demonstrate well performance, the persistent issue of statistical heterogeneity, such as skewed dataset distribution among sensors within the traffic network, especially for active mode, remains a significant concern in traffic prediction. For instance, sensors located in urban centers or major roadways may collect substantial amounts of active mode traffic flow data, while those situated in rural or suburban areas may capture only minimal data. This disparity in active mode data collection leads to statistical heterogeneity issues among sensors in the active mode transportation network. To investigate the intrinsic spatial heterogeneity of urban traffic, Cheng et al. [22] introduced an adaptive spatial-temporal k-nearest neighbor model (adaptive-STKNN) for short-term traffic prediction. This model addresses the spatial heterogeneity of urban traffic by incorporating adaptive spatial neighbors, time windows, spatial-temporal weights, and other relevant parameters. Ku et al. [23] introduced an attention mechanism utilizing transfer entropy (TE) to quantify intricate and asymmetric spatial-temporal correlations within traffic data, facilitating their integration into a spatial-temporal graph neural network for training and prediction tasks. Ji et al. [24] proposed a novel framework called spatial-temporal self supervised learning (ST-SSL) for predicting traffic flow, taking into account both spatial and temporal heterogeneity. Existing studies primarily address heterogeneity within motorized traffic networks, overlooking the unique challenges posed by active mode traffic. Active mode traffic is highly sensitive to the surrounding environment, leading to a significantly skewed and heterogeneous distribution of traffic data across network sensors. Understanding these spatially heterogeneous patterns in active mode traffic remains complex due to the intricate nature of traffic systems and the challenges associated with collecting comprehensive data on external influencing factors.

5.2.2 Federated learning

To overcome the challenges of data privacy, potential raw data leakage, and the sparse distribution of active mode sensors in active mode traffic prediction, Federated learning, introduced by McMahan [5], has emerged as a promising paradigm. This approach allows raw data to be stored locally and only model updates are shared, minimizing the need for direct communication between clients and providing an effective data security mechanism [25]. Federated Learning is gaining increasing attention in the transportation field due to its ability to address data security concerns. For example, to obtain an accurate traffic prediction while preserving traffic data security, Liu et al. [6] proposed a federated learning-based gated recurrent unit neural network algorithm (FedGRU) designed for traffic flow prediction. Through case studies conducted on a real-world dataset, the results demonstrate that the proposed model achieves precise traffic predictions while keeping data security. Despite the successful applications of federated learning in the transportation domain, its adoption for traffic graph data remains limited. Effectively learning representations from complex graph-structured data is essential in transportation, as the spatial structure of graph data closely mirrors the complexities of real-world traffic networks [26]. In this context, federated learning approaches based on GNNs have gained traction for traffic prediction. Zhang et al. [27] introduced a clustering-based hierarchical and two-step-optimized federated learning framework(CTFL) to overcome the enormous communication overhead when using GNN-based models for traffic speed prediction tasks. The case studies demonstrate that the proposed frameworks work well in training efficiency and prediction accuracy. Yuan et al. [28] proposed a federated deep learning based on the spatial-temporal long and short-term networks (FedSTN) algorithm to predict traffic flow. The local traffic flow prediction model is deployed on an edge computing server, and an additive homomorphic encryption approach is implemented to ensure data security during information sharing. However, the studies mentioned above without considering the interconnection of networks between clients or without including the hidden correlations of the graph. In contrast, Zhang et al. [29] recognized the importance of connectivity among different local networks in federated learning framework, constructing random connections among them. However, this approach falls short of comprehensively capturing the spatial relationship of the traffic network between clients, random connections are not able to represent correct correlations among different local networks. In addition, the sparse data distribution in active mode traffic makes capturing the spatial relationships within the traffic network even more critical. Therefore, effectively modeling complex spatial-temporal dependencies for active mode traffic prediction within a federated learning framework remains a challenging and open problem.

5.3 Methodology

This section begins by defining the problem of federated learning-based active mode traffic prediction. Subsequently, the proposed global model HSTGCN for the federated learning framework is introduced to capture heterogeneous spatial and temporal dependencies within the traffic network. Ultimately, a FedGSAM aggregation mechanism is proposed for the central server of federated learning framework to aggregate the spatial information uploaded from clients.

5.3.1 Definition and problem statement

In this paper, our primary focus is on addressing two key problems. The first problem involves the development of a multi-step accurate traffic prediction model.

Table 5.1 MAJOR NOTATIONS

Notation	Meaning
$\mathbf{G}$	The global network.
N	The number of nodes in the entire graph.
$\mathbf{X}$	The historical traffic flow data.
$\hat{\mathbf{x}}$	The predicting traffic flow.
$\mathbf{A}$	The topological information of the entire transportation network.
$Dis_d^{i,j}$	The distant weight of the graph.
C	The local clients.

The second problem centers around exploring federated learning framework capable of achieving high-performance active-mode traffic prediction while maintaining local storage of raw traffic data. In addition to addressing data security concerns, such a framework offers the potential to enrich the diversity of traffic patterns by aggregating model parameters from diverse clients. This capability is particularly valuable for active mode traffic prediction in scenarios with sparse data distribution. The investigation aims to strike a balance between enhancing the accuracy of active mode traffic prediction models and preserving the integrity and security of raw data.

(1) Multi-step traffic flow prediction

To simplify computations while capturing significant traffic patterns, we define the traffic network as an undirected graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{A})$, where $\mathbf{V}$ is the set of N nodes representing sensors recording traffic flow on the road network; $\mathbf{E}$ is a set of edges; and $\mathbf{A} \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ is the adjacency matrix of $\mathbf{G}$ with N nodes. The traffic flow $\mathbf{X} \in \mathbb{R}^{N\times F}$ observed at each sensor represents the feature of the graph network $\mathbf{G}$, where F is the dimension of the feature of each node. $\mathbf{X}_t \in \mathbb{R}^{\mathbb{N}}$ denotes the traffic flow at time t, and $\mathbf{X} = (\mathbf{X}_{t-k+1}, \mathbf{X}_{t-k+2}, ..., \mathbf{X}_t)$ represents the historical traffic flow data with k time steps. The research problem is to determine a function $f(\cdot)$ for predicting the next T time steps of traffic flow $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_{t+1}, \hat{\mathbf{x}}_{t+2}, ..., \hat{\mathbf{x}}_{t+T})$, that is:

$$(\hat{\mathbf{x}}_{t+1}, \hat{\mathbf{x}}_{t+2}, ..., \hat{\mathbf{x}}_{t+T}) = f_\theta((\mathbf{X}_{t-k+1}, ..., \mathbf{X}_t), \mathbf{A}) \tag{5.1}$$

where θ represents the learnable parameters of the function. Our approach entails utilizing the latest one hour of historical data as input for our model to predict the traffic flow for the subsequent hour.

(2) Federated learning framework in traffic prediction

To ensure the security of traffic data and enhance the diversity of active mode traffic patterns, we establish the federated learning framework-based model for predicting active mode traffic flow without the need to gather all the raw data together. In this context, $\mathbf{G}$ denotes the global network encompassing the entire transportation network of a given area, $\mathbf{A}$ represents the topological information of the entire transportation network. This area is subdivided into several local areas, each represented by subnetworks $\mathbf{G} = \mathbf{G}_1, \mathbf{G}_2, ..., \mathbf{G}_s$, and these subnetworks are overseen by distinct organizations $\mathbf{C} = \mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_s$, respectively. The traffic flow data collected by each organization in their local area is denoted as $D_s = (\mathbf{X}_s, \mathbf{A}_s)$, where $\mathbf{X}_s$ is the historical traffic flow data in local area, $\mathbf{A}_s$ is the topological information of the local networks. Specifically, in this paper, we assume that the organizations do not have overlapping areas and data. It means that for any subnetwork $\mathbf{G}_i$ and $\mathbf{G}_j$, $\mathbf{G}_i \cap \mathbf{G}_j = \varnothing$. We aim to develop a robust global graph-based traffic prediction model by leveraging contributions from different organizations without the necessity of sharing raw traffic flow data.

In spatial-temporal traffic predictions, accounting for the mutual influence of traffic flow among sensors is crucial for accurate predictions. To establish a high-performing federated learning-based traffic prediction model, in addition to aggregating the model parameters w in the central server, it is crucial to properly investigate the spatial correlations among local clients to get a global overview of the spatial correlations $\mathbf{A}_l$ of the global traffic network. Therefore, our objective is to investigate the graph spatial information aggregation mechanism $Aggre(\cdot)$ for central server to obtain global spatial correlations while keeping traffic flow data and geographic topological information locally. That is:

$$\mathbf{A}_l^r = Aggre(\mathbf{A}_{l,c_1}^r, \mathbf{A}_{l,c_2}^r, ..., \mathbf{A}_{l,c_s}^r) \tag{5.2}$$

where $\mathbf{A}_l^r$ denotes the spatial correlations of the global network at the central server's communication round r with the clients, $\mathbf{A}_{l,c_s}^r$ represents the spatial correlations uploaded from local client $\mathbf{C}_s$ at communication round r.

5.3.2 Heterogeneous spatial-temporal graph neural network

For precise active mode traffic prediction, we devise a statistical heterogeneous spatial-temporal traffic prediction model as illustrated in **Fig. 5.1**, which includes three module layers. This approach enables the modeling of intricate relationships of spatial and temporal characteristics. Specifically, we employ multi-head self-attention [17] to capture time series traffic pattern of active mode traffic flow and GCN [30] with a 2Dconv-based global multi-head self-attention mechanism to capture complex heterogeneous spatial traffic pattern correlations among sensors on the active mode traffic network. To facilitate deep learning, we incorporate residual connections and layer normalization between each module.

(1) Temporal dependency capture module

By considering the periodic features of active mode traffic flow, it is essential to capture both global and local traffic flow patterns for accurate traffic prediction. To achieve this, we employ the multi-head self-attention mechanism, which is capable of capturing these global and local traffic dynamics, as illustrated in Equation (5.3)

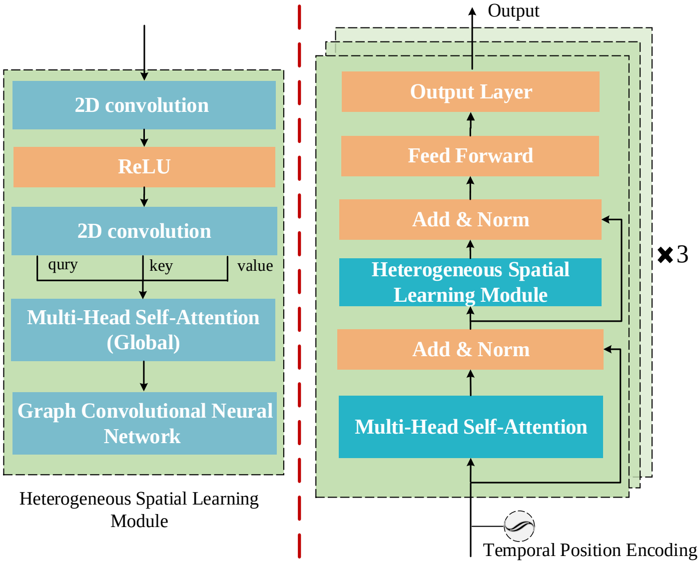


Figure 5.1 Heterogeneous spatial-temporal traffic prediction model. The global model of central server.

and Equation (5.4). This mechanism allows the model to simultaneously evaluate the significance of various traffic flows at different time slots within the input sequence. Consequently, providing a comprehensive understanding of dynamic active mode traffic flow patterns.

$$Multi - head(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = Concat(head_1, ..., head_h) \tag{5.3}$$

where $head_i$ denotes a self-attention mechanism, specifically a Scaled Dot-Product Attention [17] as illustrated in Equation (5.5). The input traffic flow sequence $\mathbf{X}$ for prediction undergoes projection onto distinct learned linear subspaces, resulting in the acquisition of the query $\mathbf{Q}$, key $\mathbf{K}$, and value $\mathbf{V}$ for the attention mechanism, respectively.

$$head_i = attention(\mathbf{X}_i\mathbf{W}_i^Q, \mathbf{X}_i\mathbf{W}_i^k, \mathbf{X}_i\mathbf{W}_i^V) \tag{5.4}$$

let $\mathbf{Q}_i = \mathbf{X}_i\mathbf{W}_i^Q$, $\mathbf{K}_i = \mathbf{X}_i\mathbf{W}_i^K$, and $\mathbf{V}_i = \mathbf{X}_i\mathbf{W}_i^V$, where $\mathbf{W}_i^Q$, $\mathbf{W}_i^K$, and $\mathbf{W}_i^V$ denote the learnable weights of the linear projection layers.

$$attention(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = softmax(\frac{\mathbf{Q}_i\mathbf{K}_i^T}{\sqrt{d_m}})\mathbf{V}_i \tag{5.5}$$

where d_m is the scaling factor, which is used to balance the complexity and the learning ability of the model.

Nevertheless, traffic flow data constitutes a time series, highlighting the crucial importance of the temporal dimension within the input sequence for precise predictions. While the multi-head self-attention module excels at capturing local and global dependencies among input elements, it lacks an explicit design for modeling sequence order similar to recurrent or convolutional neural networks. To address this, ensuring the incorporation of temporal order information in traffic data is crucial. In our approach, we achieve this by integrating fixed positional encodings [17] with the input embeddings. This strategic combination empowers the multi-head self-attention model to effectively leverage the temporal order information embedded in traffic data. The corresponding equations outlining this integration are presented below:

$$P_{(pt,2dim)} = sin(pt/10000^{2dim/d_m}) \tag{5.6}$$

$$P_{(pt,2dim+1)} = cos(pt/10000^{2dim/d_m}) \tag{5.7}$$

where pt represents the position index of the input sequence, while dim corresponds to the dim_{th} dimension of the positional encoding vector. Additionally, d_m denotes the dimension of the positional encoding, aligning with the dimension of input embeddings to enable element-wise addition.

(2) Spatial dependency capture module

The spatial-based GCN [30] is a neural network specifically designed to process non-Euclidean graph-structured data. This specialized architecture allows it to effectively operate and analyze spatial relationships embedded within the provided graph structure. By aggregating features from neighboring nodes, the GCN learns intricate feature representations, incorporating both local and global spatial information within the graph. This characteristic proves beneficial for uncovering hidden correlations within spatial data, enhancing the network's capability to capture hidden patterns and relationships.

In this paper, the application of GCN involves representing the traffic network as a graph, where nodes signify sensors, and edges denote the correlations between the data collected by them. Each node is associated with features of their corresponding historical traffic flow. The spatial-based GCN operates by performing a series of graph convolution on the historical traffic flow features with an adjacency matrix $\mathbf{A}$, which updates the feature representations of each node based on the features of its neighbors. The output of the spatial-based GCN is a set of new feature representations for each node. The operation formula is shown in Equation (5.8):

$$GCN(\mathbf{X}) = ReLU(\hat{\mathbf{D}}^{-1}\mathbf{A}\hat{\mathbf{D}}^{-1}\mathbf{X}W) \tag{5.8}$$

where $\mathbf{A}$ is the graph adjacency matrix; $\hat{\mathbf{D}}$ is a diagonal matrix where the i_{th} element on the diagonal $\hat{\mathbf{D}}_{ii}$ represents the degree of the i_{th} node, this is calculated as the sum of the elements in the i_{th} row of the adjacency matrix $\mathbf{A}$: $\hat{\mathbf{D}}_{ii} = \sum_j \mathbf{A}_{ij}$; $\mathbf{W}$ is the trainable weight matrix; $ReLU(\cdot)$ is an activation function.

We represent sensors as nodes in a graph and apply the Dijkstra algorithm [31] to search the shortest path between sensors in accordance with the road network. The topological structure of the network is constructed based on the shortest paths. Recognizing that traffic patterns are influenced by neighboring sensors, we introduce a weighted connection between sensors using a distance-based adjacency matrix $\mathbf{A}_d$. This involves applying a Gaussian kernel weighting function, with weights determined by shortest path distances, as described in [32].

$$Dis_d^{i,j} = \begin{cases} \exp(-\frac{|\mathrm{dist}(i,j)|^2}{2\hat{\theta}^2}), & \text{if} dist(i,j) < H \\ 0, & \text{else.} \end{cases} \tag{5.9}$$

where $Dis_d^{i,j}$ denotes the distant weight of the graph, with $|\mathrm{dist}(i,j)|$ representing the Euclidean distance between node i and node j. Here, $\hat{\theta}$ signifies the standard deviation of distances, and H serves as the threshold.

While GCNs excel in capturing static graph structures and node relationships, they lack the ability to capture global traffic patterns among sensors with statistical heterogeneity in their data distributions. Understanding the spatial correlations of global traffic patterns is crucial in transportation, as sensors collected active mode traffic flow data in different locations may exhibit distinct global traffic patterns. For example, sensors in the city center may capture significant traffic flow, while suburban sensors might only record limited flow, leading to a skewed data distribution. **Fig. 5.2** demonstrates significant statistical heterogeneity among sensors at identical time slots. For instance, the traffic patterns recorded by GRT07_0024 and RDH03_RMN17A exhibit notable differences. To capture these heterogeneous patterns, we use 2Dconv combined with ReLU activation to capture traffic pattern of all sensors. This approach allows us to analyze and identify key traffic features across different sensor locations, even when their data varies significantly. The output serve as the projected query, key, and value in a global multi-head self-attention mechanism for traffic pattern correlations capturing among sensors.

(3) Residual connection and feedforward networks

During the training of deep learning neural networks, particularly those with numerous layers, the gradients of the loss function concerning the parameters may diminish significantly as they propagate backward through the layers. Consequently, the earlier layers' weights may cease to learn effectively due to their gradients approaching zero. To mitigate the vanishing gradient problem and enhance model performance, we introduce residual connections between each layer. This mechanism enables the input of a specific layer to bypass one or more layers, directly adding it

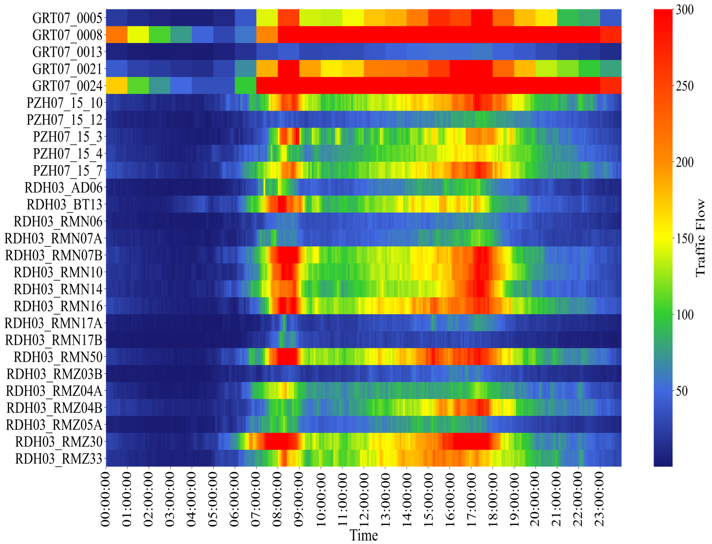


Figure 5.2 Average traffic flow distribution during a day. Traffic patterns across different sensors exhibit significant heterogeneity during the same time period, particularly during peak hours, as observed for sensors GRT07_0024 and RDH03_RMN17A.

to the subsequent layer's output, as demonstrated in Equation (5.10). Additionally, we integrate a fully connected feedforward network [17] into each encoder layer. This incorporation empowers the model to capture intricate input-output relationships and introduce nonlinearity, as depicted in Equation (5.11).

$$\mathbf{X}^{l+1} = layer(\mathbf{X}^l) + \mathbf{X}^l \tag{5.10}$$

$$FeedForward(\mathbf{X}) = ReLU(\mathbf{X}\mathbf{W}_0 + b_0)\mathbf{W}_1 + b_1 \tag{5.11}$$

where $layer(\mathbf{X}^l)$ represents the output of layer l, $\mathbf{X}^l$ denotes the input of layer l, and $\mathbf{X}^{l+1}$ signifies the output after the residual connection. The function $layer(\cdot)$ denotes the operation responsible for capturing temporal or spatial information.

(4) Multi-step traffic prediction

In this paper, we implement multi-step traffic prediction, providing a more comprehensive view of the future. In the reference section of this model, to mitigate the issue of error accumulation in traffic prediction, we have incorporated a fully connected layer within this module. This layer is specifically designed to make predictions for traffic flow across multiple future time steps simultaneously, without relying on previous prediction values.

5.3.3 Federated learning framework

In the federated learning framework, two primary entities are involved: local clients and the central server. Local clients, representing various data owners such as governments, individuals, and companies, maintain isolated datasets. Meanwhile, the central server acts as a reliable third-party cloud data center, selecting a subset of local clients for collaborative training and facilitating the updating of a global model using uploaded parameters.

In active mode traffic flow prediction scenarios, the potential spatial correlations among subnetworks in different clients are often overlooked. However, capturing these spatial correlations is crucial for accurate traffic prediction, as they are a significant feature of traffic flow patterns. To address this issue, this paper introduces a federated learning framework tailored for spatial-temporal traffic prediction. The architecture of this framework is depicted in **Fig. 5.3**.

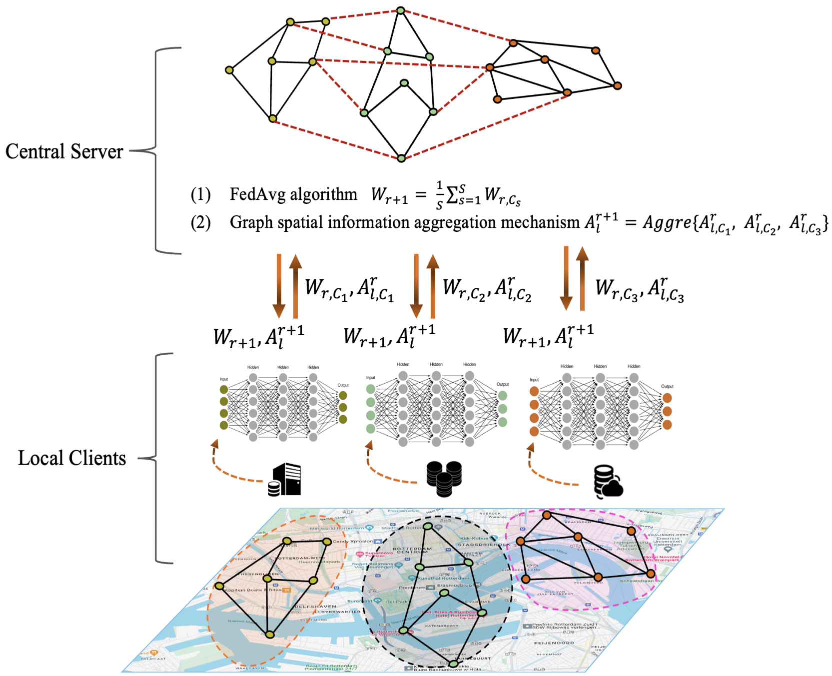


Figure 5.3 Federated learning framework. Local clients transmit model parameters to the central server for updates.

(1) Hidden spatial correlations learning

Geographical proximity is a standard metric used to illustrate spatial correlations among sensors in traffic networks. However, geographical proximity alone may not fully capture the complex spatial relationships among sensors. For instance, sensors located far apart yet in zones with similar land use or at the same road level can exhibit similar traffic patterns. Conversely, sensors in close proximity may not necessarily share similar traffic patterns. Therefore, uncovering these hidden spatial relationships within traffic networks becomes essential for comprehensively understanding global spatial correlations.

In this study, our aim is to reveal global spatial correlations among subnetworks within a federated learning framework using active mode sensor data from each client, all without the necessity of uploading raw data from each client. Specifically, we employ a learnable global adjacency matrix to explore these hidden spatial correlations based solely on the local data from each client. To accomplish this, we utilize two embedding dictionaries with learnable parameters, denoted as $\mathbf{E}_1$ and $\mathbf{E}_2 \in \mathbb{R}^{N\times P}$, which are instrumental in deriving spatial dependency weights among nodes using the traffic data of each local client [33]. These dictionaries play a crucial role in learning the global adjacency matrix within the traffic network, as elaborated below.

$$\mathbf{A}_l = ReLU(\mathbf{E}_1\mathbf{E}_2^T) \tag{5.12}$$

where P is the hidden dimensions of each node. $\mathbf{E}_1$ is the source node embedding dictionary, $\mathbf{E}_2$ is the target node embedding dictionary.

(2) Aggregation mechanism

In the federated learning framework, the central server is tasked with aggregating the updated parameters from local clients. Within the FEDHSTGCN model, we incorporate the FedAvg algorithm [5] as the aggregation mechanism to gather model parameters from diverse local clients. Given that the uploaded learnable adjacency matrices are parameters learned in each local client, we propose a FedGSAM aggregation mechanism to obtain global learnable adjacency matrix.

1) FedAvg algorithm: The FedAvg algorithm is a significant aggregation mechanism in the federated learning framework, known for its ability to improve model training efficiency and alleviate communication overhead. This is achieved by allowing each device to iterate through local updates multiple times on its own data before the aggregation step.

2) FedGSAM aggregation mechanism: To capture global spatial information across the entire traffic network, we integrate a global learnable adjacency matrix into the global model of the federated learning framework. This integration enables the model to discern inter-spatial correlations based on the data contributed by all local clients and explore the hidden spatial correlations of each client. Subsequently, we employ the proposed FedGSAM as shown in Equation (5.13) and Equation (5.14) to aggregate the learnable adjacency matrices received from local clients.

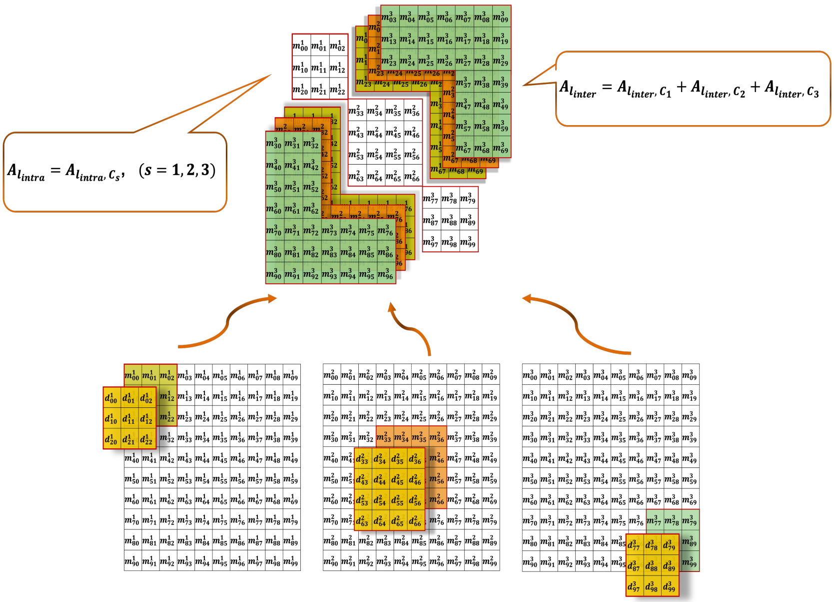


Figure 5.4 Matrix aggregation mechanism. Global spatial information update in central server

Since each local client possesses data exclusively from the sensors within its specific network, the global learnable adjacency matrix can accurately capture the spatial correlations relevant to the nodes within each client's network. Therefore, we extract these pertinent spatial correlations for aggregation separately. Specifically, the intra-spatial correlations matrix can only be learned from the data of each individual client; contributions from other clients would introduce noise. Consequently, the intra-spatial correlations matrix of each client is uploaded to the corresponding location in the global adjacency matrix, representing the global spatial correlations of that section and disregarding the spatial correlations learned by other clients, as shown in Equation (5.13). Additionally, each client can potentially contribute to capture the intercorrelations among clients. Therefore, the inter-adjacency matrix sections are derived by averaging the cumulative values of all inter-adjacency matrix sections contributed by the clients, as illustrated in Equation (5.14). This aggregation process is illustrated in **Fig. 5.4**.

$$\mathbf{A}_{l_{intra}} = \mathbf{A}_{l_{intra}, C_s}, (s = 1, 2, 3) \tag{5.13}$$

$$\mathbf{A}_{l_{inter}} = \frac{1}{S} \sum_{s=1}^{S} \mathbf{A}_{l_{inter}, C_s} \tag{5.14}$$

where S represents the number of selected clients each round, s denotes the label of an individual client. $\mathbf{A}_{l_{intra},\mathbf{C}_s}$ represents the updated portion of the learnable adjacency matrix of client $\mathbf{C}_s$, consisting solely the adjacency matrix of the corresponding nodes within client $\mathbf{C}_s$. The term $\mathbf{A}_{l_{inter},\mathbf{C}_s}$ denotes the updated segment of the learnable adjacency matrix from client s. $\mathbf{A}_{l_{inter}}$ represents the spatial correlations among clients.

(3) Learning process of FEDHSTGCN

In FEDHSTGCN framework, we employ both the FedAvg algorithm and our novel FedGSAM mechanism to aggregate client model parameters at the central server. The global HSTGCN model, updated through this aggregation process, is then redistributed to clients for federated training. This learning cycle consists of four key steps:

Algorithm 3: FedHSTGCN Algorithm.

Input: Local clients $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_s\}$; The global epochs, E_r; The learnable global adjacency matrix, $\mathbf{A}_l$; The local epochs, E_e; The learning rate, η; The size of the local mini-batch, B_z; The gradient optimizer for HSTGCN, $f(\cdot)$.
Output: Parameter $\mathbf{W}$
Central Server;
Initialize the parameters of global model $\mathbf{W}_0$;
Broadcast $\mathbf{A}_l$ and global model to the local clients;
foreach *round* $r = 1, 2, ..., r \in E_r$ **do**
  foreach *local client* $\mathbf{C}_s \in \mathbf{C}$ **do**
    $\mathbf{W}_{r+1}$, $\mathbf{A}_l^{r+1} \leftarrow$ LocalModelUpdate ($\mathbf{C}_s$, $\mathbf{A}_{l,C_s}^r$, $\mathbf{W}_{r,C_s}$);
  end
  $\mathbf{W}_{r+1} \leftarrow \frac{1}{S}\sum_{s=1}^{S} \mathbf{W}_{r,C_s}$;
  $\mathbf{A}_l^{r+1} \leftarrow Aggre\{\mathbf{A}_{l,C_1}^r, \mathbf{A}_{l,C_2}^r, ..., \mathbf{A}_{l,C_s}^r\}$;
end
LocalModelUpdate ($\mathbf{C}_s$, $\mathbf{A}_{l,C_s}^r$, $\mathbf{W}_{r,C_s}$);
$B_z \leftarrow$ (divide input $\mathbf{X}$ into mini-batches with a size of B_z);
foreach *epoch* $e = 1, 2, 3, ..., e \in E_e$ **do**
  foreach *mini-batch* $b_z = 1, 2, 3, ..., b_z \in B_z$ **do**
    $\mathbf{W}_{r,C_s} \leftarrow \mathbf{W}_{r,C_s} - \eta * f(\mathbf{A}_{l,C_s}^r, \mathbf{W}_{r,C_s})$;
  end
end
return $\mathbf{W}_{r,C_s}$ *and* $\mathbf{A}_{l,C_s}^r$ *to the central server.*

1. The central server broadcasts the initialized global model HSTGCN and the learnable global adjacency matrix $\mathbf{A}_l$ to the selected clients.

2. The local clients engage in training the received global model and refining

the learnable global adjacency matrix based on their individual datasets. Specifically, the adjacency matrix within the local clients is constructed by merging the learnable global adjacency matrix $\mathbf{A}_{l,\mathbf{C}s}$ with the local distance-based adjacency matrix $\mathbf{A}_{d,\mathbf{C}_s}$. It is to be noted that we align the dimensions of the distance-based adjacency matrix with the size of the learnable global adjacency matrix using zero-padding techniques [29]. Following this, the local clients transmit the updated global models and learnable global adjacency matrices back to the central server.

3. The central server utilizes the FedAvg algorithm to aggregate model parameters uploaded from local clients and employs FedGSAM to aggregate the learnable adjacency matrices from local clients, as depicted in Equation (5.13) and Equation (5.14). Given that the learnable global adjacency matrix serves as parameters for the global model, upon updating the model parameters through the FedAvg algorithm, we substitute the learnable global adjacency matrix of the global model with the aggregated outcome. Consequently, the central server disseminates the updated global models and learnable adjacency matrices to the chosen local clients for the subsequent round of training.

4. Iterate through steps 2) and 3) until the designated number of global epochs is achieved. The global model exhibiting the lowest validation loss will be chosen for subsequent testing.

5.4 Experiments

5.4.1 Dataset description

To evaluate the effectiveness of our proposed model, we conducted experiments using publicly accessible bicycle traffic flow data from the National Road Traffic Data Portal of the Netherlands. This dataset was collected by loop detector sensors in Rotterdam as shown in **Fig. 5.5** and includes data from 27 sensors recorded between September 1, 2022, and November 30, 2022. This dataset includes geographic information about the sensor stations.

The traffic flow dataset is aggregated into 5-minute intervals, resulting in 12 data points per hour. To enhance model convergence and training stability, we apply Z-score standardization to normalize the input dataset, rescaling the data so that it has a mean of 0 and a standard deviation of 1. Our objective is to conduct multi-step traffic prediction by utilizing one-hour historical data (consisting of 12 data points) to forecast the upcoming traffic flow for the subsequent 60, 30, and 15 minutes. The dataset is sequentially partitioned into training, validation, and test sets, with a split ratio of 6:2:2. Initially, the distance-based adjacency matrix serves as the input adjacency matrix.

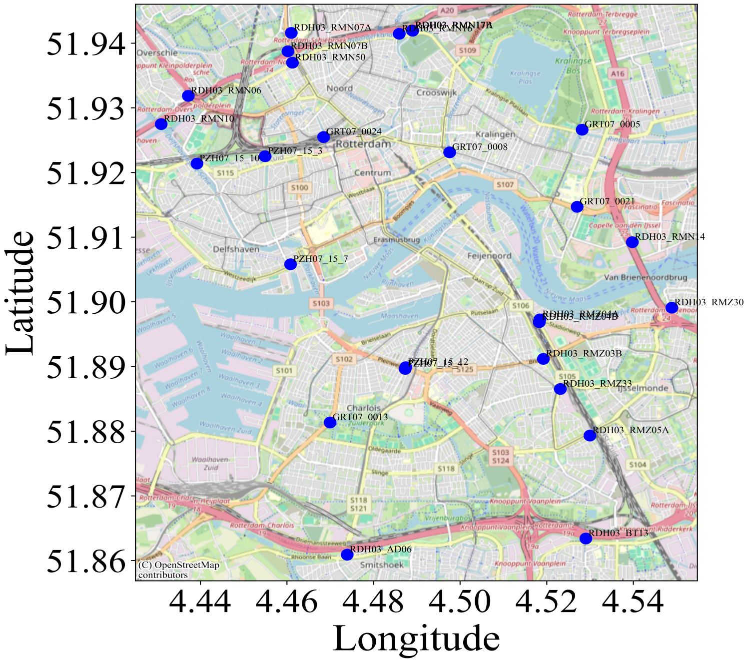


Figure 5.5 Sensor distribution. Sensors from Rotterdam

In the FEDHSTGCN model, we simulate diverse clients within the federated learning framework by configuring multiple local clients. To avoid the influence of imbalanced datasets between different clients, we partition the sensors into three clients, ensuring that all clients handle approximately equal amounts of traffic flow. For each communication round, all clients are involved.

5.4.2 Evaluation metrics

In this study, we evaluated the prediction performance of the proposed traffic flow prediction model using metrics including mean absolute error (MAE), root mean squared error (RMSE) and weighted absolute percentage error (WAPE). We present the expressions for computing these metrics as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{5.15}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2} \tag{5.16}$$

$$WAPE = \frac{\sum_{i=1}^{n}|\hat{y_i} - y_i|}{\sum_{i=1}^{n}|y_i|} \tag{5.17}$$

where $\hat{y_i}$ and y_i represent the predicted value and the ground truth, respectively. n is the amount of sample data.

5.4.3 Baseline models

To evaluate the effectiveness of the proposed FEDHSTGCN model, we compare its performance with various conventional traffic prediction models and state-of-the-art spatial-temporal traffic prediction models. Below, we provide an overview of the baseline models used for comparison.

- HA: The Historical Average (HA) model predicts traffic flow by taking the average value of historical data.

- SVR: Support Vector Regression (SVR) [34] is a regression technique that extends Support Vector Machines (SVM) to regression tasks, focusing on minimizing the discrepancy between predicted values and a defined error margin.

- LSTM: Long short-term memory (LSTM) [35] represents a distinct iteration of recurrent neural networks (RNNs), purposefully to capture and model long-term dependencies within sequential data.

- STGCN: The spatio-temporal graph convolutional network (STGCN) [36] is introduced to address the challenge of time series traffic prediction by leveraging intricate spatial-temporal correlations.

- ASTGNN: The attention-based spatial-temporal graph convolutional network (ASTGCN) [37] is formulated to enhance the prediction of highly nonlinear and intricate traffic flow. This is achieved through the fusion of the attention mechanism with spatial-temporal convolution.

- STMFGNN: Spatial-Temporal Multifactor Fusion Graph Neural Network (STMFGNN)[38] combines dynamic and static graphs through parallel convolution, using a gated fusion module to learn spatial dependencies. It integrates gated tanh convolutions, multi-scale receptive fields, and gated recurrent units for temporal modeling, enabling comprehensive traffic flow prediction.

- FLASTGCN: The federated learning-based attention-based spatial-temporal graph convolutional network (FLASTGCN) is the integration of the attention-based spatial-temporal graph convolutional network (ASTGCN) [37] and the FedAvg algorithm [5].

5.4.4 Experimental settings

The experiments were conducted on Google Colab, a cloud-based Python environment. The computational setup consisted of a Tesla A100 GPU with CUDA version 12.0, alongside an Intel(R) Core(TM) i9-9900KS CPU operating at a clock speed of 4 GHz. All deep learning models were implemented using the PyTorch framework in Python, with optimization performed using the Adam optimizer. Hyperparameters for the models were carefully tuned using a validation set. Specifically, for the FEDHSTGCN model, we set the global model dimension d_{model} to 64, the learning rate is between 0.001 and 0.0001, and the batch size to 64.

5.4.5 Experimental results and discussion

The performance comparison between the baseline models and the proposed approach is presented in **Table 5.2**. This table showcases the prediction error across various prediction horizons, including the next 15 minutes, 30 minutes, and 60 minutes. Notably, the HSTGCN model demonstrates superior performance over all baseline models in terms of MAE, RMSE, and WAPE on Rotterdam traffic flow dataset. The performance of the proposed FEDHSTGCN is slightly compromised compared to its centralized counterpart HSTGCN due to the lack of access to raw data. This suggests that despite the challenges posed by federated learning, the proposed FEDHSTGCN model exhibits promising capabilities to effectively leverage spatial-temporal graph information for accurate active mode traffic flow prediction.

Table 5.2 PERFORMANCE COMPARISON OF MULTI-STEP TRAFFIC PREDICTION. A LOWER ERROR INDICATES BETTER PERFORMANCE.

Baselines	**15 min**			**30 min**			**60 min**		
	MAE	RMSE	WAPE	MAE	RMSE	WAPE	MAE	RMSE	WAPE
HA	71.42	112.38	75.20%	71.42	112.39	75.22%	71.44	112.40	75.26%
SVR	32.91	60.83	37.59%	28.78	53.68	32.86%	26.33	49.21	30.05%
LSTM	25.66	48.36	29.29%	28.10	53.80	32.08%	32.69	63.72	37.34%
STGCN	22.80	40.30	26.03%	24.12	43.30	27.54%	26.36	48.27	30.11%
ASTGCN	23.17	41.07	26.46%	24.53	43.84	28.00%	26.79	48.00	30.60%
STMFGNN	22.34	37.41	25.51%	23.02	38.55	26.29%	24.22	40.73	27.66%
FLASTGCN	26.60	51.09	30.36%	28.80	55.76	32.88%	33.08	64.26	37.79%
HSTGCN	**20.78**	**36.37**	**23.72%**	**21.63**	**37.72**	**24.71%**	**23.18**	**39.97**	**26.47%**
FedHSTGCN	**23.70**	**40.89**	**27.05%**	**25.16**	**43.30**	**28.72%**	**27.50**	**47.31**	**31.41%**

(1) Prediction performance of baselines

Specifically, in traffic time series prediction, traditional HA shows poor performance when contrasted to other deep learning baselines. As HA predicts the state of traffic in historical traffic data by averaging past observations, it can only capture the linear dependencies within the traffic data. In most situation, active mode traffic flow is intricate nonlinear as it is influenced by diverse factors such as road conditions, weather, events, and human behavior. Deep learning methods because their ability of capturing nonlinear dependencies have been applied in predicting traffic state. SVR and LSTM exhibit the ability in handling both linear and nonlinear patterns in time series data, enabling them to capture intricate dependencies within traffic data. However, both of their performance falls short when compared to other deep learning baselines. SVR and LSTM only focus on temporal features of traffic data, whereas active mode traffic state prediction depends not only on the historical traffic data at the same position but also on the influence of surrounding traffic. Therefore, capturing spatial correlations among sensors is crucial for accurate predictions. In this context, STGCN, ASTGCN and STMFGNN leverage GNNs to consider traffic flow spatial correlations, thus well-perform in traffic predictions.

(2) Prediction performance of the proposed HSTGCN

While the baselines STGCN, ASTGCN and STMFGNN have demonstrated considerable success in capturing the spatial dependencies of active mode traffic data by considering the connections between nodes for traffic prediction, existing traffic prediction models may not effectively adapt to imbalances in the spatial distribution of data among sensors in the transportation network. Therefore, to explore and enhance the traffic prediction model to better handle skewed spatial distributions, we propose HSTGCN model to capture the heterogeneous spatial and temporal traffic patterns. The performance of the prediction model HSTGCN surpasses all baselines. Specifically, the proposed model demonstrated improvements in WAPE over STMFGNN by 7.0%, 6.0%, and 4.3% for prediction horizons of 15 minutes, 30 minutes, and 60 minutes, respectively. This outcome highlights the capability of HSTGCN effectively in capturing the spatial heterogeneity of active mode traffic data.

(3) Prediction performance of the proposed FEDHSTGCN

To strike a balance between accurate active-mode traffic prediction, data sparsity distribution and security, we have applied the federated learning framework to the HSTGCN traffic prediction model to get a FEDHSTGCN model for active-mode traffic prediction. Specifically, while the performance of FEDHSTGCN shows a decline of 12.85%, 14.84%, and 19.03% in WAPE for prediction horizons of 15 minutes, 30 minutes, and 60 minutes, respectively, when compared to its centralized counterpart, HSTGCN, the baseline FLASTGCN experiences a decline of 12.31%, 13.96%, and 15.73% in WAPE for prediction horizons of 15 minutes, 30 minutes, and 60 minutes, respectively, by comparing with its centralized model. Despite

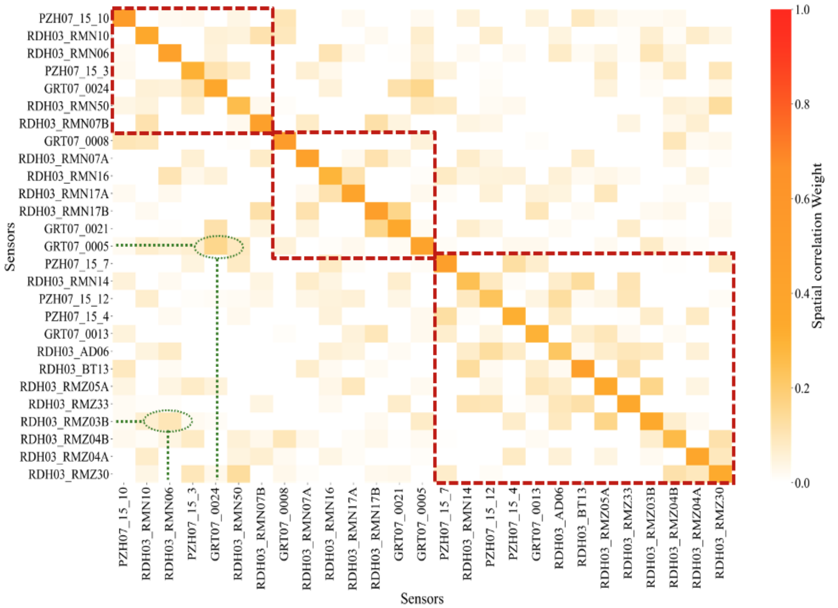


Figure 5.6 Spatial correlation weight. The intensity of the color indicates the strength of the correlations among the sensors.

this degradation, FedHSTGCN still achieves high accuracy predictions, nearly on par with the centralized model ASTGCN, while also maintaining the security of locally stored raw data. This effectively strikes a balance between data security and prediction accuracy.

Despite this degradation, FedHSTGCN still achieves high accuracy predictions, nearly on par with the centralized model STGCN, while also maintaining the security of locally stored raw data. This effectively strikes a balance between data security and prediction accuracy while enriching traffic pattern diversity in data-scarce contexts by integrating data from various sources.

Fig. 5.6 illustrates the hidden spatial correlations within the global traffic network, as captured by the proposed FedGSAM model. The figure highlights how different sensors in the network exhibit varying degrees of correlation, which reflects their interactions and dependencies. For example, sensors GRT07_0005 and GRT07_0024, as well as RDH03_RMZ03B and RDH03_RMN06, show strong spatial correlations. This is further confirmed by the traffic patterns of these sensor pairs, as shown in **Fig. 5.7** subplots (a) and (b) depict the traffic patterns of GRT07_0005 and GRT07_0024, while (c) and (d) show the patterns for RDH03_RMZ03B and RDH03_RMN06. The traffic flow between GRT07_0005 and GRT07_0024 exhibits clear peak hours, indicating a more consistent traffic pattern. In contrast, the patterns between RDH03_RMZ03B and RDH03_RMN06 show greater variability, with less

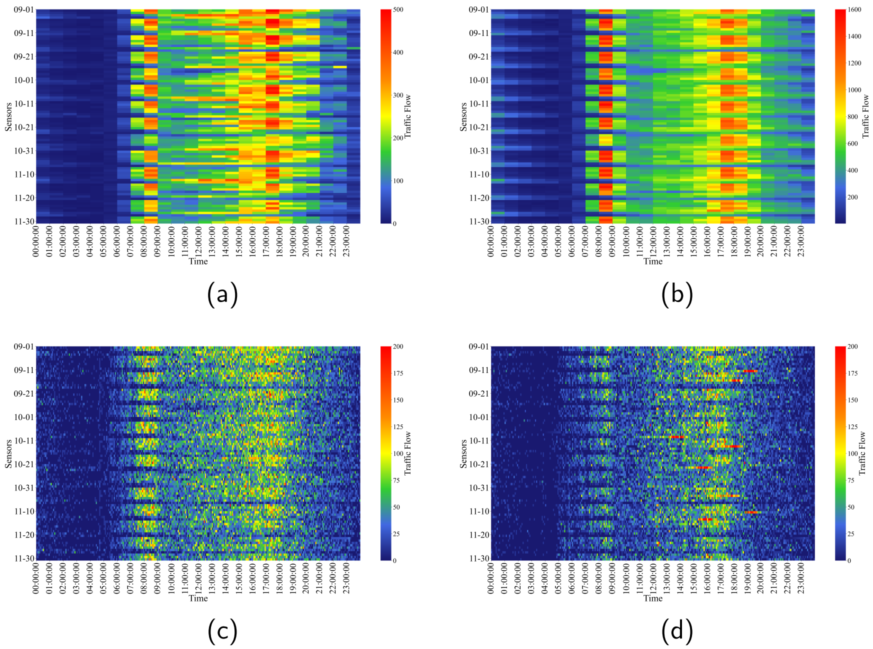


Figure 5.7 Similar traffic patterns among sensors. (a) GRT07_0005. (b) GRT07_0024. (c) RDH03_RMN06. (d) RDH03_RMZ03B. (a) and (b), and (c) and (d), respectively, show similar traffic patterns, with the model accurately identifying strong spatial correlations between the sensors.

distinct peak hours. This demonstrates the model's ability to capture different types of spatial correlations across sensors.

5.4.6 Ablation study

In our ablation study as shown in **Fig. 5.8**, we dissected the contributions of different components in our HSTGCN model and the FedHSTGCN model.

(1) The effect of different variants in HSTGCN

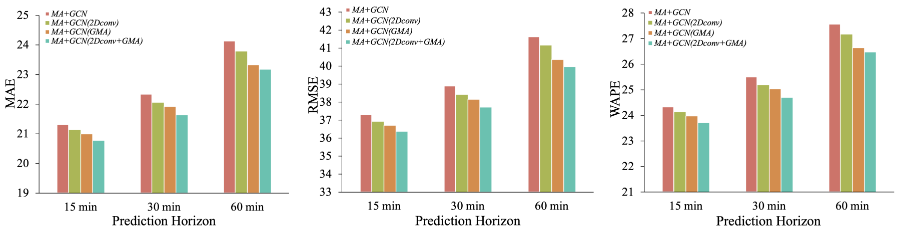
The variants of the HSTGCN model consist of four types: $MA + GCN$, $MA + GCN(SMA)$, $MA + GCN(2Dconv)$, and $MA + GCN(2Dconv + SMA)$. These variants are explained as follows:

$MA + GCN$**:** Multi-head self-attention with a graph neural network (GCN).

$MA + GCN(GMA)$**:** Multi-head self-attention integrated with a global multi-head self-attention-based GCN.

$MA + GCN(2Dconv)$: Multi-head self-attention with a 2Dconv-based GCN.

$MA + GCN(2Dconv + GMA)$: Multi-head self-attention alongside a 2Dconv-based global multi-head self-attention GCN.



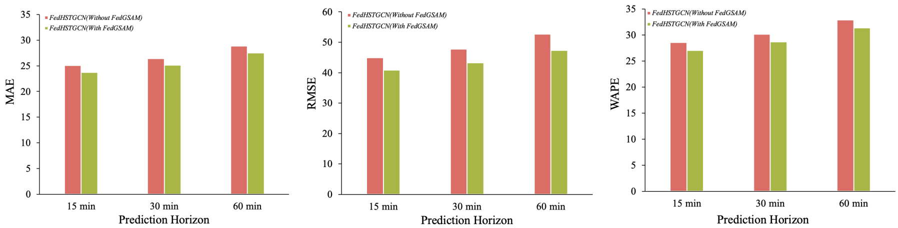
(a) (b) (c)

Figure 5.8 Ablation study of HSTGCN. (a) MAE. (b) RMSE. (c) WAPE.

Our aim is to assess the ability of a 2Dconv-based global multi-head self-attention GCN in capturing heterogeneous global traffic patterns among sensors. Using $MA + GCN$ as the baseline, we compare it with $MA + GCN(GMA)$ to gauge the global multi-head self-attention's effectiveness in capturing correlations among sensors. Additionally, $MA + GCN(2Dconv)$ is used to evaluate the 2Dconv's ability to capture traffic patterns of each sensor. $MA + GCN(2Dconv + GMA)$ investigates the collaborative effect of 2Dconv and global multi-head self-attention. Results show that the collaboration of these two components outperforms others, effectively capturing the heterogeneous traffic patterns among sensors.

(2) The effect of FedGSAM component in FEDHSTGCN

Furthermore, we conducted experiments to evaluate the impact of the proposed FedGSAM aggregation mechanism on the performance of FEDHSTGCN. The variations include FEDHSTGCN with and without the FedGSAM component. In **Fig. 5.9**, the training results suggest that incorporating the hidden correlations of the traffic networks positively impacts the performance of the FEDHSTGCN model.



(a) (b) (c)

Figure 5.9 Ablation study of FEDHSTGCN. (a) MAE. (b) RMSE. (c) WAPE.

5.5 Discussion and Conclusion

In this paper, we aim to investigate the impact of heterogeneous skewed active mode traffic data on the accuracy of traffic prediction and to evaluate the feasibility of constructing a high-performing traffic prediction model with distributed training data while ensuring data security and enriching traffic pattern diversity. Experimental results demonstrate that our model effectively captures the statistical heterogeneous relations among sensors and the proposed FedGSAM is capable of understanding the correlations between different regions or segments within a transportation network. By capturing heterogeneity among sensors and spatial correlations among clients, our model provides valuable understanding of traffic patterns and how traffic conditions propagate within the global traffic network, leading to more accurate predictions of traffic flow. This improved understanding enhances our ability to anticipate congestion hotspots and traffic fluctuations, enhances active mode traffic prediction with sparsely distributed datasets, and ensures the security of raw data. Ultimately, it contributes to the optimization of active mode transportation systems in urban environments. In real-world deployment, federated nature of the framework ensures scalability and adaptability to diverse urban environments, as local clients can join or leave the system dynamically. The framework can be integrated into existing traffic management infrastructures, such as smart city platforms, by deploying the central server in a cloud environment and connecting it to local traffic sensors or edge devices.

While the proposed model exhibits strong performance in traffic prediction compared to baseline models, there are areas that merit further enhancement. For example, augmenting the number of sensors to create more intricate client subnetworks could bolster model robustness. Moreover, the efficacy of the model may be influenced by the number of clients involved. In our future research, we aim to develop a robust data security traffic prediction model that accommodates a diverse range of clients with complex subnets. Additionally, we plan to incorporate multiple traffic modes, as interactions between different modes within the network can significantly impact overall performance.

Acknowledgements

This research is supported by the China Scholarship Council. This support is gratefully acknowledged. The contribution of Hoogendoorn is sponsored via EU's Horizon Europe research and innovation program under Grant No. 101093051 EMERALDS.

Disclosure statement

The authors report there are no competing interests to declare.

References

[1] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. “The graph neural network model”. In: *IEEE transactions on neural networks* 20.1 (2008), pp. 61–80.

[2] L. R. Medsker, L. Jain, *et al.* “Recurrent neural networks”. In: *Design and applications* 5.64-67 (2001), p. 2.

[3] X. Zhao, A. Chang, A. D. Sarma, H. Zheng, and B. Y. Zhao. “On the embeddability of random walk distances”. In: *Proceedings of the VLDB Endowment* 6.14 (2013), pp. 1690–1701.

[4] Y. Liang, G. Huang, and Z. Zhao. “Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach”. In: *Transportation research part C: emerging technologies* 140 (2022), p. 103731.

[5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. “Communication-efficient learning of deep networks from decentralized data”. In: *Artificial intelligence and statistics*. PMLR. 2017, pp. 1273–1282.

[6] Y. Liu, J. James, J. Kang, D. Niyato, and S. Zhang. “Privacy-preserving traffic flow prediction: A federated learning approach”. In: *IEEE Internet of Things Journal* 7.8 (2020), pp. 7751–7763.

[7] Y. Qi, M. S. Hossain, J. Nie, and X. Li. “Privacy-preserving blockchain-based federated learning for traffic flow prediction”. In: *Future Generation Computer Systems* 117 (2021), pp. 328–337.

[8] Y. Liu, S. Zhang, C. Zhang, and J. James. “Fedgru: Privacy-preserving traffic flow prediction via federated learning”. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2020, pp. 1–6.

[9] S. S. Sepasgozar and S. Pierre. “Fed-ntp: A federated learning algorithm for network traffic prediction in vanet”. In: *IEEE Access* 10 (2022), pp. 119607–119616.

[10] J. Lin and Y. Ban. “Complex network topology of transportation systems”. In: *Transport reviews* 33.6 (2013), pp. 658–685.

[11] X. Wang, Y. Ma, Y. Wang, W. Jin, X. Wang, J. Tang, C. Jia, and J. Yu. “Traffic flow prediction via spatial temporal graph neural network”. In: *Proceedings of the web conference 2020*. 2020, pp. 1082–1092.

[12] A. Ali, Y. Zhu, and M. Zakarya. “Exploiting dynamic spatio-temporal correlations for citywide traffic flow prediction using attention based neural networks”. In: *Information Sciences* 577 (2021), pp. 852–870.

[13] A. Ali, Y. Zhu, and M. Zakarya. “Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction”. In: *Neural networks* 145 (2022), pp. 233–247.

[14] N. Awan, A. Ali, F. Khan, M. Zakarya, R. Alturki, M. Kundi, M. D. Alshehri, and M. Haleem. "Modeling dynamic spatio-temporal correlations for urban traffic flows prediction". In: *IEEE Access* 9 (2021), pp. 26502–26511.

[15] M. Zakarya, L. Gillam, A. A. Khan, O. Rana, and R. Buyya. "ApMove: A service migration technique for connected and autonomous vehicles". In: *IEEE Internet of Things Journal* (2024).

[16] A. Ali, Y. Zhu, and M. Zakarya. "A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing". In: *Multimedia Tools and Applications* 80.20 (2021), pp. 31401–31433.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[18] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager. "Temporal convolutional networks for action segmentation and detection". In: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 156–165.

[19] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.* "Recent advances in convolutional neural networks". In: *Pattern recognition* 77 (2018), pp. 354–377.

[20] Z. Li, G. Xiong, Y. Chen, Y. Lv, B. Hu, F. Zhu, and F.-Y. Wang. "A hybrid deep learning approach with GCN and LSTM for traffic flow prediction". In: *2019 IEEE intelligent transportation systems conference (ITSC)*. IEEE. 2019, pp. 1929–1933.

[21] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong. "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting". In: *IEEE Transactions on Knowledge and Data Engineering* 34.11 (2021), pp. 5415–5428.

[22] S. Cheng, F. Lu, P. Peng, and S. Wu. "Short-term traffic forecasting: An adaptive ST-KNN model that considers spatial heterogeneity". In: *Computers, Environment and Urban Systems* 71 (2018), pp. 186–198.

[23] Y. Ku, C. Guo, K. Zhang, Y. Cui, H. Shu, Y. Yang, and L. Peng. "Toward Directed Spatiotemporal Graph: A New Idea for Heterogeneous Traffic Prediction". In: *IEEE Intelligent Transportation Systems Magazine* (2023).

[24] J. Ji, J. Wang, C. Huang, J. Wu, B. Xu, Z. Wu, J. Zhang, and Y. Zheng. "Spatio-temporal self-supervised learning for traffic flow prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 37. 4. 2023, pp. 4356–4364.

[25] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi. "An overview of IoT sensor data processing, fusion, and analysis techniques". In: *Sensors* 20.21 (2020), p. 6076.

[26] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin. "Deep learning on traffic prediction: Methods, analysis, and future directions". In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2021), pp. 4927–4943.

[27] C. Zhang, S. Zhang, S. Yu, and J. James. "Graph-based traffic forecasting via communication-efficient federated learning". In: *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2022, pp. 2041–2046.

[28] X. Yuan, J. Chen, J. Yang, N. Zhang, T. Yang, T. Han, and A. Taherkordi. "Fedstn: Graph representation driven federated learning for edge computing enabled urban traffic flow prediction". In: *IEEE Transactions on Intelligent Transportation Systems* 24.8 (2022), pp. 8738–8748.

[29] C. Zhang, S. Zhang, J. James, and S. Yu. "FASTGNN: A topological information protected federated learning approach for traffic speed forecasting". In: *IEEE Transactions on Industrial Informatics* 17.12 (2021), pp. 8464–8474.

[30] T. N. Kipf and M. Welling. "Semi-supervised classification with graph convolutional networks". In: *arXiv preprint arXiv:1609.02907* (2016).

[31] N. A. Ojekudo and N. P. Akpan. "Anapplication of Dijkstra's Algorithm to shortest route problem". In: *IOSR Journal of Mathematics (IOSR-JM)* 13.3 (2017), p. 14.

[32] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains". In: *IEEE signal processing magazine* 30.3 (2013), pp. 83–98.

[33] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang. "Graph wavenet for deep spatial-temporal graph modeling". In: *arXiv preprint arXiv:1906.00121* (2019).

[34] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. "Support vector regression machines". In: *Advances in neural information processing systems* 9 (1996).

[35] A. Graves and A. Graves. "Long short-term memory". In: *Supervised sequence labelling with recurrent neural networks* (2012), pp. 37–45.

[36] B. Yu, H. Yin, and Z. Zhu. "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting". In: *arXiv preprint arXiv:1709.04875* (2017).

[37] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan. "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 922–929.

[38] H. Jia, Z. Yu, Y. Chen, and H. Xia. "Spatial-temporal multi-factor fusion graph neural network for traffic prediction". In: *Applied Intelligence* 54.19 (2024), pp. 9464–9487.

Chapter 6
Conclusions and Perspectives

This chapter provides the conclusions and perspectives derived from this research. Section 6.1 summarizes the main findings, offering a comprehensive overview of the key outcomes. Section 6.2 presents the overall conclusion of the study, linking the various research implications for science. Section 6.3 discusses the practical implications of the work. Finally, Section 6.4 offers reflections on the thesis, and presents recommendations for future work, highlighting potential directions for advancing the field.

6.1 Key Findings

In this section, we briefly discuss the main findings of the research in relation to the research questions proposed.

Question 1: To what extent can a dynamic heterogeneous spatial-temporal traffic prediction model accurately capture spatial correlations specific by considering external influence to bicycle traffic? [Chapter 2]

Our analysis identified three main types of spatial correlations: predefined distance-based graph correlations, parameter-based adaptive graph correlations, and dynamic attention-based data correlations. While each of these methods captures spatial dependencies from different perspectives, they commonly overlook the influence of external environmental factors, such as weather, which can significantly impact bicycle traffic dynamics.

The integration of weather information with traffic data in data-driven spatial correlation modeling indicates that weather conditions can influence the spatial dependencies within traffic networks. Experimental results show that weather changes exhibit a lagging effect on these spatial correlations. Specifically, incorporating weather data with a 20-minute lag yields the best traffic prediction performance, highlighting the temporal delay in how environmental factors impact traffic dynamics.

The heterogeneous spatial correlation fusion approach demonstrates that combining multiple types of graph-based spatial correlations yields a more comprehensive representation of the spatial dynamics in bicycle traffic compared to using individual correlation types. This indicates that the fusion process effectively integrates heterogeneous spatial information without introducing negative interference, thereby validating its effectiveness for spatial correlation aggregation.

Question 2: How can a pre-trained LLM-based framework be applied for bicycle traffic prediction that performs effectively across full-sample and limited-sample conditions? [Chapter 3]

This study identifies a promising solution for bicycle traffic prediction in both full-sample and limited-sample data scenarios through the use of a pre-trained LLM-based framework. Specifically, we find that it is essential to accurately extract the embeddings of bicycle traffic patterns prior to input. This step enables the pre-trained LLM to effectively interpret the patterns in a sequence-like format that preserves the underlying spatial and temporal dependencies inherent in bicycle traffic patterns.

In this study, the results show that incorporating the lagged impact of weather conditions and leveraging semantic geographic features enables more effective capture of spatial-temporal patterns in bicycle traffic. By dynamically modeling these influences, the proposed approach generates richer numerical representations of traffic behavior. Experimental results show that the model consistently outperforms baseline methods in predictive performance.

Experimental results of few-shot prediction reveal that the proposed bicycle traffic prediction approach based on a pre-trained LLM-based framework, significantly outperforms baseline models even when trained on only 10% of the available data. This highlights the framework's strong generalization capability and robustness in limited-sample scenarios.

Question 3: What are the opportunities in developing an effective spatial-temporal traffic prediction model for bicycle traffic in data-scarce scenarios by leveraging knowledge from multi-source traffic data collected across diverse urban environments? [Chapter 4]

To solve this question, we utilized bicycle traffic data from six Dutch cities to explore knowledge transfer from data-rich to data-scarce cities. Our findings indicate that the use of data from a single source city can result in negative transfer, largely due to significant differences in traffic patterns and the complexity of bicycle infrastructure across cities. To mitigate this, a cluster-based modeling approach was employed, which groups similar traffic patterns across cities. This approach effectively reduces the risk of negative transfer by addressing both structural and behavioral differences in bicycle traffic.

According to multi-source transfer learning experiments, we found that incorporating knowledge from multiple source cities significantly enhances traffic prediction performance in the target city. By adaptively integrating information from different traffic pattern clusters, the approach flexibly captures relevant knowledge while accounting for the diversity of traffic behaviors. Compared to simpler fusion strategies, this adaptive cooperation yields superior predictive performance.

Our findings show that the proposed multi-source transfer learning approach remains robust even as the available training data is gradually reduced from 100% to 10%. In contrast to models without transfer learning, it maintains superior performance under increasingly limited data conditions. This demonstrates the potential of transfer learning as an effective solution for data-scarce scenarios in bicycle traffic prediction.

Question 4: What challenges arise in managing the heterogeneity of bicycle traffic patterns to develop accurate spatial-temporal traffic prediction models while ensuring that raw data remain stored locally? [Chapter 5]

This study finds that accounting for statistical heterogeneity in traffic patterns across sensors is essential for accurate prediction of bicycle traffic flow in bicycle transportation networks. The proposed HSTGCN model effectively captures complex, heterogeneous spatial correlations among sensors.

To address the challenges of distributed data, a federated learning framework is employed. Integrating spatial correlations among clients significantly improves network-wide prediction performance. Notably, the proposed FedGSAM aggregation mechanism demonstrates strong capability in aggregating spatially relevant information during training, enabling the model to effectively learn interspatial correlations across the traffic network without compromising data security.

6.2 Overall Conclusion

This thesis has addressed three key research topics to advance bicycle traffic prediction. First, it examined how external factors, such as weather conditions, influence traffic dynamics and affect the predictive performance of bicycle traffic models. Second, it developed solutions for data-scarce scenarios leveraging knowledge learned from other datasets, enabling effective prediction. Third, this thesis explored data security concerns and proposed a distributed learning framework that enables collaborative model training across multiple data sources without requiring centralized data collection while maintaining strong predictive performance. The overall conclusions to these topics are as follows:

- Accurate identification and integration of both static and dynamic spatial correlations in bicycle traffic is critical to improving prediction performance. Static geographic features, such as land-use features and road network topology, provide essential information for modeling spatial dependencies. Properly accounting for the influence of these features strengthens the representation of spatial correlations and enhances model accuracy. In addition, dynamic external factors, such as weather conditions, introduce temporal variability into bicycle traffic flows. These factors often exert immediate and delayed effects on traffic behavior. Using this information effectively, through careful temporal alignment and fusion with traffic data to capture spatial correlations, could enhance the model's ability to capture bicycle dynamics spatial correlations and reduces prediction errors.

- leveraging pre-trained models that capture knowledge from related or even unrelated sequential datasets provides a promising solution for bicycle traffic prediction in data-limited settings. By exploiting similarities between pre-trained knowledge and bicycle traffic data, fine-tuning the models on bicycle-specific datasets can yield accurate predictive performance. In particular, when adapting models pre-trained on unrelated domains such as text, where sequential patterns are also captured, extracting meaningful spatial-temporal embeddings from the bicycle data becomes essential. In contrast, when transferring from related domains, such as traffic data, the key challenge lies in effectively leveraging shared patterns while avoiding negative transfer. These findings highlight the potential of advanced transfer learning approaches to enable robust and reliable bicycle traffic prediction in real-world urban environments, even under conditions of limited data availability.

- Accurate bicycle traffic prediction can be achieved through distributed training without direct access to raw data, thereby addressing critical data security concerns in spatial-temporal modeling. The key contribution lies in showing that the proposed distributed learning can preserve predictive performance with minimal spatial correlations information loss compared to centralized training. This work advances the application of distributed learning in transportation systems and highlights its potential as a secure and practical approach for traffic prediction.

In conclusion, this thesis advances the development of AI-based approaches for bicycle traffic prediction, providing a support for applications that depend on reliable traffic monitoring. By addressing key challenges including external sensitivity, data scarcity, and data security, this work not only demonstrates the feasibility of accurate prediction in complex settings but also adds useful insights to the broader field of intelligent transportation systems.

6.3 Practical Implications

The approaches developed in this thesis aim to provide practical and adaptable solutions for bicycle traffic prediction across a range of real-world scenarios. To highlight the relevance and applicability of our work beyond theoretical contributions, this section outlines the key practical implications in three areas: (1) urban planning and traffic management, (2) real-world application use cases, and (3) deployment considerations for practice. These insights are intended to guide policymakers, mobility service providers, and system developers in leveraging the proposed methods for more efficient, data-driven decision-making in bicycle transportation systems.

First, the main goals of this thesis is to develop accurate spatial-temporal prediction models for bicycle traffic. Accurate prediction of traffic dynamics plays a critical role in supporting policymakers and traffic managers in making data-informed decisions that enhance the efficiency, safety, and reliability of active transportation systems. For example, real-time and reliable traffic predictions can inform infrastructure planning by identifying high-demand areas, optimizing the placement and expansion of bicycle lanes, and guiding the allocation of resources during peak hours on weekdays. From the perspective of individual users, such as cyclists, access to timely and accurate traffic information can support more informed travel decisions. Cyclists can adjust their routes based on predicted congestion or adverse weather conditions, reducing delays, enhancing safety, and improving overall travel experience. These applications demonstrate how accurate bicycle traffic prediction can directly support both system-level planning and individual-level decision-making.

Second, the proposed approaches are directly aligned with real-world challenges in this domain. In particular, the developed models offer practical solutions for cities facing data scarcity, dynamic environmental influences, and data security concerns. For instance, in cities where limited bicycle traffic data is available, models such as the bicycle traffic prediction framework based on pre-trained LLMs (BiSTLLM) and the multi-source transfer learning spatial-temporal graph neural network (MultiTLST-

GCN) can be effectively applied. These models are designed to perform well even in data-scarce environments by leveraging pre-trained knowledge or transferring shared patterns from data-rich cities. Moreover, when adverse weather conditions significantly impact traffic patterns, the spatial modeling components of these methods can be replaced or enhanced using the spatial component of the dynamic attention-based spatial-temporal graph convolution network model (DyASTGCN), which explicitly accounts for dynamic weather-related spatial correlations. As a practical example, a small city lacking sufficient data to train a deep learning model might encounter issues like overfitting and poor generalization. In such cases, the proposed transfer learning-based models can provide robust traffic predictions by borrowing knowledge from similar urban contexts with richer data. Additionally, when data security is a concern, as is often the case with multiple data collectors or institutions reluctant to share raw data, the heterogeneous spatial-temporal graph neural network based on federated learning (FedHSTGCN) offers a viable distributed learning solution. This framework enables model training across multiple decentralized data sources, allowing a third party to aggregate a global prediction model without accessing any raw data. This ensures data security while still achieving high prediction performance.

In real-world deployment, beyond evaluating whether a proposed model effectively addresses the prediction problem, it is equally important to consider the computational and operational costs associated with running the model. Large and complex models may achieve high accuracy, but often require significant computational resources, storage, and processing power, which may not be feasible or cost-effective for many users or city governments. The proposed pre-trained LLM-based or transfer learning-based approaches for bicycle traffic prediction in this thesis could solve this problem effectively. Specifically, these models do not require training from scratch, which significantly reduces computational cost and resource consumption. Moreover, because they are supported by prelearned knowledge, they can achieve competitive performance with much fewer training data compared to conventional deep learning models. This is particularly advantageous for cities with limited data availability, where training a high-performing model from scratch would otherwise be infeasible due to the risk of overfitting and poor generalization. For distributed learning scenarios, where training takes place across multiple devices or institutions without centralizing raw data, the computational load on a central server is minimized. However, the resource limitations at each local client still matter. In such settings, lightweight LLM-based or transfer learning-based global models are preferred, as they reduce the training cost on each client while maintaining accuracy and preserving data security.

6.4 Discussion and Recommendations

This thesis presents a range of approaches aimed at achieving accurate spatial-temporal prediction of bicycle traffic. While the proposed methods demonstrate promising results across various data conditions and urban contexts, several important challenges and open questions remain. Developing an effective and robust spatial-temporal traffic prediction model for bicycle traffic continues to require further investigation.

In the development of the DyASTGCN model for accurate bicycle traffic prediction, we introduced a data-driven approach to capture spatial correlations by incorporating the influence of weather conditions. Experimental results reveal that weather impacts on bicycle traffic exhibit a clear lagging effect, highlighting the importance of aligning external factors with traffic dynamics. However, one limitation of this analysis is the granularity mismatch between the datasets. While the traffic data is recorded at five-minute intervals, the weather data is only available at an hourly resolution. This discrepancy may limit the accuracy of capturing the true temporal influence of weather on traffic patterns. Future research would benefit from higher-resolution weather data to better model these delayed effects. Additionally, although dynamically capturing spatial correlations across the traffic network enhances model accuracy and expressiveness, it also significantly increases computational complexity. The DyASTGCN model requires substantial GPU resources and long training times, which can hinder its practical deployment in real-time or resource-constrained settings. Therefore, future work should explore more computationally efficient methods for learning dynamic spatial dependencies, such as pretrained LLM, approaches that strike a balance between predictive performance and resource efficiency, making them more suitable for real-world applications.

As part of this thesis, the BiSTLLM model was proposed to leverage a pre-trained large language model (LLM) to effectively capture the unique temporal characteristics of bicycle traffic time series. While our approach includes a temporal alignment module to align weather data with traffic observations, given their shared time-varying nature and lagging effect, the subsequent fusion of all feature embeddings treats each factor with equal importance. This uniform treatment does not fully reflect the varying influence that different external conditions exert on bicycle traffic. For example, weather conditions may have a more significant effect on traffic fluctuations than static geographical attributes. As a result, the current simple concatenation method may fall short in capturing the nuanced dependencies between features. A more advanced fusion strategy, such as an attention-based or learnable weighting mechanism, is recommended to better integrate heterogeneous inputs and enhance predictive performance. Additionally, our fine-tuning process for the pre-trained LLM relied on a grid search to determine the optimal number of layers for bicycle traffic data. While effective, this method is computationally expensive and time-consuming. Future work should explore more efficient strategies for architecture adaptation.

In this thesis, the MultiTLSTGCN model was developed to address bicycle traffic prediction using data collected from multiple cities. The core idea behind this model is to reduce negative transfer by clustering traffic sensors into groups with similar patterns before applying transfer learning. Experimental results demonstrate that this clustering strategy effectively reduces prediction error, indicating that it mitigates the risk of transferring irrelevant knowledge between dissimilar traffic environments. However, the current analysis relies primarily on empirical observations rather than theoretical justification. Incorporating causal analysis in future research could strengthen the understanding of how and why clustering reduces negative transfer, thereby providing a more robust foundation for the model's effectiveness. Another limitation of this approach lies in the method used for clustering. We utilized the average traffic pattern of each sensor as input for clustering. While this offers a simplified representation, averaging may obscure critical variations in traffic behavior over time. This simplification may affect the accuracy of cluster assignments. Future work could explore more expressive representations that better capture the representative patterns of each sensor. Additionally, the training data from source cities covers a relatively limited spatial region, often characterized by similar land use and infrastructure. This restricts the spatial diversity of the model's input and may constrain its generalizability to broader urban contexts. Expanding the sensor coverage to include a wider range of city types and land use patterns is recommended to improve the robustness and transferability of the model across diverse urban settings.

The FedHSTGCN model was developed by integrating federated learning with a spatial-temporal prediction framework tailored to bicycle traffic, with the primary goal of preserving data security by ensuring that raw data remains localized. This approach addresses data security concerns while enabling distributed model training across multiple data holders. However, current implementation does not reflect a real-world deployment. Due to hardware constraints, all federated training experiments were simulated on a single machine, significantly limiting scalability and computational efficiency. As a result, the number of participating clients in the experiments was restricted to only three, which reduces the model's robustness and limits its generalizability to practical applications involving larger and more diverse networks. Future work should aim to include more clients in the federated setting to evaluate the model's scalability and performance under more realistic distributed conditions. Another important consideration is the heterogeneity across clients. While the model currently accounts for statistical spatial heterogeneity among sensors within each individual client, it does not explicitly address inter-client heterogeneity, such as variations in network structure, sensor coverage, or regional traffic patterns. Investigating this broader heterogeneity is critical to improving the model's adaptability and robustness in real-world deployments. Moreover, the current setup does not allow for evaluating scenarios where only a subset of clients participate in each training round due to varying availability and willingness. Assessing the model's performance under such partial participation conditions is an important for future research, as it would provide insights into the stability of the federated training process in dynamic environments.

In summary, while the approaches proposed in this thesis demonstrate strong performance, there remain opportunities for further improvement. These include enhancing the quality of external data, reducing the computational cost of certain methods, providing stronger causal explanations for the observed results and developing a robust distributed framework to adapt to different structure setting. Looking ahead, several research directions is worth further exploration: (1) Spatial-temporal bicycle traffic prediction with small-size datasets, which is critical for making advanced models accessible to institutions with limited data or computational resources; (2) Explainable AI for bicycle traffic analysis, which is essential to transform prediction models from opaque black boxes into transparent, interpretable, and trustworthy tools for city planners and decision-makers. (3) Federated learning–based active-mode multi-agent trajectory prediction, which could be applied in connected autonomous vehicles and robotics to predict the future trajectories of surrounding agents, including cyclists and pedestrians, in real time, enabling proactive conflict avoidance.

Summary

With growing awareness of the environmental impacts caused by emissions from motorized traffic, sustainable mobility has gained significant attention in recent years. Among various low-carbon transportation modes, cycling has emerged as one of the most widely used traffic modes, particularly in countries like the Netherlands, where bicycles play an important role in daily commuting. However, the increasing use of bicycles has also introduced new challenges. For example, during weekday peak hours, the dense flow of bicycle traffic can lead to travel delays and safety risks increasing. As a result, the need to monitor and manage bicycle traffic has become increasingly important. Among various monitoring tools, traffic prediction plays a critical role by enabling the estimation of near-future traffic conditions across the network. Accurate and timely predictions not only help travelers make informed decisions, such as choosing optimal travel routes to avoid congestion, but also support policymakers and traffic managers in developing data-driven strategies to mitigate delays, enhance safety, and optimize the allocation of infrastructure and resources.

In recent years, the increasing availability of large-scale traffic data and advances in computational power have facilitated the widespread application of data-driven methods, particularly deep learning techniques, in the field of traffic prediction. These models are well-suited to capturing the complex temporal and spatial dynamics of traffic flow. For instance, recurrent neural networks (RNNs) and their variants have been extensively used to model the temporal evolution of traffic patterns. In parallel, spatial dependencies have been addressed using different architectures depending on the data structure: convolutional neural networks (CNNs) are often applied to grid-based data representations, while graph convolutional networks (GCNs) are well-suited to model non-Euclidean and irregular sensor networks. A popular and effective direction in recent research has been the development of hybrid models that jointly learn spatial and temporal dependencies, combining the strengths of these approaches for more accurate traffic prediction. Substantial research efforts have been devoted to optimizing such spatial-temporal architectures.

However, most existing deep learning approaches for traffic prediction have been developed and evaluated using publicly available datasets, which are primarily collected from highway systems or dock-based bike-sharing platforms. These datasets are often less representative of real-world, free-flowing active mode traffic such as bicycle traffic. Few studies specifically address the unique characteristics of certain traffic modes, particularly those that are highly sensitive to external environmental conditions. For instance, bicycle traffic is generally more unpredictable than motorized highway traffic

or controlled bike-share systems, as it is strongly influenced by factors such as weather, infrastructure quality, and route flexibility. Cyclists may alter their travel behavior depending on whether it is raining, or may choose unmonitored shortcuts and paths not covered by sensor networks, leading to incomplete or noisy data collection. Moreover, the availability and quality of bicycle traffic data are often limited, making it difficult to train high-performance models. In addition to data scarcity, data security and privacy present further constraints. In many cases, data collectors or organizations may be unwilling to share raw traffic data due to competitive concerns, which limits the possibility of centralized model training. Therefore, this thesis highlights the need to address not only the environmental sensitivity of bicycle traffic but also the dual challenges of limited data availability and data-sharing restrictions. By tackling these issues, the research aims to improve the accuracy and robustness of short-term traffic prediction models for active transportation systems.

To address the aforementioned scientific gaps, this thesis explores four key research directions. First, we investigate the dynamic and heterogeneous spatial correlations within active mode traffic networks by incorporating weather influences and designing an effective fusion mechanism. This enables a more accurate representation of spatial dependencies and provides insight into how external environmental factors, such as weather, impact bicycle traffic dynamics, laying the foundation for subsequent predictive modeling. Second, we develop a method to extract spatial-temporal embeddings from active mode traffic data and fine-tune a pre-trained LLM for traffic prediction under both full-data and limited-data scenarios. This offers a viable solution for improving predictive accuracy when data availability is scarce. Third, we propose a transfer learning framework to enhance traffic prediction in data-poor cities by leveraging knowledge from multiple data-rich urban areas. This approach demonstrates how shared patterns across cities can be transferred to improve performance in data-scarcity regions. Finally, we introduce a distributed learning framework that preserves data security by enabling collaborative model training without the need to exchange raw data. This provides a practical solution to the growing concerns around data security in transportation systems. The following sections summarize the results of each of these studies in detail.

Dynamic Spatial-temporal Model for Active Mode Prediction

In this study, we examine the dynamic spatial correlations in active mode traffic networks by incorporating the influence of weather conditions into the prediction model. Active mode traffic, such as bicycle traffic, is highly sensitive to external environmental factors, which can introduce uncertainty into traffic patterns. Our experiments reveal that weather conditions exhibit a lagging effect on spatial correlations, indicating that their impact on traffic is delayed rather than immediate. Since the spatial dependencies in active mode networks can be influenced by various factors, such as road structure, traffic characteristics, and land use, we propose a fused approach to represent the spatial correlations more accurately. By combining different spatial correlation graphs, it provides a more comprehensive and representative spatial correlation of the active mode network. This leads to improved prediction performance and offers deeper insight into the underlying spatial dynamics of active transportation systems.

LLM-Based Framework for Bicycle Traffic Prediction

In this study, we address the challenge of predicting bicycle traffic in scenarios where data is limited, sensor coverage is sparse, and traffic patterns are highly sensitive to external factors such as weather. Inspired by the strong sequential modeling capabilities of LLMs, we develop a prediction framework that requires less data while maintaining high accuracy. To do this, we first extract spatial-temporal features of bicycle traffic and design a contextual feature module that incorporates external influences including weather and geographical context to capture the uncertainties inherent in active mode traffic. Our experiments, conducted across multiple cities and under both high-data and low-data conditions, show that the proposed model outperforms existing state-of-the-art methods. These results demonstrate that LLM-based models can effectively support accurate traffic prediction even with limited data and computational resources.

Transfer Learning for Bicycle Traffic Prediction

In scenarios where data from multiple source cities are available but the target city lacks sufficient data, transferring learned traffic-related knowledge from source cities can offer a practical and effective solution for active mode traffic prediction. This study proposes a multi-source transfer learning framework that mitigates the risk of negative transfer by clustering traffic patterns and reconstructing spatial correlations to ensure better alignment between cities. The adaptive transfer mechanism further enhances performance by dynamically weighing the contribution of each source city based on its relevance to the target. Experimental results demonstrate that this method consistently outperforms non-transfer baselines, even under various levels of data scarcity in the target city. These findings highlight the robustness and effectiveness of the proposed approach in enabling accurate bicycle traffic prediction in data-limited urban environments.

Federated Spatial-Temporal Learning for Active Traffic Prediction

With growing concerns around data security and the high cost of storing and exchanging large volumes of data, distributed training has emerged as a promising approach for traffic prediction. In this study, we address key challenges associated with federated learning for active mode traffic prediction. First, we tackle sensor-level heterogeneity, which can hinder the model's ability to accurately capture consistent traffic patterns across locations. To enable effective distributed learning, we develop a federated learning framework incorporating a novel global spatial aggregation mechanism (FedGSAM), which minimizes information loss typically associated with decentralized training. Experimental results demonstrate that our proposed approach significantly mitigates the performance degradation commonly observed in federated learning settings, while ensuring that raw data remains stored locally. This offers a security and resource-efficient solution for traffic prediction across distributed environments.

Conclusion and Implications

In summary, this thesis focuses on developing spatial-temporal traffic prediction approaches specifically tailored for active modes of transportation, such as cycling, under a variety of real-world conditions. To address the core research question, we propose and evaluate a series of models that consider key challenges related to active mode traffic: high sensitivity to external environmental conditions (e.g., weather), limited and sparse data availability, and growing concerns around data security. Through the integration of advanced deep learning techniques including dynamic spatial modeling, pre-trained language models, transfer learning, and federated learning, we provide robust, scalable solutions that improve prediction accuracy while maintaining practicality for deployment in diverse urban settings.

The findings and methodologies developed in this thesis offer several important practical implications for the implementation of active mode traffic prediction systems. First, accurate short-term traffic predictions can empower travelers, particularly cyclists, with timely network information that enables better route planning, helping to reduce delays and avoid risk, especially during adverse conditions. From a decision-making perspective, predictive insights can support urban planners and traffic managers in optimizing infrastructure design, allocating resources more effectively, and improving the overall efficiency and safety of active transportation systems. Furthermore, the proposed models are particularly useful in scenarios where traffic data is scarce or unevenly distributed, as they are designed to perform well with limited input data. In addition, for contexts where data security is a concern and raw traffic data cannot be shared due to competitive or regulatory reasons, the federated learning-based approach offers a viable solution by enabling decentralized model training without compromising data security.

Samenvatting

Met de groeiende bewustwording van de milieueffecten veroorzaakt door de uitstoot van gemotoriseerd verkeer, heeft duurzame mobiliteit de afgelopen jaren aanzienlijk aan aandacht gewonnen. Onder de verschillende koolstofarme vervoerswijzen is fietsen uitgegroeid tot een van de meest gebruikte vormen van verkeer, vooral in landen zoals Nederland, waar fietsen een belangrijke rol speelt in het dagelijkse woon-werkverkeer. Echter, het toenemende gebruik van fietsen heeft ook nieuwe uitdagingen geïntroduceerd. Bijvoorbeeld, tijdens de spitsuren op weekdagen kan de dichte stroom van fietsverkeer leiden tot vertragingen en een toename van veiligheidsrisico's. Als gevolg hiervan is de behoefte aan monitoring en beheer van fietsverkeer steeds belangrijker geworden. Onder de verschillende monitoringtools speelt verkeersvoorspelling een cruciale rol door het mogelijk te maken om de verkeerssituatie in de nabije toekomst binnen het netwerk te schatten. Nauwkeurige en tijdige voorspellingen helpen niet alleen reizigers om weloverwogen beslissingen te nemen, zoals het kiezen van optimale reismogelijkheden om congestie te vermijden, maar ondersteunen ook beleidsmakers en verkeersbeheerders bij het ontwikkelen van datagestuurde strategieën om vertragingen te beperken, de veiligheid te vergroten en de inzet van infrastructuur en middelen te optimaliseren.

In de afgelopen jaren hebben de toenemende beschikbaarheid van grootschalige verkeersgegevens en de vooruitgang in rekenkracht de brede toepassing van datagedreven methoden, met name deep learning-technieken, in het domein van verkeersvoorspelling mogelijk gemaakt. Deze modellen zijn bijzonder geschikt voor het vastleggen van de complexe temporele en ruimtelijke dynamiek van verkeersstromen. Zo worden recurrente neurale netwerken (RNN's) en hun varianten veelvuldig gebruikt om de temporele evolutie van verkeerspatronen te modelleren. Tegelijkertijd worden ruimtelijke afhankelijkheden aangepakt met behulp van verschillende architecturen, afhankelijk van de datastructuur: convolutionele neurale netwerken (CNN's) worden vaak toegepast op rastergebaseerde gegevensrepresentaties, terwijl graaf-convolutionele netwerken (GCN's) beter geschikt zijn voor het modelleren van niet-Euclidische en onregelmatige sensornetwerken. Een populaire en effectieve richting in recent onderzoek is de ontwikkeling van hybride modellen die gelijktijdig ruimtelijke en temporele afhankelijkheden leren, waarbij de sterke punten van beide benaderingen worden gecombineerd voor een nauwkeurigere verkeersvoorspelling. Er is veel onderzoek gedaan naar het optimaliseren van dergelijke ruimtelijk-temporele architecturen.

De meeste bestaande deep learning-benaderingen voor verkeersvoorspelling zijn echter ontwikkeld en geëvalueerd op basis van openbaar beschikbare datasets, die voornamelijk afkomstig zijn van snelwegsystemen of dock-gebaseerde deelfietssystemen. Deze datasets zijn vaak minder representatief voor realistisch, vrijstromend actief verkeer, zoals fietsverkeer. Slechts weinig studies richten zich specifiek op de unieke kenmerken van bepaalde vervoersmodi, met name die welke sterk gevoelig zijn voor externe omgevingsfactoren. Fietsverkeer is bijvoorbeeld over het algemeen minder voorspelbaar dan gemotoriseerd verkeer op snelwegen of verkeer binnen gecontroleerde deelfietssystemen, aangezien het sterk wordt beïnvloed door factoren zoals het weer, de kwaliteit van de infrastructuur en routeflexibiliteit. Fietsers kunnen hun reisgedrag aanpassen afhankelijk van weersomstandigheden, zoals regen, of kiezen voor ongecontroleerde sluiproutes en paden die niet door sensornetwerken worden gedekt, wat leidt tot onvolledige of ruisgevoelige gegevensverzameling. Bovendien is de beschikbaarheid en kwaliteit van fietsverkeersgegevens vaak beperkt, wat het moeilijk maakt om hoogwaardige voorspellingsmodellen te trainen. Naast gegevensschaarste vormen ook gegevensbeveiliging en privacy extra beperkingen. In veel gevallen zijn gegevensverzamelaars of organisaties niet bereid om ruwe verkeersgegevens te delen vanwege concurrentieoverwegingen, wat de mogelijkheid tot gecentraliseerde modeltraining beperkt. Daarom benadrukt deze thesis de noodzaak om niet alleen de omgevingsgevoeligheid van fietsverkeer aan te pakken, maar ook de dubbele uitdaging van beperkte gegevensbeschikbaarheid en beperkingen in gegevensdeling. Door deze kwesties te behandelen, beoogt het onderzoek de nauwkeurigheid en robuustheid van kortetermijnverkeersvoorspellingen voor actieve transportsystemen te verbeteren.

Om de eerder genoemde wetenschappelijke lacunes aan te pakken, verkent deze thesis vier belangrijke onderzoekrichtingen. Ten eerste onderzoeken we de dynamische en heterogene ruimtelijke correlaties binnen netwerken van actieve vervoersmodi door weersinvloeden te integreren en een effectief fusiemechanisme te ontwerpen. Dit maakt een nauwkeurigere weergave van ruimtelijke afhankelijkheden mogelijk en biedt inzicht in hoe externe omgevingsfactoren, zoals het weer, de dynamiek van fietsverkeer beïnvloeden, waarmee een basis wordt gelegd voor daaropvolgende voorspellende modellering. Ten tweede ontwikkelen we een methode om ruimtelijk-temporele embeddings uit actieve verkeersgegevens te extraheren en een voorgetraind LLM-model fijn af te stemmen voor verkeersvoorspelling in zowel volledige als beperkte datascenario's. Dit biedt een haalbare oplossing om de voorspellende nauwkeurigheid te verbeteren wanneer de beschikbaarheid van gegevens schaars is. Ten derde stellen we een transfer learning-framework voor om verkeersvoorspellingen in data-arme steden te verbeteren door kennis over te dragen vanuit meerdere data-rijke stedelijke gebieden. Deze aanpak toont aan hoe gedeelde patronen tussen steden kunnen worden benut om de prestaties in regio's met beperkte gegevens te verbeteren. Tot slot introduceren we een gedistribueerd leerframework dat de gegevensbeveiliging waarborgt door samenwerking in modeltraining mogelijk te maken zonder dat ruwe data hoeft te worden uitgewisseld. Dit biedt een praktische oplossing voor de groeiende zorgen rondom gegevensbeveiliging in transportsystemen. De volgende secties vatten de resultaten van elk van deze studies in detail samen.

Dynamisch ruimtelijk-temporeel verkeersvoorspelling model

In deze studie onderzoeken we de dynamische ruimtelijke correlaties binnen netwerken van actieve vervoerswijzen door de invloed van weersomstandigheden te integreren in het voorspellingsmodel. Actief vervoer, zoals fietsverkeer, is zeer gevoelig voor externe omgevingsfactoren, die onzekerheid in verkeerspatronen kunnen veroorzaken. Onze experimenten tonen aan dat weersomstandigheden een vertraagd effect hebben op ruimtelijke correlaties, wat aangeeft dat hun invloed op het verkeer niet direct maar met enige vertraging optreedt. Aangezien de ruimtelijke afhankelijkheden binnen netwerken van actieve vervoerswijzen beïnvloed kunnen worden door verschillende factoren, zoals de weginfrastructuur, verkeerskenmerken en grondgebruik, stellen we een gefuseerde aanpak voor om de ruimtelijke correlaties nauwkeuriger te representeren. Door verschillende grafieken van ruimtelijke correlaties te combineren, ontstaat een meer omvattende en representatieve weergave van de ruimtelijke correlaties in het actieve vervoersnetwerk. Dit leidt tot verbeterde voorspellingsprestaties en biedt diepgaandere inzichten in de onderliggende ruimtelijke dynamiek van actieve transportsystemen.

Een fietsverkeersvoorspellingsraamwerk gebaseerd op voorgetrainde grote taalmodellen

In deze studie pakken we de uitdaging aan van het voorspellen van fietsverkeer in scenario's met beperkte data, beperkte sensordekking en verkeerspatronen die sterk gevoelig zijn voor externe factoren zoals het weer. Geïnspireerd door de sterke sequentiële modelleringsmogelijkheden van grote taalmodellen (LLM's), ontwikkelen we een voorspellingskader dat minder data vereist terwijl het toch hoge nauwkeurigheid behoudt. Hiervoor extraheren we eerst ruimtelijk-temporale kenmerken van fietsverkeer en ontwerpen we een contextueel feature-module die externe invloeden, waaronder weer en geografische context, integreert om de inherente onzekerheden in actief vervoer vast te leggen. Onze experimenten, uitgevoerd in meerdere steden en onder zowel data-rijke als data-arme omstandigheden, tonen aan dat het voorgestelde model beter presteert dan bestaande state-of-the-art methoden. Deze resultaten laten zien dat LLM-gebaseerde modellen effectieve ondersteuning kunnen bieden voor nauwkeurige verkeersvoorspellingen, zelfs bij beperkte data en rekenkracht.

Een voorspellingsmethode voor actief verkeersverkeer met behulp van overdracht van kennis uit meerdere bronnen

In scenario's waarbij data beschikbaar is uit meerdere bronsteden, maar de doelsstad onvoldoende gegevens heeft, kan het overdragen van geleerde verkeersgerelateerde kennis van de bronsteden een praktische en effectieve oplossing bieden voor de voorspelling van actief vervoer. Deze studie stelt een multi-source transfer learning raamwerk voor dat het risico op negatieve overdracht vermindert door verkeerspatronen te clusteren en ruimtelijke correlaties te reconstrueren om een betere afstemming tussen steden te waarborgen. Het adaptieve overdrachtsmechanisme verbetert de prestaties verder door dynamisch het gewicht van elke bronstad te bepalen op basis van de relevantie voor de doelsstad. Experimentele resultaten tonen aan dat deze methode consequent beter presteert dan modellen zonder transfer, zelfs bij verschillende niveaus van dataschaarste in de doelsstad. Deze bevindingen benadrukken de robuustheid en effectiviteit van de voorgestelde aanpak bij het mogelijk maken van nauwkeurige fietsverkeersvoorspellingen in stedelijke omgevingen met beperkte data.

Een federated learning-gebaseerde benadering voor heterogene verkeersvoorspelling in actieve vervoersmodi

Met toenemende zorgen over datasecurity en de hoge kosten van het opslaan en uitwisselen van grote hoeveelheden data, is gedistribueerde training naar voren gekomen als een veelbelovende aanpak voor verkeersvoorspelling. In deze studie pakken we de belangrijkste uitdagingen aan die gepaard gaan met federated learning voor de voorspelling van actief vervoer. Ten eerste behandelen we heterogeniteit op sensorniveau, wat het vermogen van het model kan belemmeren om consistente verkeerspatronen op verschillende locaties nauwkeurig vast te leggen. Om effectieve gedistribueerde learning mogelijk te maken, ontwikkelen we een federated learning raamwerk dat een nieuw globaal ruimtelijk aggregatiemechanisme (FedGSAM) integreert, dat het informatieverlies, dat typisch is bij gedecentraliseerde training, minimaliseert. Experimentele resultaten tonen aan dat onze voorgestelde aanpak de prestatievermindering die vaak wordt waargenomen bij federated learning aanzienlijk vermindert, terwijl de ruwe data lokaal opgeslagen blijft. Dit biedt een veilige en hulpbronnenefficiënte oplossing voor verkeersvoorspelling in gedistribueerde omgevingen.

Conclusie en Implicaties

In het kort richt deze thesis zich op het ontwikkelen van ruimtelijk-temporele verkeersvoorspellingsmethoden die specifiek zijn afgestemd op actieve vervoerswijzen, zoals fietsen, onder diverse reële omstandigheden. Om de kernonderzoeksvraag te beantwoorden, stellen we een reeks modellen voor en evalueren deze, waarbij we rekening houden met belangrijke uitdagingen met betrekking tot actief verkeersverkeer: de hoge gevoeligheid voor externe omgevingsfactoren (bijvoorbeeld het weer), beperkte en verspreide beschikbaarheid van data, en toenemende zorgen over dataveiligheid. Door de integratie van geavanceerde deep learning-technieken, waaronder dynamische ruimtelijke modellering, voorgetrainde taalmodellen, transfer learning en federated learning, bieden we robuuste en schaalbare oplossingen die de voorspellingsnauwkeurigheid verbeteren en tegelijkertijd praktisch inzetbaar zijn in diverse stedelijke omgevingen.

De bevindingen en methodologieën die in deze thesis zijn ontwikkeld, bieden verschillende belangrijke praktische implicaties voor de implementatie van verkeersvoorspellingssystemen voor actieve vervoerswijzen. Ten eerste kunnen nauwkeurige kortetermijnvoorspellingen reizigers, met name fietsers, voorzien van tijdige netwerkinformatie die hen in staat stelt hun route beter te plannen, waardoor vertragingen en risico's, vooral onder ongunstige omstandigheden, kunnen worden verminderd. Vanuit een besluitvormingsperspectief kunnen voorspellende inzichten stedelijke planners en verkeersbeheerders ondersteunen bij het optimaliseren van infrastructuurontwerp, het effectiever toewijzen van middelen en het verbeteren van de algehele efficiëntie en veiligheid van systemen voor actief vervoer. Bovendien zijn de voorgestelde modellen bijzonder nuttig in scenario's waarin verkeersdata schaars of ongelijk verdeeld is, om-dat ze ontworpen zijn om goed te presteren met beperkte invoergegevens. Daarnaast biedt de op federated learning gebaseerde aanpak een haalbare oplossing voor situaties waarin databeveiliging een zorg is en ruwe verkeersdata vanwege concurrentie- of regelgevingsredenen niet gedeeld kan worden, doordat het gedecentraliseerde modeltraining mogelijk maakt zonder concessies te doen aan de databeveiliging.

Acknowledgements

It feels almost unreal that I have reached the final stage of my PhD journey. Looking back, it all began nearly four years ago on a rainy morning, when the entire city was under lockdown due to COVID-19. The path was long, yet it passed in a blur of late nights, breakthroughs, and setbacks. Now, as I put together the fragments of data and memory, I see the things they form: not just a thesis, but a testament to perseverance. And I could never have reached this point alone, it was the unwavering support, patience, and encouragement of those who walked beside me that carried me through. To every person who listened, advised, or simply believed in me when I doubted myself.

First and foremost, I would like to express my deepest gratitude to my promotor Prof.dr.ir. Serge Hoogendoorn. Thank you for giving me this opportunity to pursue my research in Delft. I am grateful for your guidance, patience, and unwavering support throughout this journey. Your belief has been a constant source of motivation. Your insights and ideas have always inspired me, pushing my work to new heights. I also deeply appreciate your prompt responses and dedication, even during your busy schedule. None of this would have been possible without your mentorship. Meanwhile, I would also like to extend my sincere gratitude to my supervisor, Dr.ir. Dorine Duives. Thank you for joining my supervisory team and for your invaluable contributions to my research. I am deeply grateful for your encouragement and support throughout this process. It has been a true pleasure working with you, your expertise and insightful feedback have been crucial in shaping this work.

I would also like to express my sincere gratitude to my collaborators, Dr. ir. Panchamy Krishnakumari and Dr. Megha Khosla. Panchamy, thank you for your guidance during the first two years of my PhD journey. Your patience and support were invaluable as I navigated the challenges of this process. I deeply appreciate how you stood by me through every difficulty, your presence gave me the strength to persevere. Our collaborations were always inspiring, and I truly valued your insightful ideas and thoughtful feedback. Megha, thank you for your support during my PhD. Your helpful comments and constructive feedback consistently pushed my work forward. I am truly grateful for all the assistance you provided along the way.

I would also like to thank the members of my dissertation committee, Prof.dr.ir. J.W.C. van Lint, Prof. F. VITI, Prof. M. Saberi Kalaee, Prof. S.V. Ukkusuri, Dr.ir. W. Daamen, and Prof.dr.ir. B. van Arem for their valuable feedback and insightful comments, which have greatly enhanced the quality of my dissertation.

I am deeply grateful to the faculty, staff, and fellow PhD researchers in the Department of Transport and Planning, as well as those from other departments, who became part of my journey. You have transformed my PhD experience into something truly special. From our gym sessions and fun filled Zumba, Yoga, Xcore, stock and house dance classes, to our lively board game nights and shared meals exploring delicious cuisines, these moments brought joy and balance to my life. I especially cherish our wide-ranging conversations in the corridors, where we discussed everything from history, poetry, and music to entertainment and, of course, research! These treasured memories made my PhD journey truly meaningful. Thank you for being such wonderful companions on this challenging yet rewarding path.

A special thanks to our department secretaries for keeping my PhD journey running smoothly. I am so grateful for your help with meeting schedules, your quick answers to all my questions.

Throughout my academic journey, I have been fortunate to learn from many extraordinary teachers who have done more than share knowledge, they have profoundly shaped my thinking and working style. Among them, I am especially grateful to my master's supervisor, Prof. L.P. Fu. Thank you for your support and guidance. You were always there when I needed help.

Finally, I wish to express my deepest gratitude to my beloved family. Your unconditional love and steadfast support have been my anchor throughout this journey. You have been my constant motivation, my source of courage, and my reason to persevere. Every challenge was easier to face knowing you were behind me. I could never have reached this milestone without you.

Xiamei Wen
Delft, 31 July 2025

About the Author

Xiamei Wen was born in Guangxi Province, China, where she completed her primary, middle, and high school education. She received her Bachelor's degree in Transportation Engineering from Chang'an University in Xi'an, China, followed by a Master of Science degree in Intelligent Transportation Engineering from Wuhan University of Technology in Wuhan, China, in 2021.

In late 2021, she began her Ph.D. journey at the Department of Transport and Planning at Delft University of Technology, the Netherlands. Her research focuses on developing accurate, Artificial Intelligence(AI)-based spatial-temporal models for predicting active mode traffic.

List of publications

Journal Articles

1. **Wen, X.**, Krishnakumari, P. and Hoogendoorn, S., 2025. Dynamic Spatial-Temporal Graph Convolutional Neural Networks Approach for Active Mode Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*. - published

2. **Wen, X.** Duives, D. and Hoogendoorn, S. A Bicycle Traffic Prediction Framework Using Pretrained Large Language Models. - under review, R1 revision

3. **Wen, X.**, Khosla, M. and Hoogendoorn, S., 2025. Multi-Source Transfer Learning with Spatial-Temporal Graph Neural Network for Short-term Bicycle Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems* - published

4. **Wen, X.**, Khosla, M. and Hoogendoorn, S. Federated Learning-based Heterogeneous Spatial-Temporal Graph Neural Network for Active Mode Traffic Prediction. - under review, R1 revision

Peer-reviewed Conference Contribution

1. **Wen, X.**, Krishnakumari, P. and Hoogendoorn, S., 2023, September. A Data Protection Method for Short-Term Traffic Prediction with Applications to Network Active Mode Operations. *In 2023 IEEE 26th International Conference on Intelligent Transportation System (ITSC)*, pp. 2953-2958.

2. **Wen, X.**, Khosla, M. and Hoogendoorn, S. 2024, September. A Federated Learning-based Traffic Prediction Approach with Graph Spatial Information Aggregation Mechanism, *Conference in Emerging Technologies in Transportation Systems (TRC-30).*

3. **Wen, X.**, Khosla, M. and Hoogendoorn, S. 2025, January. A Federated Learning-based Traffic Prediction Approach for Short-term Traffic Flow Prediction, *In 104th Annual Meeting of the Transportation Research Board (TRB 2025).*

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 400 titles, see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Wen, X., *Data-Driven Spatial-Temporal Modeling for Bicycle Traffic Prediction*, T2026/7, March 2026, TRAIL Thesis Series, the Netherlands

Wang, Z., *Optimising Performance of Automatic Train Operation on Railway Networks*, T2026/6, March 2026, TRAIL Thesis Series, the Netherlands

Hadi, A.H., *DEM Modelling of Multi-Component Segregation in the Blast Furnace Charging System*, T2026/5, February 2026, TRAIL Thesis Series, the Netherlands

Farhani, M., *Demand Management Strategies for Operations of Shared Mobility Services*, T2026/4, February 2026, TRAIL Thesis Series, the Netherlands

Yao, X., *Driving Heterogeneity in Traffic Flow Theory: An action-based framework for identification, modelling, and simulation*, T2026/3, January 2026, TRAIL Thesis Series, the Netherlands

Versluis, N.D., *Optimising Railway Traffic Management under Radio-Based Distance-to-Go Signalling*, T2026/2, January 2026, TRAIL Thesis Series, the Netherlands

Jiao, Y., *Proactive Collision Risk Quantification in Multi-directional Traffic Interactions*, T2026/1, January 2026, TRAIL Thesis Series, the Netherlands

Asadi, M., *Accessibility and Road Safety: Integration of road safety in accessibility evaluation*, T2025/20, November 2025, TRAIL Thesis Series, the Netherlands

Akse, R., *Understanding and untangling the uncertainty knot: How to catalyse decision-making in mobility innovations*, T2025/19, November 2025, TRAIL Thesis Series, the Netherlands

Führer, K., *Participatory Decision-making under Deep Uncertainty: Modelling mobility transitions*, T2025/18, November 2025, TRAIL Thesis Series, the Netherlands

Picco, A., *Monitoring and Feedback in Driving*, T2025/17, October 2025, TRAIL Thesis Series, the Netherlands

Cebeci, M.S., *Behaviour of Prosumers in Last-mile Logistics: The case of crowdshipping*, T2025/16, September 2025, TRAIL Thesis Series, the Netherlands

Kuijpers, A., *Enabling Inter-Organizational Collaboration Through Platforms: The role of trust*, T2025/15, September 2025, TRAIL Thesis Series, the Netherlands

Song, R., *Human-MASS Interaction in Decision-Making for Safety and Efficiency in Mixed Waterborne Transport Systems*, T2025/14, June 2025, TRAIL Thesis Series, the Netherlands

Destyanto, A.R., *A Method for Evaluating Port Resilience in an Archipelago*, T2025/13, June 2025, TRAIL Thesis Series, the Netherlands

Karademir, C., *Synchronized Two-echelon Routing Problems: Exact and approximate methods for multimodal city logistics*, T2025/12, May 2025, TRAIL Thesis Series, the Netherlands

Vial, A., *Eyes in Motion: A new traffic sensing paradigm for pedestrians and cyclists*, T2025/11, May 2025, TRAIL Thesis Series, the Netherlands

Chen, Q., *Towards Mechanical Intelligence in Soft Robotics: Model-based design of mechanically intelligent structures*, T2025/10, April 2025, TRAIL Thesis Series, the Netherlands

Eftekhar, Z., *Exploring the Spatial and Temporal Patterns in Travel Demand: A data-driven approach*, T2025/9, June 2025, TRAIL Thesis Series, the Netherlands

Reddy, N., *Human Driving Behavior when Interacting with Automated Vehicles and the Implications on Traffic Efficiency*, T2025/8, May 2025, TRAIL Thesis Series, the Netherlands

Durand, A., *Lost in Digitalisation? Navigating public transport in the digital era*, T2025/7, May 2025, TRAIL Thesis Series, the Netherlands

Dong, Y., *Safe, Efficient, and Socially Compliant Automated Driving in Mixed Traffic: Sensing, Anomaly Detection, Planning and Control*, T2025/6, May 2025, TRAIL Thesis Series, the Netherlands

Droffelaar, I.S. van, *Simulation-optimization for Fugitive Interception*, T2025/5, May 2025, TRAIL Thesis Series, the Netherlands

Fan, Q., *Fleet Management Optimisation for Ride-hailing Services: from mixed traffic to fully automated environments*, T2025/4, April 2025, TRAIL Thesis Series, the Netherlands

Hagen, L. van der,*Machine Learning for Time Slot Management in Grocery Delivery*, T2025/3, March 2025, TRAIL Thesis Series, the Netherlands

Schilt, I.M. van, *Reconstructing Illicit Supply Chains with Sparse Data: a simulation approach*, T2025/2, January 2025, TRAIL Thesis Series, the Netherlands

Ruijter, A.J.F. de, *Two-Sided Dynamics in Ridesourcing Markets*, T2025/1, January 2025, TRAIL Thesis Series, the Netherlands

Fang, P., *Development of an Effective Modelling Method for the Local Mechanical Analysis of Submarine Power Cables*, T2024/17, December 2024, TRAIL Thesis Series, the Netherlands

TRAIL

Summary

With increasing awareness of the environmental impacts of motorized transport, low-carbon mobility modes such as cycling are gaining importance. This dissertation investigates bicycle traffic dynamics and develops data-driven spatial-temporal prediction models under varying environmental and operational conditions. The findings aim to support policymakers, mobility providers, and system developers in making more efficient and evidence-based decisions for bicycle transportation systems.

About the Author

Xiamei Wen conducted her PhD in the Transport & Planning Department at Delft University of Technology. She holds a Master's degree in Intelligent Transportation Engineering. Her current research focuses on developing accurate, Artificial Intelligence(AI)-based spatial-temporal models for predicting active mode traffic.

TRAIL Research School ISBN 978-90-5584-383-1

Erasmus University Rotterdam Radboud University rijksuniversiteit groningen
TU Delft Delft University of Technology UNIVERSITY OF TWENTE. TU/e Technische Universiteit Eindhoven University of Technology