

Review

Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities

Maryam Shaygan^{a,1}, Collin Meese^{a,1}, Wanxin Li^b, Xiaoliang (George) Zhao^c, Mark Nejad^{a,*}

^a Department of Civil and Environmental Engineering, University of Delaware, Newark, DE 19716, United States

^b Department of Communications and Networking, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China

^c BlueHalo, Rockville, MD 20855, United States



ARTICLE INFO

Keywords:

Traffic prediction
Artificial intelligence
Intelligent transportation systems
Traffic data
Deep learning
Survey

ABSTRACT

Traffic prediction plays a crucial role in alleviating traffic congestion which represents a critical problem globally, resulting in negative consequences such as lost hours of additional travel time and increased fuel consumption. Integrating emerging technologies into transportation systems provides opportunities for improving traffic prediction significantly and brings about new research problems. In order to lay the foundation for understanding the open research challenges in traffic prediction, this survey aims to provide a comprehensive overview of traffic prediction methodologies. Specifically, we focus on the recent advances and emerging research opportunities in Artificial Intelligence (AI)-based traffic prediction methods, due to their recent success and potential in traffic prediction, with an emphasis on multivariate traffic time series modeling. We first provide a list and explanation of the various data types and resources used in the literature. Next, the essential data preprocessing methods within the traffic prediction context are categorized, and the prediction methods and applications are subsequently summarized. Lastly, we present primary research challenges in traffic prediction and discuss some directions for future research.

1. Introduction

Traffic congestion is a critical problem with increasingly adverse impacts globally. The consequences are widespread, including increased accident rates, less reliable travel times, additional fuel consumption, excessive air pollution, and deterioration of societal health. For example, the 2019 Urban Mobility Report estimates that congestion in the United States resulted in a yearly 8.8 billion hours of additional travel time and 3.3 billion gallons of increased fuel consumption, totaling a cost of about \$179 billion (Schrank et al., 2019). Thus, it becomes critical to optimally utilize transportation infrastructure capacity to reduce congestion, especially in dense urban areas.

Accurate traffic prediction significantly improves network capacity utilization while also helping alleviate congestion by empowering traffic management centers (TMCs) and road operators to control traffic more effectively. In Table 1, we provide some practical examples of projects that focus on improving traffic management or deploying traffic predictors to facilitate real-world applications.

* Corresponding author.

E-mail addresses: mshaygan@udel.edu (M. Shaygan), cmeese@udel.edu (C. Meese), wanxin.li@xjtlu.edu.cn (W. Li), Xiaoliang.Zhao@bluehalo.com (X. Zhao), nejad@udel.edu (M. Nejad).

¹ These authors contributed equally to this work.

Table 1
Sample ITS-based urban traffic management projects.

Project	Objectives	Project Sponsor	Year of start & completion
Artificial Intelligence and ML enhanced ^a Integrated Transportation Management System (AI-ITMS) Deployment Program	–Adopting new technologies such as AI and ML to achieve better transportation management results & AI and ML enhanced integrated transportation	Federal Highway Administration, US	2020-Ongoing
An Evaluation of the Valley Metro–Waymo ^b Automated Vehicle RideChoice Mobility on Demand Demonstration	–Understand the potential behavioral impacts of AV mobility on demand services	Research and Innovative Technology Administration, US	2019–2020
Mobility on Demand (MOD) Sandbox ^c Demonstration: (TriMet) OpenTripPlanner (OTP) Shared-Use Mobility	–Improves transit trip planning	Research and Innovative Technology Administration, US	2017–2019
Mobility on Demand (MOD) Sandbox ^d Demonstration: Vermont Agency of Transportation (VTrans) OpenTripPlanner	–Provide an advanced statewide trip planner to present flexible transit services	Research and Innovative Technology Administration, US	2017–2019
Traffic Signal Control in the Transition ^e from Manual to Automated Driving	–Present how existing traffic signal controllers can cope with hybrid traffic environment	Dutch Traffic and Transport Laboratory	2021–2021
CAVITE-LAGUNA (CALA) Expressway ^f Project	–Reduce traffic congestion	Department of Public Works and Highways, Philippines	2019-Ongoing
Keys COAST Connected and Automated ^g Vehicle (CAV) Project	–Improve mobility and safety performance by establishing connectivity between various modes and a connected vehicle traffic signal system (CVTSS)	Office of Research and Development, State of Florida, Department of Transportation	2020-Ongoing
5G for Connected and Automated Road ^h Mobility in the European Union	–Build a 5G-enabled corridor to conduct cross-border trials and will deploy a mixture of 5G micro-/macro-cells for ubiquitous C-V2X connectivity	European Union	2018–2022
End-to-end Hardware Implementation of ⁱ Artificial Neural Networks for Edge Computing in Autonomous Vehicles	–Develop an alternative for DL tasks to reduce space-, power- and cost-consuming while also allowing for the advancement of AV technologies	European Union	2019–2020

^a<https://deldot.gov/Programs/itms/index.shtml?dc=projects>.

^b<https://www.transit.dot.gov/research-innovation/evaluation-valley-metro-waymo-automated-vehicle-ridechoice-mobility-demand>.

^c<https://www.transit.dot.gov/sites/fta.dot.gov/files/2021-06/FTA-Report-No-0197.pdf>.

^d<https://www.transit.dot.gov/research-innovation/mobility-demand-mod-sandbox-demonstration-vermont-agency-transportation-vtrans>.

^ehttps://www.its-edulab.nl/wp-content/uploads/MSc_Thesis_Femke_van_Giessen.pdf.

^f<https://www.dpwh.gov.ph/dpwh/PPP/projs/cala>.

^g<https://sunguide.info/its-program/projects/>.

^h<https://cordis.europa.eu/project/id/825012>.

ⁱ<https://cordis.europa.eu/project/id/849921>.

In addition, other applications such as route guidance and navigation systems (e.g., Google Maps,² Waze,³ MapQuest,⁴ Apple Maps,⁵ TomTom GO Navigation,⁶ INRIX,⁷ Traffic Spotter⁸) can also leverage traffic prediction methods to provide travelers with more accurate information in real-time and more quickly alleviate, or even prevent, congestion. Therefore, developing more intelligent and robust traffic prediction methodologies will be a crucial approach to relieving traffic congestion.

There are two primary classifications for traffic congestion: recurrent and non-recurrent. While recurrent congestion generally occurs when the road capacity is insufficient to accommodate the existing vehicle volume, non-recurrent congestion mainly results from accidents, disabled vehicles, inclement weather, work zones, and special events. A primary problem in traffic forecasting is accurately predicting the outcome of non-recurrent traffic events, which account for about 50% of all traffic congestion according to the Federal Highway Administration (FHWA) (FHWA, 2021). Thus, traffic prediction during non-recurrent events is a critical research area that needs more attention. Despite this, most existing studies focus on short-term and long-term recurrent traffic prediction problems, especially during rush hours.

² <https://www.google.com/maps>

³ <https://www.waze.com/>

⁴ <https://maps.questfordirections.com/>

⁵ <https://www.apple.com/maps/>

⁶ https://www.tomtom.com/en_us/

⁷ <https://inrix.com/>

⁸ <http://www.trafficspotter.com/>

In short-term traffic forecasting, the prediction horizon usually ranges from seconds to multiple hours ahead of time and utilizes both historical and current traffic information (Qu et al., 2019), making it applicable for traffic control route planning, etc. However, the precise definition of long-term forecasting, to differentiate it from short-term approaches, is largely ambiguous, and many researchers classify short-term and long-term forecasts differently (Manibardo et al., 2021). More specifically, some papers differentiate the two approaches solely based upon a quantitative threshold of the prediction horizon (Abdulhai et al., 2002; Chrobok et al., 2004; Bogaerts et al., 2020). In contrast, others assume short-term relates to the number of forecasting steps output by the model (Zhang et al., 2014; Cai et al., 2016; Ma et al., 2020c). However, both classification approaches result in ambiguous definitions for some models (e.g., multi-output models).

Moreover, it is imperative to examine emerging technologies, such as autonomous vehicles (AVs), Internet of Things (IoT), novel machine learning techniques, and secure wireless connectivity and information sharing schemes, as they have the potential to impact how we predict and manage traffic drastically. For example, blockchain is a secure-by-design network technology that has recently been explored as a way to manage decentralized traffic data for modeling traffic prediction states, such as congestion probability estimation (Hassija et al., 2020). Additionally, new machine learning concepts, including federated learning (FL), have attracted significant attention in traffic flow prediction applications (Meese et al., 2022). Specifically, their ability to train new models in a privacy-preserving and decentralized manner presents a more efficient and secure method for utilizing the enormous amounts of existing traffic data (Liu et al., 2020e). Furthermore, AVs can transform existing mobility and congestion patterns, laying the foundation for sustainable smart cities. In particular, AVs can better cope with many of the current problems (e.g., congestion) experienced on the roads today thanks to the real-time Vehicle-to-everything (V2X) communication with critical infrastructure and stakeholders, such as vehicle to traffic control centers, Road Side Units (RSUs), and other vehicles (Miglani and Kumar, 2019). However, this reality only becomes possible if AVs become both common and affordable for society (Litman, 2020). Additionally, there still exists a large knowledge gap regarding the impacts of these types of vehicles within mixed traffic scenarios (Di and Shi, 2021).

Most of the new AI methods in this review seek to leverage DL approaches' strong non-linear modeling capabilities to capture the complex spatio-temporal relationships between the multivariate traffic time series. Over the years, research has uncovered many unique properties of multivariate traffic time series, compared to other multivariate time series (e.g., product recommendations), such as non-stationary relationships, global spatial correlations with strong locality considerations, high temporal variability across multiple time-frames (e.g., recent, daily-periodic, seasonal), and also recurrent, non-recurrent, and external network impacts (e.g., accidents, weather, events). Throughout the paper, we refer to these models and their associated applications as multivariate traffic prediction (MTP) approaches.

This survey provides an introduction to, and extensive review of, the recent advances and emerging opportunities in AI-based traffic prediction methodologies. To the best of our knowledge, this is the first survey to provide an exhaustive review of both historical and recent AI-based traffic prediction approaches and associated data, discussed within the context of the common traffic prediction states and their practical applications, as well as the existing challenges and necessary future research directions. Thus, we believe this survey will be of paramount interest to researchers seeking a survey paper within the context of the cutting-edge Deep Learning (DL) approaches.

Concretely, we focus the literature review on the following broad research questions:

- How are different data type resources and classifications utilized within the literature for traffic prediction?
- Which preprocessing methods are applicable and demonstrated to be most effective for traffic prediction applications?
- How are varying prediction models and methodologies leveraged within the literature; and how do they apply to various traffic prediction problems?
- What are the fundamentally critical open research issues and challenges presenting barriers in existing traffic prediction research; and which of the emerging technologies are best suited for addressing these problems?

The remainder of this paper is organized as follows. Section 2 provides an overview of traffic prediction survey papers, comparing our work to the existing surveys. Section 3 summarizes the most commonly used traffic prediction research datasets found in the literature. Section 4 discusses the traffic prediction methods, including classical and machine learning methods. Section 5 introduces some common traffic prediction states and associated applications. Next, we discuss the open research challenges and list promising future research directions in Section 6. Finally, we conclude the paper in Section 7. Table 2 has summarized all acronyms used in this survey.

2. Prediction aspects assessed in previous reviews

This section outlines related survey papers on traffic prediction and briefly discusses their varying perspectives and approaches. Vlahogianni et al. (2004) contributed a comprehensive survey on traditional short-term traffic prediction methods and their applications. A decade later, the same authors discussed ten significant challenges for newly developed short-term traffic prediction approaches (Vlahogianni et al., 2014). Later, Barros et al. (2015) explored data-driven and model-driven approaches for short-term traffic prediction to highlight their capabilities. Li and Shahabi (2018) also suggested some possible future research directions after a review of several short-term traffic prediction techniques. Besides, Nagy and Simon (2018) summarized the data resources and provided an overview of existing Machine Learning (ML) traffic prediction methods. In the same year, Ermagun and Levinson (2018) conducted a comprehensive survey and examined long-term traffic forecasting literature to identify critical spatial components and their role in traffic forecasting. The authors also proposed possible future research directions for improving

Table 2
List of acronyms.

Acro.	Description	Acro.	Description
AE	Autoencoder	GB	Gradient Boosting
AI	Artificial Intelligence	HA	Historical Average
ANN	Artificial Neural Network	HMM	Hidden Markov Model
API	Application Programming	IoT	Internet of the Things
ARIMA	Autoregressive Integrated Moving Average	IoV	Internet of Vehicles
AV	Autonomous Vehicle	ITS	Intelligent Transportation System
AVI	Automatic Vehicle Identification	KF	Kalman Filter
BN	Bayesian Networks	kNN	K-nearest Neighbors
BP	backpropagation	LSTM	Long short-term Memory
CAV	Connected Autonomous vehicle	ML	Machine Learning
CFCO	Cellular Floating Car Data	MLP	Multi-Layer Perceptron
CNN	Convolutional Neural Network	MTP	Multivariate Traffic Prediction
DBM	Deep Boltzmann Machine	NN	Neural Network
DBN	Deep Belief Network	OD	Origin–Destination
DL	Deep Learning	POD	Points of Dispensing
DNN	Deep Neural Network	RBM	Restricted Boltzmann Machine
DT	Decision Tree	RF	Random Forest
ETC	Electronic Toll Collection	RNN	Recurrent Neural Network
FCD	Floating Car Data	RSU	Road Side Unit
FCN	Fully-Connected Neural Network	STTP	Short-term Traffic Prediction
FHWA	Federal Highway Administration	SAE	Stacked Autoencoder
FL	Federated Learning	SVR	Support Vector Regression
FNN	Feedforward Neural Network (FFNN)	TCN	Temporal Convolution Network
GCN	Graph Convolutional Network	TMC	Traffic Management Center
GMM	Gaussian Mixture Model	VAR	Vector Auto-Regressive
GPS	Global Positioning Systems	WNN	Wavelet Neural Network
GRU	Gated Recurrent Unit	XGBoost	Extreme Gradient Boosting

long-term traffic prediction. Meanwhile, [Angarita-Zapata et al. \(2019\)](#) proposed a new taxonomy categorizing the traffic forecasting regression problem in terms of both ML modeling and traffic specifications, focusing on the core characteristics that may alter the complexity of modeling traffic forecasting regression problems.

In addition, both [Boukerche and Wang \(2020a\)](#) and [Lana et al. \(2018\)](#) reviewed models that leverage classical approaches and some early deep learning methods. Other relevant surveys attempted to elucidate how different deep learning models are utilized in various transportation applications, including [Wang et al. \(2019e\)](#) and [Do et al. \(2019\)](#). More recently, [Akhtar and Moridpour \(2021\)](#) compared and contrasted some popular machine learning techniques for predicting traffic congestion. Furthermore, recurrent neural network, convolutional neural network, and feedforward neural network models for traffic prediction were reviewed by [Tedjopurnomo et al. \(2020\)](#). In contrast, [Ye et al. \(2020\)](#) instead concentrated on outlining graph-based deep learning architectures from the literature, emphasizing applications in the general traffic domain.

Urban flow prediction using spatio-temporal data was systematically outlined by [Xie et al. \(2020\)](#), and some traditional machine learning and deep learning models were discussed within the context of the urban flow prediction application. Later, [Gobezie and Fufa \(2020\)](#) reviewed robust traffic flow prediction concerning different approaches, development tools, and benchmark performance evaluation metrics. However, [Lee et al. \(2021\)](#) also suggested possible directions for future work in deep neural network (DNN)-based short-term traffic prediction (STTP) by reviewing input data representation methods, DNN approaches application domains, and dataset availability.

In another work, [Yin et al. \(2021\)](#) summarized various deep learning methods and collected and organized a comprehensive list of publicly available traffic datasets to help future research. Additionally, the authors conducted experiments to investigate the performance of the outlined methods. Meanwhile, [Shi et al. \(2019, 2021\)](#) focused on reviewing the hybrid deep learning methods, detailing their advantages while highlighting their ability to increase the accuracy of traffic prediction and better capture dependency in multi-dimensional studies. Later, [Yuan and Li \(2021\)](#) focused on surveying traffic data types based on their spatial and temporal dimensions and data preprocessing techniques. Specifically, traffic classification, generation, and forecasting are discussed to show how the existing methods can address traffic prediction challenges.

Moreover, [Wang et al. \(2020a\)](#) reviewed DL models' success in automatic feature representation learning and their ability to be widely applied in various spatio-temporal data mining (STDM) tasks such as predictive learning, anomaly detection, and classification. Another related survey, [Manibardo et al. \(2021\)](#), indicated why DL may not be the best modeling technique for every short-term traffic prediction application, which unveils some caveats unconsidered to date that the community in prospective studies should address. Lastly, [Laña et al. \(2021\)](#) identified the main actionability gaps in the data-based modeling workflow and how big data in ITS can be used to learn and adapt data-driven models for efficiently operating ITS assets, systems, and processes.

Recently, with the ongoing advancements in ITS and autonomous vehicle technologies, some efforts have been made to explore and design traffic prediction methods concentrated on autonomous vehicle environments. [Nguyen et al. \(2018\)](#) focused on reviewing not only the recent DL studies in traffic data processing, such as transportation network representation, traffic flow forecasting, traffic signal control, automatic vehicle detection, traffic incident processing, and travel demand prediction, but also the effects and

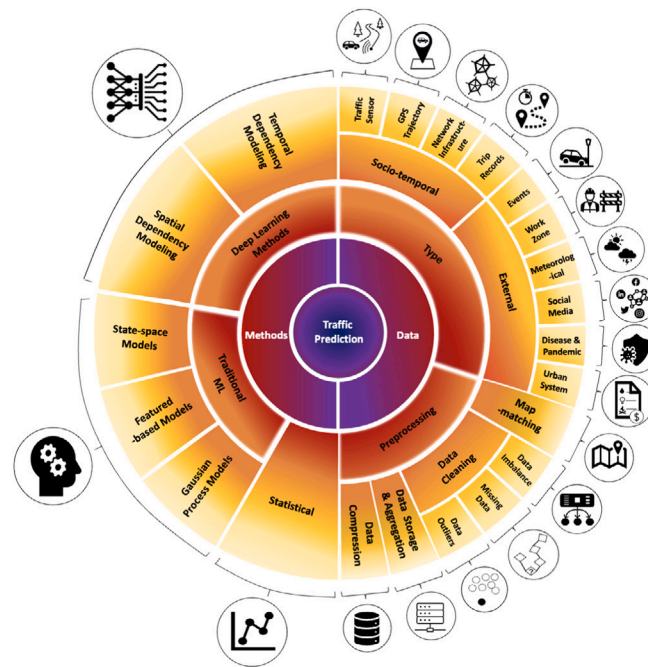


Fig. 1. The structure of traffic prediction based on the literature.

challenges of the autonomous driving and driver behavior on traffic prediction. In [Kumar and Saha \(2020\)](#), the authors reviewed some recent traffic prediction research within the self-driving application context. A comprehensive survey on DL applications and practical approaches for traffic prediction in autonomous vehicles is also provided in [Miglani and Kumar \(2019\)](#), where the main unresolved challenges in this area are presented to help guide future research.

While many related traffic prediction survey papers exist, they are generally focused on specific technical areas or subsets of the methods due to the density of approaches within the literature. On the other hand, not all surveys relate the recent advances to practical applications and future research directions. Besides, these surveys are also segmented across various discipline-specific journals and publication venues, making it challenging for readers to assess the current state of traffic prediction methods altogether. As the literature continues to become more diverse and interdisciplinary, it is necessary to provide a review that considers the literature comprehensively. Consequently, the primary contribution of this survey paper is to provide a systematic and extensive review of the existing traffic prediction literature as a whole, focusing on highlighting the recent advances and emerging research opportunities in Artificial Intelligence (AI)-based traffic prediction methods. To this end, we categorize the structure of the existing traffic prediction literature as shown in [Fig. 1](#). Additionally, to aid the reader in distinguishing our specific contributions from the myriad of existing surveys, we compare our survey to the related works in [Table 3](#).

3. Traffic data

This section extensively outlines and categorizes the varying data types and existing public data resources available for traffic prediction research. We also discuss data resolution and the different approaches in the literature for preprocessing the raw data into valuable and actionable information.

3.1. Data types and resources

The first step in accurate traffic prediction is acquiring high-quality data. Generally, the data collection operation is conducted by various stakeholders, including government organizations, municipalities, researchers, and commercial companies, on different time scales, utilizing various techniques. Government agencies and municipalities often measure road traffic for planning future roads, operation and maintenance, and analysis of accident risks and environmental impacts. The timescale of interest is often days, months, or years, but measurements are sometimes collected and used in real-time for traffic control and incident detection. On the other hand, commercial companies such as Google, Waze,⁹ Here,¹⁰ TomTom, and INRIX, collect and provide traffic information

⁹ <https://www.waze.com/>

¹⁰ <https://www.here.com/platform/traffic-solutions/real-time-traffic-information>

Table 3
Comparison with related survey articles.

Criteria		Vlahogianni et al. (2004)	Vlahogianni et al. (2014)	Barros et al. (2015)	Nguyen et al. (2018)	Li and Shahabi (2018)	Nagy and Simon (2018)	Lana et al. (2018)	Ermagan and Levinson (2018)	Migliani and Kumar (2019)	Wang et al. (2019)	Do et al. (2019)	Shi et al. (2019)	Angarita-Zapata et al. (2019)	Boukerche and Wang (2020a)	Tedjopurnomo et al. (2020)	Ye et al. (2020)	Xie et al. (2020)	Wang et al. (2020)	Gobezie and Fufa (2020)	Akhtar and Mordipour (2021)	Lee et al. (2021)	Yin et al. (2021)	Shi et al. (2021)	Yuan and Li (2021)	Mambrado et al. (2021)	Laña et al. (2021)	Current Survey	
Type	Spatio-Temporal	✓	✓	✓	×	×	✓	✓	~	✓	✓	✓	✓	✓	~	✓	✓	✓	✓	✓	✓	~	✓	×	✓	✓	✓	✓	
	Trajectory	~	✓	✓	×	×	✓	✓	~	✓	✓	✓	✓	✓	~	✓	✓	✓	✓	✓	✓	~	✓	×	✓	✓	✓	✓	
	Trip Records	~	×	~	×	×	×	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	×	×	×	×	×	
	Network Infrastructure	×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓	×	×	×	×	×	×	×	×	×	×	×	~	✓
Data	Events	×	~	~	×	×	~	✓	×	✓	~	~	✓	×	~	~	~	~	✓	~	~	~	✓	×	✓	✓	✓	✓	
	Meteorological	×	~	✓	×	×	✓	✓	×	✓	~	~	✓	✓	~	~	~	~	✓	~	~	~	✓	×	✓	✓	✓	✓	
	Social Media	×	×	~	×	×	×	✓	×	✓	~	~	×	×	×	×	×	×	×	×	~	×	×	×	~	~	✓	✓	
	Disease & Pandemics	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓
	Work Zone	×	×	~	×	×	~	×	×	✓	×	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓
	Urban System	×	×	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	✓
Resolution	~	~	~	×	×	~	~	~	~	~	~	×	✓	~	✓	✓	×	~	✓	~	✓	~	×	~	✓	~	✓	✓	
Prepro-cessing	Map Matching	×	×	×	×	×	×	×	×	×	×	×	✓	×	~	~	×	×	~	×	×	×	×	×	✓	×	✓	✓	
	Data Cleaning	×	~	×	×	×	×	×	×	×	~	×	×	✓	✓	~	×	✓	×	×	×	×	×	×	✓	×	✓	✓	
	Data Storage & Aggregation	×	×	×	×	×	×	×	×	~	×	×	×	~	~	~	×	✓	×	×	×	×	×	×	✓	×	✓	✓	
	Data Compression	×	×	×	×	×	×	×	×	×	×	×	×	✓	✓	~	×	✓	×	×	×	×	×	×	✓	×	✓	✓	
Anomaly Detection& Mitigation	×	×	×	×	×	×	~	×	~	~	~	×	×	~	×	×	×	✓	~	~	~	~	×	✓	~	~	✓		
Mixed Traffic Transportation Networks	×	×	×	×	×	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	×	×	×	×	×	✓	
Limited Data Accessibility	×	~	~	×	×	~	✓	×	✓	×	×	×	×	~	~	✓	~	✓	×	~	×	~	✓	×	~	✓	~	✓	
Benchmarking Traffic Prediction Datasources	×	×	×	×	×	×	✓	×	✓	×	×	×	~	~	✓	✓	✓	✓	×	×	×	~	✓	×	~	✓	✓	✓	
External Data Restrictions	×	✓	~	×	~	✓	✓	×	✓	✓	✓	×	~	~	✓	✓	✓	×	×	×	×	✓	✓	×	~	✓	✓	✓	
Multi-source Data	×	×	×	×	×	✓	✓	×	✓	✓	~	×	~	×	×	×	~	×	~	×	×	✓	✓	×	✓	✓	✓	✓	
Data Collection Area	✓	✓	×	×	×	~	✓	✓	~	×	~	×	~	×	~	✓	×	×	×	×	×	~	×	×	×	~	✓	✓	
Method-ology	Statistical	✓	✓	✓	×	~	✓	✓	✓	~	✓	×	×	✓	×	×	×	×	×	×	✓	×	✓	×	✓	✓	~	✓	
	Traditional ML	✓	✓	✓	×	~	✓	✓	✓	~	✓	~	×	✓	✓	~	×	✓	×	✓	✓	~	✓	~	✓	✓	~	✓	
	DL	×	×	×	~	~	~	~	×	✓	✓	✓	~	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	~	✓	✓	~	✓	
	Limitations	×	×	~	~	~	~	×	×	~	~	~	~	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	~	✓	✓	~	✓	
Evaluation Matrices	×	×	~	×	~	×	~	×	~	~	~	×	×	~	~	×	✓	×	✓	✓	×	✓	×	×	×	×	✓		
Long-term Traffic Prediction	×	×	~	~	✓	×	✓	×	✓	×	×	×	×	×	×	×	×	×	×	×	×	✓	×	×	~	✓	✓		
Real-time Prediction	×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓	×	×	×	×	×	✓	×	~	~	~	✓	✓		
Model Selection, Improve & Combine and Testing	~	✓	~	~	~	✓	✓	~	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Applications	×	×	~	~	~	×	×	×	~	✓	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	✓	
Prediction States	~	~	~	~	×	~	×	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	✓	
Emergence of AV&CAV	×	×	×	~	×	×	~	×	✓	~	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓	
FL Approach	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	✓	
Blockchain-enabled Traffic Prediction	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	~	✓	
Actionability	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓	✓	×

✓ indicates topic is covered, × indicates that topic is not covered and ~ indicates that topic is partially covered.

to drivers (and in some cases also to businesses, urban planners, and road authorities). However, some of the collected data is not accessible since it is maintained privately. That being said, many publicly available datasets are of interest to researchers. We direct the reader to a recent survey conducted by [Jiang and Luo \(2022a\)](#) for a list of open datasets. Specifically, the data types used within the literature can be categorized into two areas based on their various characteristics and associated resources: (1) the spatio-temporal sequence data; and (2) the external data.

3.1.1. Spatio-temporal sequence data

The most commonly used spatio-temporal, open-access datasets within the literature are summarized based on their characteristics in [Table 4](#). In [Table 5](#), we outline some recent research works utilizing these datasets. More specifically, we summarize the available spatio-temporal sequence data commonly used in traffic forecasting into the following categories:

- **Fixed Position Sensor Data.** A primary method for collecting traffic data leverages fixed-position sensors, which detect nearby vehicles at a specific point on the roadway. Inductive-loop detector, camera, radar, and light detecting and ranging (LIDAR) are standard sensors used to collect data. These sensors are usually limited in spatial coverage and measure traffic volume, speed, occupancy, and other traffic flow parameters. However, the main reason for this type of data's popularity is its availability and compatibility with deep neural network models; this data usually does not require significant transformation steps and can be used in its collected format. For example, PEMS¹¹ and its subsets (PeMS-BAY¹², PeMSD3, PeMSD4, PeMSD7 and PeMSD8) are widely considered as the most used datasets within the literature. PEMS contains data from all major metropolitan areas of California from 2001 to 2019. Another frequently used dataset is METR-LA,¹³ which contains traffic speed and volume data collected from the highways of Los Angeles County in 2012 from March 1st to June 30th, aggregated in 5-min intervals. Additionally, LOOP¹⁴ and Los-loop¹⁵ are the other popular traffic sensor data that gathered data from Seattle and Los Angeles freeways, respectively. Furthermore, Electronic Toll Collection (ETC) systems, with their multitude of fixed-position sensors, have also been leveraged to collect traffic data in some studies ([Chen et al., 2020b](#)).
- **Trajectory Data.** This type of data is generally collected using probe vehicle techniques designed for dynamic real-time data collection. In addition, with the emergence of IoT and AVs, mobile sensors can provide more detailed information ([Migliani and Kumar, 2019](#); [Lippi et al., 2013](#)). These systems, known as moving sensors, can be utilized for travel time data collection and assist in real-time applications, including routine monitoring of traffic operations, anomaly event detection, and route guidance. Based on the literature, the ITS probe vehicle data collection systems, such as Automatic Vehicle Identification (AVI) systems ([Wang et al., 2019a](#)), Cellular Geo-location systems, and Global Positioning Systems (GPS) ([Li et al., 2020e](#)) help improve traffic prediction by quantifying the exact trajectory and motion patterns of nearby vehicles, bicycles, and pedestrians. Additionally, they can help identify road segment connections where limited or no fixed sensors are deployed ([Nagy and Simon, 2018](#); [Hu et al., 2021](#)). Floating Car Data (FCD) is another subset of trajectory which can be generated by smartphones or vehicles equipped with GNSS receivers ([Elleuch et al., 2020](#)). When the data is collected by cellular phones using the cellular network, it is generally referred to as Cellular Floating Car Data (CFCF). On the other hand, the data generated by GNSS receivers are classified as GNSS probe data. Notably, most GNSS probe data is collected by fleet vehicles (e.g., taxi services, public transport services, private fleet companies).

Notably, there are some operational differences between moving and fixed sensors. While stationary sensors capture spatio-temporal traffic data without any confusion regarding the location of vehicles, due to their fixed detection range, these sensors can be less reliable, and operation may be interrupted dynamically. Thus, it is vital to consider the possibility of outage periods when working with large datasets collected from fixed sensor deployments. Meanwhile, the shared and privately-owned moving sensors do not incur general maintenance costs (e.g., weather damage, part failure, accidents, vandalism) associated with fixed sensors, providing a cost-benefit. However, while a single fixed sensor on the roadway can collect all the data, moving sensors require a high proportion of vehicles or pedestrians to collect and provide data to represent the approximate ground-truth flow and all the possible trajectories ([Nagy and Simon, 2018](#)). Moreover, it becomes imperative to preprocess the GPS trajectory data since measurement or interpolation errors related to sampling rates can cause uncertainties.

The available public trajectory data, which are provided by several probe-data vendors, such as INRIX, HERE, TomTom, NAVTEQ¹⁶ and TrafficCast¹⁷, can be divided into various categories. Some of them are recognized as Taxi Data, which contains many collected taxicab trajectories derived from GPS data. T-Drive,¹⁸ TaxiBJ,¹⁹ SZ-taxi,²⁰ NYC taxi²¹ and TaxiCD²² are some of

¹¹ <https://pems.dot.ca.gov/>

¹² <https://github.com/liyaguang/DCRNN>

¹³ <https://github.com/liyaguang/DCRNN>

¹⁴ <https://github.com/zhiyongc/Seattle-Loop-Data>

¹⁵ <https://github.com/lehaifeng/T-GCN/tree/master/data>

¹⁶ <https://www.here.com/navteq>

¹⁷ <https://www.iteris.com/TCI-redirect>

¹⁸ <https://www.microsoft.com/en-us/research/publication/t-drive-driving-directions-based-on-taxi-trajectories/>

¹⁹ <https://github.com/lucktroy/DeepST/tree/master/data/TaxiBJ>

²⁰ <https://github.com/lehaifeng/T-GCN>

²¹ <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

²² <https://js.dclab.run/v2/cmptDetail.html?id=175>

the well-known datasets used for traffic prediction. Ridehailing data gathered by some companies, such as Uber²³ and Didi²⁴, represents another type of GPS data which contains ride-hailing trip requests with their travel time, pick-up, and drop-off locations. Thus, this type of data is often applied for traffic demand prediction problems. Furthermore, Bus Data (e.g., NYC BUS,²⁵ CHI BUS²⁶), Bike Data (e.g., CHI-Bike²⁷, DCBike,²⁸ NYC-Bike²⁹) and Subway Data (e.g. SHMetro³⁰, HZMetro³¹) represent other types of trajectory data frequently used in the current traffic prediction literature.

- **Network Infrastructure Data.** Network infrastructure data (e.g., GIS data) is becoming increasingly available within open-access traffic datasets due to its importance in representing both spatial and temporal correlations (Liao et al., 2018a; Pan et al., 2019a; Peng et al., 2020b; James, 2021). More specifically, in applications such as map matching, network infrastructure data can be leveraged to map trajectory data sequences to actual positions on the roadway infrastructure, improving traffic prediction capabilities (Atif et al., 2020; James, 2021). This data includes information about the topology of roadway, subway, and bus networks. Much of the available network infrastructure data is open-access and can be obtained from government agencies (e.g., FHWA) or extracted from applications such as Open Street Map (OSM).³² As one of the most common open-access map databases, OSM provides information about road networks worldwide. Notably, this map data is collected by hand from volunteers performing systematic ground surveys using tools such as a handheld GPS unit, a notebook, a digital camera, or a voice recorder. In addition to OSM, A-map is a popular smartphone-based navigation app utilized in recent studies to collect network infrastructure data (Zhang et al., 2019a).
- **Trip Records Data.** The trip information, such as departures and arrivals times, dates, and locations, is necessary for traffic speed and demand prediction. Trip record data can be collected in many ways and from various sources, including taxis, ride-hailing services, e-scooters, bicycles, buses, and navigation systems. Another method for collecting trip record data is extracting it from existing datasets that may include trip data in forms other than the traditional origin–destination pairs. Much of this data can be automatically collected by the AFC (Automatic Fare Collection) system in subways and buses. In addition, trip record data can elucidate how drivers' characteristics influence their driving patterns, preferred route choices, and other driving-related decisions. When the collected information is labeled with additional driver-related attributes, such as driver ID, age, and gender, it is possible to improve prediction accuracy and provide personalized predictions (Zou et al., 2020).

3.1.2. External data

In addition to spatio-temporal sequence data, the prediction models must also be adaptable and responsive to the dynamic traffic and road environment changes, making it necessary to consider external data. Specifically, external data refers to related factors that impact traffic dynamics and roadway conditions that are not directly present within the collected spatio-temporal sequence data (e.g., meteorological and social media data). Problematically, this data is sometimes difficult to collect in practice or is unavailable. The existing studies leveraging external data are presented as part of Table 6. Below, we provide an outline of the most common types of external data and their importance in producing accurate traffic prediction models:

- **Event Traffic Data.** This type of data represents unexpected and non-recurrent traffic events on roads, such as accidents, sports events, and activities that may affect traffic flow and cause congestion. This data can be found in some reports, such as traffic accident reports. Notably, abnormal event data can also be found in some collected traffic datasets. Still, it may be difficult to distinguish it from other events and recurrent congestion without referencing other external data (e.g., accident). Event traffic data can also include trajectory data that is considered an anomaly because it appears with low frequency from an origin point to a destination point. Generally, anomaly trajectory data would be removed from the dataset during the cleaning process or repaired using various techniques (Liu et al., 2021) including activity sequence occurrence probability (Sani et al., 2019). However, this type of data can be helpful for specific applications, such as for detecting abnormal criminal behaviors in taxi drivers (Shen et al., 2015; Wang et al., 2018c).
- **Work Zone Data.** Construction activities on the roadways can disrupt expected traffic flows and create challenges for predicting traffic accurately. For example, work zones generally result in modifications to the existing roadway structure, such as reducing or eliminating lanes, shifting lanes, removing shoulders, using temporary signals, and reducing speed limits. As a result, additional traffic congestion becomes unavoidable in many cases, and accurate traffic prediction during these events is critical to mitigating the negative impacts.

²³ <https://github.com/fivethirtyeight/uber-tlc-foil-response>

²⁴ <https://outreach.didichuxing.com/research/opendata/>

²⁵ https://new.mta.info/system_modernization/bus-network

²⁶ <http://www.gbis.go.kr/>

²⁷ <https://www.divvybikes.com/system-data>

²⁸ <https://www.capitalbikeshare.com/system-data>

²⁹ <https://www.citibikenyc.com/system-data>

³⁰ <https://github.com/ivechan/PVCGNZ>

³¹ <https://github.com/ivechan/PVCGN>

³² <https://www.openstreetmap.org/#map=4/38.01/-95.84>

Table 4
Commonly used and large-scale real-world datasets in traffic prediction.

Dataset	Area	Spatial coverage	Time		
PeMS	California, USA	PeMS3	-358 sensors	9/1/2018~11/30/2018	
		PeMS4	-3848 sensors on 29 roads	1/1/2018 ~2/28/2018	
		PeMS7(M)	-228 stations	May and June, 2012	
		PeMS7(L)	-1026 stations	May and June, 2012	
		PeMS7	-883 sensor stations	7/1/2016~8/31/2016	
		PeMS8	-1979 sensors on 8 roads	7/1/2016~8/31/2016 7/1/2016~8/31/2016	
		PeMS-SF	-440 days of traffic data -963 sensors in the San Francisco	1/1/2008~3/30/2009	
		PeMS-BAY	-325 sensors	1/1/2017~6/30/2017	
METR-LA	Los Angeles, USA	-207 sensors on the highways	3/1/2012~6/30/2012		
LOOP	Seattle, USA	-323 sensor stations on four connected freeways: I-5, I-405, I-90 & SR-520	2015		
Los-loop	Los Angeles, USA	-207 sensors on the highways	3/1/2012~3/7/2012		
Q-Traffic	Beijing, China	-15073 road segments -Approximately 73891 km (6th, ring road) from the Baidu Map	4/1/2017~5/31/2017		
TOPIS	Seoul, South Korea	-1153 sensors and over 70000 taxis	1/2014 ~		
ENG-HW	Three cities, Britain	-Inter-city highways	2006~2014		
WebTRIS	England	-All motorways and "A" roads	4/2015 ~		
HK	Hong Kong, China	-Four regions in Hong Kong	12/28/2015 ~		
DRIVENet	Seattle, USA	-323 stations, 85 miles	2011		
Travel Time Index data	Shenzhen, Suzhou, Jinan, and Haikou	-Include city-level, district-level, and road-level	1/1/2018 ~ 12/31/2018		
DiDi	Chengdu, China	DiDiChengdu	-5476 geographical blocks	11/1/2016~11/30/2016	
		DiDiTTIChengdu	(each block size = 1 km ²)	2018	
		DiDiXi'an	Xi'an, China	-Contain DiDi Express and DiDi	10/2016~11/2016 2016.
		DiDiHaikou	Haikou, China	Premier drivers	5/1/2017~10/31/2017
		Didi Chuxing GAIA Initiative	JiNan & XiAn, China	-Total sample number of the two datasets is 52286 each -561 road segments for JiNan -792 road segments for XiAn	10/2017 ~
Uber	USA	-Over 18.8 million pick-up	4/1/2014~9/30/2014 1/1/2015~6/30/2015		
T-Drive	Beijing, China	-10,357 taxis	2/1/2015~6/2/2015		
SHSpeed	Shanghai, China	-for 156 urban road segments	4/1/2015~4/30/2015		
TaxiBJ	Beijing, China	-Contains inflow and outflow data	7/1/2013 ~10/30/2013		
		-Over 34000 taxicab GPS data	3/1/2014 ~ 6/30/2014		
		-Include meteorology data	3/1/2015 ~6/30/2015 11/1/2015~4/10/2016		
SZ-taxi	Shenzhen, China	-156 major roads of Luohu district	1/1/2015~1/31/2015		
TaxiCD	Chengdu, China	-1.4 billion GPS records from 14,864 taxis	8/3/2014~8/30/2014		
		-Consists of a taxi ID, latitude, longitude, an indicator of whether the taxi is occupied, and a timestamp			
NYC Taxi	New York City	-Including Bronx, Brooklyn, Manhattan Queen, Staten Island	2009~2018		

(continued on next page)

- **Meteorological Data.** Broadly, meteorological data includes multiple weather factors such as temperature and humidity, wind speed and strength, precipitation (e.g., rain, snow, hail), and severe weather (e.g., storm, tornado). For example, meteorological factors, such as sun glare, can negatively impact roadway congestion by obstructing the perception ability of both drivers and sensors, leading to roadway accidents that result in unexpected congestion (Taghipour et al., 2020). The impacts of meteorological events on traffic flow and roadway conditions (e.g., dry, wet, frost, icy) have become increasingly important in

Table 4 (continued).

Dataset	Area	Spatial coverage	Time
Porto	Porto, Portugal	-442 taxis, 420000 trajectories, 16735m*14389m	7/1/2013~6/30/2014
BikeDC	Washington, USA	-Over 500 stations across 6 jurisdictions -4300 bikes	20/9/2010 ~
NYC Bike	New York City	-800 stations -13000 bikes	5/7/2013 ~
CHI Bike	Chicago, USA	-580 stations across chicagoland, 5800 bikes	6/27/2013~
NYC Bus	New York, USA	-All public buses	2014
CHI Bus	Chicago, USA	-Buses on arterial streets in real time, -1250 road segments covering 300 miles	8/2/2011~5/3/2018
Chengdu	Chengdu, China	-14864 taxis, 9737557 trajectories	8/2014
Shanghai	Shanghai, China	-1.5 thousand main roads	3/1/2015~4/31/2015
Beijing	Beijing, China	-More than 2 million trajectories	3/1/2016~7/31/2016
SHMetro	Shanghai, China	-811.8 million transaction records of metro -288 metro stations and 958 physical edges	7/1/2016~9/30/2016
HZMetro	Hangzhou, China	-80 metro stations and 248 physical edges	1/2019

recent years with the advent of autonomous vehicles, which must have the appropriate sensors and configurations to develop safe driving strategies during periods of inclement weather (Huang et al., 2021b).

- **Social Media Data.** Online social media data (e.g., Tweets), also known as crowdsourced data, provide rich information for short- and long-term traffic prediction. Facebook, Twitter, and Snapchat are some of the online platforms that have seen remarkable usage in the last few years for communication, news reporting, and advertising events (Essien et al., 2020). Since these social media platforms provide real-time data retrieval using application programming interfaces (APIs), many researchers have leveraged social media data to understand better the relationships between people, their decisions, and the urban transportation systems (Ni et al., 2014; Yao and Qian, 2021). However, while social media data is generally available free of charge, it must be mined and can be unreliable due to human factors (Cvetek et al., 2021).
- **Disease Spread and Pandemic Data.** The worldwide COVID-19 pandemic has attracted much traffic prediction attention due to drastic changes in societal movement patterns. Pandemic events can significantly impact the prediction performance of existing models. More research is needed to investigate and quantify the effects of a non-recurrent worldwide pandemic on short-term and long-term traffic prediction (Macioszek and Kurek, 2021). Notably, public health agencies generally employ Points of Dispensing (PODs) during a large-scale biological emergency event for distributing medical countermeasures to affected populations. While the public health community has studied the clinical operations aspect of PODs, little attention has been paid to how the transportation system can support POD access and how the choice of POD location impacts roadway traffic. For example, to accurately model the traffic impacts of PODs operations, it is necessary to consider expected demand concerning the parking capacity of a POD in conjunction with the nearby road traffic capacity at any given time.
- **Urban Systems Data.** There may be a strong correlation between utility systems usage and transportation patterns within the urban environment. For example, high electricity usage in a neighborhood may indicate that most residents are at home and not utilizing the transportation network. Thus, the spatio-temporal usage patterns of a community can be identified by connecting and analyzing user demand data from urban systems, including transportation, energy, water, and building infrastructure. Despite this intuition, little research has been done in this area. However, a recent study exploring the correlation between electricity usage and roadway congestion revealed that electricity usage data could be leveraged to make reliable predictions regarding the beginning of peak period highway congestion, highlighting the promise of considering urban systems data for improving traffic prediction performance (Zhang and Qian, 2018).

3.2. Data resolution

For training traffic prediction models, data resolution is an important consideration. Presently, data collection technologies allow acquiring traffic data at a variety of resolutions to match the needs for traffic management and control applications. Specifically, defining the appropriate data resolution is a critical issue, especially in data-driven algorithms, because it affects the quality of spatio-temporal dependencies, which are learnable from the underlying data. This section outlines and describes the resolution of two types of data commonly used in traffic forecasting: temporal and spatial data resolutions.

3.2.1. Temporal resolution

Data resolution is strongly related to the forecasting horizon and step size. The horizon describes the forecasting time window (e.g., one day, one week, one month). At the same time, the step size denotes the frequency of forecasts within the horizon (e.g., every

Table 5
Detailed classification of current traffic prediction surveys based on the spatio-temporal Sequence Data.

Data Category	Ref.	Dataset	
Fixed Position Sensor Data	Song et al. (2020)	PeMS3	
	Shi et al. (2020), Bai et al. (2020), Guo et al. (2019) Li et al. (2018c), Huang et al. (2020), Song et al. (2020)	PeMS4	
	Yu et al. (2018), Zhang et al. (2021b) Zhang et al. (2020a), Sen et al. (2019)	PeMS7(M)	
	Song et al. (2020), Li et al. (2018c), Huang et al. (2020)	PeMS7	
	Shi et al. (2020), Bai et al. (2020), Guo et al. (2019) Huang et al. (2020), Song et al. (2020)	PeMS8	
	Zhao and Zhang (2018)	PeMS-SF	
	Wu et al. (2019), Zheng et al. (2020a), Shang (2020) Sun et al. (2021), Zhang et al. (2020b), Lu et al. (2020) Jin et al. (2021), Oreshkin et al. (2021), He et al. (2020) Boukerche and Wang (2020b), Cui et al. (2020b) Wang et al. (2020b), Kong et al. (2020), Cai et al. (2020a) Wu et al. (2020c), Pan et al. (2020), Zhou et al. (2020) Chen et al. (2020a), Fu et al. (2016) Wang et al. (2020c), Huang et al. (2019) Zhang et al. (2020a), Park et al. (2020) Chen et al. (2019), Li et al. (2018c)	PeMS-BAY	
	Li et al. (2018c), Wu et al. (2019) Jin et al. (2021), Shang (2020), He et al. (2020) Chen et al. (2019, 2020a), Zhou et al. (2020) Zhang et al. (2020a,b), Jiang and Luo (2022b) Sun et al. (2021), Lu et al. (2020) Chen et al. (2020c), Boukerche and Wang (2020b) Oreshkin et al. (2021), Kong et al. (2020) Cai et al. (2020a), Pan et al. (2020), Wang et al. (2020c) Wu et al. (2020c), Wang et al. (2020b), Wu et al. (2020a) Yang et al. (2019), Huang et al. (2019), Park et al. (2020) Zhang et al. (2018b), Pan et al. (2019a), Cui et al. (2020b)	METR-LA	
	Liao et al. (2018a)	Q-Traffic	
	Shin and Yoon (2020)	TOPIS	
	Zhao et al. (2018, 2019c)	ENG-HW	
	He et al. (2018)	HK	
	Zhao et al. (2019b), Qu et al. (2019)	DRIVENet	
	Zhang et al. (2010)	Travel Time Index data	
	Ride-hailing Data	Lin et al. (2019), Jiang et al. (2021) Zhao et al. (2020), Faghih et al. (2019)	DiDi Uber
	Taxi Data	Pan et al. (2019a)	T-Drive
		Li et al. (2020e), Bai et al. (2019)	TaxiBJ
		Zhao et al. (2019a)	SZ-taxi
		Ye et al. (2019), Wang et al. (2020e)	NYC Taxi
		Karimzadeh et al. (2021)	Porto
Bike Data	Bai et al. (2019), Wang et al. (2020e)	NYC Bike	
	Chai et al. (2018)	CHI Bike	
Bus Data	Wang et al. (2021c)	NYC Bus	
	Jang et al. (2018)	CHI Bus	
Subway Data	Liu et al. (2020a)	SHMetro	
	Liu et al. (2020a)	HZMetro	
Network Infrastructure Data	James (2021) Hou et al. (2021)	OSM A-map	

five minutes, every hour, every day). For example, a prediction model may predict daily traffic for an entire month. In other words, determining the most suitable forecasting interval (step) relates to the type of ITS application for which the algorithms are to be integrated. On the one hand, a significant prediction time horizon is preferred to provide sufficient time for control and reactions. On the other hand, a small prediction error is desirable for accurately informing the response of both TMCs and individual users.

Table 6
Applying external datasets in traffic prediction studies.

Ref.	External Data Type	Characteristics	Time
Yao and Qian (2021)	Event Traffic Data	-Use PennDOT RCRS dataset including 2,696 traffic incidents on Pittsburgh	2014
	Meteorological data	-Use Weather underground dataset collected in Pittsburgh International Airport	NA
	Social Media Data	-Use twitter dataset contains 1,782,636 tweets -Include date/time, text, user ID, language, latitude and longitude (if available), user profile location, etc.	1/23/2014~12/31/2014
Liu et al. (2019b)	Meteorological Data	-Use Wunderground dataset collected from Central Park and Manhattan Stations	NA
Liu et al. (2020d)	Meteorological data	-Collect in NYC, USA	1/1/2016~6/30/2016
	Event Data	-contains 22 weekdays, 8 weekends and holidays	1/11/2016~30/11/2016
Yao et al. (2018)	Event Traffic Data	-Collect from Didi Chuxing dataset	2/1/2017~3/26/2017
	Meteorological Data		
Li et al. (2021a)	Event Traffic Data	-Collect in England	1/7/2018~ 28/1/2020
Taghipour et al. (2020)	Meteorological Data	-Include weather conditions provided by the National Weather Service (NWS) for 94 different states in US -Include sun glare data based on sun's angle in relation to cars, solar elevation, solar azimuth, driving direction, and slope	4/2017~9/2017
	Event Traffic Data	-Contain Incident, stadium events, parades, and road races data collected by Illinois Department of Transportation (IDOT) -Include time and geographical locations of the accidents and events, severity of accident and type of events	
	Work Zone Data	-Use Chicago highways network data collected by Illinois Department of Transportation (IDOT)	
Chen et al. (2018)	Social Media Data	-Collect nearly forty thousand traffic relevant microblogs on china by searching Sina Weibo with key words	NA
Lin et al. (2017)	Social Media Data	-Collect 10000 tweets via the Twitter REST search API from Washington D.C. & Philadelphia	6/1/2013~3/31/2014
Zhang and Qian (2018)	Urban System Data	-Household-level electricity usage data from 322 household in Austin, Texas -Collect from midnight to 6 am for 251 weekdays	2014
Macioszek and Kurek (2021)	Disease Spread and Pandemic Data	-COVID19 cases reports from Poland	2019~2020

Notably, there exists a trade-off between the time horizon and prediction accuracy: a model trained to predict traffic over a small time horizon with a short step size may produce better accuracy for fine-grained prediction, but performance will degrade as the time horizon or step size increases.

Reviewing the literature reveals that most prior works on traffic prediction aggregate data over a specific time interval, generally no less than 5-min. However, this interval may not serve well for informing dynamic operational decisions (e.g., intelligent signal timing systems) that require traffic information in real-time at much shorter intervals. Regarding higher-resolution data, the increasing development of IoT technologies, in conjunction with improved data processing and storage techniques, enables high resolution data to be collected and stored by many traffic systems. Thus, it is possible to leverage the information obtained from high resolution data to predict traffic more effectively. Nevertheless, accurate traffic prediction based on high resolution data is still challenging because this type of data generally exhibits strong fluctuations. This lack of 'balance' in high resolution data presents a challenge to forecasting problems (particularly when applying traditional variants of time series techniques): sparsely represented data classes tend to be overlooked by conventional prediction models, especially when the imbalance is pronounced. In essence, the lower frequency labels are treated as an outlier (Li et al., 2020b) and excluded from the data before model training and evaluation.

3.2.2. Spatial resolution

Within the literature, spatial data resolution generally ranges from a few hundred meters to several kilometers. Different data sources can have varying spatial resolutions. For example, the spatial resolution of FCDs is generally high (data is collected on average every 100 m) compared to fixed-position detector data with a typical spacing of 1–3 km between detector positions (data collection points). Due to the high resolution, FCD can be plotted directly to illustrate the traffic dynamics. At the same time, the

fixed-position detector data must be interpolated to reconstruct the spatio-temporal network dynamics during post-processing. Thus, FCD data is generally more helpful in elucidating spatial correlations within the data because it encompasses information about the entire stretch of a roadway link in contrast to isolated and fixed-position detectors (Kessler et al., 2018). Moreover, FCD can provide traffic information in areas of the network where no sensors are deployed.

While the level of spatial resolution is customized in most traffic prediction studies, some datasets introduced specific geographic units for spatial resolution. For example, INRIX architecture utilizes TMC-Segments and XD-Segments as the basis for defining road sections to report speed and incident data. TMC location codes were initially conceived of as points on the road network, typically assigned at significant decision points, interchanges, or intersections to describe locations of traffic incidents (e.g., accidents, construction, traffic slowdowns) in a specific format independent of map vendor. The INRIX XD Segments are similar to TMC Segments in that they delineate a particular section of roadway; however, they address some of the limitations of TMC Segments, such as limited road coverage and segment overlapping and gapping, as well as segment resolution. Besides, XD Segments are defined and maintained solely by INRIX. Therefore, INRIX can create XD Segments for sections of roads or new roads that do not yet have defined TMC location codes or TMC Segments. XD Segments also have a maximum length of 1.5 miles, providing a more consistent, unambiguous, and granular definition of road segments compared to TMC Segments, which vary significantly in length depending on the distance between neighboring TMC location codes (INRIX, 2022; FHWA, 2019; Ahsani et al., 2019).

3.3. Data preprocessing

To effectively predict traffic and solve problems, collected raw data needs to be preprocessed to remove the irrelevant, redundant, noisy, and unreliable data, thus preventing misleading results while simultaneously producing more accurate insights. In general, data preprocessing is considered a preliminary step for enhancing raw data quality and employs varying techniques such as map-matching, data cleaning, data storage and aggregation, and data compression. In this subsection, we categorize and discuss the different preprocessing methods within the existing literature.

3.3.1. Map-matching

One of the most common preprocessing methods is Map-matching. This process assigns the latitude/longitude coordinates within the data to positions on the geographical road networks. Notably, it becomes necessary to apply practical map-matching algorithms to the raw positional data to produce meaningful results, and as such, many studies have been conducted on this problem. While Quddus et al. (2007), Wei et al. (2013) and Zheng (2015) introduced various map matching categorizations, from a technical perspective, Chao et al. (2020) proposed four new classifications of the existing methods: similarity models, state transition models, candidate-evolving models, and scoring models. Later, Yuan and Li (2021) provided a different categorization, arranging the methods into five groups based on the geometric information, topological structure of the road networks, and global optimal matching features of the data: Point-Distance, Path-Distance, Probability-Based, Model-Based, and Learning-Based.

Although Map-matching helps clarify and resolve uncertainties with traffic prediction data, inaccurate measurements and low sampling rates for collected data bring about quality issues, resulting in challenges for map matching approaches (Bhowmick and Narvekar, 2018; Cui et al., 2021). For example, attempting to solve the problem of match uncertainty caused by varying map densities, Quddus and Washington (2015) proposed a new shortest path and vehicle trajectory aided map-matching (stMM) algorithm to match low-frequency GPS data on a road map. In more recent work, Singh et al. (2020) applied a novel genetic algorithm to evaluate both sparse and dense GPS data for map matching. DMM, a fast map matching framework for cellular data that uses a recurrent neural network (RNN), has also been proposed to find the most likely road path given a sequence of cell towers (Singh et al., 2020). This method contrasts the commonly used Hidden Markov Models (HMM), which incur heavy computational overhead and scale poorly on large datasets.

3.3.2. Data cleaning

Accurate and efficient traffic prediction relies on clean spatio-temporal data. During data cleaning, the corrupted, poorly formatted, duplicate, or incomplete data is fixed or removed from the dataset. Notably, data cleaning techniques vary according to the underlying problem and type of data, but they can be broadly grouped into three categories:

- **Missing Data.** Hardware failures, software bugs, and human errors (Yuan and Li, 2021) can cause missing data, which reduces the power of a model. Thus, more attention should be paid to precisely estimating and interpolating the missing data. Missing data challenges can be divided into two categories: short-period missing values and long-period missing values. Short-period missing values range include outages ranging from 1 s to about 5 min and are chiefly caused by unsteady equipment or a cluttered environment. Temporal smoothness is typically employed to deduce the missing values since the surrounding temporal information is rich in these situations. On the other hand, long-period missing values, which can last hours, or even days, are principally caused by system failures. Due to the lack of surrounding contextual information, it is typical for the mean or mode of the dataset (concerning the time of day) to be used as a substitution for long-term missing values. However, the prediction error is huge in this situation because of the scarcity of surrounding temporal information (Tian et al., 2018). Various imputation methods have been proposed to solve this problem while retaining the total sample size by replacing missing values with substituted values rather than eliminating observations or variables with missing data. Cheng and Lu (2017) introduced a two-step spatio-temporal missing data reconstruction (ST-2SMR) method to consider the missing patterns

by leveraging a coarse-grained interpolation method. The authors then used a dynamic sliding window selection approach to identify which sample data was most relevant for fine-grained interpolation.

A more recent approach to data imputation was proposed by Chan et al. (2021), where Fuzzy Neural Network optimization is leveraged to perform robust missing data imputation in spatio-temporal and multi-dimensional datasets, effectively improving imputation accuracy. Notably, the imputation method was applied to the problem of dynamic vehicle rerouting, and the experimental results demonstrate the effectiveness of the imputed data. Meanwhile, in a different study by Wang and Mao (2018), missing data is estimated by the Optimum Closed Cut (OCC) method, utilizing both the road topological information and the spatio-temporal correlation among road traffic.

For the imputation of long-period consecutive missing values, Ma et al. (2020b) proposed a new methodology, namely transferred long short-term memory-based iterative estimation (TLSTM-IE). In this method, an LSTM model first learns from a complete data sequence of similar distribution to the sequence with missing values. After that, the model employs transfer learning to estimate the missing values for the incomplete dataset iteratively. In this way, the missing values can be intelligently calculated based on the observations of similar time series, in contrast to naive imputation approaches, preserving the underlying long-term data trends. A year later, Kwon et al. (2021) introduced another data imputation algorithm based on the M-LSTM architecture with parameter transfer in V2I networks. This algorithm can adopt an LSTM network to learn the representation of time series traffic data and proposes a multilayered LSTM (M-LSTM) network with parameter transfer for efficient imputation. Specifically, the parameters trained in one layer can be transferred to the following layers, reducing the training and imputing time compared to the prior approaches. Therefore, the proposed algorithm can ensure reliable input for ITS applications leveraging real-time sensor data, even when the sensor is offline.

Another missing data problem results when no historical trajectory data exists: commonly referred to as the cold start problem. Notably, it is challenging to establish predictors without historical data. For example, sensors may be continuously added and removed daily, making the cold start problem a practical consideration in growing transportation networks. In one of the earliest works, Gao et al. (2012) proposed a geo-social correlation (gSCorr) model based on social information to alleviate the cold starts problem in location prediction models. More recently, Tu et al. (2019) proposed a transfer learning-based generative model for providing personalized location recommendations, which transfers user interest and location features from app usage data to provide location recommendations in the absence of historical data. This method builds a user-location matrix, user-app matrix, and location-app matrix from the data, enabling the model to learn the latent vectors of users, locations, and apps jointly. Meanwhile, Gong et al. (2020) handled the cold-start challenge for passenger flow prediction by mining supplementary information from urban statistical data to guide the learning process. Lastly, Magalhaes et al. (2021) proposed the global and cluster-based approaches, where models are trained on data coming from all the sensors in the network (or groups of similar sensors, in the cluster-based case) to build resilient predictive functions. However, among these two approaches, the global-based approach naturally solves the cold start problem because it relies on a single prediction function independent of specific sensors.

- **Data outliers.** Outliers and erroneous data create severe problems for data analysis and can result from a myriad of problems, including human mistakes, rare exceptions, mislabels, poor calibration, malfunctioning devices, and other data-gathering issues. While each situation is unique, developing methodologies to identify and repair the outliers and erroneous data becomes an essential and challenging task. In the literature, various techniques have been used to detect outliers. Some studies have applied distance-based approaches, which leverage neighborhood computation (Wang et al., 2018c; Tran et al., 2020) to classify and exclude outlier points. Density-based approaches have also been explored, where the density of each trajectory is computed, and low-density values are excluded as outliers (Tang and He, 2017; Duggimpudi et al., 2019; Riahi-Madvar et al., 2021; Abid et al., 2021). Pattern-mining is another methodology where outliers are classified by examining the varied correlations between trajectories (Homayoun et al., 2017; Cai et al., 2020b). On the other hand, some researchers have also directly applied machine learning approaches to learning the threshold for outlier classification when training on the trajectory data (Tian et al., 2020; Belhadi et al., 2021).

After an outlier has been identified, it is crucial to understand the cause of the outlier data and why it exists. Specifically, it is desirable to determine if outliers were caused by anomalous network conditions or equipment failure. How to distinguish erroneous data from abnormal traffic conditions is a challenging task (Chen et al., 2010). However, some methods and data mining techniques can help cope with this problem. In recent work, Kullberg et al. (2021) tried to detect outliers resulting from anomalous behavior using statistical hypothesis tests such as T-score and n-step prediction. Meanwhile, Van Zoest et al. (2021) leveraged knowledge on the variable of interest and its correlation with other factors to determine whether the detected outliers result from non-recurrent events or equipment errors.

While some studies focus on detecting and removing outlier data, others have attempted to repair detected outliers in various ways. For example, Sani et al. (2019) first caught outlier behavior based on the occurrence probability of an activity sequence concerning the surrounding contextual behavior. After that, the authors replaced the identified outlier activity sequences with one of higher probability within the same contextual behavior space. Meanwhile, Liu et al. (2021) introduced TsClean, a uniform framework for identifying and repairing outliers in IoT data detected by various algorithms. Conversely, some research has demonstrated that outlier data can be helpful if analyzed separately from the primary data instead of being repaired or removed (John, 2021).

- **Data imbalance.** One of the most prevalent problems in datasets is the imbalance of data distribution or labels. While easy to identify in smaller sets, it becomes challenging for machine learning algorithms to identify these rare cases in the large datasets emerging in recent years. Recently, Peng et al. (2020a) attempted to solve the issue of imbalanced data by optimizing

the imbalanced classification algorithm concerning the output level, data level, and algorithm level. In another work, [Elamrani Abou Elassad et al. \(2020\)](#) employed a Synthetic Minority Oversampling Technique (SMOTE) to rebalance the training sets and cope with a bias towards the majority class.

3.3.3. Data storage and aggregation

The massively increased volume of spatio-temporal data collected by various IoT devices and sensors within the traffic network (e.g., mobile phones, RSUs, traffic cameras) has exposed a need for improved data storage and aggregation systems. While the increased volume of data undoubtedly improves traffic prediction accuracy, it comes with considerable computation, transmission, and storage costs. Furthermore, the complex deep learning models used in the literature for modern traffic prediction generally require the data to be aggregated to a centralized location (e.g., a data center), presenting additional costs. In an effort to address these drawbacks, some recent studies have focused on leveraging various distributed systems, like blockchain networks, to store and shard the data efficiently for use in various ITS applications ([Li et al., 2019a, 2020c](#); [Hassija et al., 2020](#)). FL is also a new approach to coping with these issues, where devices in the network train prediction models locally and only distribute the model parameters instead of the underlying data ([Liu et al., 2020e](#)). That being said, these novel concepts need substantially more research to improve our understanding of their benefits within modern traffic prediction applications.

3.3.4. Data compression

Data compression is another preprocessing method that aims to reduce the size of datasets while preserving the most critical insights. Different online and offline algorithms are proposed in the literature to compress data from large and diverse urban networks, with negligible loss of information and efficient representation of all types of data. The offline methodologies, which can be categorized as either simplification-based ([Birmingham and Lee, 2017](#); [Zhang et al., 2018a](#)) or road network-based methods ([Yang et al., 2017](#); [Li et al., 2021b](#)), are applied to the raw trajectory data in order to eliminate some irrelevant data points and increase overall compression quality. On the other hand, the online methods focus on compressing trajectory data in a timely fashion by using window-based ([Gao et al., 2018](#); [Roy et al., 2021](#)) or moving attribute-based methodologies ([Al-Hussaeni et al., 2018](#); [Zhang et al., 2021c](#)).

4. Traffic prediction methods and models

The traffic forecasting problem represents a spatio-temporal time series prediction problem, where the input contains a traffic variable (e.g., flow or speed) represented in one or more time series, and the output is a forecast of the future conditions. The prediction equation for a single time series can be represented as ([Tedjopurnomo et al., 2020](#)):

$$y_{t+T'} = f([X_{t-T+1}, X_{t-T}, \dots, X_t]) \quad (1)$$

where y_t is the predicted traffic at time t and X_t is the input denoting the attributes at the t th timestamp. T indicates the input sequence length, i.e., how many time steps of historical traffic data are used as the input. T' represents the prediction horizon, while f is an arbitrary function that calculates the traffic prediction based on the input data and depends on the actual model used.

The field of traffic prediction has existed for almost six decades and covers a wide array of methodologies which can be divided into two main categories. Generally, traffic forecasting models are either univariate, where the input data consists of a single time series from one individual sensor, or multivariate, where the time series of multiple sensors throughout the network are considered within a single framework ([Nagy and Simon, 2018](#)). Multivariate time series forecasting methods inherently assume interdependencies among variables. In other words, each variable depends not only on its historical values but also on the historical values of other variables. Today, most models are multivariate and attempt to capture the complex spatio-temporal relationships between the traffic data collected across the transportation network.

Most notably, the spatio-temporal relationship is fundamental to traffic forecasting. Many approaches for precise traffic prediction consider the topological structure of the urban road network (spatial) and the dynamic network changes over time (temporal). However, only recently has the joint spatio-temporal relationship been considered (e.g., the impact of historical trends from other networked nodes on the future state of a given node). The resulting data-driven traffic forecasting methodologies can be broadly categorized into three groups: (1) statistical methods, (2) traditional machine learning, and (3) deep learning. Statistical methods are especially suitable for smaller datasets, benefiting from a clear and simplified computational structure compared to more advanced machine learning approaches. However, traditional machine learning algorithms are better suited for capturing complex non-linear correlations and processing high-dimensional data inherent in transportation-related data.

Meanwhile, increased computational capabilities and data access have resulted in the popularity of deep learning-based models, which have highly complex structures and have been shown to outperform many of the widely used traditional approaches assuming sufficient data. That being said, given the importance of jointly capturing the localized and network-wide spatio-temporal relationships and the varying trade-offs of the individual models, hybrid approaches combining multiple models have become popular amongst researchers in recent years ([Gu et al., 2019](#); [Hajirahimi and Khashei, 2019](#); [Alsolami et al., 2020](#)). Generally, the hybrid approaches follow an ensemble-based design [Dietterich \(2000\)](#), where predictions from multiple models are weighted and fused to produce the final prediction. Recently in traffic forecasting, the new Bayesian combination method (NBCM) proposed by [Wang et al. \(2014a\)](#) has been demonstrated to make efficient use of diverse sub-predictors. Alternative hybrid approaches seek to connect different DL architectures in sequence, such that the hidden representation learned and output by one network can be used as input to a subsequent model ([Li et al., 2021a](#)). In these designs, a convolution-focused model (e.g., CNN, GCN) extracts essential

features from the input data with efficiently deep architecture and subsequently inputs the encoded features into an RNN-based gated design to capture the temporal dependencies. We discuss the most interesting hybrid approaches in the existing literature in future subsections.

In the remainder of this subsection, we provide a model-driven review of the existing literature on multivariate traffic forecasting models, highlighting their scientific contributions and unique approaches, which improved our understanding and ability to forecast traffic.

4.1. Statistical methods

Statistical models, which treat forecasting as a regression problem, are widely applied to short-term traffic prediction (Yang et al., 2020) due to their clear computational structure and strong theoretical interpretation ability. For the non-learning approaches, such as k-nearest-neighbors (KNN) (Zhang et al., 2013; Harrou et al., 2020), historical average (HA), and the conventional vector autoregressive (VAR) models, considering the effects of both the downstream and upstream traffic generally improves prediction performance.

Conversely, non-linear regression (Högberg, 1976), average algorithm, smoothing, Bayesian networks (BN), and Kalman filtering (KF) are common parametric techniques for analyzing and forecasting time series using observed ground truth data. In addition, the autoregressive integrated moving average (ARIMA) (Williams and Hoel, 2003), and its variants, are among the most consolidated approaches and have been frequently applied to traffic prediction problems (Williams and Hoel, 2003; Shekhar and Williams, 2007; Li et al., 2012; Moreira-Matias et al., 2013; Lippi et al., 2013; Wagner-Muns et al., 2017). ARIMA, an extended form of the autoregressive-moving average (ARMA) model, is specially used for future points prediction in time series data. Other variants of ARIMA, including Seasonal ARIMA (SARIMA), ARIMA with explanatory variable (ARIMAX), Kohonen ARIMA (KARIMA), and Vector ARIMA also exist in the literature and have been shown to improve traffic flow prediction accuracy.

Although suitable for smaller datasets with only a few time series having short observation periods, due to a simplified and transparent computational structure, the statistical methods generally lack the complexity to consider the spatio-temporal relationship and instead focus exclusively on temporal time series data. Thus, the prediction capabilities of these statistical models are limited for the traffic prediction application due to their inability to account for the inherent spatial dependencies. To overcome this challenge, some studies developed extended time series-based approaches, where the extension took into account spatial and temporal interactions in a new manner, such as ST-ARIMA (Min et al., 2009, 2010) and VARMA (Min and Wynter, 2011). Since these approaches use explicit parameterization functions with strong assumptions during the modeling process, they are unsuitable for simulating real traffic application scenarios (Cheng et al., 2018). In contrast, non-parametric spatio-temporal models such as ST-KNNs (Wu et al., 2014), without any requirement for prior knowledge or explicit mathematical expression, can easily achieve satisfactory portability and comparatively greater prediction accuracy. However, these statistical methods are ineffective at processing complex (high-dimensional) and dynamic time series data. For example, ST-KNNs, which use distance functions and correlation coefficients to identify spatial neighbors and measure the temporal interaction by only considering the temporal closeness of traffic, cannot fully reflect the essential features of road traffic.

Despite the drawbacks of classical statistical approaches when processing the highly-dimensional ITS data, some hybrid methods integrating classical and DL-based methods have emerged to overcome the existing challenges (Li et al., 2017). For example, Li et al. (2019b) recognized the strong ability of state space and hidden Markov models for capturing the importance of exogenous variables (e.g., weather, special events). However, these methods are not suitable for modeling long-term dependencies or non-linear relationships, two key considerations when forecasting traffic. Accordingly, the authors relax the first-order Markov assumption to include a sequence of deterministic states generated by an RNN-based model, effectively capturing the long-term dependencies. A data-driven sparse Bayesian learning approach (SPL) is proposed to determine the relevant importance of features at the current time step, pruning away any low-impact or redundant features while mitigating overfitting to exogenous variables. Consequently, the model can process many exogenous features while selectively considering the most critical subset of external data at the current moment. For MTP, this is a promising approach for training a model to capture the sparsely represented but high-intensity exogenous variables (e.g., a rare accident on a bridge or a monthly concert event).

In summary, exogenous variables are important considerations for traffic prediction to improve our models. However, the case study in this paper is limited and only considers exogenous temporal variables, such as whether it is a workday or peak period. More research is needed to see if this approach to considering exogenous variables can be further improved and integrated into the cutting-edge MTP models.

4.2. Traditional machine learning

In addition to the classic statistical methods, ML-based methods are becoming increasingly popular in traffic prediction. Compared to classical methods, ML models can provide more robust generalization capabilities while also being able to learn more complex relationships and better adapt to the changing conditions of the traffic network (Yang et al., 2020).

As shown in Table 7, traditional ML approaches can be segmented into three categories: feature-based methods, Gaussian process models and state-space models (Yin et al., 2021). Feature-based methods are used to train regression models for some practical traffic prediction problems using human-engineered traffic features (Lin et al., 2017; Li et al., 2018a; Guan et al., 2018; Tang et al., 2018; Kowalek et al., 2019). While the feature-based models are easy to implement, Gaussian process approaches utilize multiple kernel functions to characterize the inner features of traffic data, incorporating both spatial and temporal correlations

Table 7
Traditional ML traffic prediction methods.

Characteristic	Model	Advantage	Disadvantage
Process high-dimensional data and capture complex non-linear relationships.	Feature-based Models	<ul style="list-style-type: none"> -Developing a regression model with human-engineered features -Easy-to-implement method 	-Limited and Inconsistent performance across various applications due to reduced complexity and the human-engineered features
	Gaussian process Models	<ul style="list-style-type: none"> -Use various kernel functions to model the inner characteristics of traffic data -Effective and feasible in traffic prediction based models -Gets spatial and temporal relationship simultaneously 	<ul style="list-style-type: none"> -Hard to process large amount of data -Have higher computational load and storage pressure than feature-based models
	State-space Models	<ul style="list-style-type: none"> -Assume that the observations are generated by Markovian hidden states. -Can naturally represent the uncertainty of the system and better capture the latent structure of the spatio-temporal data. 	-The nonlinearity of these models is restricted, and they are not always suitable for simulating complex and dynamic traffic data.

simultaneously (Matthews et al., 2018; Salinas et al., 2019; Liu et al., 2020c). Although research has demonstrated Gaussian methods to be useful and practical for traffic prediction, they have higher computational complexity ($O(N^3)$) and data storage requirements ($O(N^3)$) than feature-based models (Datta et al., 2016). This presents scalability issues when working with large datasets, which are increasingly common in traffic prediction applications.

Meanwhile, state-space models presume that the observations are derived from Markovian hidden states. State-space models excel at capturing latent data structures and can naturally model the uncertainty of a system, which is a desirable property for traffic prediction applications. However, these models have difficulty generalizing non-linear relationships (Tan et al., 2016; Duan et al., 2018; Shin and Sunwoo, 2018; Ishibashi et al., 2018; Gong et al., 2018), and they are not always the best choice for modeling complex and dynamic traffic data, especially for long-term forecasting.

4.3. Deep learning for multivariate traffic prediction

Deep learning (DL) models have become increasingly popular in recent years, and research has demonstrated their strong applicability for MTP and capturing non-linearity (Lai et al., 2018; Shih et al., 2019). Previously, deep learning approaches were widely considered impractical due to their increased data, computation, and storage requirements compared to classical methods. However, stemming from a desire to utilize better the enormous quantities of data generated in the previous two decades, scientists have demonstrated that DL methods have the potential to significantly outperform the widely used approaches if given sufficient resources. This subsection introduces the most popular and widely used DL models within the literature, focusing on the traffic prediction application.

4.3.1. Basic DNN models

- **MLP.** The multi-Layer perceptrons (MLP), sometimes called vanilla neural networks, are a feed-forward neural network (FNN) class generally having more than three layers of fully-connected neurons. A defining characteristic of these networks is the absence of loops, meaning that neurons form connections between layers but never connect within a layer. The information flows from the input layer to the output layer, hence the name feed-forward. MLP uses the backpropagation training technique, a supervised learning algorithm that learns the desired output from varied input data. Like RNN, exploding and vanishing gradients present challenges and can produce poor results when training on large MLP networks. To overcome this challenge, a resilient backpropagation algorithm, Rprop, was proposed to update the learning rate for each neuron connection dynamically. Later, Oliveira et al. (2016) showcased the benefits of Rprop in network traffic prediction by comparing the model with standard MLP, RNN, and stacked autoencoder (SAE). Additionally, Song et al. (2017) compared the MLP-based traffic prediction model with a CNN-based approach, and the results showed CNN outperforming MLP models. Notably, MLP has issues reflecting local dependencies in the traffic data and is also more prone to noise than other approaches. While initially popular for traffic prediction due to its ability to generalize the nonlinear relationships in traffic data, MLP is falling out of use in favor of more powerful DL models better suited for traffic prediction (e.g., RNN and CNN).
- **AE.** Autoencoder (AE) methods are a variety of artificial neural networks that learn an efficient encoding of unlabeled data in an unsupervised manner. In traffic prediction, AEs are primarily leveraged to impute or effectively compress data (Ghosh et al., 2017; Wang et al., 2018a), reducing its dimensionality while retaining the most vital information. The network design is relatively simple at the highest level, with two main portions: (1) an encoder that compresses the input (x) into fewer bits; and (2) a decoder that takes the sparse representation of the input (x') and outputs the original value. More specifically, AEs (Liou et al., 2014) are based upon unsupervised learning (unlabeled training example) techniques, having a deliberately placed bottleneck between the encoder and decoder for learning a compressed representation of the input. Moreover, AEs comprise the same number of input and output units along with one or more hidden layers (Fig. 2). Eqs. (2) and

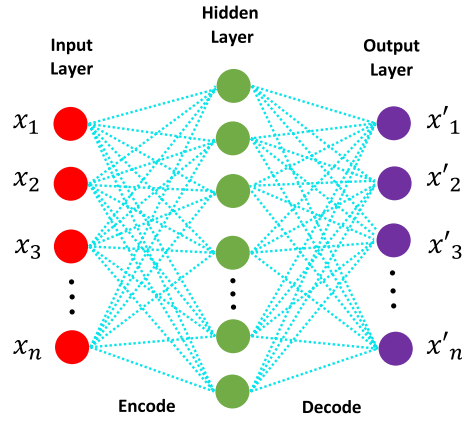


Fig. 2. Structure of AE.

(3) represent the structure of AEs with a single hidden layer where f and g are encoding and decoding functions, respectively; b is the bias value; W and W' are encoding and decoding weight matrix. AEs are utilized for tasks such as dimensionality reduction, classification, and anomaly detection because they generate similar input at the output layer. In particular, the bottleneck is an essential feature of AE that restricts the amount of data allowed to traverse the network. By learning to recreate the input from a compressed representation, AEs learn the most fundamental aspects of sparse input material. Besides, AEs use backpropagation (BP) techniques to train an approximation identity function that generates a target value comparable to the input. However, they demand a pretraining stage that may lead to a vanishing gradient problem (Mahmud et al., 2018). The following example presents a simple introduction to the internal workflow of an AE. First, let us assume a set of training samples $x_1, x_2, x_3, \dots, x_n$. After passing through the AE, the first input x_i is encoded to hidden form $y(x)$ using Eq. (2) and then decoded back to x'_i using Eq. (3). The training procedure aims to minimize the reconstruction error $\delta(x, g(f(x))) = \delta(x, x')$, i.e., difference between input and re-constructed output. That being said, if the count of neurons in hidden layers is the same or greater than the input and output layers, then the model behaves the same as an identity function (e.g., it just copies the inputs). Therefore, determining the ideal number of hidden units is a challenging and application-specific task requiring careful fine-tuning through trial and error (Palm, 2012).

$$y = f(Wx + b) \quad (2)$$

$$x' = g(W'y + b') \quad (3)$$

For MTP applications, AEs can be applied in multiple ways, including for data imputation (Ghosh et al., 2017), data compression (Wang et al., 2018a), and feature extraction (Liu et al., 2019c). Specifically for traffic prediction, AEs are typically stacked in sequence to hierarchically extract the most important features from the traffic data. Referred to as Stacked AEs (SAEs), the hidden representation from the first encoder is passed to the subsequent AE layer, and this process can continue for a variable number of layers. The dimensionality is reduced as the input data passes through each layer, and only the most important spatio-temporal features are passed to the next layer. In this way, SAEs present another method for performing hierarchical feature extraction, in contrast to convolutional or gated recurrent network architectures (Wang and Boukerche, 2020). However, deep SAEs require a large volume of highly-dimensional input data for training, which must be preprocessed, and suffer from long training times before using the output features to train the final prediction model or layer. Furthermore, due to the black-box nature of NNs and the hidden representation output by the encoder, the features learned by the SAE are not well understood.

- **RBM & DBN.** Restricted Boltzmann Machines (RBMs) are stochastic neural networks that learn in an unsupervised fashion and are widely used to model binary variables in the literature (Fig. 3). Different from the traditional neural network models, RBM joins the feature learning part based on the original multilayer neural network. The feature learning portion imitates the human brain to handle data signal classification. The specific operation is to increase the partial connection of the convolution layer and dimension layer in front of the original fully connected network layer. In simple terms, the traditional shallow neural network projections step is from the feature mapping to the value, and the characters are artificially selected. RBM projections step is from the signal to the feature and then to the value. The network freely chooses the data characteristics. For example, Ma et al. (2015) applied a stacked RBM to extract temporal and spatial correlations from traffic flow data for traffic congestion forecasting. Notably, RBMs effectively extract representative characteristics from traffic flow data and capture the interaction between variables. Additionally, multiple variations of the RBM have been utilized in the literature, including Deep Belief Network (DBN) and Deep Boltzmann Machine (DBM). DBN combines several RBMs to capture low-pass network

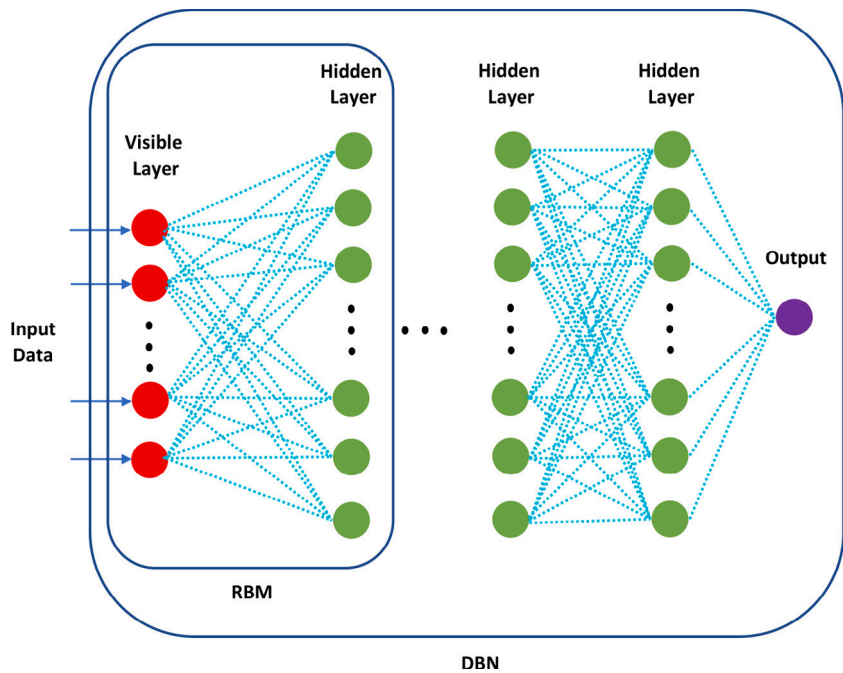


Fig. 3. Structure RBM & DBN.

traffic components and elucidate the underlying data's long-range dependencies. More specifically, DBN can learn internal representations from the input data (higher layers capture complex statistical structures) and make inferences quickly (hidden layers state computation).

Although DBN can effectively learn the object features in an unsupervised manner, this model has difficulty mapping the complex relationships between the elements in the traffic system. To overcome this challenge, Bao et al. (2021) introduced an improved DBN by integrating SVR with DBN to map the complex relationship between the elements in the traffic system. On the other hand, DBM is a modified Boltzmann Machine that is unrestricted, containing more hidden layers and directionless connections between nodes. DBM provides additional advantages resulting from its top-down feedback loop, allowing higher-level information to be leveraged for resolving uncertainty regarding basic or intermediate-level features.

4.3.2. CNN

Convolutional neural networks (CNNs) are one of the most representative supervised deep learning approaches for modeling spatial or temporal correlations. CNN architecture contains four main layers: convolutional, ReLU (activation), pooling, and fully connected. Moreover, some studies denote the input and output as two additional layers positioned at the beginning and end of the network, respectively. During the training process, the model first extracts features from image data into convolution layers, which are then convolved using a filter with learnable weights. Notably, different models and applications require varying numbers of convolutional layers. Following the convolution layer, the ReLU layer uses an activation function to set the negative values in the incoming data to 0. Next, in the pooling layer, the dimensionality of the input is reduced by subsampling (pooling) layers through a feature map size reduction method. This layer also helps to reduce the number of parameters by discarding unimportant information while keeping the most marginal and critical parameters. Lastly, the input is flattened and fed into fully connected layers to map the input to the required output dimension. The bundle of convolution layer and pooling layer can repeat multiple times to generate a deeper neural network. For example, Fig. 4 shows a CNN architecture with two convolution layers and two pooling layers. With more layers in the CNN network, higher-level information is hierarchically captured. The drawback is that more layers add additional parameters, which may lead to over-fitting and increased training times.

For MTP applications (e.g., flow or speed), CNNs provide one method for capturing the localized spatio-temporal dependencies between links or nodes within the roadway network. The intuition behind applying CNN to MTP problems revolves around the convolution and pooling operations, which collectively learn localized features from the input data. Notably, these features are discovered in an unsupervised manner spontaneously from the raw input, enabling the model to understand previously unknown localized dependencies. Since traffic time series data from nodes in close proximity along the same link are likely to be related, multiple works have demonstrated the effectiveness of leveraging convolution-based DL models for extracting localized spatial dependencies from the multivariate time series data (Wu et al., 2018).

In one of the first works to explore using CNN for MTP, Song et al. (2017) compared two versions of the well-researched MLP architecture with a CNN-based model for predicting traffic speed across four directly connected roadway links within Seoul.

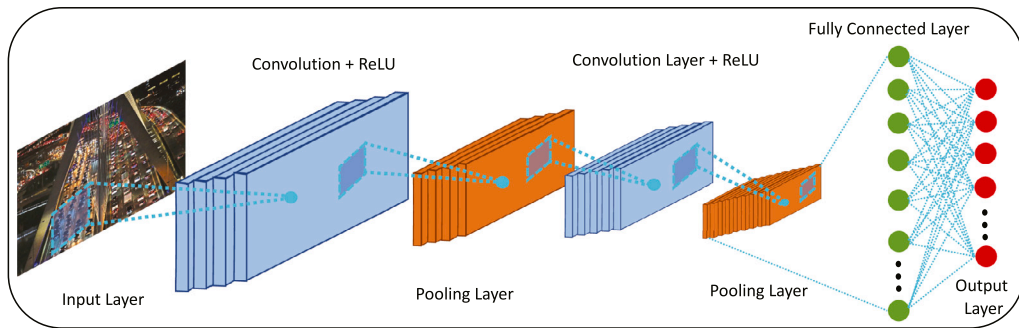


Fig. 4. Structure of CNN.

In this design, a CNN model is developed for each link to capture the intra-link node dependencies, while a fifth CNN models the temporal features. Lastly, each model's output is aggregated to produce the final prediction. The case study demonstrated that the CNN approach could better represent the localized spatial dependencies between nodes within the same link, improving prediction accuracy by upwards of 6% compared to the state-of-the-art MLP approaches. However, the CNN model used in this paper only performs convolution on the subset of nodes within a single link, ignoring the inter-link dependencies. Expanding the CNN model's spatial dimension presents a trade-off between performance and computational demand. Capturing the long-range interlink dependencies between distant infrastructure (e.g., the relationship of flows between a suburban freeway and the CBD/downtown area) requires many convolutional layers and increases the network depth.

Later, in an alternative approach by Wu et al. (2018), a hybrid model is proposed by combining CNN with LSTM to capture the spatial and temporal traffic trends separately. Specifically, the CNN model extracts the spatial features of the traffic time series, while the LSTM model extracts the short- and long-term temporal patterns. Motivated by related work in computer vision, Springenberg et al. (2014), the authors decide to perform full convolution within their CNN model by removing the pooling layers. Notably, this contrasts with previous work where the conventional CNN design is applied. Since the primary purpose of pooling is to reduce the size of feature maps, removing the pooling layers can help preserve the underlying structure of the roadway network, which would otherwise be distorted. Interestingly, related work in crowd flow prediction also reported improved prediction accuracy when using a full convolutional network (Zhang et al., 2017).

Building upon the earlier approaches, Liu et al. (2019a) proposed DeepTRP: a modified convolutional model for MTP forecasting, which also removed the pooling layer. In the case study, the authors take a big data approach to regional traffic forecasting, where an area consisting of 17,017 road segments is divided into a 4×8 grid of 32 regions. In contrast to Wu et al. (2018), DeepTRP performs 2D convolution on a 2D matrix representation of the traffic data while pre-processing the data into hourly, daily, and weekly groups. By using 2D convolution, the inter-regional relationships can be better learned during model training, whereas Wu et al. (2018) outlined that 1D convolution is best suited for small study areas consisting of only a few road segments. However, 2D convolution and deeper convolutional networks are more challenging to train, can be prone to exploding gradients, and run the risk of overfitting. To mitigate against this, Liu et al. (2019a) proposed stacking a variable number of residual units between the first and last convolutional layers, a method previously successful in related areas such as image recognition (He et al., 2016). Notably, their sensitivity analysis revealed that increasing the number of residual units beyond two harmed performance and that more fine-grained partitioning of regions also increased prediction error. For future work, experimentation with a significantly deeper convolution-based model is necessary to assess whether additional convolutions can better capture the complex and fine-grained regional dependencies and allow for a finer-grained regional representation.

Additionally, other works have identified the importance of considering regional dependencies within the existing models. Notably, the standard convolutional methods proposed in many works operate on a fine-grained representation of the transportation network, focusing only on the spatio-temporal relationships between directly adjacent nodes (e.g., neighborhoods). Recognizing this deficiency, Pan et al. (2019b) proposed a matrix factorization method for extracting the high-level, regional relationships from multivariate traffic flow data. The most notable contribution of this work is the design approach taken by the authors, who ensure their proposed regional feature extraction method is compatible with the existing state-of-the-art finer-grained methods (e.g., Zhang et al., 2017). In future research, it would be interesting to experiment with a combined weighted prediction approach, where the regional predictions can be weighted and combined with the outputs of existing cutting-edge but finer-grained methods.

Another modification to the traditional CNN structure proposed the convolutional LSTM (Conv-LSTM) to jointly consider the spatio-temporal features within the multivariate traffic time series data (Zheng et al., 2020b). The primary contribution of this work is to combine the CNN with the LSTM network sequentially, enabling the model to jointly consider the short- and long-term spatio-temporal trends from a single set of input data. Prior hybrid approaches, including Wu et al. (2018) treated the spatial and temporal features separately, which may overlook some of the intricate relationships between the spatial location of the infrastructure and the period of interest (e.g., the impacts of rush hour CBD congestion on the traffic in residential areas). Using the convolutional layers' output as input to the LSTM network, the spatial relationships between nearby flow sequences are captured concerning changing temporal network dynamics, such as differences in weekday and weekend traffic patterns. Case study results demonstrate

that the Conv-LSTM model can outperform the standalone LSTM and SVM models, while Dai et al. (2020) also utilized a Conv-LSTM for traffic forecasting with great success. In Dai et al. (2020), the performance of a hybrid model combining the Conv-LSTM and GCN architectures is explored, incorporating cutting-edge convolutional methods in both the temporal and spatial feature extraction modules. The comparison highlighted that combining both models yielded about a 4% increase in RMSE accuracy compared to using a Conv-LSTM or GCN alone. Consequently, it is vital to consider specialized approaches for extracting the study area's spatial and temporal features.

While the works mentioned above apply more traditional CNN model structures to MTP problems, related research proposed modified convolution-based networks optimized for time series forecasting. Notably, the seminal work of Lea et al. (2016), who first introduced the temporal convolution network (TCN) for capturing global patterns across high-dimensional time series with diverse scales in the action segmentation field, has inspired innovation. Before this work, the primary method involved using two high-level DL models, usually a CNN and an RNN, to capture local and global patterns. In contrast, the proposed TCN method provides a unified approach for hierarchically extracting temporal features at the local, intermediate, and global levels using a combination of 1D convolutions, pooling, and channel-wise normalization.

Notably, the 1D Convolution layer in TCN has two differences compared to standard CNNs: (1) Causality, which implies that each layer output is produced looking only at the most recent historical samples, and (2) Dilation, which ensures that each filter only examines a subset of historical samples. More simply, rather than performing convolution on a contiguous time window of input samples, dilation inserts a fixed step between convolution inputs, thus increasing the receptive field while keeping the number of parameters low (Burrello et al., 2020). Besides, the 1D convolution-based design improves training time significantly for time series modeling compared to RNN-based models due to the activation functions being computed hierarchically instead of sequentially, as in RNN. While the TCN was designed for action segmentation, it solved a shared problem with MTP models: how to jointly capture the low-level and localized spatio-temporal features with the high-level temporal trends (e.g., seasonality).

More recent work has attempted to improve the original TCN design and apply it to MTP forecasting. Sen et al. (2019) proposed DeepGLO: a TCN-based hybrid forecasting model designed for a large-scale MTP problem, including time series from 228 traffic sensors. An essential contribution of this work is the ability of DeepGLO to consider the global network patterns during prediction. In contrast, alternative approaches (e.g., CNN, RNN) generally focus on only the local past data (e.g., the past data from a specific detector only), despite being trained on the entire set of time series. DeepGLO thinks globally but acts locally by using a temporally regularized matrix factorization (TRMF) model, normalized by a TCN, to output features representing the global trends. Notably, the regularization process by the TCN is required to capture non-linear relationships, as the standard TRMF approach can only describe linear trends. The output global-level features are then used as covariates for another TCN model, producing a final model that jointly considers the historical local and global trends during prediction. Case study results demonstrate the effectiveness of DeepGLO for MTP compared to the existing cutting-edge methods, including the spatio-temporal graph convolutional network (STGCN), a competing approach discussed later. However, DeepGLO is not directly compared to other competing hybrid models, making cross-comparison difficult.

In summary, research has demonstrated the effectiveness of CNN-based models for capturing complex spatio-temporal dynamics in multivariate traffic time series data. Both 1D and 2D convolution show promise for identifying relationships between adjacent time series at different spatial granularity; however, they require different input representations and varying sequentially stacked layers and filters. Most recently, with great success, cutting-edge approaches have integrated the convolutional operation into new hybrid models, such as Conv-LSTM, TCN, and DeepGLO. However, it is crucial to consider that the traditional convolution operation is designed for data in Euclidean space, which can distort the underlying roadway network structure and destroy critical spatial relationships. Furthermore, the pooling operation traditionally included in the CNN model further distorts the road network through subsampling, and multiple research works found success by using fully convolutional approaches instead. That being said, cross-comparison between the various methods within the literature is challenging because the studies use varying datasets and compare their methods to different baselines. Standardized baseline datasets for other MTP applications and a standardized set of baseline models for comparison with new approaches would greatly enrich the existing literature.

4.3.3. RNN

Recurrent neural networks (RNN) represent an evolution of the FNN model, designed to improve performance for time series data. The conventional FNN ANN designs do not perform well when modeling sequences and time series as they lack a memory component. In contrast, RNNs work well with sequential and time series data due to adding a memory cell. More specifically, RNN models process input sequentially and maintain an internal memory by constantly updating a set of hidden states. In this way, RNN-based models consider the dependency between adjacent time steps, whereas traditional FNNs treat time steps independently. During prediction, the input at the current timestep is taken along with the output from the previous timestep as depicted in Fig. 5(a). In other words, the RNN learns by updating a state s_t depending on the prior output at state $t - 1$ and the current input x_t at time t . The RNN network is represented by the Eq. (4) and (5), where w_i is the weight matrix from the input to the hidden layer, w_R is the recurrent weight matrix, and w_o is the weights associated with the hidden to the output layer. Memory is realized by configuring each hidden neuron with a feedback loop that passes the current output as the input during the next step. Due to the RNN network's use of loops, the original Backpropagation (BP) approach is ineffective. Therefore, the network is trained using Backpropagation Through Time (BPTT) (Werbos, 1990). BPTT is a version of normal BP in which gradient descent is applied to an unrolled network. While unfolding all timesteps, the error of the current time step is backpropagated to the first timestep in BPTT. Besides, BPTT can be expensive for high-dimensional time series analyses with several timesteps.

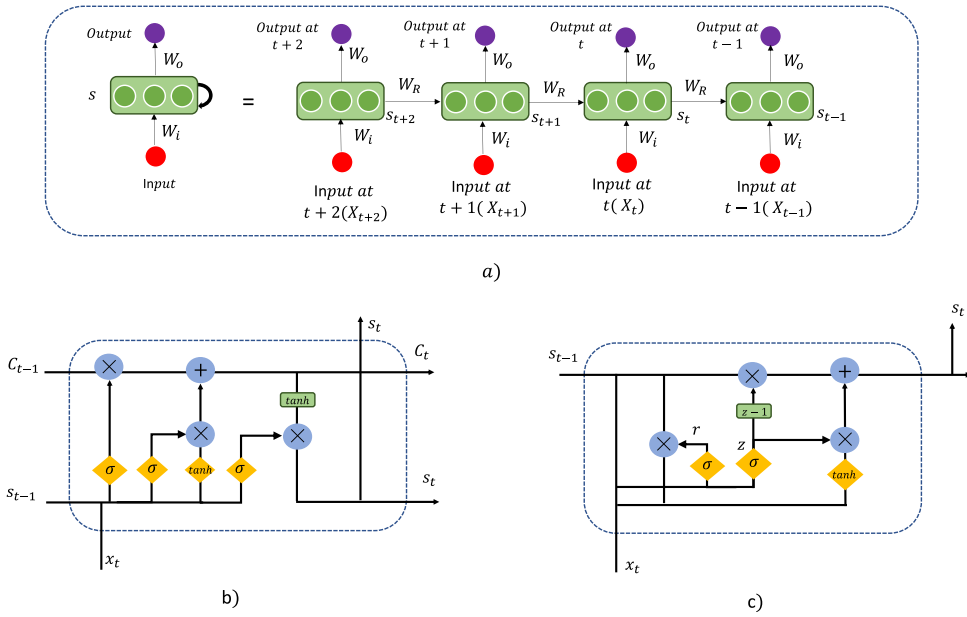


Fig. 5. Structure of (a) RNN, (b) LSTM and (c) GRU.

$$s_t = f_s(w_i x_t + w_R s_{t-1} + b_s) \tag{4}$$

$$y_t = f_i(w_o s_t + b_y) \tag{5}$$

The first proposed RNN was a two-layer, fully-connected neural network with a feedback loop in the hidden layer, with the addition of the feedback loop being the primary contribution. However, this simple design becomes prone to gradient vanishing or exploding when training large networks due to the backpropagation involved, presenting problems when working with extended multivariate traffic time series. More specifically, adding more hidden layers in the network results in more derivatives that must be multiplied during each network pass. As a result, if the products are large, the gradient can explode in large networks and result in arithmetic overload. On the other hand, if the derivatives are very small, the continued multiplication will result in a vanishing gradient and potential arithmetic underflow. Summarily, for conventional RNNs, the effect of past input on output decays exponentially as it cycles around the recurrent connection. Hence, RNNs can only deal with the problem of sequence handling for short-term dependencies and fail when modeling long-term dependencies due to the vanishing gradient problem. The literature has proposed a new class of Gated RNN with multiple design approaches for various applications to address this problem. The most common variations of Gated RNN are the Long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Cho et al., 2014) models.

For MTP applications, LSTM is favored over conventional RNN due to its ability to remember long-range dependencies selectively. Consequently, the LSTM network can improve modeling performance for time series containing both long- and short-term dependencies. Similar to the conventional RNN, the LSTM NN includes an input layer, a recurrent hidden layer, and an output layer. However, it treats the hidden layer as a memory unit that helps remember the longer-term dependencies. The hidden layer comprises a series of memory blocks called cells, each with a self-connection that enables retaining a dynamic temporal state. In addition, each block consists of three multiplicative gates, i.e., one input gate, one output gate, and one forget gate, as shown in Fig. 5(b). The multiplicative gate enables cells to keep and retrieve information over a long period, which helps mitigate the vanishing gradient problem. Each cell transfers cell state and hidden state to the immediate next cell. Here, the input gate controls whether the memory cell is updated or not. In contrast, the forget gate indicates whether the memory cell should be set to 0, i.e., the LSTM network no longer requires the information. This way, the LSTM model can selectively discriminate between useful information as conditions change over time.

Moreover, a primary difference between LSTM and RNN is that the former uses a forget gate to control the state of a cell and ensure that it does not degrade, further improving the long-range dependency modeling performance. The Output gate controls if the current information of the memory cell is to be made visible or not, i.e., it generates output for the LSTM network, which can be an interpretable prediction or a hidden state. In cases where the output is a hidden state (e.g., latent representation), the hidden state is generally used as the input features for another layer or model, such as an MLP or fully connected layer, to produce the final prediction output (Ziat et al., 2017). In other words, the hidden representation can be interpreted as a compressed representation of the input data containing the most important features. Consequently, if the hidden representation is learned, such as in the

RNN-based STNN model proposed in Ziat et al. (2017), it can provide one method for extracting the most important features from the stochastic and noisy traffic data. Eqs. (6)–(10) depict the working of the LSTM network:

$$i = \sigma(W_i x_t + U_i s_{t-1} + V_i C_{t-1} + b_i) \quad (6)$$

$$f = \sigma(W_f x_t + U_f s_{t-1} + V_f C_{t-1} + b_f) \quad (7)$$

$$o = \sigma(W_o x_t + U_o s_{t-1} + V_o C_{t-1} + b_o) \quad (8)$$

$$C_t = f * C_{t-1} + i - g(W_c x_t + W_c s_{t-1} + b_c) \quad (9)$$

$$s_t = o * h(C_t) \quad (10)$$

where i , f , o represent input, forget, and output gates activation vectors, respectively; and C_t is the cell state vector, while h_t is the output vector. W , U , and V are weight matrices.

To address the issue of vanishing gradient that the LSTM network experienced, Chung et al. (2014) introduced the GRU model. Compared to LSTM, GRU training is more efficient due to a smaller number of learnable parameters, as there is no separate memory block. Despite having less complexity, GRU has been demonstrated to perform similarly to LSTM for most MTP applications. The reset gate and the update gate of the memory cell of a GRU unit determine what information should be sent to create the output. As shown in Fig. 5(c), unlike LSTM, GRU merges the forget gate and input gate into a reset gate. The reset gate functions similarly to the forget gate of LSTM in that it determines whether or not to integrate present and past information. In contrast, the update gate defines how much prior knowledge should be retained. The following equations define the GRU NN's inner workings (Migilani and Kumar, 2019):

$$z = \sigma(x_t U_z + s_{t-1} W_z) \quad (11)$$

$$r = \sigma(x_t U_r + s_{t-1} W_r) \quad (12)$$

$$h = \tanh(x_t U_h + (s_{t-1} * r) W_h) \quad (13)$$

$$s_t = (1 - z) * s_{t-1} + z * h \quad (14)$$

where, x_t indicates input vector, z shows update gate vector, r is the reset gate vector, s_t indicates the output vector and W , U are weight parameter matrices. σ is the sigmoid function that outputs a number between 0 and 1 where 1 signifies that all information gets to pass through the cell state, whereas 0 signifies no information is passed through. However, a \tanh function outputs value a between -1 and 1 .

In the recent literature, LSTM and GRU have played essential roles in improving our MTP capabilities and understanding. With the emergence of ITS, data generation has increased dramatically, and many traffic time series datasets are highly dimensional (e.g., one dimension for each detector in the network). Adding a memory cell within the LSTM and GRU enables the models to learn the long-term temporal dependencies in the data, even if they are sparsely represented, making them highly applicable to MTP applications. Notably, in recent approaches, LSTM and GRU are seldom used alone; instead, researchers propose unique ways to incorporate them into hybrid designs (Dai et al., 2020).

In Luo et al. (2019) study, a modified KNN-LSTM architecture was designed to optimize the LSTM-based model for traffic flow prediction. Notably, the KNN algorithm extracts the spatio-temporal correlations between time series (sensing stations) within the network by computing the similarity between the measurements at each station. The most novel contribution of the paper is the use of a rank-exponent method for weighting the predictions from the K-nearest stations to produce the final prediction at any test station. Case study results reveal that integrating KNN and weighted rank-exponent methods with the 2-layer LSTM improved prediction performance compared to existing ML and DL approaches, including standalone LSTM. However, the success of the proposed method is heavily dependent on similarity in the magnitude and distribution of time series between different stations. If the traffic stream between stations exhibits high variance, the weighted rank-exponent process could produce inconsistent results when predicting flows for some stations.

During the same year, Zhao et al. (2019a) proposed an alternative RNN-based hybrid MTP model by combining GRU with GCN to create the T-GCN architecture. In the novel approach, a T-GCN cell is designed, which performs the graph convolution on input data before feeding it through a GRU-inspired series of gates (e.g., reset and update gates) to produce the final prediction. The GRU was chosen over LSTM due to its similar performance but reduced complexity (Chung et al., 2014), having fewer parameters to train. Graph convolution was used to extract the spatial features in favor of euclidean convolution, as graphs have been demonstrated better to represent the transportation network's link and node structure. Notable experimental results showcase that the T-GCN design can outperform the predictions of other popular models, including the standalone GCN and GRU, highlighting the importance of jointly considering the spatio-temporal trends in traffic data. Interestingly, the model predicts poorly at the local minima/maxima, which could be a byproduct of the smooth filters applied during graph convolution. Moreover, the small changes resulting from smoothing

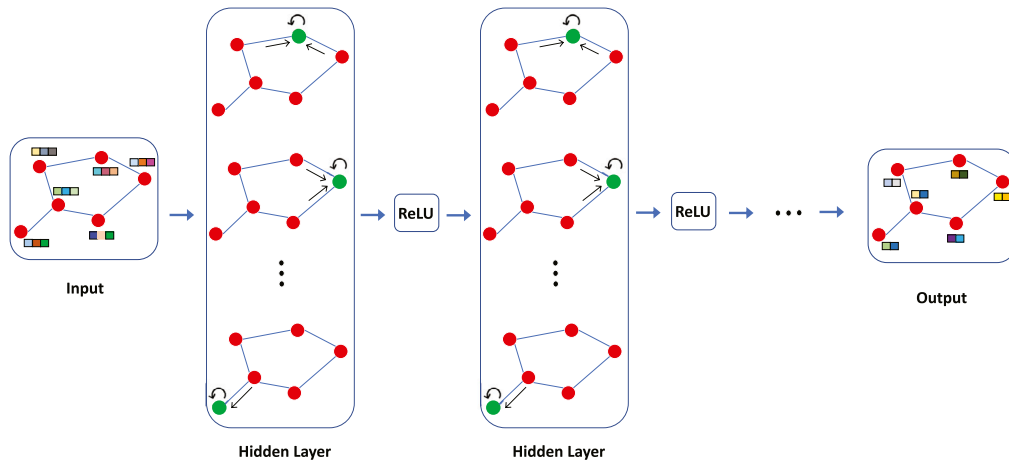


Fig. 6. Structure of GCN.

are likely incapable of mimicking the harsh and stochastic changes in inter-link dependencies within the transportation network, especially in peak periods.

Another work by Dai et al. (2019) introduced an alternative approach for mining the spatial traffic features for RNN-based models, in contrast to KNN or GCN. The proposed Spatio-temporal feature selection algorithm (STFSFA) uses the Pearson correlation coefficient to extract the most relevant traffic features for the current prediction task. The approach is similar to the KNN feature mining in the study of Luo et al. (2019), except the method for computing similarity between time series at each station differs. STFSFA is an exciting contribution to MTP models because it offers a simple and easy-to-interpret method for spatio-temporal feature generation. Notably, the STFSFA approach could be used with various RNN-based models to improve feature extraction and capture significant traffic flow trends. The temporal component of STFSFA determines the ideal input length by selecting a subset of historical input with a strong correlation to the current period. At the same time, the spatial component selects the six most highly correlated detectors from the set of time series. After running STFSFA, the input data matrix is constructed using the optimal subset of input data from each selected time series. Next, the GRU model is used to predict future traffic flow. Compared to standalone GRU and CNN models, the STFSFA combined with GRU produced the best predictions regardless of the time horizon (e.g., 5 min, 15 min). However, the study area for this paper was only a single road segment, and the effectiveness of STFSFA on a larger region of the roadway network has not been explored. At the same time, cross-comparison between this approach and the aforementioned competing designs is not meaningful because the studies use different datasets in their experiments.

In summary, RNN models have demonstrated promise for modeling the temporal characteristics of the multivariate traffic time series data. While earlier approaches primarily used LSTM-based RNN architectures in their designs, multiple works have highlighted the competitive performance of GRU-based models and prefer them for their reduced complexity and faster training time. Moreover, the most recent cutting-edge MTP approaches integrate RNN-based designs within more sophisticated, hybrid model architectures designed to consider spatio-temporal traffic trends jointly. The motivation behind applying RNN-based architectures to MTP stems from their strong ability to capture short- (e.g., daily periodicity) and long-term (e.g., seasonality) trends within the time series data. However, recent research has demonstrated that hybrid approaches are becoming necessary to jointly consider the spatial relations between links and nodes within the road network, such as the impacts of heavy congestion in the CBD on the downstream residential district. The spatial relationships also have an inherent temporal component, as the effects of an event at one location of the network need time to propagate, further highlighting the importance of jointly considering the spatio-temporal features. Thankfully, scholars have proposed multiple promising hybrid approaches for enhancing the RNN-based models with spatial capabilities, including nonparametric methods (e.g., KNN), graph-based convolution (e.g., GCN), convolutional RNNs (e.g., Conv-LSTM), and traditional statistics (e.g., Pearson correlation coefficient). That being said, despite each study demonstrating the superiority of their proposed method compared to some existing popularized methods, the datasets used in experiments differ, and the models are not directly cross-compared. Consequently, future research is needed to elucidate if an ideal model architecture exists for MTP applications.

4.3.4. GCN

The Graph convolution network (GCN) (Kipf and Welling, 2016) is an evolution of the CNN model designed to perform convolution on structured graphical input, in contrast to the euclidean-structured input data of traditional convolution. As outlined in the CNN section, the convolution operation is one well-researched method for extracting the spatial features of a roadway network in MTP models. As shown in Fig. 6, during convolution, the model learns many filters for detecting interlink relationships (features) concerning the traffic flow or speed, improving prediction performance by accounting for the inherent time-dependencies between related links. However, modeling the roadway network in euclidean space necessitates a grid-based representation, distorting meaningful connections between nodes and limiting the effectiveness of the models. For example, two adjacent roadways may lie very close in euclidean space but be unconnected and exhibit vastly different characteristics, which limits the effectiveness

of traditional image-based convolutional approaches for MTP models. Consequently, scientists have begun experimenting with graphical convolution in their MTP models as an alternative to the widely used euclidean convolution.

Prior to the introduction of GCN, Gori et al. (2005) and Scarselli et al. (2008) identified the lack of a NN architecture capable of processing graphical input, proposing the graph neural network (GNN) as a solution. Succinctly, GNN is a form of RNN that learns a matrix-based representation of the input graph while encoding each node with a set of learned weights relating to the dependencies between directly connected nodes. The intuition behind the approach is that the nodes within the graph represent objects, while the links express the relationship between objects. Interestingly, this design directly maps to the practical interpretation of the node and link structure in roadway networks; however, this method still relies on a euclidean representation of the graph. More recently, Bruna et al. (2014) generalized the popular CNN model to operate directly on graph-based analogs as input, combining spectral graph theory with DNNs. The results show that their approach can significantly reduce the required parameters while providing faster forward propagation for input data with location-based features. This approach was later improved by proposing a method for computing fast localized convolutional filters (Defferrard et al., 2016) and was further simplified to make GCN practical for modeling high-dimensional time series (Kipf and Welling, 2016).

Despite the noticeable applicability of the aforementioned graph-based convolutional models to MTP applications, they were sparsely considered by researchers until the emergence of GCN (Kipf and Welling, 2016). The primary contribution of this work is to simplify graph convolutions with a localized first order approximation of the spectral graph convolution approach in Bruna et al. (2014), further reducing the number of parameters and improving learning speed. Specifically, convolution is localized to consider only the K -nearest nodes resulting in truncated filters represented as K -order polynomials. By truncating the representation, many filters can be stacked sequentially, enabling deeper networks with better approximation ability which can still be trained efficiently. While this seminal work laid the foundation for many future MTP works, the original GCN operates on undirected graphs which cannot represent the complex traffic diffusion in transportation networks (e.g., upstream and downstream traffic dependencies, directed roadways). Moreover, the approximations leveraged in GCN to reduce computational complexity implicitly assumes equal locality which attributes an equal relationship between all edges connected to a node. In practice, this assumption does not hold because roadways are designed with different classifications, capacities, and goals despite being interconnected, limiting the performance of the original GCN for MTP applications.

While the earlier iterations of GCN excelled at capturing the spatial dependencies between interconnected nodes, they failed to jointly consider the temporal features of structured time series data. Prior MTP approaches proposed to mine the spatio-temporal features through embedding convolutional operations within RNN structures (Shi et al., 2015), but are unable to fully exploit the spatial features of the roadway due to a fixed euclidean input representation. Moreover, they are computationally heavy to train due to the iterative propagation inherent to RNNs. To improve upon these drawbacks, the seminal work of Yu et al. (2018) proposed the spatio-temporal graph convolutional network (STGCN) to jointly capture the spatio-temporal features without using an RNN cell. STGCN achieves this by introducing the spatio-temporal convolutional block (ST-Conv block), which sandwiches a GCN layer between two gated and fully convolutional temporal layers (Gehring et al., 2017) to capture temporal and spatial features, respectively. In this way, graph convolution is combined with 1D convolution to jointly capture the spatial and temporal trends, respectively. Due to the simpler internal structure and lack of dependency constraints within CNN, STGCN achieves parallelization over the input and is faster to train compared to convolutional RNNs. Moreover, the ST-Conv blocks can be stacked sequentially to build deep models and facilitate MTP in large and complex transportation networks. The authors also conduct a case study on the PeMSD7 dataset, demonstrating that STGCN can be trained about 14 times faster while having nearly 95% less parameters compared to the RNN-based alternatives. These unique properties of STGCN laid the foundation for many of the recent advancements in GCN-based MTP models.

In an alternative approach, the diffusion convolutional recurrent network (DCRNN) is proposed to improve the spatio-temporal modeling capabilities of spectral graph convolution. This model generalized the graph convolution approach to operate on both directed and undirected graphs (Li et al., 2018c). The key contribution of this work is the proposed diffusion convolution operation, which relates the traffic flow stream to a diffusion process and can effectively capture stochastic and noisy data trends. More specifically, the diffusion process is represented by a bidirectional random walk of the graph, where the stationary distribution of the Markov process is estimated using a weighted combination of a finite number of random graph walks. In contrast to earlier works, the authors consider the directionality of the network graph and propose both a forward and reverse direction diffusion process to capture the influence of upstream and downstream flows on the sensor of interest.

Moreover, to account for the temporal relationships, the authors integrate diffusion convolution into the GRU cell by replacing the traditional matrix multiplication operation with diffusion convolution. The modified GRU layers are then stacked in an encoder-decoder architecture to create DCRNN. Notably, the authors also propose to use a sequence learning approach (Sutskever et al., 2014) with scheduled sampling (Bengio et al., 2015) to improve the model performance for long-term forecasting (e.g., 1 h in the future) by minimizing error propagation when expanding the prediction horizon. Experimental results measure the benefits of DCRNN by comparing diffusion convolution to the prior spectral graph convolution approach, highlighting that considering the directionality of traffic through diffusion convolution along a directed graph considerably improves the model prediction accuracy.

Subsequent research attempted to address the shortcomings of GCN for MTP by proposing modifications to the original GCN structure. Wu et al. (2019) extended the capabilities of GCN by proposing Graph WaveNet: a spatial-based graph convolution method combining graph convolution with dilated casual (1D) convolution. Graph Wavenet dynamically learns the adjacency matrix without requiring an explicit and fixed graphical structure, resulting in two primary benefits: (1) nodes interdependence can be updated dynamically as network conditions change; and (2) hidden influences between nodes without direct connections can be elucidated by the model. Moreover, this approach can dynamically incorporate the addition of new time series as the network topology changes

(e.g., addition or removal of sensors and links), presenting the opportunity to learn an online MTP model from evolving live traffic data streams. Notably, Graph WaveNet employs stacked dilated 1D causal convolutions, enabling the creation of deep MTP models capable of processing increasingly long traffic sequences without the vanishing or exploding gradient problem inherent to RNNs. Experiments highlight that Graph WaveNet performs better than STGCN and DCRNN on multivariate datasets with more than 200 sensors, and it can train five times faster than DCRNN but is about twice as slow as STGCN. Graph Wavenet also achieves the fastest inference time of all aforementioned approaches because it outputs twelve predictions each run, while the alternatives depend on historical predictions for inference at larger time horizons.

Similarly, Yu et al. (2020) also proposed incorporating a dynamic adjacency matrix into the GCN architecture; however, this approach learns multiple adjacency matrices from the input data corresponding to specific time intervals in the past. Likewise to diffusion convolution, the intuition behind this approach is to allow the model to imitate the natural phenomenon of traffic propagation, where the traffic correlations between any two nodes are time-dependent and propagate based on direct upstream and downstream connections. Most notably, the authors encode additional features into each node, including the capacity and length of each road segment, demonstrating that incorporating prior domain knowledge into the deep learning models can improve performance.

Since then, Song et al. (2020) further improved graphical approaches for MTP by directly considering an often overlooked spatio-temporal relationship between nodes: the impact of neighboring nodes on the future state of their neighbors. The authors define three high-level relationships the models should consider: (1) the influence of a node on its neighbors at the current time-step (spatial dependency); (2) the impact of a node on itself in the future time-step (temporal correlation); and (3) the impact of neighboring nodes on the future state of a given node (spatio-temporal correlation). Previous works considered spatial dependencies and temporal correlations separately. Specifically, the output features from the spatial modeling component (graph convolution) are used as input to a gated temporal unit (e.g., GRU or LSTM design), thereby indirectly accounting for the spatio-temporal dependencies. In the approach by Song et al. (2020), a new model denoted Spatial–Temporal Synchronous Graph Convolutional Networks (STSGCN) is proposed to consider all three relationships for each node directly. One of the primary contributions of this approach is a unique localized spatial–temporal graph construction, where the adjacency matrix is modified to include a self-connection between all nodes in the current period with its future state in the subsequent time-step. In this way, the impact of neighbors on a given node in the future time step can be represented and learned by the model. Compared to Wu et al. (2019) and other cutting-edge designs, which consider only the first two relationships, STSGCN demonstrated higher prediction performance across four subsets of the PEMS data. Interestingly, this work also demonstrated the effectiveness of using the gated linear unit (GLU) activation function over the conventional ReLU due to its increased number of parameters and ability to better model the resulting complex dependencies.

Most recently, Guo et al. (2021a) proposed a modified GCN-based model, the hierarchical graph convolution network (HGCN), to investigate the benefits of considering higher-level, regional relationships during high-dimensional MTP. Intuitively, the transportation network contains macro and micro layers, where the microlayers represent the low-level node and link connections. The macro layer represents hot traffic regions (e.g., CBDs, churches, schools, and residential areas). Previous DL-based studies ignore the macro layer, representing the transportation network with only the micro-layer and focusing on the localized relationships between neighboring nodes. Notably, this fine-grained network representation and focus on localized convolution to extract spatial features could obfuscate more significant trends, especially when the distance between interdependent regions in the macro layer is considerable. HGCN can simultaneously model trends at the micro and macro layers by representing the region and road networks separately. GCN-based convolution is applied to both graphs, and the resulting regional features are merged with the localized road segment features to generate the final prediction. While an extensive case study verifies that considering both the micro and macro traffic trends can improve prediction performance compared to existing baselines, more research is needed to determine the best macro-level representation for MTP applications. In this initial approach, a simple spectral clustering method is proposed; however, employing more complex and optimized processes for detecting the hot regions in a network could improve performance.

In summation, GCN models have been demonstrated as a practical component in MTP modeling, specifically for extracting the spatial relationships within multivariate time series. The seminal work by Kipf and Welling (2016) proposed the first GCN to generalize the CNN model to operate on arbitrarily structured graphs, in contrast to Euclidean space. For MTP applications, graphs present an intuitive representation of the spatial relations within the transportation network, whereas a grid-based euclidean representation can destroy the complex underlying structure. Early applications of GCN combine its robust spatial feature extraction with fully convolutional or gated RNN-based layers in sequence to jointly capture the spatio-temporal trends. However, some significant trends can be lost due to the sequential design. Furthermore, early attempts leverage a static adjacency matrix that does not map well to the real-time phenomena of traffic propagation, where the relationship between any two road segments is dynamic. Recent works have proposed multiple methods for learning dynamic adjacency matrices, including Song et al. (2020) and Wu et al. (2019), which improves the prediction performance of GCN-based models. Additionally, it has been demonstrated that considering both the macro and micro network layer relationships during modeling can improve the prediction results, but determining the best macro-level representation of the roadway network is still an open research problem.

4.3.5. WNN

A wavelet neural network (WNN) is an integrated network combining wavelet transformations and ANNs, which was introduced by Zhang and Benveniste (1992) to solve the convergence performance problem. In summary, while the authors could not theoretically guarantee that their approach would converge successfully, they demonstrated significant experimental improvement compared to the old approach (which typically diverged while their approach consistently converged). In general, WNN aims to

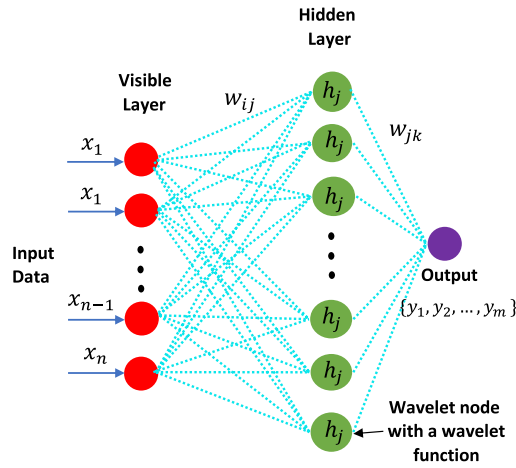


Fig. 7. Structure of WNN.

find a group of wavelets in feature space that can reflect complex relationships within the original signal. WNNs were created as a tool for function approximation, allowing them to solve concerns with traditional NNs' weak convergence.

As shown in Fig. 7, along with an input and output layer, WNN consists of one hidden layer that utilizes a non-linear wavelet basis function as an activation function. Notably, adding scale and shift factors improves WNN's approximation results and provides better accuracy. The equations defining WNN are expressed below: assume input and output layer has n and m neurons respectively, i.e., x_1, x_2, \dots, x_n are inputs and $y_1, y_2, y_3, \dots, y_m$ are outputs of NN and w_{ij}, w_{jk} are connection weights. With input $x_i (i = 1, 2, \dots, n)$, the output of hidden layer is given as follows:

$$h(j) = z_j \left[\frac{\sum_{i=1}^n W_{ij} x_i - b_j}{a_j} \right] \quad j = 1, 2, \dots, p \tag{15}$$

where $h(j)$ is the output of the j th node in the hidden layer, w_{ij} is the weight between the input and hidden layer of the network, a_j is the scaling factor of the wavelet basis function, b_j is the translation factor of the wavelet basis function z_j , z_j is the wavelet activation function, and p represents the number of hidden layer nodes. WNN uses a gradient descent algorithm to adjust connection weights, shift and scale factor. Generally, the Morlet Wavelet basis function is selected as the activation function for the hidden layer, which extracts the magnitude and phase information of the analyzed signal as shown in Eq. (16).

$$y = \cos(1.75x) e^{-x^2/2} \tag{16}$$

Finally, the output of WNN can be represented as:

$$y(k) = \sum_{j=1}^p W_{jk} h(j) \quad k = 1, 2, \dots, m \tag{17}$$

where w_{jk} represents the connection weight of the hidden layer to the output layer, $h(j)$ is the output of the j th hidden, and m represents the number of output layer nodes.

Given the inherent similarities between time series and signals, recent research has experimented with the WNN for predicting short-term traffic flow. WNNs can be a desirable architecture for MTP applications because of their strong nonlinear processing power, self-organization, self-adaptation, and learning ability. However, the choice of parameter optimization algorithm during training is a crucial consideration. Like many DL models, the traditional WNN structure uses a stochastic gradient descent algorithm for learning the weights. This method is easily trapped in the local extremum, resulting in slow convergence and low prediction accuracy. To address these challenges for flow prediction in ITS, Chen et al. (2020d) proposed an improved particle swarm optimization (I-PSO) algorithm to address the slow convergence rate and the local optimality problems of stochastic gradient descent. Compared to the standard WNN, or WNN with traditional PSO, WNN with I-PSO can achieve the highest accuracy while converging the fastest. However, the authors note that the proposed approach is ineffective during non-recurrent network events and does not consider external factors such as weather, presenting much room for future research.

Parallel research by Shen (2020) applied the harmony search algorithm to WNN, denoted HS-WNN, as an alternative optimization method for learning the WNN weights. While particle swarm optimization imitates the social behavior of flocking or herding biological species (e.g., fish, birds), the harmony search optimization heuristic is inspired by the improvisation process of jazz musicians. Notably, the case studies in both papers demonstrate the improvements of alternative optimization approaches on the convergence speed and optimal solution reliability, compared to stochastic gradient descent, for WNN-based MTP models. However, the datasets used in both papers are poorly documented, making cross-comparison difficult and leaving the readers wondering which optimization approach is the best for traffic prediction.

4.3.6. Attention-based

Attention-based methods for DL first appeared in machine translation, where they were demonstrated as an effective measure for overcoming degraded performance in AEs when working with long input sequences (Bahdanau et al., 2015). With the emergence of ITS and big data processing, transportation networks are generating a vast amount of time series data across a diverse set of geospatial sensing locations within the network. However, many of the cutting-edge short-term MTP approaches focus only on the near-term (e.g., hourly, daily, or weekly) dependencies due to limitations with the existing architectures. For example, GCN only considers relationships between the K -nearest nodes, while LSTM and GRU suffer from vanishing gradient problems when dealing with longer sequences. Attention-based mechanisms are an exciting method for overcoming these challenges by enabling the model to learn a set of high-impact features indicating the relative importance of each subset of input for the current prediction interval.

Recently, attention-based methods have been applied to various MTP models to improve their spatio-temporal feature extraction capabilities and learning efficiency. Wu et al. (2020b) proposed integrating an attention mechanism within a bi-directional LSTM architecture for improving MTP for high-dimensional time series. In this approach, the attention mechanism can dynamically learn the relative importance of each adjacent time series on the series to be predicted, capturing the most impactful spatio-temporal dependencies. Wavelet decomposition is also used to decompose each time series into a low-frequency and noise series, where the low-frequency series represents the important high-level trends with the stochastic noise removed. Focusing on the input's low-frequency representation makes learning the high-dimensional time series traffic data more efficient. At the same time, the prediction accuracy is preserved because the attention mechanism dynamically weights the time series in order of importance.

Complementary work around the same time was performed by Liu et al. (2019c), where attention was integrated into two bi-directional LSTM models to capture the spatial and temporal traffic features in high-dimensional time series data. A unique contribution of this work is the proposed preprocessing module, which can cluster the time series data into groups with similar distributions before model training, improving the convergence time and prediction accuracy. Compared with Wu et al. (2020b), the case studies in both papers demonstrate the benefits of incorporating an attention mechanism into LSTM-based MTP models, notably for high-dimensional time series modeling. In this way, the model can learn new hyperparameters relating the importance of input segments to the current prediction task, in contrast to condensing input into a fixed-length representation regardless of size. That being said, the papers conduct experiments using different datasets and study areas, so cross-comparison becomes challenging.

While the previous works focus on integrating attention mechanisms into RNN-based designs, Guo et al. (2019) proposed an attention-based convolutional model (ASTGCN), which employs both graph-based and standard convolution with attention mechanisms for capturing the spatio-temporal traffic trends. The authors improve upon prior approaches to dynamically learning the GCN adjacency matrix, combining an attention mechanism with the GCN. More specifically, the attention method learns a separate hyperparameter matrix indicating the relative importance of each node to the current prediction and time, which is used to weight the adjacency matrix dynamically. Consequently, the latent dynamic and time-dependent relationships between time series are effectively considered. An extensive case study demonstrates the effectiveness of this approach for dynamically assessing the temporal traffic dependencies within the GCN framework.

More recently, Shi et al. (2020) leveraged attention mechanisms to further improve our understanding of high-dimensional MTP modeling by integrating attention and a modified LSTM design to overcome the vanishing gradient problem for long-range time series modeling. Notably, a recurrent structure with skip connections is proposed to alter the traditional LSTM design and capture pair-wise long-range dependencies between distant time series without a vanishing gradient. A spatial attention mechanism in the encoder learns the impact of each node on the entire graph, giving the most consideration to the small subset of nodes with the most significant effect on the current time-step. Moreover, the proposed approach is directly compared to Guo et al. (2019), demonstrating that the skip-connections with attention can effectively consider the long-range and pair-wise dependencies better than a GCN-based architecture. However, the proposed model has almost twice as many parameters as ASTGCN, requiring more data and processing power to achieve better results.

In an alternative approach to improving the reliability of long-term MTP, Zheng et al. (2020a) proposed the graph multi-attention network (GMAN) by integrating attention mechanisms into an encoder–decoder architecture. One of the most notable contributions of GMAN is the transform attention mechanism, which is an attention layer between the encoder and decoder. It is responsible for modeling the direct relationship between the future and previous time steps, enabling the model to consider only the most important historical features while limiting error propagation. In other words, it transforms the encoder output using an attention mechanism prior to feeding the output into the decoder. Moreover, GMAN stacks multiple spatio-temporal attention blocks within the encoder and decoder, forming a deep network capable of extracting the traffic trends in large sensor networks. Within each spatio-temporal block, the spatial attention mechanism dynamically updates the pair-wise correlation between vertices with respect to the current time step, while the temporal attention mechanism dynamically weights the impact of previous and future time steps on the current step. Notably, the input data must be first transformed via a spatio-temporal embedding process to incorporate the time-dependent graph data into the attention mechanisms. Experimental results highlight the benefits of the transform attention mechanism, with GMAN outperforming both DCRNN and Graph Wavenet for prediction horizons of 30 and 60 min. However, GMAN performed worse than Graph Wavenet for shorter (15 min) predictions while requiring more time for training and inference.

In summary, DL models are widely applicable to MTP applications due to their superior ability to capture stochastic and non-linear relationships in traffic data. Recent research has proposed many efficient hierarchical feature extraction approaches, including spectral and spatial graph-based convolution methods, attention mechanisms, diffusion convolution, dilated convolution, and RNN-based methods. The performance of each approach is comparable in terms of prediction accuracy at different time horizons and computational complexity, and both must be considered when selecting the ideal modeling approach. In Table 8, we compare the above-discussed models for reference. In the following subsection, we review the limitations of Deep Learning-based methods.

Table 8
Comparison of DL architectures.

Model	Description	Advantage	Disadvantage	Example
MLP	-a simple ANN -Consists of three layers:input layer, layer to features from the input, output layer -Layers are composed of several neurons	-Simpler and more interpretable than mother methods -Slow convergence -Occupies a large amount of memory	-time series data requires much preprocessing -Outperformed by cutting-edge DNN designs	Song et al. (2017)
AE	-Consists of three parts:encoder (converts the input features into an abstraction (code)) , code & decoder blocks (Try to reconstruct the input features) -Use some non-linear hidden layers to reduce the input features	-Extracts the most useful and high-level features -Works with big and unlabeled data	-Multiple forward passes create a long training and preprocessing time -Hard to find the code layer size	Ghosh et al. (2017) Wang et al. (2018a) Liu et al. (2019c) Wang and Boukerche (2020)
RBM&DBN	-DBN is stacked by several RBN -Take advantage of the greedy learning process to initialize the model parameters and then fine-tunes the whole model using the label	-Effectively extracts representative data characteristics -DBN: fast inference time	-Difficulty mapping complex relationships between network elements -High computational cost	Ma et al. (2015) Bao et al. (2021)
CNN	-Consists of a number of convolution and pooling layers (to reduce the dimensions of features and find useful patterns) followed by a fully connected layers (used for classification.)	-Adept at capturing spatial dependencies -Extracts relevant features -Have high competitive performance	-Long training time is needed -Sequence and time series data are not supported	Lea et al. (2016), Zhang et al. (2017); Song et al. (2017), Wu et al. (2018); Wu et al. (2018), Liu et al. (2019a); Sen et al. (2019), Liu et al. (2019a); Zheng et al. (2020b), Dai et al. (2020)
RNN	-Consists of numerous successive recurrent layers, which are sequentially modeled in order to map the sequence with other sequences	-Strong performance on time series data	-Inability to consider spatial dependencies -Exploding and vanishing gradient problem	Wang et al. (2018b) Zhang et al. (2017)
LSTM	-An extension of RNN -Created as the solution to short-term memory -Has internal mechanisms called gates (forget gate, input gate, and output gate) to learn which data in a sequence is important to keep or throw away	-Memory cell enables better performance on long sequences -Good for sequential data	-Longest training time -High model complexity -High computational cost	Luo et al. (2019) Fu et al. (2016) Wu et al. (2020b) Liu et al. (2019c) Li et al. (2021a)
GRU	-Similar to LSTM but has fewer parameters -Has two gates (update gate & reset gate)	-Similar performance but faster training time compared to LSTM -Good for sequential data	-Less efficient in accuracy than LSTM	Zhao et al. (2019a) Dai et al. (2019)
GCN	-A spectral-based convolutional GNNs -Defined by introducing filters from graph signal processing	-Exploits graph-based structure of transportation networks -Effectively captures complex spatial dependencies in large networks	-Relies on human-defined topological structure	Wu et al. (2019) Yu et al. (2020) Song et al. (2020) Guo et al. (2021a)
Attention	-Consists of an encoder network (to map input acoustic vectors into a higherlevel representation), an attention model (to summarize the output of the encoder based on the current state of the decoder), & a decoder network (to model an output distribution over the next target conditioned on the sequence of previous predictions)	-Provide strong results in presence of long input sequences -More interpretable than the other DL models because of the attention weights -Demonstrates outstanding results in NLP models	-Increase training time of long input sequences data by adding more weight parameters to the model	Bahdanau et al. (2015) Wu et al. (2020b) Guo et al. (2019) Liu et al. (2019c) Shi et al. (2020)
WNN	-Consist of one hidden layer that utilized a non-linear wavelet basis function as an activation function	-Time Frequency localization of wavelet -Self-learning -Fault-tolerance	-Can be bias towards local minima and maxima	Chen et al. (2020d) Shen (2020)

4.3.7. Limitations of the deep learning-based methods

In comparison with classical methods, the literature highlights many advantages of leveraging deep learning models for MTP applications. Despite this, the emerging deep learning approaches have some notable limitations:

- Computational complexity and efficiency.** Compared to other machine learning methods, deep learning models have increasingly complex data structures and require higher computing power for training. Recently, researchers have expressed concerns over the high resource requirements of cutting-edge deep learning models across various application areas. To better understand and quantify the computational limits of existing deep learning methods, [Thompson et al. \(2020\)](#) examined 1058 articles on deep learning methods across five application areas (image classification, object detection, questioning, named entity recognition, and machine translation). They identified a clear trend in the literature, highlighting that advances in accuracy heavily rely on simultaneous advances in computing power (p -value < 0.01). Specifically, the authors leverage statistical learning theory to estimate that computational requirements scale as a fourth-order polynomial, at a minimum, concerning performance (but practically, the scaling can be demonstrated to be worse). One explanation for this relationship is that NNs overparameterize the network by design because it has been shown that performance is significantly improved when NNs have more parameters than available training instances ([Soltanolkotabi et al., 2018](#)). For example, one of the best-performing image recognition NNs, NoisyStudent, has 480 million parameters trained on only 1.2 million data points. Notably, the authors conclude that the current trajectory of deep learning advances will quickly become economically, technically, and environmentally unsustainable due to the relationship between advancement and computation requirements. On the bright side, DL for traffic prediction has yet to reach the same computational ceiling as the application areas studied in [Thompson et al. \(2020\)](#). There is a need for re-examining the existing DL-based approaches for traffic prediction from a scalability perspective. The statistical learning analysis highlights that there is much room for improving the efficiency of the current methods by reducing computational complexity and developing high-performing but lightweight deep learning architectures. In recent works, some researchers have attempted to improve the efficiency of NNs by leveraging optimization to search for the hyperparameter set, which maximizes accuracy while minimizing computation ([Pham et al., 2018](#)). On the other hand, another group of approaches, denoted model pruning strategies, seeks to compress an already trained network and reduce inference computation costs while maintaining good performance ([Deng et al., 2020](#)), but this does little to improve efficiency from a model training perspective. Notably, if the pruned model representations can be leveraged to train new models with sparser parameter counts, the training efficiency can also potentially be increased. Lastly, exploring alternative model architectures to DL for solving traffic prediction problems may be necessary if increases in efficiency are not achievable. In summation, computation requirements for improving DL approaches are rapidly outpacing advances in hardware or efficiency, highlighting a need for further research to enhance DL efficiency.
- Optimal model choice.** For MTP applications, the ideal model can capture the three critical trends: (1) the influence of a node on its neighbors at the current time-step (spatial dependency); (2) the impact of a node on itself in the future time-step (temporal correlation); and (3) the impact of neighboring nodes on the future state of a given node (spatio-temporal correlation) ([Song et al., 2020](#)). While the appropriate model selection always involves application- and data-specific considerations, the literature is mainly undecided on the best approach for multivariate speed or flow forecasting, and new techniques are frequently proposed. Moreover, in addition to the trends above, the ideal model should also capture both the short- and long-term dependencies in the spatial (e.g., dependencies between distant but related nodes such as CBD and residential) and temporal (e.g., recent, daily-periodic, seasonal) areas. Existing works have revealed specific architectures optimized for accomplishing a subset of the modeling goals, but no unified architecture has emerged as the best for all considerations. Furthermore, after deciding on your model architecture, determining the best model parameters (including how many hidden layers, the number of hidden nodes, desired learning rate, hyperparameters, choice of activation function, evaluation method, etc.) is challenging and generally requires a brute-force approach. Encouragingly, some recent research has proposed and implemented optimization methods that search for the ideal hyperparameter set given dynamic application requirements ([Balaprakash et al., 2018](#); [Andonie, 2019](#)).
- Lack of interpretability.** DL algorithms are generally considered "black box" models with limited practical interpretability compared to traditional approaches. However, traffic forecasting models are leveraged in many high-stakes decision-making applications (e.g., signal timing and planning, route guidance), where interpretability of the resulting predictions is key to understanding why specific approaches perform better or worse than others. At the same time, many DL algorithms lack the ability to quantify the uncertainty of their prediction output due to the dynamic nature of the prediction errors and resulting uncertainty when modeling stochastic and noisy traffic data. Under these dynamic conditions, the conventional approaches for measuring uncertainty (e.g., Kalman or Bayes filters) may fail due to the assumption of a fixed covariance matrix. Moreover, models in high-stakes applications are generally regarded as useless unless their decision-making process can be verified by a domain expert ([Yu et al., 2020](#)). Consequently, the full potential of DL approaches is not realized in critical decision-making systems due to the potential for catastrophic failures. Recent research has proposed new methods for adapting DL models to learn a deep uncertainty covariance matrix dynamically while also considering both the epistemic (uncertainty of model parameters) and aleatoric (noise within the data) uncertainties ([Russell and Reale, 2021](#)). However, the work makes many assumptions, including assuming the aleatoric and epistemic uncertainties follow a multivariate Gaussian distribution, which may not hold for highly nonlinear systems where error distributions are multimodal. Moreover, the measurement uncertainty is assumed to be uncorrelated in time between measurements. Both assumptions limit the applicability of the proposed approach for MTP models, where the inherently stochastic data may impair the performance, necessitating further research.

In the traffic prediction domain, some studies have also proposed variations of DL architectures that focus on improving the interpretability of the prediction results. For example, Wang et al. (2019a) enhanced the interpretability of traffic speed forecasting by proposing a bi-directional and path-based LSTM model for urban network speed prediction. Precisely, the deep learning network is mapped to the physical roadway network resulting in interpretable output from the hidden layer using visualization and qualitative analysis. However, more research is needed to explain better how the newly interpretable information can elucidate insights that drive future improvements. In another related work, Wang et al. (2020d) introduced an integrated approach for short-term traffic flow prediction combining SARIMA and Group Method of Data Handling (GMDH) to improve our understanding of the causality between the historical spatio-temporal traffic information and future conditions. Their interpretability analysis concluded that the causality could be interpreted as a polynomial function of historical temporal data and spatial traffic information. The approach demonstrates high prediction accuracy compared to existing and less interpretable methods. That being said, the generalizability of this approach has yet to be explored in other traffic prediction applications. Further research is needed to improve the interpretability of DL-based systems and elucidate methods for leveraging the new insights to enhance the models' predictive capabilities further.

- **Limited transferability.** The cutting-edge traffic forecasting approaches are generally network-specific, rendering them useless for predicting traffic outside the specific context in which they were trained. More specifically, the spatial relations learned by the model depend on the topological structure of the sensor network. In other words, two urban road networks are unlikely to be highly similar in structure. It is not efficient to train a network-specific prediction model for every transportation network. Thus, methods for improving model generalizability (transferability) interest the research community greatly. Recently, transfer learning is an emerging ML-based research area to apply knowledge gained while training one model to speed up the training time of a new model. For example, traffic prediction transfer learning can leverage a Seattle-based traffic flow model to help train a flow prediction model for the Boston transportation network. However, transfer learning is a new research area, and little literature exists on transfer learning approaches for traffic prediction problems.

In one very recent work, Mallick et al. (2021) proposed a transfer learning approach for diffusion convolution recurrent neural networks (DCRNN), one of the state-of-the-art model architectures for highway traffic forecasting. DCRNN cannot perform transfer learning in its traditional form because it inherently learns location-specific traffic patterns. In the proposed approach, traffic is realized as a function of network connectivity and temporal patterns, enabling generalization to unseen networks with varying spatial representations. The experimental results demonstrate that the approach can successfully perform transfer learning between the Los Angeles and San Francisco transportation networks. However, the study focuses only on highway networks and does not consider external data sources in its analysis. In an alternative approach to cross-city transfer learning, RegionTrans is proposed to effectively transfer knowledge learned from a data-rich city to a data-scarce city (Wang et al., 2019b). Notably, this approach relies on matching a data-scarce city with a data-rich city exhibiting a similar network topology, limiting practical applicability. While these recent works demonstrate progress in applying transfer learning to traffic prediction problems, additional research is needed to refine the approaches and analyze their benefits compared to existing ones.

4.3.8. Deep learning evaluation methods

Deep learning models can leverage various parameters to evaluate their traffic prediction accuracy, and the most notable are summarized in Table 9. These parameters generally represent different methods for calculating the prediction accuracy or error. Among these methods, MSE, MAPE, and RMSE are the most commonly used in the literature and are negatively correlated with the training data size. Specifically, as more instances are added to the training set, the model prediction error should decrease, resulting in more accurate prediction capabilities.

That being said, it is also essential to consider the problems of overfitting and underfitting when deciding the training set size and desired number of features. At this moment, more research is needed to elucidate the overfitting problem within the context of deep learning models, which challenge our conventional understanding due to their strong generalization ability even when they appear to be overfitting the training data. Until recently, the community assumed that a model with an increasing number of parameters and training data would lead to poor generalization performance on testing data; however, new research into deep learning highlights that this is not always the case in practice (Belkin et al., 2019).

5. Common traffic prediction states and associated applications

Traffic prediction approaches focus on forecasting network-based metrics (e.g., flows, speeds, travel time, demand) at varying scopes based on the dynamic spatio-temporal network conditions. In addition to the previous model-focused literature discussion, the existing traffic prediction methods can also be categorized based on the prediction problem, namely, what is being forecasted. In this section, we outline and describe some of the literature's most frequent and relevant classes of prediction problems. We also categorize recent traffic prediction studies according to their respective application task, area, data type, the model used, and other descriptive dimensions in Table 10.

Table 9
Deep learning evaluation methods.

Evaluation Parameter	Formula	Characteristics
Mean Absolute Error (MAE)	$\frac{1}{n} \sum_{i=1}^n \hat{y}_i - y_i $	
Root Mean Square Error (RMSE)	$[\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2]^{\frac{1}{2}}$	
Mean Square Error (MSE)	$\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$	
Mean Relative Error (MRE)	$\frac{1}{n} \sum_{i=1}^n \frac{ \hat{y}_i - y_i }{y_i}$	
Absolute Percentage Error (APE)	$\frac{\hat{y}_i - y_i}{y_i}$	-Measure the prediction error -The smaller the value is, the better the prediction effect
Variance Absolute Percentage Error (VAPE)	$Var(\frac{\hat{y}_i - y_i}{y_i}) \times 100\%$	
Normalized Root Mean Square Error (NRMSE)	$\frac{[\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2]^{\frac{1}{2}}}{\max(y_i) - \min(y_i)}$	
Mean Absolute Percentage Error (MAPE)	$\frac{100\%}{n} \sum_{i=1}^n \frac{\hat{y}_i - y_i}{y_i} $	
Symmetric Mean Absolute Percent Error (SMAPE)	$\frac{2}{n} \sum_{i=1}^n \frac{ \hat{y}_i - y_i }{\hat{y}_i + y_i} * 100\%$	
Coefficient of Determination (R^2)	$1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2}$	-Calculate the correlation coefficient, which measures the ability of the predicted result to represent the actual data
Explained Variance Score (νar)	$1 - \frac{Var(Y - \hat{Y})}{Var(Y)}$	-The larger the value is, the better the prediction effect
Accuracy	$1 - \frac{\ Y - \hat{Y}\ _F}{\ Y\ _F}$	-Detect the prediction precision -The larger the value is, the better the prediction

\hat{y}_i : The i th predicted value, y_i : The i th target output, \hat{Y} : Set of \hat{y}_i , Y : Set of y_i , \bar{Y} : Average of Y , n : the total number of samples.

5.1. Multivariate traffic flow and speed forecasting

The traffic forecasting problem represents a spatio-temporal time series prediction problem, where the input contains a traffic variable (e.g., flow or speed) represented in one or more time series, and the output is a forecast of the future conditions. For the most part, multivariate and short-term traffic forecasting literature focuses on the microscopic level by analyzing and predicting traffic conditions (e.g., flow or speed) at low resolutions, such as at the detector or link levels. However, a growing body of literature has shifted its focus to predicting traffic flows at the regional level. In regional flow prediction, the traffic network is segmented into a collection of regions, and the higher-level features and relationships between areas (e.g., residential and CBD) are considered. Notably, this can provide benefits of reduced computation overhead compared to predicting regional traffic flow using a combination of fine-grained models.

5.1.1. Regional flow

In practice, regional flow prediction models are paramount to the application of urban planning. Stakeholders rely on high-level regional insights about the transportation network to plan for future improvements. For example, predicting the traffic flow at entrances to a regional expressway system could be helpful for planners seeking to ensure vehicles enter the system safely and effectively (Gao et al., 2021). Two primary studies attempt to identify the impacts of weather on regional traffic flow concerning region-specific factors (e.g., age, income, road density, hotel density, attraction density) (Ding et al., 2015, 2017). In particular, inclement weather was demonstrated to have a considerable impact on traffic. At the same time, region-specific features, such as age, also are shown to influence the travel behavior within a region during inclement weather. However, this work is still in its infancy, and the authors express the need to consider the impact of additional external factors on the region-specific traffic flow in future work.

Building upon this research, Liu et al. (2019a) presented the first work to predict traffic congestion at the region level. The proposed approach employs a CNN to extract the spatial features from the traffic data. After that, a novel Traffic State Index (TSI) metric is proposed to measure regional traffic conditions and classify traffic data into three categories. These are then used to capture hourly, daily, and weekly traffic patterns in three residual NNs. Lastly, the outputs of the NNs are carefully synthesized to produce the region-level traffic flow prediction. The authors validate their method with a comparative analysis against five alternative methods, highlighting its superiority. However, the prediction accuracy degrades as the number of regions within the partitioned network increases, presenting the need for future research into deeper and more complex NNs for modeling regional traffic flow. For future work, clustering regions with similar properties has the potential to help overcome the aforementioned challenge in more extensive networks. Still, this approach has yet to be explored for regional flow prediction. Interestingly, related work on taxi demand prediction proposes considering the heterogeneity between different regions in their analysis, which could serve as a basis for future work in regional flow prediction (Zhang et al., 2021d).

5.1.2. Network flow

In contrast to analyzing traffic at the region level, estimating and attributing traffic flow through each roadway and intersection within a transportation network is the focus of the network flow forecasting state. With the emergence of ITS, network flow models can enable more proactive and intelligent traffic management while providing input for real-time route guidance and advanced traveler information systems. However, forecasting network flow is an exceedingly challenging problem due to the intricate spatial relationships between links in the roadway network, the consideration of intersections and traffic control devices, and the volatility and uncertainty of vehicle flow over time. Two primary concerns for network flow forecasting that impact modeling challenges are the scope of the study area and the prediction horizon. Increasing the prediction horizon or expanding the study area to forecast traffic flow on an increased number of links requires more historical data, elevates computational requirements, and demands more complex modeling approaches.

Accordingly, the earlier approaches leveraged statistical ARIMA-based models for short-term network flow prediction across only a single arterial roadway (Kumar and Vanajakshi, 2015; Kumar, 2017) or a small subset of the urban network (e.g., 30 links) (Cai et al., 2016). At the time of publication, these approaches outperformed the existing baselines but also presented some drawbacks to the ARIMA approach. Specifically, ARIMA-based methods struggle to handle complex, non-stationary time series data, instead assuming the data is stationary (ideal stationary assumption). Consequently, the capability of ARIMA models to predict traffic during periods where the underlying relationships are more complex, such as during peak periods, is limited.

In an attempt to improve upon these drawbacks, more recent research has leveraged DL-based modeling approaches for predicting network flow at greater scopes. Wu et al. (2018) proposed a novel hybrid model, combining fully-connected NNs, RNNs, and CNNs, to jointly mine the underlying spatio-temporal features from the traffic data and predict network flow within a 45-min time horizon. The scope of this study included 33 detectors along a major interstate highway. In an alternative approach, Ma et al. (2020a) introduced a combination of statistical (ARIMA) with deep learning (MLP) models to predict traffic flows across a single arterial highway with data from 44 double-loop detectors. Although both studies present improved prediction results compared to the existing baselines, their study area's scope raises questions about generalizability: it is unknown whether these approaches will perform well in large-scale urban transportation networks.

To improve large-scale network traffic flow prediction, Zhang et al. (2019b) proposed integrating an attention-based mechanism in conjunction with the CNN and RNN networks to predict traffic flow under an adversarial learning framework with a study area consisting of 1,250 road segments. Despite the improved performance of this approach, leveraging a CNN to extract spatial features from the roadway requires sampling traffic data in regular grids for input, destroying the underlying spatial structure of the network. Most recently, researchers have experimented with alternative methods for spatial feature extraction in traffic flow prediction approaches, including graph-based models. Zhao et al. (2019a) integrated GCN and GRU for network flow prediction and conducted a case study using data from 156 major roads and 207 sensors across two urban transportation networks. The approach outperformed five existing baselines (HA, ARIMA, SVR, GCN, GRU). It demonstrated its superior ability to capture the relationship between network topology characteristics concerning the dynamic temporal traffic state. That being said, the performance analysis highlights that this approach performs poorly near local minima/maxima, presenting room for future research.

Later, Guo et al. (2020) identified that existing graph-based methods learn the spatial representation using simple, empirical spatial graphs in contrast to a data driven approach, which may hinder performance. To explore the benefits of building the graph in a data-driven way, the authors propose an optimized graph convolution recurrent neural network, which develops an optimized graph that identifies the latent association between road segments directly arising from traffic data. In their performance analysis consisting of data from 370 roadway segments, the authors prove their method outperforms existing baselines for prediction horizons of 15, 30, and 120 min ahead; however, it underperforms when predicting 60 min ahead. Due to the black-box nature and limited interpretability of DL models, elucidating the model's shortcomings in this scenario is challenging, requiring further research.

5.1.3. Speed

While some studies focus on predicting traffic flow, related approaches focus on speed instead, and the literature contains a diverse set of methods for predicting traffic speed. Practically, traffic speed prediction and analysis can aid in dynamic congestion detection and mitigation applications. More specifically, speed information at the segment level, when compared to the predicted (expected) speed, can potentially reveal problematic congestion in real-time and inform the mitigation actions of TMCs. While most of the early studies only considered speed with respect to time (Vanajakshi and Rilett, 2004; Guo and Williams, 2010), there is a recent ongoing effort to incorporate the underlying spatial dependencies into the models (Dai et al., 2020; Zhou et al., 2022).

However, despite the rich body of literature on traffic speed prediction, most existing studies have limited scopes and focus on forecasting speed for a particular roadway topology by learning from historical data collected from only a few specific roads (e.g., PEMS data). Furthermore, few studies focus on speed prediction in this network-wide context due to the increased complexity of modeling urban road networks. To fill this research gap, Liu et al. (2019c) introduced a novel deep learning architecture, leveraging a temporal clustering and hierarchical attention (TCHA) mechanism to capture the latent spatio-temporal correlations in traffic data dynamically. Notably, this method elucidated numerous correlations between the target road and the nearby roads, which can be leveraged to improve prediction results.

In another recent related work, a novel vertical stack of two LSTM models was applied by Wang et al. (2019a), in conjunction with a network path segmentation algorithm, to effectively capture and elucidate both spatial and temporal correlations from network-wide high dimensional traffic data. Compared to the previous studies, this work naturally captures domain-related knowledge of transportation systems and presents realistic physical meaning; mainly, the output of the hidden layer is interpretable through visualization and qualitative analysis. Later, to leverage graph-based DNNs to improve network-wide spatial feature extraction, Cui

et al. (2020a) introduced a graph wavelet gated recurrent (GWGR) neural network for capturing localized characteristics/features from the complicated geometric or topological structure of transportation networks. Specifically, the new feature extraction process enables increased flexibility by removing the need for an explicitly given human-defined topological network structure, such as the case in GCN.

Most recently, James (2021) proposed a real-time citywide traffic speed prediction model leveraging a new geometric deep learning approach. This design accounts for the complex structure of the urban road network directly in the learning process. Additionally, to further increase the model's ability to process massive transportation networks, this study suggested a tailor-made feature design and training technique combined with an attention mechanism. This design performs multi-step-ahead short-term traffic speed prediction while considering the correlations between prediction variables through improving basic regression trees. The result provides more information and better predicts the longer-term trend of traffic speed in large urban networks compared to a single-step approach.

While both aforementioned approaches improve our understanding of feature extraction and improve prediction accuracy compared to the existing baselines, there is still much room for improving the reliability and adaptability of the associated approaches. In particular, the methods extract network-specific spatial features from the underlying transportation data, which results in poor performance when attempting to generalize the models to other networks. Moreover, these models generally are trained on heavily preprocessed data where anomalies have been regularized or removed, limiting their prediction accuracy during non-recurrent network congestion or anomaly events. Further research is needed to improve upon these drawbacks, and methods such as transfer learning present promising avenues for future research.

5.2. Travel time forecasting

In addition to flow and speed forecasting, travel time estimation represents another related transportation application that can benefit from DL-based MTP approaches. The existing literature can be segmented into two broad categories: (1) path travel time approaches, which focus on predicting the travel time for a predefined path or road segment; and (2) OD travel time estimation, where the model infers the driven path and outputs a prediction solely using the OD pair and timestamp as inference input. Below we describe both approaches in more detail.

5.2.1. Path travel time

In the path travel time prediction state, the prediction models take specific route(s) as input (derived from GPS trajectory data) and output the estimated travel time. Accurately estimating travel time for a given route is an important consideration for many transportation applications, including route planning (e.g., Google Maps), traffic monitoring, and taxi/delivery dispatching (e.g., Uber, Grubhub, Lyft). However, path travel time estimation is challenging due to the complex spatial correlations, external factors, and temporal dependencies in transportation networks.

In general, there are two primary methods for estimating path travel time: (1) individual, where the entire path is segmented into smaller chunks, and the travel time for each portion is estimated and summed together (Pan et al., 2012; Yang et al., 2013; Wang et al., 2014b, 2016); or (2) collective, where the entire path is considered and a single overall travel time estimate is generated (Jenelius and Koutsopoulos, 2013). The problem becomes more straightforward in the individual method by decomposing the path into smaller sub-problems. Still, these methods fail to capture the complex dependencies between the sub-paths resulting from intersections, traffic lights, and other network infrastructure considerations. On the other hand, the collective method can successfully capture and consider the intricate dependencies along the entire path but produce unreliable estimates for longer paths due to probe data sparsity (e.g., some paths are traveled considerably less than others). Consequently, the most recent research attempts to rectify these issues by proposing collective methods using DL models, which can better learn the underlying road network structure and improve the travel time prediction accuracy and reliability for longer trips (Wang et al., 2018d; Li and Liu, 2020). These works demonstrate that considering the network topology within the travel time prediction models can significantly improve accuracy and reliability, especially for longer trips. That being said, these studies fail to consider external factors (e.g., weather, special events, accidents, construction) in their analysis, which can further improve the prediction accuracy.

Most recently, two research works have identified this literature gap and integrated external factors into their modeling approach with great success (Chen et al., 2020b; Li et al., 2021d). Despite this, the cited works generally only consider a subset of external effects, such as just weather and planned special events, instead of considering all contributing factors. Moreover, unpredictable external events (e.g., accidents) and their impacts on prediction performance are not analyzed, presenting a future research direction. At the same time, the existing works primarily leverage trajectory data collected from taxis in their training and analysis of the models. However, taxi drivers are more experienced than traditional roadway users. They may not make mobility and routing decisions the same way, potentially hindering the generalization capability of the models for other drivers. Lastly, path travel time approaches can also be impeded by probe data density, as the methods need to infer the path taken by drivers between consecutive trajectory data points. When the time between data collection increases, the vehicle may have traversed multiple links and intersections between successive data points, making it challenging to infer the chosen path accurately. To resolve this challenge, some researchers have attempted to propose origin–destination (OD) travel time models, which can predict the travel time given only an origin and destination pair without inferring the path. We discuss the OD travel time state in more detail below.

5.2.2. OD travel time

In contrast to the path travel time prediction state, OD travel time seeks to leverage vehicle trajectory data to predict total trip travel time without requiring the exact path traversed by a given vehicle. The OD travel time prediction approach provides the benefits of reducing computation time while eliminating inference errors when using sparse trajectory data. Practically, the OD travel time state is of importance to the same applications as path travel time (e.g., route planning and traffic monitoring). Similar to path travel time, existing research has demonstrated that considering the spatial dependencies between links within the transportation network can improve prediction accuracy (Jenelius and Koutsopoulos, 2013). Other work has also improved prediction accuracy by leveraging the travel time of previous trips in their models to learn an encoding of OD and trajectory pairs for different departure times (Yuan et al., 2020).

Moreover, considering external factors, including meteorological data, time-of-day, and historical travel time within the models also improves the predictive performance (Li et al., 2018b; Xu et al., 2018; Wang et al., 2018b). Notably, these studies have some drawbacks requiring future research. First, the generalizability of the models for predicting travel time for general traffic scenarios, in contrast to opportunistic probe vehicles (e.g., taxis), is not considered. Second, existing approaches assume the driver will always take the shortest distance path between origin and destination. This assumption is not always practical, as network conditions impact the travel time and chosen route (sometimes, the shortest distance path is not always the fastest).

Furthermore, these models generally consider limited modality data or leverage a single model without considering the one-sidedness of such a design. Consequently, a recent study proposed a combination of gradient boosting decision trees with a DNN to improve prediction accuracy while incorporating external factors such as meteorological data (Zou et al., 2020). While the case study proves this approach outperforms the existing baselines, the authors highlight room for further improvement. Specifically, additional external factors, utilizing more fine-grained weather data, and considering traffic flow data in conjunction with the trajectory data are not explored.

Additionally, following the emerging trend of learning the network topology as a graph, Wang et al. (2021d) proposed a combination of GCN and GRU to improve the spatial representation and better capture the dynamic spatio-temporal network conditions for OD travel time prediction. A case study is conducted, and the proposed method can reduce the MAPE error by upwards of 1.78% compared to the state-of-the-art baselines. However, the model takes considerable time to train and perform inference, even on strong hardware, presenting room for improving the efficiency of this approach. This method also does not consider external data in its model, which has improved prediction accuracy for OD travel time models. It would be interesting to see if considering external factors and their impacts could further enhance the performance of this approach. For interested researchers, we provide the following simple baseline from the recent literature to assist in conducting a comparative analysis of future work (Wang et al., 2019c).

5.3. Travel demand forecasting

Accurate travel demand forecasting is essential to practical transportation applications such as dispatching and scheduling mobility and delivery services (e.g., Taxi, Uber, GrubHub). However, predicting travel demand is challenging given that the overall roadway demand is decentralized and results from many areas (e.g., demand for conventional taxi service, ride-hailing applications, service delivery, and non-recurrent events). In general, travel demand forecasting can be split into two subcategories: region-level demand prediction and OD demand prediction. In literature, there exists a large historical body of work on region-level demand prediction (Davis et al., 2016; Liao et al., 2018b; Yao et al., 2018; Liu et al., 2020d). The progression of the methods within the literature for the region-level travel demand forecasting state is similar to the other states: the earlier works focus on leveraging statistical models (Davis et al., 2016), while the subsequent studies prove that combinations of DL-based models outperform the earlier statistical approaches (Liao et al., 2018b; Yao et al., 2018).

Later, to improve prediction performance in large-scale urban transportation networks, Liu et al. (2020d) demonstrated the effectiveness of integrating a novel attention-based mechanism with DL-based models, outperforming the existing baselines. In an alternative approach to better capture the exogenous dependencies and underlying correlations of travel demand, Guo and Zhang (2020) introduced a novel residual spatio-temporal network (RSTN) model. This approach combines CNN and LSTM while reformulating the prediction problem to learn a residual function based on the travel density in each temporal epoch. The resulting model is easier to train, and the case study demonstrated its improved accuracy compared to other state-of-the-art baseline approaches. Moreover, at the multi-region level, Tang et al. (2021) recently proposed a multi-community spatio-temporal graph convolutional network (MC STGCN) to overcome the challenge of capturing inter-region demand correlations of complex urban road network structures with irregular shapes and arrangements.

Despite much research attention to region-level demand prediction, few attempts have been made to explore the origin-destination (OD) demand prediction problem (Wang et al., 2017, 2019d; Liu et al., 2019b). One problem with the existing studies is that they either do not suitably account for the non-Euclidian demand structure in the data or only capture the pair-wise relationships using undirected and static graphs. Additionally, it is significantly more difficult to perform OD-based demand forecasting compared to the region-level due to the challenge of capturing the spatio-temporal correlation between two different OD pairs. Attempting to overcome this challenge, Ke et al. (2021) proposed OD graphs for characterizing non-Euclidian pair-wise geographical and semantic connections between various OD pairs and demonstrated that their design outperforms the baseline algorithms by a significant margin. In a different approach, Zhang et al. (2021a) captured the Euclidean and non-Euclidean spatial-temporal relationships of OD graphs using a novel Dynamic Node-Edge Attention Network (DNEAT) model, which incorporates a temporal node-edge attention mechanism. Experimental results demonstrated that the model outperformed six baselines across two different datasets while showing increased resilience to sparse training data.

Table 10
A summary of most recent traffic prediction studies.

Author	Area	Dataset (Type)	Data Interval (min)	External Data	Application Task	Model	Evaluated Algorithm	Comparative Model
Tran et al. (2022) ^{S,T}	U	Haiphong (GPS)	10	×	Speed	LSTM	RMSE, MAE, MDAE	CNN, AR, ARMA, ARIMA, SARIMAX, PROPHET, MLP, SES, HWES
Zafar et al. (2022) ^{S,T}	U	FCD(GPS) OSM ETA	15	✓	Speed	LSTM-GRU	RMSE, MAE, MAPE	KNN, XGBoost, ANN, Linear Regressor, MLP
Jian et al. (2022) ^{S,T}	F	PeMSD4, PeMSD8 PeMS-BAY(Detector)	15,30,60	×	Speed	ST-DWGRU	RMSE, MAE, MAP	HA, ARIMA, STGCN, DCRNN, GWN, ASTGCN, LSGCN
Jin et al. (2022) ^{S,T}	U	AutoNavi	15	×	Speed	PL-WGAN	MAE, RMSE, ACC, R ² , VAR	HA, ARIMA, SVR, GRU, GCN, ST-GCN, T-GCN, A3T-GCN
Lee and Rhee (2022) ^{S,T}	F	SeoulTaxi(GPS)	5	×	Speed	DDP-GCN	MAPE, MAE, RMSE	HA, VAR, LSVR, ARIMA, STGCN FC-LSTM, DCRNN
Lu et al. (2022) ^{S,T}	E F	Guangzhou(Detectors) PEMS-DB(Detectors)	10 5	×	Speed	DENN	RMSE, MAE, R ² , ACC	HA, ARIMA, SVR, LSTM, FDL, STNN, BTF, GLA, T-GCN
Zhao et al. (2022) ^{S,T}	F H	METR-LA, PEMS-BAY, PEMA-S (Detectors)	5	×	Speed	ADSTGCN	RMSE, MAE, MAPE	HA, ARIMA, FNN, FC-LSTM, DCRNN, STGCN, SLCNN, GMAN, STAWnet
Ma et al. (2022) ^{S,T}	U	Hangzhou(Detectors)	5	×	Network Flow	STFSA-CNN-GRU	MAPE, MAE, RMSE	ARIMA, SVR, CNN, RNN, LSTM, GRU
Wang et al. (2022) ^{S,T}	F	PeMSD4, PeMSD8 (Detectors)	5	×	Network Flow	ASTGAT	MAE, RMSE	HA, ARIMA, VAR, LSTM, GRU, STGCN, GeoMan, ASTGCN(12)
Aljuaydi et al. (2022) ^{S,T}	F	MRWA WebEOC	1 15	✓	Regional Flow	1-D CNN LSTM	MAE, RMSE	MLP
Sun et al. (2021) ^{S,T}	H U	METR-LA(Detectors) PeMS-BAY(Detectors)	NA	×	Network Flow	GST+GAT	MAE, RMSE, MAPE	ARIMA, SVR, LSTM, STGCN, ASTGCN, STANN, GMAN
Li et al. (2021c) ^{S,T}	F	NDW-Netherlands (Detectors)	1-5	×	Network Flow	DGCN	MAE, MAPE, RMSE	HA, KNN, GAT, DCNN, DCRNN, STGCN, Graph Wavenet
Wang et al. (2021d) ^{S,T}	U	Chengdu Taxi (GPS) Xi'an Taxi (GPS)	10	✓	OD Travel Time	GraphTTE	MAPE, RMSE, MAE	Average, GBDT, DeepTTE, GTTE-D, GraphTTE, GTTE-S, DeepOD
Li et al. (2021d) ^T	E	OPENITS(Detectors)	5 1	✓	Path Travel Time	soft set theory model	MAE, MAPE, R ² , RMSPE	BPR
James (2021) ^{S,T}	U	NavInfo(Trajectory) OSM	5	×	Speed	GA ²	MAPE	HA, AK, SVR, B-LSTM, DCRNN, T-GCN, A3TGCN
Guo et al. (2021a) ^{S,T}	U	Didi Chuxing GAIA Initiative (GPS)	10	×	Speed	HGCN HGCN-WH HGCN-WDF	MAE, MAPE, RMSE	HA, OTSGGCN, LSTM, GRU, GatedSTGCN, GCRN, ARIMA, OGCRNN, GWNET
Li et al. (2021a) ^T	H	England	15	✓	Network Flow	W-CNN-LSTM	RMSE, MAE, R ²	LSTM, MLP, ARIMA, CNN-LSTM
Ke et al. (2021) ^{S,T}	U	NYC Taxi & Limousine Commission(GPS)	30	×	Travel Demand	ST-ED-RMGC	RMSE, MAE, MAPE	HA, GBDT, RF, XGB, LASSO, MLP, LSTM, Spatial LSTM, MGC, ED-MGC, RMGC
Zhang et al. (2021a) ^{S,T}	U U U	Didi-Chuxing(GPS) NYC Taxi &Limousine Commission (GPS) OSM	10 15	×	Travel Demand	DNEAT	RMSE, MAPE	HA, LSTM, TCN, DySAT ConvLSTM, ST-GCN
Ma et al. (2020a) ^{S,T}	H	Tel Aviv (Detectors)	60	×	Network Flow	NN-ARIMA	MAPE, MSE	NN, ARIMA, MSVR, MSVR-ARIMA
Zou et al. (2020) ^T	U U	Shanghai Taxi(GPS) Chengdu Taxi (GPS)	20s-30s	✓	OD Travel Time	TTE-Ensemble	MAE, MAPE, MARE	LRM, GBDT, DNN
Yuan et al. (2020) ^{S,T}	U	Chengdu Taxi (GPS) Xi'an Taxi (GPS) Beijing Taxi (GPS) OSM	3s 3s 1	✓	OD Travel Time	DeepOD	MAE, MAPE, MARE	TEMP, LR, GMB, STNN, MURAT
Xu et al. (2020) ^{S,T}	U	GeoLife(Trajectory)	1s-5s	×	OD Travel Time	MTLM	MAE, MAPE, RMSE	GBDT, MlpTTE, Deeptravel, DeepTTE
(Li and Liu, 2020) ^S	U	Sydney Taxi (GPS)	NA	×	Path Travel Time	SFM		FM, FFM, CFM
Chen et al. (2020b) ^T	F	Taiwan(ETC)	5	✓	Path Travel Time	GB	MAPE, MAE, RMSE	Random Forest
Guo and Zhang (2020) ^{S,T}	U U	DiDi ChuXing (GPS) NYS Taxi & Limousine Commission (GPS)	30 30	×	Travel Demand	FCNs+CNN +LSTM	SMAPE, MAE, RMSE	STN, LSTM-C, ARIMA, ST-ResNet, FCL-Net FCL-Net-v
Kuang et al. (2020) ^{S,T}	U U	DiDi-Chengdu(GPS) Chengdu Bus (GPS)	NA	×	Regional Flow	ResNet+TCN	RMSE	HA, ARIMA, XGBoost, SAE, GRU, CNN-LSTM, ST-ResNet, LSTM
Guo et al. (2020) ^{S,T}	U U U	D.C(Detectors) Philadelphia(Detectors) PeMSD4(Detectors)	5 5 5	×	Network Flow	GCN+GRU	MAE, MAPE, MSE	HA, SVR, Random Forest, ARIMA, FNN, GRU, GCGRU
Song et al. (2020) ^{S,T}	U	PEMS3, PEMS4, PEMS7, PEMS08(Detectors)	5	×	Network Flow	STSGCM	MAE, MAPE, RMSE	VAR, SVR, LSTM, DCRNN, STGCN, ASTGCN(r), STG2Seq, Graph WaveNet
Dai et al. (2020) ^{S,T}	U	Ninjing (video surveillance devices)	15, 30, 45, 60, 75, 90	×	Speed	ConvLSTM +GCN	RMSE	GCN, ConvLSTM, TGCN
Cui et al. (2020a) ^{S,T}	F U	Loop (Detectors) NPMRDS (Probe veh.)	5 5	×	Speed	GWGR	MAE, MAPE, RMSE	ARIMA, SVR, FNN, LSTM, LSGC + LSTM, SGC + LSTM STGCN, TGCLSTM,

(continued on next page)

Table 10 (continued).

Author	Area	Dataset (Type)	Data Interval (min)	External Data	Application Task	Model	Evaluated Algorithm	Comparative Model
Wang et al. (2019a) ^{S,T}	U	Xuancheng(AVI)	5	×	Speed	Bi-LSTM NN	MAE	PBDL-CP, PBDL-RP, PBDL-CP CNN, LSTM NN, ANN, KNN
Bogaerts et al. (2020) ^{S,T}	U	DiDiXi'an (GPS) DiDiChengdu (GPS)	5	×	Speed	GraphCNN+LSTM+ FNNN	RMSE, MSE, MAE, MAPE	k-NN, LSTM, SVM
Liu et al. (2019b) ^{S,T}	U	NYC Taxi & Limousine Commission (GPS)	30	✓	Travel Demand	CSTN	OD-MAPE, O-MAPE	HA, LR, XGBoost, MLP, ConvLSTM, ST-ResNet
Wang et al. (2019d) ^{S,T}	U U	UCAR-Beijing(GPS) DiDi-Chengdu(GPS)	60	✓	Travel Demand	GEML	RMSE, SMAPE	HA, LSTM, LSTNet, GCRN
Liu et al. (2019a) ^{S,T}	U	OSM	15	×	Regional Flow	DeepRTP	RMSE	HisAve, HisAve-w, ARIMA, SARIMA, VAR
Zhao et al. (2019a) ^{S,T}	U H	SZ-taxi(GPS) LOS-loop(Detectors)	15 5	×	Network Flow	GCN+GRU	MAE, RMSE, R ² , var, Accuracy	HA, ARIMA, SVR, GCN, GRU
Zhang et al. (2019b) ^{S,T}	U U	CTA buses (GPS) PeMS(Detectors)	10 NA	×	Network Flow	GAN+CNN +LSTM	MAE, MRE, RMSE	ARIMA, SVR, CNN-SVR , the scalable deep traffic flow neural networks, DCRNN
Qu et al. (2019) ^T	F	DRIVENET (Detectors)	2,10,15, 30,60	✓	Network Flow	DNN	MAPE	CM
Liu et al. (2019c) ^{S,T}	R	PeMS (Detectors)	5	×	Speed	TCHA	MAE, MAPE, RMSE	SVR, SAE, LSTM, GRU, hierarchical attention
Wu et al. (2019) ^{S,T}	H U	METR-LA(Detectors) PEMS-BAY(Detectors)	NA	×	Speed	Graph WaveNet	MAE, RMSE, MAPE	ARIMA, FC-LSTM , WaveNet, DCRNN, GGRU, GGRU, STGCN

S: Spatial Dependency, T: Temporal Dependency, F: Freeway(s), E: Expressway(s), H: Highway(s), U: Urban Arterial/Road Networks.

6. Open research issues and future research directions

Section 5 outlines the various traffic prediction states in the literature and their associated applications. However, applying the developed models and assessing their real-world benefits presents open research challenges. Consequently, the existing literature generally concludes its performance analysis after quantifying the model’s predictive accuracy. In other words, it fails to assess the real-world benefits of their approach for facilitating practical applications (e.g., vehicle routing). As shown in Fig. 8, we highlight some of the open research areas and future challenges in DL-based traffic prediction and elucidate how the common traffic prediction states and practical problems influence future research directions in this section.

6.1. Traffic prediction for connected and autonomous vehicles

Reliable traffic prediction methods will be a key enabler of autonomous vehicle applications within the IoV to ensure efficient vehicle routing within the transportation network. AV driving systems can revolutionize societal mobility and provide widespread

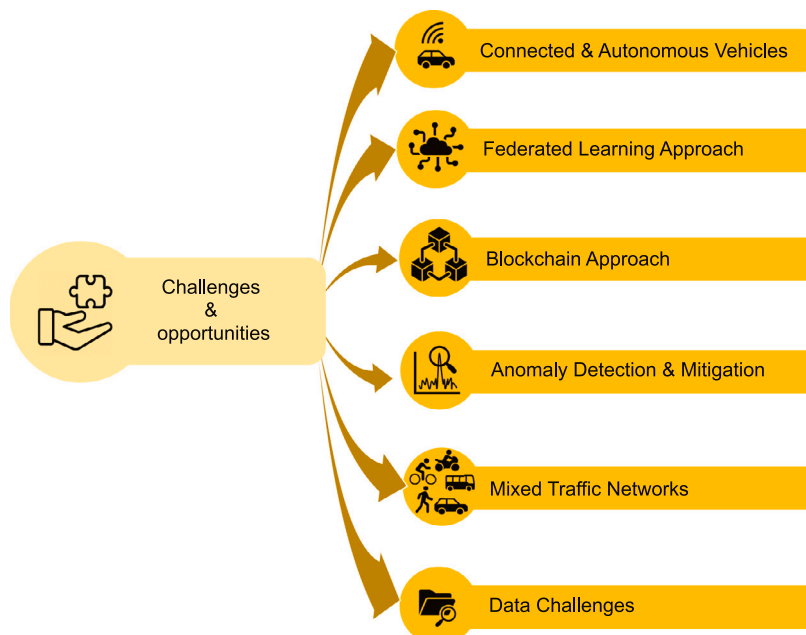


Fig. 8. Schematic overview of identified challenges and suggested research opportunities.

benefits, including better utilization of existing roadway infrastructure, increased access to various mobility services, reduced negative environmental impact from vehicle usage, and improved safety. AVs are expected to eventually eliminate driver ignorance in favor of dynamic and data-driven intelligent decision systems that adapt their approach to the surrounding environmental conditions. Consequently, new applications with the potential to reduce traffic congestion will become possible, including adaptive cruise control and connected and autonomous traffic control systems (Zhong et al., 2020a). In this work, a simulation analysis elucidates that introducing CAVs into the mixed traffic environment will considerably impact the behavior of human drivers (e.g., increasing lane change frequency). Notably, the performance of these AV-based systems will become inherently dependent on the available traffic prediction capabilities (Miglani and Kumar, 2019; Zhong et al., 2020b).

However, the decentralized nature of the IoV environment, where data is collected and stored on a diverse set of heterogeneous devices (e.g., CAV, base station, RSU, sensor, edge device), does not match well with the existing cutting-edge traffic prediction methods due to their centralized nature. For example, network-scale traffic prediction has been studied extensively as a way to predict the traffic for all links within a complex urban transportation network simultaneously to support vehicle routing and traffic management decisions (Cui et al., 2020a). Still, the resulting models require immense amounts of traffic data to be collected throughout the network and aggregated to a central server (e.g., cloud) for processing and model training. This design presents privacy concerns in the decentralized IoV environment, where most trajectory data will be crowd-sourced and collected by Connected Autonomous vehicles (CAVs) owned by multiple stakeholders, in contrast to a single party (e.g., TMC or DoT) collecting processing all of the traffic data.

Moreover, latency and bandwidth for excessive data transfer present additional challenges, and it is unlikely that it will be efficient to transfer and store all connected vehicle and sensor data in the centralized cloud. Additionally, energy consumption for onboard model training within CAVs presents a critical concern. Energy-efficient computation offloading strategies for ITS networks is an open research area receiving considerable attention (Li et al., 2020a). Notably, the existing research reveals that an optimization strategy considering the computational resources of RSUs, edge devices, and mobile base stations is a promising direction. However, this system will need to interface and account for the capacity of existing mobile edge computing networks and electrical grid infrastructure, presenting a complex challenge for future research.

Consequently, additional research into privacy-preserving and decentralized traffic prediction models will be necessary to support the deployment of CAVs and associated applications (e.g., ridesharing and vehicle loaning) within the emerging IoV-based transportation networks. Recently, FL has been proposed as one method for achieving decentralized and privacy-preserving traffic prediction models (Liu et al., 2020b); however, the approach is still new and requires significant improvements before the proposed models are practical. Below, we provide more details on FL and its potential to revolutionize traffic prediction within the IoV.

6.2. Federated learning for traffic prediction

FL is an emerging approach to privacy-preserving and distributed machine learning that can potentially revolutionize the future of traffic prediction within the IoV. In FL, participating devices work collaboratively to train a shared prediction model iteratively in a series of rounds. During each round, the traffic collection device will share just the learned parameters (gradients) with the other devices. Specifically, each device uses its own collected data to train a localized model. After training, devices will periodically transmit the local model state to a parameter server, where an algorithm (e.g., Federated Averaging) is used to generate a new global set of parameters from the contributions of each participant. After that, the global parameter set is distributed back to each device, and the process repeats until a convergence condition is achieved. A comprehensive survey on FL is provided in Du et al. (2020) for interested readers.

Regarding traffic prediction in the IoV, FL can train a more comprehensive traffic prediction model and continuously update it while preserving the privacy of crowd-sourced CAV data and alleviating the need to transmit and store vast quantities of distributed data. Recently, there have been a few attempts to propose FL-based models for traffic prediction; notably, federated versions of the GRU-based traffic prediction model are presented in Liu et al. (2020b) and Qi et al. (2021). However, the simulations conducted in these studies are not entirely practical as they appear to simulate FL on identical data shards distributed amongst a set of participants. In the practical IoV environment, devices are heterogeneous, and participants are autonomous, resulting in a data skew between participants. Notably, problems of data imbalance, missing classes, missing features, and missing values can limit the practical implementation of FL-based traffic prediction models. When these problems arise, the resulting models perform poorly compared to the existing approaches. As data skew is likely to occur in the distributed IoV environment, further research will be necessary to account for data skew amongst participants and improve the performance of the proposed traffic prediction FL models under this constraint.

In addition to data skew, it has yet to be discovered whether the best-performing traffic prediction model designs can be trained using Federated Learning. As we learned from previous sections, the revolutionary models within the literature are hybrid approaches operating on high-dimensional multivariate time series sensor data. They leverage DL methods (e.g., convolution) to extract the input data's spatial, temporal, and spatio-temporal trends. However, doing so requires an aggregated and pre-processed data structure built upon data sharing. This observation raises the question of whether it is possible to train an online FL-based traffic prediction model that can provide suitable accuracy compared to the centralized approaches. While probe trajectory data collected from mobile devices or CAVs will include some private and identifiable information, the privacy of ground truth traffic sensor data is not a practical consideration. Consequently, one potential option is to relax the data sharing assumption for localized groups or neighborhoods of sensors. In this way, a more complex model benefiting from feature extraction of multivariate sensor

data can be realized while simultaneously benefiting from the distributed edge computation and reduced communication overhead provided by FL.

In summary, due to the distributed and heterogeneous nature of data collection in IoV, FL-based models are a promising method for leveraging the benefits of edge computing within the IoV during model training for various MTP applications. Other related disciplines with similar data and network structures, including industrial IoT (Lu et al., 2019), have demonstrated the effectiveness of FL approaches in distributed networks having properties of strong temporal and spatial locality. Since traffic time series data and the IoV closely resemble the industrial IoT environment, we believe FL designs have strong applicability to MTP applications.

6.3. Blockchain-enabled traffic prediction

Existing traffic prediction methods based on centralized machine learning need to gather raw data for model training, which involves serious privacy exposure risks when the data is collected by heterogeneous IoV devices owned by multiple stakeholders. Due to the inherent immutability and decentralization qualities of blockchain networks, this technology has enabled considerable advances in traffic prediction capabilities during the last few years. For accurate and efficient traffic jam probability estimation, Hassija et al. (2020) proposed a neural network-based smart contract to be deployed onto the blockchain network. Qi et al. (2021) proposed a consortium blockchain-based FL framework is proposed to enable decentralized, reliable, and secure FL without a centralized model coordinator. In addition, Shahbazi and Byun (2021) introduced a blockchain-based framework for traffic demand service prediction. Despite much progress recently, blockchain technology has yet to be fully formed or standardized in the traffic prediction context. For example, existing popular blockchain networks (e.g., Bitcoin and Ethereum) require high energy consumption, provide low transaction throughput, and have storage and bandwidth scalability problems.

Like FL, blockchain has strong applicability to the IoV and MTP applications because of its inherent decentralized network structure with fault tolerance, strong immutability, and pluggable security protocols (Wang et al., 2021a). Advancements in edge computing have demonstrated the benefits of data locality (e.g., having data and computation exist close to the user or point of access) (Carlini, 2016). For MTP within the IoV, blockchain can provide an edge solution for distributing computation (e.g., model training and inference), storing, and sharing data. In contrast to distributed databases, blockchain networks support smart contracts, which are chunks of code capable of controlling access to computation and storage resources. This way, a programmable interface enables fine-grained security controls and supports distributed applications. However, the additional functionality comes at the cost of added resource consumption for nodes managing the blockchain network, compared to a distributed database approach. Consequently, if distributed and edge-based data storage and retrieval are the only desired operations, MTP applications may benefit from the more lightweight approach combining local peer-to-peer gossip networks and distributed databases.

That being said, smart contracts present an exciting method for enabling distributing computation within the IoV and have the potential to support much future distributed IoV applications (e.g., ridesharing) (Li et al., 2020d). However, concerns of excessive energy consumption, low throughput, and high resource consumption associated with the existing blockchain networks necessitate further research to improve the scalability and efficiency of these distributed networks. Thankfully, much research is being performed to address these challenges using novel approaches such as zero-knowledge proof transaction offloading (Ashur and Dhooghe, 2018; Panait and Olimid, 2020; Ben-Sasson et al., 2019), and lighter-weight consensus schemes (Guo et al., 2021b) designed for location-dependent networks (e.g., IoV).

6.4. Anomaly traffic event detection and mitigation

Non-recurrent anomaly events, such as accidents, work zones, weather, and special events, drastically reduce the capacity of the transportation network, creating traffic congestion. According to the Federal Highway Administration (FHWA), it was estimated that about 50% of traffic network congestion results from non-recurrent events (FHWA, 2021). Resolving non-recurrent congestion and returning the network to total capacity requires aggressive action and management by the TMC and associated stakeholders (e.g., police and emergency services). Example response efforts include timely communication to travelers, dynamic rerouting of existing traffic, and removing any obstructions from the traveled way. However, coordinating a response effort and communicating updates to travelers takes time. Consequently, developing methods for rapid anomaly detection is an exciting and essential research area for the transportation community.

While anomaly detection has been studied more extensively in wireless networks, the literature for anomaly detection in transportation networks using traffic data is small. In one related work, Lu et al. (2009), novel path anomaly detection algorithms leveraging KNNs are proposed for urban-scale transportation networks. The authors conducted a case study and concluded the classification accuracy to be around 90%. However, the case study was conducted on synthetic path data and not ground truth sensor data and may not be capable of detecting abnormal non-recurrent congestion in real-time. More recently, Hassan et al. (2019) proposed a method for detecting spatio-temporal anomalies in ITS leveraging traffic sensor data with a 5-min resolution. Experiments are conducted to determine the classification precision and optimal parameters for historical data length and training window, with the optimal parameter set having around 70% classification precision and a running time of less than one second. F1 score is used to account for precision and recall in the performance analysis, demonstrating a trade-off between classification precision and recall as the historical data length and sliding training window sizes increase. However, the authors' dataset consists exclusively of highway ramp detectors, and it is uncertain if this method will operate satisfactorily on links of different functional classes. Moreover, the performance of various network sizes is not analyzed. Further research will be necessary to elucidate the effects of varying network topology and spatial data granularity for anomaly detection in ITS.

Few existing works notably leverage DL models to classify anomaly traffic congestion, providing a potential future research direction. In more recent work, [Davis et al. \(2020\)](#) proposed an end-to-end deep learning-based anomaly detection method for transportation networks. The authors leverage an LSTM model with a customized extreme value theory-based objective function to classify anomaly events across multiple datasets, including speed, travel time, and taxi demand data. Compared to the existing statistical, ML, and hybrid DL approaches, the proposed approach has the best F1 score for classification across most datasets. The authors also outline additional avenues for future work, such as exploring different objective functions, testing the method on other datasets, and proposing strategies for identifying and quantifying the factors that cause the anomaly.

Moreover, it would be interesting to test different combinations of DL models for classifying anomaly events. LSTM and GCN model combinations have demonstrated considerable success in predicting traffic flow and speed in small-scale and network-wide contexts using low-resolution temporal data sequences (e.g., 5 min). A GCN model could extract important spatial features from the network topology for classifying anomaly events in real-time, while the online LSTM model processes recent temporal data sequences to classify the current state as recurrent or non-recurrent. However, the training time can be long in large-scale transportation networks with fine-grained data resolution. Furthermore, these models have degraded performance when predicting under perturbation (e.g., non-recurrent network events). To address these shortcomings, research into a hybrid and real-time system integrating models optimized for non-recurrent event identification and prediction under perturbation, to be used dynamically with the recurrent prediction models, presents an interesting future direction.

Prediction under perturbation presents another problem relating to the response effort for resolving non-recurrent traffic congestion: efficiently rerouting the existing traffic. Practically, rerouting travelers requires selecting less optimal routes, and the response effort may temporarily change the network topology, hindering the performance of route guidance applications leveraging existing DL models. The challenge of predicting under perturbation can be partially attributed to the fact that most literature focuses on the recurrent traffic prediction problem and trains their models with preprocessed data where anomaly events have been normalized or removed. Some researchers have recently challenged the community to rethink how we treat anomaly data, illustrating its potential to elucidate new insights when analyzed separately from the recurrent data ([John, 2021](#)). Currently, benchmark datasets containing a sufficient density of anomaly events are challenging to find. A standardized benchmark dataset for this application is of sufficient need to the research community. In the interim, some potential datasets for future experimentation found in the existing literature include travel time, vehicle occupancy, and traffic speed data from the Twin Cities Metro Area.³³

6.5. Traffic prediction in mixed traffic transportation networks

Automobile-related traffic congestion is a significant problem in transportation networks that adversely impacts travel time reliability. The consensus in the transportation community over the past two decades is that we cannot build ourselves out of congestion ([Downs, 2004](#)), emphasizing the need for improving access to multi-modal travel. Consequently, increased emphasis on building multi-modal mobility systems, in contrast to personal vehicle infrastructure, has resulted in a mixed traffic environment, especially in urban areas. One of the most promising applications of traffic prediction for relieving congestion in mixed traffic transportation networks relates to OD travel time prediction problems: scheduling complete trips. A complete trip can be defined as an origin to destination trip encompassing multiple modes (e.g., walking, bicycle, metro, automobile).

Efficiently scheduling complete trips given an origin and destination pair encompasses two essential functions: multi-modal and time-dependent traffic prediction and shortest path algorithms. Ideally, reliable traffic prediction models trained on historical data should inform the shortest path algorithms. The body of literature on multi-modal shortest path algorithms is extensive and includes [Li et al. \(2010\)](#), [Zhou et al. \(2008\)](#) and [Bielli et al. \(2006\)](#). However, these works make many assumptions regarding individual decision-making, transit schedules, and travel time reliability. Moreover, they generally only consider a subset of modes and exclude non-automated modes such as walking and bicycling. They also do not employ DL models for predicting traffic for consideration when scheduling paths. It would be beneficial to experiment with integrating reliable multi-modal OD and path travel time DL models to inform the time-dependent shortest path algorithms and expand their multi-modal decision frame. A system for scheduling complete trips while accounting for individual considerations can alleviate automobile congestion and provide more reliable and cost-effective travel opportunities while expanding societal access to mobility.

That being said, mixed traffic environments are the least represented in the literature due to a limited understanding of complex and unknown interactions and relationships between human vehicles, autonomous vehicles, bicycles, pedestrians, public transportation, etc. More research into traffic prediction within the mixed traffic environment will be required to improve our understanding of the diverse interactions and spatio-temporal correlations. Data availability presents one barrier to future research in this area, as some transportation networks have limited spatial data coverage of pedestrian facilities. At the same time, datasets for some modes, such as pedestrian volumes, are also limited in availability. Existing datasets also generally contain data from a single mode, and constructing a multi-modal dataset for benchmarking various approaches would benefit future research. Moreover, an interdisciplinary effort between traffic engineers, behavioral scientists, and computer scientists will likely be necessary to account for the human decision-making process during modeling. In summary, it is currently unknown if DL models can support a multi-modal traffic prediction application, as the existing models primarily focus on traffic prediction of a single mode.

³³ <https://github.com/numenta/NAB/tree/master/data>

6.6. Data challenges in traffic prediction

In this subsection, we outline some existing challenges in traffic prediction research related to input data for training DL models.

- **Limited Data Accessibility and Synthetic data.** The increasing number of sensors deployed within the traffic network in recent years has led to high volumes of collected data in some geographic areas; however, accessing this data for modeling purposes can be challenging in practice. Much of this data is collected by private companies (e.g., Uber, DiDi), as well as government organizations (e.g., TMC and Department of Transportation (DoT)), and is not always made publicly available. Moreover, if the desired data exists but is collected by a private data broker (e.g., INRIX), the acquisition cost can be quite high, restricting access for some researchers. In other cases, the data used in an experiment is provided by a partnership or grant with a local or federal organization and is not readily available to the public for results comparison and validation against other approaches. As a potential solution to data accessibility problems, some researchers have experimented with methods for generating artificial traffic data to augment their existing datasets or to create entirely new datasets that mimic the distribution of the ground truth data. Recently, the GAN attention-based neural network model has shown promise in this area, and many attempts have been made to generate actionable spatio-temporal data for future modeling efforts (Gao et al., 2022). GAN-based techniques are widely applicable to improving traffic prediction methods where data sparsity is a concern. However, only a few recent research works focus on the traffic data imputation problem in the literature (Huang et al., 2021a; Kazemi and Meidani, 2021).
- **Benchmarking Traffic Prediction.** A primary need in future traffic prediction research is a comprehensive database of benchmark datasets for comparative analysis of the proposed methods. As more traffic data become available each year, the diversity of datasets leveraged within the literature for analyzing the performance of proposed approaches has become quite diverse. For example, Table 4 provides a list of 40 of the most commonly used large-scale ground truth datasets leveraged for traffic prediction DL models within the literature. The high variability between datasets used in existing case studies is notable because the performance of any prediction model is dependent not only on the model architecture but also on the underlying data characteristics (e.g., distribution, density, quality, features). As a result, cross-comparison of results between approaches trained and evaluated on different datasets does not provide much actionable insight for real-world stakeholders seeking to implement DL models for practical applications. We believe further standardization of traffic datasets within the literature will elucidate new insights into the practical performance trade-offs of the vast array of traffic prediction approaches within the existing literature. Some researchers have recently attempted to pave the way towards standardization by providing open-sourced platforms for accessing diverse sets of unified traffic data, such as LibCity (Wang et al., 2021b), laying the foundation for future work in the area of standardization.
- **External Data Restrictions.** Many works have demonstrated the benefits of considering external data features in their prediction models (e.g., meteorological data, pandemic data, event data, social media data, and work zone data), as illustrated in Table 6. However, the literature on external data-based traffic prediction is somewhat limited due to the challenges of synthesizing the external data with the ground truth traffic data. This issue is also discussed in another survey (Tedjopurnomo et al., 2020), where the authors suggest first establishing a benchmark dataset of ground truth traffic data with sufficient spatio-temporal coverage and data density. After that, the traffic data can be concatenated with supplementary data if needed. Lastly, each temporal data instance can be augmented to include the necessary geographical external data, such as weather and accident information. This way, a single comprehensive benchmark dataset is built from possibly disjoint spatio-temporal and external datasets. We believe this is a sound and practical approach to building external datasets for benchmarking. A benchmark dataset in this format would greatly aid future researchers in comparative analysis, enabling the elucidation of underlying spatio-temporal relationships between the traffic and external data.
- **Multi-source Data.** We define multi-source data as a comprehensive dataset combining multiple types of ground truth traffic data, such as flow and speed data, with external data sources (e.g., weather, social media, and event data). Most existing traffic prediction research generally focuses on experimenting with a single dataset for their proposed models (e.g., speed data, flow data, trip data). Consequently, the influence of other related factors on the overall prediction outcome is not always thoroughly investigated. Furthermore, non-linear interactions among different data series are generally not considered and are treated as separate features. In the future, when attempting to forecast traffic conditions, we should consider adapting traditional approaches to fuse information from multiple sources and varying datasets to improve our understanding of the underlying correlations between the various transportation data types. However, finding effective methods for fusing the multitudes of raw data from varying sources into one comprehensive and overarching model becomes a challenging and open research problem. Besides, high-dimensional features create increasingly high computational costs for training, so it also becomes necessary to determine the feature importance and only extract those that will be most beneficial to the application goals. Attention-based DL approaches have recently found success in the literature in quantifying the relevance of various features. SAE DL models can also help to reduce feature counts in multiple applications where the feature set is large (e.g., network-wide traffic prediction) by hierarchically extracting the most important trends and discarding the other information. GCN networks have also demonstrated some success translating traffic data into a graph-based representation capable of capturing and encoding the spatial relationships within the traffic network at the node and link level. We believe further research into these approaches will be fundamental to improving the predictive performance of the resulting traffic prediction models.

- **Data Collection Area.** The distinct topological structures of various transportation networks are a primary consideration when training data-driven traffic prediction models. In the cutting-edge approaches for applications such as network-wide traffic prediction, GCN and other graph-based methods are leveraged to capture the spatial properties of the transportation network and translate them to actionable features for temporal prediction models such as LSTM. Consequently, the cutting-edge approaches perform well when predicting traffic within the network they were trained in but tend to fail when attempting to predict outside of this context. Theoretically, we could train specialized prediction models for each subset of the transportation network, but this is not the most efficient approach. Practically, we seek to answer whether these existing models can be adapted to work in other transportation networks or if we can leverage the current models to speed up the training of newer models.

Recently, transfer learning has emerged as a deep learning technique for rapidly building new models by leveraging the wisdom of existing models. However, transfer learning literature for the traffic prediction application is minimal. We believe transfer learning methods will become increasingly important in the future of traffic prediction as a means of solving the adaptability problem of large-scale traffic prediction models. In addition to adaptability, the data collection area also influences the focus of recent studies. Generally, the widely used traffic datasets come from four main collection areas: regional urban transportation networks, expressways, freeways, and highways. Currently, most studies focus on utilizing available freeway, highway, or expressway data to achieve high accuracy. At the same time, traffic prediction in urban arterial networks is more challenging in practice due to the increasingly complicated network structure and the existence of signalized intersections (Stathopoulos and Karlaftis, 2003; Advani et al., 2020). Notably, less emphasis has been applied to the urban arterial prediction problem. That being said, some current works focus on broadened decision frames that consider not only a single corridor or expressway but also the impacts of neighboring links and intersections within the urban network (Yu et al., 2017), while Advani et al. (2020) improved the representation of the various vehicle categories found in urban systems, such as cars, bicycles, pedestrians, and public transit vehicles within the models. However, these advances lead to more complex prediction models (e.g., complex hybrid DNNs), which require more data and significantly increased time to train effectively.

7. Conclusion

Although a variety of traffic forecast methods have been proposed in the literature, ranging from statistical modeling to machine learning, traffic forecasting remains a challenging task due to the complex relationships between the data, external factors, non-linearity, and the inherent spatial and temporal dependencies. While deep learning models have made great strides in improving the accuracy and scale of traffic prediction methods, there is still much work to be done. In this paper, we conduct a broad survey on the current state of traffic prediction research. More specifically, we first comprehensively outline the various data types used in traffic prediction and their limitations, and then we survey the existing work on preprocessing this traffic data. Next, we summarize the classical and current traffic prediction methods and provide a taxonomy. Additionally, we survey different traffic prediction tasks and analyze the existing approaches within the literature, focusing on multivariate traffic time series modeling applications. Lastly, some significant challenges, as well as future research directions, are discussed. This paper aims to provide a suitable reference for readers to quickly get up-to-speed regarding the current state of traffic prediction and provide a strong reference for guiding future research work in this field.

CRedit authorship contribution statement

Maryam Shaygan: Investigation, Writing – original draft. **Collin Meese:** Investigation, Writing – original draft. **Wanxin Li:** Review & editing. **Xiaoliang (George) Zhao:** Review & editing. **Mark Nejad:** Supervision, Review & editing.

Acknowledgments

This research is supported in part by a Federal Highway Administration grant: Artificial Intelligence Enhanced Integrated Transportation Management System, 2020–2023. The authors thank Gene Donaldson (DelDOT TMC) for helpful discussions and insights.

References

- Abdulhai, B., Porwal, H., Recker, W., 2002. Short-term traffic flow prediction using neuro-genetic algorithms. *ITS J.-Intell. Transp. Syst. J.* 7 (1), 3–41.
- Abid, A., El Khediri, S., Kachouri, A., 2021. Improved approaches for density-based outlier detection in wireless sensor networks. *Computing* 1–18.
- Advani, C., Thakkar, S., Arkatkar, S., Bhaskar, A., 2020. Performance evaluation of urban arterial network using Wi-Fi sensors under heterogeneous traffic conditions. *Transp. Res. Procedia* 48, 1022–1037.
- Ahsani, V., Amin-Naseri, M., Knickerbocker, S., Sharma, A., 2019. Quantitative analysis of probe data characteristics: Coverage, speed bias and congestion detection precision. *J. Intell. Transp. Syst.* 23 (2), 103–119.
- Akhtar, M., Moridpour, S., 2021. A review of traffic congestion prediction using artificial intelligence. *J. Adv. Transp.* 2021.
- Al-Hussaeni, K., Fung, B.C., Iqbal, F., Dagher, G.G., Park, E.G., 2018. SafePath: Differentially-private publishing of passenger trajectories in transportation systems. *Comput. Netw.* 143, 126–139.
- Aljuaydi, F., Wiwatanapataphee, B., Wu, Y.H., 2022. Deep learning-based prediction models for freeway traffic flow under non-recurrent events. In: 2022 8th International Conference on Control, Decision and Information Technologies, Vol. 1. CoDIT, IEEE, pp. 815–820.
- Alsolami, B., Mehmood, R., Albeshri, A., 2020. Hybrid statistical and machine learning methods for road traffic prediction: A review and tutorial. *Smart Infrastruct. Appl.* 115–133.

- Andonie, R., 2019. Hyperparameter optimization in learning systems. *J. Membrane Comput.* 1 (4), 279–291.
- Angarita-Zapata, J.S., Masegosa, A.D., Triguero, I., 2019. A taxonomy of traffic forecasting regression problems from a supervised learning perspective. *IEEE Access* 7, 68185–68205.
- Ashur, T., Dhooghe, S., 2018. Marvellous: A STARK-friendly family of cryptographic primitives. *Cryptol. EPrint Archive*.
- Atif, Y., Kharrazi, S., Jianguo, D., Andler, S.F., 2020. Internet of things data analytics for parking availability prediction and guidance. *Trans. Emerg. Telecommun. Technol.* 31 (5), e3862.
- Bahdanau, D., Cho, K.H., Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations. ICLR 2015.
- Bai, L., Yao, L., Kanhere, S.S., Wang, X., Sheng, Q.Z., 2019. STG2seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. pp. 1981–1987.
- Bai, L., Yao, L., Li, C., Wang, X., Wang, C., 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Adv. Neural Inf. Process. Syst.* 33, 17804–17815.
- Balaprakash, P., Salim, M., Uram, T.D., Vishwanath, V., Wild, S.M., 2018. DeepHyper: Asynchronous hyperparameter search for deep neural networks. In: 2018 IEEE 25th International Conference on High Performance Computing. HiPC, IEEE, pp. 42–51.
- Bao, X., Jiang, D., Yang, X., Wang, H., 2021. An improved deep belief network for traffic prediction considering weather factors. *Alex. Eng. J.* 60 (1), 413–420.
- Barros, J., Araujo, M., Rossetti, R.J., 2015. Short-term real-time traffic prediction methods: A survey. In: 2015 International Conference on Models and Technologies for Intelligent Transportation Systems. MT-ITS, IEEE, pp. 132–139.
- Belhadi, A., Djenouri, Y., Djenouri, D., Michalak, T., Lin, J.C.W., 2021. Machine learning for identifying group trajectory outliers. *ACM Trans. Manag. Inf. Syst. (TMIS)* 12 (2), 1–25.
- Belkin, M., Hsu, D., Ma, S., Mandal, S., 2019. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proc. Natl. Acad. Sci.* 116 (32), 15849–15854.
- Ben-Sasson, E., Bentov, I., Horesh, Y., Riabzev, M., 2019. Scalable zero knowledge with no trusted setup. In: Annual International Cryptology Conference. Springer, pp. 701–732.
- Bengio, S., Vinyals, O., Jaitly, N., Shazeer, N., 2015. Scheduled sampling for sequence prediction with recurrent neural networks. *Adv. Neural Inf. Process. Syst.* 28.
- Bermingham, L., Lee, I., 2017. A framework of spatio-temporal trajectory simplification methods. *Int. J. Geogr. Inf. Sci.* 31 (6), 1128–1153.
- Bhowmick, K., Narvekar, M., 2018. Trajectory outlier detection for traffic events: A survey. In: Intelligent Computing and Information and Communication. Springer, pp. 37–46.
- Bielli, M., Boulmakoul, A., Mouncif, H., 2006. Object modeling and path computation for multimodal travel systems. *European J. Oper. Res.* 175 (3), 1705–1730.
- Bogaerts, T., Masegosa, A.D., Angarita-Zapata, J.S., Onieva, E., Hellinckx, P., 2020. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transp. Res. C* 112, 62–77.
- Boukerche, A., Wang, J., 2020a. Machine learning-based traffic prediction models for intelligent transportation systems. *Comput. Netw.* 181, 107530.
- Boukerche, A., Wang, J., 2020b. A performance modeling and analysis of a novel vehicular traffic flow prediction system using a hybrid machine learning-based model. *Ad Hoc Networks* 106, 102224.
- Bruna, J., Zaremba, W., Szlam, A., LeCun, Y., 2014. Spectral networks and deep locally connected networks on graphs. In: 2nd International Conference on Learning Representations. ICLR 2014.
- Burrello, A., Pagliari, D.J., Bartolini, A., Benini, L., Macii, E., Poncino, M., 2020. Predicting hard disk failures in data centers using temporal convolutional neural networks. In: European Conference on Parallel Processing. Springer, pp. 277–289.
- Cai, L., Janowicz, K., Mai, G., Yan, B., Zhu, R., 2020a. Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting. *Trans. in GIS* 24 (3), 736–755.
- Cai, S., Li, L., Li, Q., Li, S., Hao, S., Sun, R., 2020b. UWFP-outlier: An efficient frequent-pattern-based outlier detection method for uncertain weighted data streams. *Appl. Intell.* 50 (10), 3452–3470.
- Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., Sun, J., 2016. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transp. Res. C* 62, 21–34.
- Carlini, S., 2016. The drivers and benefits of edge computing. *Schneider Electric–Data Center Sci. Center* 8.
- Chai, D., Wang, L., Yang, Q., 2018. Bike flow prediction with multi-graph convolutional networks. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 397–400.
- Chan, R.K.C., Lim, J.M.Y., Parthiban, R., 2021. A neural network approach for traffic prediction and routing with missing data imputation for intelligent transportation system. *Expert Syst. Appl.* 171, 114573.
- Chao, P., Xu, Y., Hua, W., Zhou, X., 2020. A survey on map-matching algorithms. In: Australasian Database Conference. Springer, pp. 121–133.
- Chen, W., Chen, L., Xie, Y., Cao, W., Gao, Y., Feng, X., 2020a. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 04), pp. 3529–3536.
- Chen, C., Li, K., Teo, S.G., Zou, X., Wang, K., Wang, J., Zeng, Z., 2019. Gated residual recurrent graph neural networks for traffic prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, (no. 01), pp. 485–492.
- Chen, C.M., Liang, C.C., Chu, C.P., 2020b. Long-term travel time prediction using gradient boosting. *J. Intell. Transp. Syst.* 24 (2), 109–124.
- Chen, P., Liu, R., Aihara, K., Chen, L., 2020c. Autoreervoir computing for multistep ahead prediction based on the spatiotemporal information transformation. *Nature Commun.* 11 (1), 1–15.
- Chen, Y., Lv, Y., Wang, X., Li, L., Wang, F.Y., 2018. Detecting traffic information from social media texts with deep learning approaches. *IEEE Trans. Intell. Transp. Syst.* 20 (8), 3049–3058.
- Chen, Q., Song, Y., Zhao, J., 2020d. Short-term traffic flow prediction based on improved wavelet neural network. *Neural Comput. Appl.* 1–10.
- Chen, S., Wang, W., van Zuylen, H., 2010. A comparison of outlier detection algorithms for ITS data. *Expert Syst. Appl.* 37 (2), 1169–1178.
- Cheng, S., Lu, F., 2017. A two-step method for missing spatio-temporal data reconstruction. *ISPRS Int. J. Geo-Inf.* 6 (7), 187.
- Cheng, S., Lu, F., Peng, P., Wu, S., 2018. A spatiotemporal multi-view-based learning method for short-term traffic forecasting. *ISPRS Int. J. Geo-Inf.* 7 (6), 218.
- Cho, K., van Merriënboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: EMNLP.
- Chrobok, R., Kaumann, O., Wahle, J., Schreckenberger, M., 2004. Different methods of traffic forecast based on real data. *European J. Oper. Res.* 155 (3), 558–568.
- Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In: NIPS 2014 Workshop on Deep Learning, December 2014.
- Cui, G., Bian, W., Wang, X., 2021. Hidden Markov map matching based on trajectory segmentation with heading homogeneity. *GeoInformatica* 25 (1), 179–206.
- Cui, Z., Ke, R., Pu, Z., Ma, X., Wang, Y., 2020a. Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction. *Transp. Res. C* 115, 102620.
- Cui, Z., Lin, L., Pu, Z., Wang, Y., 2020b. Graph Markov network for traffic forecasting with missing data. *Transp. Res. C* 117, 102671.
- Cvetek, D., Muštra, M., Jelušić, N., Tišljarić, L., 2021. A survey of methods and technologies for congestion estimation based on multisource data fusion. *Appl. Sci.* 11 (5), 2306.

- Dai, F., Huang, P., Xu, X., Qi, L., Khosravi, M.R., 2020. Spatio-temporal deep learning framework for traffic speed forecasting in IoT. *IEEE Internet Things Mag.* 3 (4), 66–69.
- Dai, G., Ma, C., Xu, X., 2019. Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU. *IEEE Access* 7, 143025–143035.
- Datta, A., Banerjee, S., Finley, A.O., Gelfand, A.E., 2016. On nearest-neighbor Gaussian process models for massive spatial data. *Wiley Interdiscip. Rev. Comput. Stat.* 8 (5), 162–171.
- Davis, N., Raina, G., Jagannathan, K., 2016. A multi-level clustering approach for forecasting taxi travel demand. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems. ITSC, IEEE, pp. 223–228.
- Davis, N., Raina, G., Jagannathan, K., 2020. A framework for end-to-end deep learning-based anomaly detection in transportation networks. *Transp. Res. Interdiscipl. Perspect.* 5, 100112.
- Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* 29.
- Deng, L., Li, G., Han, S., Shi, L., Xie, Y., 2020. Model compression and hardware acceleration for neural networks: A comprehensive survey. *Proc. IEEE* 108 (4), 485–532.
- Di, X., Shi, R., 2021. A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning. *Transp. Res. C* 125, 103008.
- Dietterich, T.G., 2000. Ensemble methods in machine learning. In: *International Workshop on Multiple Classifier Systems*. Springer, pp. 1–15.
- Ding, Y., Li, Y., Deng, K., Tan, H., Yuan, M., Ni, L.M., 2015. Dissecting regional weather-traffic sensitivity throughout a city. In: 2015 IEEE International Conference on Data Mining. IEEE, pp. 739–744.
- Ding, Y., Li, Y., Deng, K., Tan, H., Yuan, M., Ni, L.M., 2017. Detecting and analyzing urban regions with high impact of weather change on transport. *IEEE Trans. Big Data* 3 (2), 126–139.
- Do, L.N., Taherifar, N., Vu, H.L., 2019. Survey of neural network-based models for short-term traffic state prediction. *Wiley Interdiscipl. Rev.: Data Min. Knowl. Discov.* 9 (1), e1285.
- Downs, A., 2004. Why traffic congestion is here to stay... and will get worse. *Access Mag.* 1 (25), 19–25.
- Du, Z., Wu, C., Yoshinaga, T., Yau, K.L.A., Ji, Y., Li, J., 2020. Federated learning for vehicular internet of things: Recent advances and open issues. *IEEE Open J. Comput. Soc.* 1, 45–61.
- Duan, P., Mao, G., Liang, W., Zhang, D., 2018. A unified spatio-temporal model for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* 20 (9), 3212–3223.
- Duggimpudi, M.B., Abbady, S., Chen, J., Raghavan, V.V., 2019. Spatio-temporal outlier detection algorithms based on computing behavioral outlieriness factor. *Data Knowl. Eng.* 122, 1–24.
- Elamrani Abou El Assad, Z., Mousannif, H., Al Moatassime, H., 2020. Class-imbalanced crash prediction based on real-time traffic and weather data: A driving simulator study. *Traffic Inj. Prev.* 21 (3), 201–208.
- Elleuch, W., Wali, A., Alimi, A.M., 2020. Neural congestion prediction system for trip modelling in heterogeneous spatio-temporal patterns. *Internat. J. Systems Sci.* 51 (8), 1373–1391.
- Ermagun, A., Levinson, D., 2018. Spatiotemporal traffic forecasting: Review and proposed directions. *Transp. Rev.* 38 (6), 786–814.
- Essien, A., Petrounias, I., Sampaio, P., Sampaio, S., 2020. A deep-learning model for urban traffic flow prediction with traffic events mined from twitter. *World Wide Web* 1–24.
- Faghhi, S., Safikhani, A., Moghimi, B., Kamga, C., 2019. Predicting short-term uber demand in new york city using spatiotemporal modeling. *J. Comput. Civ. Eng.* 33 (3), 05019002.
- FHWA, 2019. Utilizing probe-vehicle data for work zone mobility performance measurement. URL <https://rosap.ntl.bts.gov/view/dot/43540>.
- FHWA, 2021. Reducing non-recurring congestion. URL https://ops.fhwa.dot.gov/program_areas/reduce-non-cong.htm. (Online Accessed 5 October 2021).
- Fu, R., Zhang, Z., Li, L., 2016. Using LSTM and GRU neural network methods for traffic flow prediction. In: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation. YAC, IEEE, pp. 324–328.
- Gao, M., Shi, G., Li, W., 2018. Online compression algorithm of AIS trajectory data based on improved sliding window. *J. Traffic Transp. Eng.* 18 (3), 218–227.
- Gao, H., Tang, J., Liu, H., 2012. GSCorr: Modeling geo-social correlations for new check-ins on location-based social networks. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. pp. 1582–1586.
- Gao, H., Wang, Z., Yan, Z., Yu, Z., Luo, W., Yuan, L., 2021. Synchronized entry-traffic flow prediction for regional expressway system based on multidimensional tensor. *Transp. Res. Rec.* 03611981211011169.
- Gao, N., Xue, H., Shao, W., Zhao, S., Qin, K.K., Prabowo, A., Rahaman, M.S., Salim, F.D., 2022. Generative adversarial networks for spatio-temporal data: A survey. *ACM Trans. Intell. Syst. Technol.* 13 (2), 1–25.
- Gehring, J., Auli, M., Grangier, D., Yarats, D., Dauphin, Y.N., 2017. Convolutional sequence to sequence learning. In: *International Conference on Machine Learning*. PMLR, pp. 1243–1252.
- Ghosh, S., Asif, M.T., Wynter, L., 2017. Denoising autoencoders for fast real-time traffic estimation on urban road networks. In: 2017 IEEE 56th Annual Conference on Decision and Control. CDC, IEEE, pp. 6307–6312.
- Gobezie, A., Fufa, M.S., 2020. Machine learning and deep learning models for traffic flow prediction: A survey.
- Gong, Y., Li, Z., Zhang, J., Liu, W., Yi, J., 2020. Potential passenger flow prediction: A novel study for urban transportation development. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, (no. 04), pp. 4020–4027.
- Gong, Y., Li, Z., Zhang, J., Liu, W., Zheng, Y., Kirsch, C., 2018. Network-wide crowd flow prediction of sydney trains via customized online non-negative matrix factorization. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. pp. 1243–1252.
- Gori, M., Monfardini, G., Scarselli, F., 2005. A new model for learning in graph domains. In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*, Vol. 2. IEEE, pp. 729–734.
- Gu, Y., Lu, W., Xu, X., Qin, L., Shao, Z., Zhang, H., 2019. An improved Bayesian combination model for short-term traffic prediction with deep learning. *IEEE Trans. Intell. Transp. Syst.* 21 (3), 1332–1342.
- Guan, J., Wang, W., Li, W., Zhou, S., 2018. A unified framework for predicting kpis of on-demand transport services. *IEEE Access* 6, 32005–32014.
- Guo, K., Hu, Y., Qian, Z., Liu, H., Zhang, K., Sun, Y., Gao, J., Yin, B., 2020. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.*
- Guo, K., Hu, Y., Sun, Y., Qian, S., Gao, J., Yin, B., 2021a. Hierarchical graph convolution network for traffic forecasting. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, (no. 1), pp. 151–159.
- Guo, H., Li, W., Nejad, M., 2021b. A location-based and hierarchical framework for fast consensus in blockchain networks. In: 2021 4th International Conference on Hot Information-Centric Networking. HotICN, IEEE, pp. 1–6.
- Guo, S., Lin, Y., Feng, N., Song, C., Wan, H., 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, (no. 01), pp. 922–929.
- Guo, J., Williams, B.M., 2010. Real-time short-term traffic speed level forecasting and uncertainty quantification using layered Kalman filters. *Transp. Res. Rec.* 2175 (1), 28–37.
- Guo, G., Zhang, T., 2020. A residual spatio-temporal architecture for travel demand forecasting. *Transp. Res. C* 115, 102639.

- Hajirahimi, Z., Khashei, M., 2019. Hybrid structures in time series modeling and forecasting: A review. *Eng. Appl. Artif. Intell.* 86, 83–106.
- Harrou, F., Zeroual, A., Sun, Y., 2020. Traffic congestion monitoring using an improved kNN strategy. *Measurement* 156, 107534.
- Hassan, M.H., Tizghadam, A., Leon-Garcia, A., 2019. Spatio-temporal anomaly detection in intelligent transportation systems. *Procedia Comput. Sci.* 151, 852–857.
- Hassija, V., Gupta, V., Garg, S., Chamola, V., 2020. Traffic jam probability estimation based on blockchain and deep neural networks. *IEEE Trans. Intell. Transp. Syst.* 1–10.
- He, Z., Chow, C.Y., Zhang, J.D., 2018. STANN: A spatio-temporal attentive neural network for traffic prediction. *IEEE Access* 7, 4795–4806.
- He, Z., Chow, C.-Y., Zhang, J.-D., 2020. STNN: A spatio-temporal neural network for traffic predictions. *IEEE Trans. Intell. Transp. Syst.*
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Högberg, P., 1976. Estimation of parameters in models for traffic prediction: A non-linear regression approach. *Transp. Res.* 10 (4), 263–265.
- Homayoun, S., Dehghantanha, A., Ahmadzadeh, M., Hashemi, S., Khayami, R., 2017. Know abnormal, find evil: Frequent pattern mining for ransomware threat hunting and intelligence. *IEEE Trans. Emerg. Top. Comput.* 8 (2), 341–351.
- Hou, Y., Chen, J., Wen, S., 2021. The effect of the dataset on evaluating urban traffic prediction. *Alex. Eng. J.* 60 (1), 597–613.
- Hu, S., Gao, S., Wu, L., Xu, Y., Zhang, Z., Cui, H., Gong, X., 2021. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Comput. Environ. Urban Syst.* 87, 101619.
- Huang, T., Chakraborty, P., Sharma, A., 2021a. Deep convolutional generative adversarial networks for traffic data imputation encoding time series as images. *Int. J. Transp. Sci. Technol.*
- Huang, R., Huang, C., Liu, Y., Dai, G., Kong, W., 2020. LSGCN: Long short-term traffic prediction with graph convolutional networks. In: *IJCAI*. pp. 2355–2361.
- Huang, L., Leng, H., Li, X., Ren, K., Song, J., Wang, D., 2021b. A data-driven method for hybrid data assimilation with multilayer perceptron. *Big Data Res.* 23, 100179.
- Huang, Y., Weng, Y., Yu, S., Chen, X., 2019. Diffusion convolutional recurrent neural network with rank influence learning for traffic forecasting. In: *2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering. TrustCom/BigDataSE, IEEE*, pp. 678–685.
- INRIX, 2022. INRIX Segments. URL <https://docs.inrix.com/reference/getsegments/>.
- Ishibashi, K., Harada, S., Kawahara, R., 2018. Inferring latent traffic demand offered to an overloaded link with modeling qos-degradation effect. *IEICE Trans. Commun.*
- James, J., 2021. Citywide traffic speed prediction: A geometric deep learning approach. *Knowl.-Based Syst.* 212, 106592.
- Jang, S.Y., Kim, S.A., Shin, D., 2018. Integration of big data to predict transportation for smart cities. *Int. J. Archit. Environ. Engineering* 11 (10), 1473–1480.
- Jenelius, E., Koutsopoulos, H.N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transp. Res. B* 53, 64–81.
- Jian, Y., Li, J., Wei, L., Gao, L., Mao, F., 2022. Spatiotemporal DeepWalk gated recurrent neural network: A deep learning framework for traffic learning and forecasting. *J. Adv. Transp.* 2022.
- Jiang, R., Cai, Z., Wang, Z., Yang, C., Fan, Z., Chen, Q., Tsubouchi, K., Song, X., Shibasaki, R., 2021. DeepCrowd: A deep model for large-scale citywide crowd density and flow prediction. *IEEE Trans. Knowl. Data Eng.*
- Jiang, W., Luo, J., 2022a. Big data for traffic estimation and prediction: A survey of data and tools. *Appl. Syst. Innov.* 5 (1), 23.
- Jiang, W., Luo, J., 2022b. Graph neural network for traffic forecasting: A survey. *Expert Syst. Appl.* 117921.
- Jin, J., Rong, D., Zhang, T., Ji, Q., Guo, H., Lv, Y., Ma, X., Wang, F.-Y., 2022. A GAN-based short-term link traffic prediction approach for urban road networks under a parallel learning framework. *IEEE Trans. Intell. Transp. Syst.*
- Jin, C., Ruan, T., Wu, D., Xu, L., Dong, T., Chen, T., Wang, S., Du, Y., Wu, M., 2021. HetGAT: A heterogeneous graph attention network for freeway traffic speed prediction. *J. Ambient Intell. Humaniz. Comput.* 1–12.
- John, J., 2021. Outlier detection and spatial analysis algorithms. *arXiv preprint arXiv:2106.10669*.
- Karimzadeh, M., Aebi, R., de Souza, A.M., Zhao, Z., Braun, T., Sargento, S., Villas, L., 2021. Reinforcement learning-designed LSTM for trajectory and traffic flow prediction. In: *2021 IEEE Wireless Communications and Networking Conference. WCNC, IEEE*, pp. 1–6.
- Kazemi, A., Meidani, H., 2021. IGANI: Iterative generative adversarial networks for imputation with application to traffic data. *IEEE Access* 9, 112966–112977.
- Ke, J., Qin, X., Yang, H., Zheng, Z., Zhu, Z., Ye, J., 2021. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *Transp. Res. C* 122, 102858.
- Kessler, L., Huber, G., Kesting, A., Bogenberger, K., 2018. Comparing speed data from stationary detectors against floating-car data. *IFAC-PapersOnLine* 51 (9), 299–304.
- Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. *J. Int. Conf. Learn. Representations (ICLR 2017)*.
- Kong, X., Xing, W., Wei, X., Bao, P., Zhang, J., Lu, W., 2020. STGAT: Spatial-temporal graph attention networks for traffic flow forecasting. *IEEE Access* 8, 134363–134372.
- Kowalek, P., Loch-Olszewska, H., Szwabiński, J., 2019. Classification of diffusion modes in single-particle tracking data: Feature-based versus deep-learning approach. *Phys. Rev. E* 100 (3), 032410.
- Kuang, L., Hua, C., Wu, J., Yin, Y., Gao, H., 2020. Traffic volume prediction based on multi-sources GPS trajectory data by temporal convolutional network. *Mob. Netw. Appl.* 25 (4), 1405–1417.
- Kullberg, A., Skog, I., Hendeby, G., 2021. Learning motion patterns in AIS data and detecting anomalous vessel behavior. In: *2021 IEEE 24th International Conference on Information Fusion. FUSION, IEEE*, pp. 1–8.
- Kumar, S.V., 2017. Traffic flow prediction using Kalman filtering technique. *Procedia Eng.* 187, 582–587.
- Kumar, A., Saha, P., 2020. A review of deep learning models for traffic flow prediction in autonomous vehicles. In: *2020 2nd International Conference on Advances in Computing, Communication Control and Networking. ICACCCN, IEEE*, pp. 303–308.
- Kumar, S.V., Vanajakshi, L., 2015. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *Eur. Transp. Res. Rev.* 7 (3), 1–9.
- Kwon, J., Cha, C., Park, H., 2021. Multilayered LSTM with parameter transfer for vehicle speed data imputation. In: *2021 IEEE International Symposium on Circuits and Systems. ISCAS, IEEE*, pp. 1–5.
- Lai, G., Chang, W.C., Yang, Y., Liu, H., 2018. Modeling long-and short-term temporal patterns with deep neural networks. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. pp. 95–104.
- Lana, I., Del Ser, J., Velez, M., Vlahogianni, E.I., 2018. Road traffic forecasting: Recent advances and new challenges. *IEEE Intell. Transp. Syst. Mag.* 10 (2), 93–109.
- Laña, I., Sanchez-Medina, J.J., Vlahogianni, E.I., Del Ser, J., 2021. From data to actions in intelligent transportation systems: A prescription of functional requirements for model actionability. *Sensors* 21 (4), 1121.
- Lea, C., Vidal, R., Reiter, A., Hager, G.D., 2016. Temporal convolutional networks: A unified approach to action segmentation. In: *European Conference on Computer Vision*. Springer, pp. 47–54.
- Lee, K., Eo, M., Jung, E., Yoon, Y., Rhee, W., 2021. Short-term traffic prediction with deep neural networks: A survey. *IEEE Access* 9, 54739–54756.
- Lee, K., Rhee, W., 2022. DDP-GCN: Multi-graph convolutional network for spatiotemporal traffic forecasting. *Transp. Res. C* 134, 103466.
- Li, W., Cao, J., Guan, J., Zhou, S., Liang, G., So, W.K., Szczecinski, M., 2018a. A general framework for unmet demand prediction in on-demand transport services. *IEEE Trans. Intell. Transp. Syst.* 20 (8), 2820–2830.

- Li, Y., Chai, S., Ma, Z., Wang, G., 2021a. A hybrid deep learning framework for long-term traffic flow prediction. *IEEE Access* 9, 11264–11271.
- Li, T., Chen, L., Jensen, C.S., Pedersen, T.B., 2021b. TRACE: Real-time compression of streaming trajectories in road networks. *Proc. VLDB Endow.* 14 (7), 1175–1187.
- Li, X., Dang, Y., Aazam, M., Peng, X., Chen, T., Chen, C., 2020a. Energy-efficient computation offloading in vehicular edge cloud computing. *IEEE Access* 8, 37632–37644.
- Li, Y., Fu, K., Wang, Z., Shahabi, C., Ye, J., Liu, Y., 2018b. Multi-task representation learning for travel time estimation. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 1695–1704.
- Li, J., Guo, F., Wang, Y., Zhang, L., Na, X., Hu, S., 2020b. Short-term traffic prediction with deep neural networks and adaptive transfer learning. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems. ITSC, IEEE*, pp. 1–6.
- Li, G., Knoop, V.L., van Lint, H., 2021c. Multistep traffic forecasting by dynamic graph convolution: Interpretations of real-time spatial correlations. *Transp. Res. C* 128, 103185.
- Li, Z.-C., Lam, W.H., Wong, S., Sumalee, A., 2010. An activity-based approach for scheduling multimodal transit services. *Transportation* 37 (5), 751–774.
- Li, L., Liu, W., 2020. Travel time prediction on un-monitored roads: A spatial factorization machine based approach (student abstract). In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, (no. 10), pp. 13855–13856.
- Li, W., Meese, C., Guo, H., Nejad, M., 2020c. Blockchain-enabled identity verification for safe ridesharing leveraging zero-knowledge proof. In: *2020 3rd International Conference on Hot Information-Centric Networking. HotICN*, pp. 18–24.
- Li, W., Meese, C., Guo, H., Nejad, M., 2020d. Blockchain-enabled identity verification for safe ridesharing leveraging zero-knowledge proof. In: *2020 3rd International Conference on Hot Information-Centric Networking. HotICN, IEEE*, pp. 18–24.
- Li, W., Nejad, M., Zhang, R., 2019a. A blockchain-based architecture for traffic signal control systems. In: *2019 IEEE International Congress on Internet of Things. ICIOT, IEEE*, pp. 33–40.
- Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., Zhang, W., Wang, Z., 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Front. Comput. Sci.* 6 (1), 111–121.
- Li, Y., Shahabi, C., 2018. A brief overview of machine learning methods for short-term traffic forecasting and future directions. *Sigspatial Special* 10 (1), 3–9.
- Li, H., Wang, Q., Xiong, W., 2021d. New model of travel-time prediction considering weather conditions: Case study of urban expressway. *J. Transp. Eng., Part A: Syst.* 147 (3), 04020161.
- Li, L., Yan, J., Yang, X., Jin, Y., 2019b. Learning interpretable deep state space model for probabilistic time series forecasting. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. pp. 2901–2908.
- Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018c. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: *International Conference on Learning Representations*.
- Li, K.L., Zhai, C.J., Xu, J.M., 2017. Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ann. In: *2017 Chinese Automation Congress. CAC, IEEE*, pp. 2804–2807.
- Li, T., Zhang, J., Bao, K., Liang, Y., Li, Y., Zheng, Y., 2020e. Autost: Efficient neural architecture search for spatio-temporal prediction. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 794–802.
- Liao, B., Zhang, J., Wu, C., McIlwraith, D., Chen, T., Yang, S., Guo, Y., Wu, F., 2018a. Deep sequence learning with auxiliary information for traffic prediction. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 537–546.
- Liao, S., Zhou, L., Di, X., Yuan, B., Xiong, J., 2018b. Large-scale short-term urban taxi demand forecasting using deep learning. In: *2018 23rd Asia and South Pacific Design Automation Conference. ASP-DAC, IEEE*, pp. 428–433.
- Lin, L., Li, J., Chen, F., Ye, J., Huai, J., 2017. Road traffic speed prediction: A probabilistic model fusing multi-source data. *IEEE Trans. Knowl. Data Eng.* 30 (7), 1310–1323.
- Lin, F., Xu, Y., Yang, Y., Ma, H., 2019. A spatial-temporal hybrid model for short-term traffic prediction. *Math. Probl. Eng.* 2019.
- Liou, C.Y., Cheng, W.C., Liou, J.W., Liou, D.R., 2014. Autoencoder for words. *Neurocomputing* 139, 84–96.
- Lippi, M., Bertini, M., Frasconi, P., 2013. Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Trans. Intell. Transp. Syst.* 14 (2), 871–882.
- Litman, T., 2020. Autonomous vehicle implementation predictions: Implications for transport planning.
- Liu, L., Chen, J., Wu, H., Zhen, J., Li, G., Lin, L., 2020a. Physical-virtual collaboration modeling for intra-and inter-station metro ridership prediction. *IEEE Trans. Intell. Transp. Syst.*
- Liu, Z., Huang, M., Ye, Z., Wu, K., 2019a. DeepRTP: A deep spatio-temporal residual network for regional traffic prediction. In: *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks. MSN, IEEE*, pp. 291–296.
- Liu, Y., James, J., Kang, J., Niyato, D., Zhang, S., 2020b. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet Things J.* 7 (8), 7751–7763.
- Liu, H., Ong, Y.S., Shen, X., Cai, J., 2020c. When Gaussian process meets big data: A review of scalable GPs. *IEEE Trans. Neural Netw. Learn. Syst.* 31 (11), 4405–4423.
- Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W., Lin, L., 2019b. Contextualized spatial-temporal network for taxi origin-destination demand prediction. *IEEE Trans. Intell. Transp. Syst.* 20 (10), 3875–3887.
- Liu, D., Tang, L., Shen, G., Han, X., 2019c. Traffic speed prediction: An attention-based method. *Sensors* 19 (18), 3836.
- Liu, T., Wu, W., Zhu, Y., Tong, W., 2020d. Predicting taxi demands via an attention-based convolutional recurrent neural network. *Knowl.-Based Syst.* 206, 106294.
- Liu, Y., Yu, J.J.Q., Kang, J., Niyato, D., Zhang, S., 2020e. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet Things J.* 7 (8), 7751–7763.
- Liu, Z., Zhang, Y., Huang, R., Chen, Z., Song, S., Wang, J., 2021. EXPERIENCE: Algorithms and case study for explaining repairs with uniform profiles over IoT data. *J. Data Inf. Qual. (JDIQ)* 13 (3), 1–17.
- Lu, Q., Chen, F., Hancock, K., 2009. On path anomaly detection in a large transportation network. *Comput. Environ. Urban Syst.* 33 (6), 448–462.
- Lu, Y., Huang, X., Dai, Y., Maharjan, S., Zhang, Y., 2019. Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Trans. Ind. Inform.* 16 (6), 4177–4186.
- Lu, H., Huang, D., Song, Y., Jiang, D., Zhou, T., Qin, J., 2020. St-trafficnet: A spatial-temporal deep learning network for traffic forecasting. *Electronics* 9 (9), 1474.
- Lu, W., Yi, Z., Wu, R., Rui, Y., Ran, B., 2022. Traffic speed forecasting for urban roads: A deep ensemble neural network model. *Physica A* 593, 126988.
- Luo, X., Li, D., Yang, Y., Zhang, S., 2019. Spatiotemporal traffic flow prediction with KNN and LSTM. *J. Adv. Transp.* 2019.
- Ma, T., Antoniou, C., Toledo, T., 2020a. Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transp. Res. C* 111, 352–372.
- Ma, J., Cheng, J.C., Ding, Y., Lin, C., Jiang, F., Wang, M., Zhai, C., 2020b. Transfer learning for long-interval consecutive missing values imputation without external features in air pollution time series. *Adv. Eng. Inform.* 44, 101092.
- Ma, X., Yu, H., Wang, Y., Wang, Y., 2015. Large-scale transportation network congestion evolution prediction using deep learning theory. *PLoS One* 10 (3), e0119044.
- Ma, Y., Zhang, Z., Ihler, A., 2020c. Multi-lane short-term traffic forecasting with convolutional LSTM network. *IEEE Access* 8, 34629–34643.

- Ma, C., Zhao, Y., Dai, G., Xu, X., Wong, S.C., 2022. A novel STFSA-CNN-GRU hybrid model for short-term traffic speed prediction. *IEEE Trans. Intell. Transp. Syst.*
- Macioszek, E., Kurek, A., 2021. Extracting road traffic volume in the city before and during COVID-19 through video remote sensing. *Remote Sens.* 13 (12), 2329.
- Magalhaes, R.P., Lettich, F., Macedo, J.A., Nardini, F.M., Perego, R., Renso, C., Trani, R., 2021. Speed prediction in large and dynamic traffic sensor networks. *Inf. Syst.* 98, 101444.
- Mahmud, M., Kaiser, M.S., Hussain, A., Vassanelli, S., 2018. Applications of deep learning and reinforcement learning to biological data. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (6), 2063–2079.
- Mallick, T., Balaprakash, P., Rask, E., Macfarlane, J., 2021. Transfer learning with graph neural networks for short-term highway traffic forecasting. In: 2020 25th International Conference on Pattern Recognition. ICPR, IEEE, pp. 10367–10374.
- Manibardo, E.L., Laña, I., Del Ser, J., 2021. Deep learning for road traffic forecasting: Does it make a difference? *IEEE Trans. Intell. Transp. Syst.*
- Matthews, A.G.d.G., Hron, J., Rowland, M., Turner, R.E., Ghahramani, Z., 2018. Gaussian process behaviour in wide deep neural networks. In: International Conference on Learning Representations.
- Meese, C., Chen, H., Asif, S.A., Li, W., Shen, C.C., Nejad, M., 2022. BFRT: Blockchain federated learning for real-time traffic flow prediction. In: 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing. CCGrid, IEEE, pp. 317–326.
- Migliani, A., Kumar, N., 2019. Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Veh. Commun.* 20, 100184.
- Min, X., Hu, J., Chen, Q., Zhang, T., Zhang, Y., 2009. Short-term traffic flow forecasting of urban network based on dynamic STARIMA model. In: 2009 12th International IEEE Conference on Intelligent Transportation Systems. IEEE, pp. 1–6.
- Min, X., Hu, J., Zhang, Z., 2010. Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. In: 13th International IEEE Conference on Intelligent Transportation Systems. IEEE, pp. 1535–1540.
- Min, W., Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. *Transp. Res. C* 19 (4), 606–616.
- Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., Damas, L., 2013. Predicting taxi-passenger demand using streaming data. *IEEE Trans. Intell. Transp. Syst.* 14 (3), 1393–1402.
- Nagy, A.M., Simon, V., 2018. Survey on traffic prediction in smart cities. *Pervasive Mob. Comput.* 50, 148–163.
- Nguyen, H., Kieu, L.M., Wen, T., Cai, C., 2018. Deep learning methods in transportation domain: A review. *IET Intell. Transp. Syst.* 12 (9), 998–1004.
- Ni, M., He, Q., Gao, J., 2014. Using social media to predict traffic flow under special event conditions. In: The 93rd Annual Meeting of Transportation Research Board.
- Oliveira, T.P., Barbar, J.S., Soares, A.S., 2016. Computer network traffic prediction: A comparison between traditional and deep learning neural networks. *Int. J. Big Data Intell.* 3 (1), 28–37.
- Oreshkin, B.N., Amini, A., Coyle, L., Coates, M., 2021. FC-GAGA: Fully connected gated graph architecture for spatio-temporal traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, (no. 10), pp. 9233–9241.
- Palm, R.B., 2012. Prediction as a candidate for learning deep hierarchical models of data. *Techn. Univ. Denmark* 5.
- Pan, B., Demiryurek, U., Shahabi, C., 2012. Utilizing real-world transportation data for accurate traffic prediction. In: 2012 IEEE 12th International Conference on Data Mining. IEEE, pp. 595–604.
- Pan, Z., Liang, Y., Wang, W., Yu, Y., Zheng, Y., Zhang, J., 2019a. Urban traffic prediction from spatio-temporal data using deep meta learning. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1720–1730.
- Pan, Z., Wang, Z., Wang, W., Yu, Y., Zhang, J., Zheng, Y., 2019b. Matrix factorization for spatio-temporal neural networks with applications to urban flow prediction. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. pp. 2683–2691.
- Pan, Z., Zhang, W., Liang, Y., Zhang, W., Yu, Y., Zhang, J., Zheng, Y., 2020. Spatio-temporal meta learning for urban traffic prediction. *IEEE Trans. Knowl. Data Eng.*
- Panait, A.E., Olimid, R.F., 2020. On using zk-SNARKs and zk-STARKs in blockchain-based identity management. In: International Conference on Information Technology and Communications Security. Springer, pp. 130–145.
- Park, C., Lee, C., Bahng, H., Tae, Y., Jin, S., Kim, K., Ko, S., Choo, J., 2020. ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. pp. 1215–1224.
- Peng, Y., Li, C., Wang, K., Gao, Z., Yu, R., 2020a. Examining imbalanced classification algorithms in predicting real-time traffic crash risk. *Accid. Anal. Prev.* 144, 105610.
- Peng, H., Wang, H., Du, B., Bhuiyan, M.Z.A., Ma, H., Liu, J., Wang, L., Yang, Z., Du, L., Wang, S., et al., 2020b. Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Inform. Sci.* 521, 277–290.
- Pham, H., Guan, M., Zoph, B., Le, Q., Dean, J., 2018. Efficient neural architecture search via parameters sharing. In: International Conference on Machine Learning. PMLR, pp. 4095–4104.
- Qi, Y., Hossain, M.S., Nie, J., Li, X., 2021. Privacy-preserving blockchain-based federated learning for traffic flow prediction. *Future Gener. Comput. Syst.* 117, 328–337.
- Qu, L., Li, W., Li, W., Ma, D., Wang, Y., 2019. Daily long-term traffic flow forecasting based on a deep neural network. *Expert Syst. Appl.* 121, 304–312.
- Quddus, M.A., Ochieng, W.Y., Noland, R.B., 2007. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transp. Res. C* 15 (5), 312–328.
- Quddus, M., Washington, S., 2015. Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. *Transp. Res. C* 55, 328–339.
- Riahi-Madvar, M., Azirani, A.A., Nasersharif, B., Raahemi, B., 2021. A new density-based subspace selection method using mutual information for high dimensional outlier detection. *Knowl.-Based Syst.* 216, 106733.
- Roy, R., Sato, K., Bhattacharya, S., Fang, X., Joti, Y., Hatsui, T., Hiraki, T.N., Guo, J., Yu, W., 2021. Compression of time evolutionary image data through predictive deep neural networks. In: 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing. CCGrid, IEEE, pp. 41–50.
- Russell, R.L., Reale, C., 2021. Multivariate uncertainty in deep learning. *IEEE Trans. Neural Netw. Learn. Syst.*
- Salinas, D., Bohlke-Schneider, M., Callot, L., Medico, R., Gasthaus, J., 2019. High-dimensional multivariate forecasting with low-rank Gaussian copula processes. *Adv. Neural Inf. Process. Syst.* 32.
- Sani, M.F., van Zelst, S.J., van der Aalst, W.M., 2019. Repairing outlier behaviour in event logs using contextual behaviour. *Enterp. Model. Inf. Syst. Archit. (EMISAJ)* 14, 1–24.
- Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2008. The graph neural network model. *IEEE Trans. Neural Netw.* 20 (1), 61–80.
- Schrank, D., Lomax, T., Eisele, B., 2019. 2019 Urban mobility report. Texas Transportation Institute, [ONLINE]. Available: <https://Static.Tti.Tamu.Edu/Tti.Tamu.Edu/Documents/Mobility-Report-2019.Pdf>.
- Sen, R., Yu, H.F., Dhillon, I., 2019. Think globally, act locally: a deep neural network approach to high-dimensional time series forecasting. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. pp. 4837–4846.
- Shahbazi, Z., Byun, Y.-C., 2021. A framework of vehicular security and demand service prediction based on data analysis integrated with blockchain approach. *Sensors* 21 (10), 3314.
- Shang, C., 2020. End-to-end structure-aware convolutional networks on graphs.
- Shekhar, S., Williams, B.M., 2007. Adaptive seasonal time series models for forecasting short-term traffic flow. *Transp. Res. Rec.* 2024 (1), 116–125.

- Shen, G., 2020. Short-term traffic flow prediction based on harmony search algorithm optimized wavelet neural network. In: *J. Phys.: Conf. Ser.* 1682, (1), IOP Publishing, 012047.
- Shen, M., Liu, D.R., Shann, S.H., 2015. Outlier detection from vehicle trajectories to discover roaming events. *Inform. Sci.* 294, 242–254.
- Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 28.
- Shi, Y., Feng, H., Geng, X., Tang, X., Wang, Y., 2019. A survey of hybrid deep learning methods for traffic flow prediction. In: *Proceedings of the 2019 3rd International Conference on Advances in Image Processing*. pp. 133–138.
- Shi, J., Leau, Y.B., Li, K., Obit, J.H., 2021. A comprehensive review on hybrid network traffic prediction model. *Int. J. Electr. Comput. Eng.* 11 (2), 1450.
- Shi, X., Qi, H., Shen, Y., Wu, G., Yin, B., 2020. A spatial-temporal attention approach for traffic prediction. *IEEE Trans. Intell. Transp. Syst.*
- Shih, S.Y., Sun, F.K., Lee, H.y., 2019. Temporal pattern attention for multivariate time series forecasting. *Mach. Learn.* 108 (8), 1421–1441.
- Shin, J., Sunwoo, M., 2018. Vehicle speed prediction using a Markov chain with speed constraints. *IEEE Trans. Intell. Transp. Syst.* 20 (9), 3201–3211.
- Shin, Y., Yoon, Y., 2020. Incorporating dynamicity of transportation network with multi-weight traffic graph convolutional network for traffic forecasting. *IEEE Trans. Intell. Transp. Syst.*
- Singh, S., Singh, J., Sehra, S.S., 2020. Genetic-inspired map matching algorithm for real-time GPS trajectories. *Arab. J. Sci. Eng.* 45 (4), 2587–2603.
- Soltanolkotabi, M., Javanmard, A., Lee, J.D., 2018. Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *IEEE Trans. Inform. Theory* 65 (2), 742–769.
- Song, C., Lee, H., Kang, C., Lee, W., Kim, Y.B., Cha, S.W., 2017. Traffic speed prediction under weekday using convolutional neural networks concepts. In: *2017 IEEE Intelligent Vehicles Symposium. IV, IEEE*, pp. 1293–1298.
- Song, C., Lin, Y., Guo, S., Wan, H., 2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, (no. 01), pp. 914–921.
- Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2014. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*.
- Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. *Transp. Res. C* 11 (2), 121–135.
- Sun, B., Zhao, D., Shi, X., He, Y., 2021. Modeling global spatial-temporal graph attention network for traffic prediction. *IEEE Access* 9, 8581–8594.
- Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. *Adv. Neural Inf. Process. Syst.* 27.
- Taghipour, H., Parsa, A.B., Mohammadian, A.K., 2020. A dynamic approach to predict travel time in real time using data driven techniques and comprehensive data sources. *Transp. Eng.* 2, 100025.
- Tan, H., Wu, Y., Shen, B., Jin, P.J., Ran, B., 2016. Short-term traffic prediction based on dynamic tensor completion. *IEEE Trans. Intell. Transp. Syst.* 17 (8), 2123–2133.
- Tang, B., He, H., 2017. A local density-based approach for outlier detection. *Neurocomputing* 241, 171–180.
- Tang, J., Liang, J., Liu, F., Hao, J., Wang, Y., 2021. Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network. *Transp. Res. C* 124, 102951.
- Tang, L., Zhao, Y., Cabrera, J., Ma, J., Tsui, K.L., 2018. Forecasting short-term passenger flow: An empirical study on shenzhen metro. *IEEE Trans. Intell. Transp. Syst.* 20 (10), 3613–3622.
- Tedjopurnomo, D.A., Bao, Z., Zheng, B., Choudhury, F., Qin, A., 2020. A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. *IEEE Trans. Knowl. Data Eng.*
- Thompson, N.C., Greenewald, K., Lee, K., Manso, G.F., 2020. The computational limits of deep learning. *arXiv preprint arXiv:2007.05558*.
- Tian, L., Fan, Y., Li, L., Mousseau, N., 2020. Identifying flow defects in amorphous alloys using machine learning outlier detection methods. *Scr. Mater.* 186, 185–189.
- Tian, Y., Zhang, K., Li, J., Lin, X., Yang, B., 2018. LSTM-based traffic flow prediction with missing data. *Neurocomputing* 318, 297–305.
- Tran, Q.H., Fang, Y.M., Chou, T.Y., Hoang, T.-V., Wang, C.T., Vu, V.T., Ho, T.L.H., Le, Q., Chen, M.H., et al., 2022. Short-term traffic speed forecasting model for a parallel multi-lane arterial road using GPS-monitored data based on deep learning approach. *Sustainability* 14 (10), 6351.
- Tran, L., Mun, M.Y., Shahabi, C., 2020. Real-time distance-based outlier detection in data streams. *Proc. VLDB Endow.* 14 (2), 141–153.
- Tu, Z., Fan, Y., Li, Y., Chen, X., Su, L., Jin, D., 2019. From fingerprint to footprint: cold-start location recommendation by learning user interest from app data. *Proc. ACM Interact., Mob., Wearable Ubiquitous Technol.* 3 (1), 1–22.
- Van Zoest, V., Liu, X., Ngai, E., 2021. Data quality evaluation, outlier detection and missing data imputation methods for IoT in smart cities. In: *Mach. Intell. Data Anal. Sustain. Future Smart Cities*. Springer, pp. 1–18.
- Vanajakshi, L., Rilett, L.R., 2004. A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. In: *IEEE Intelligent Vehicles Symposium, 2004. IEEE*, pp. 194–199.
- Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: Overview of objectives and methods. *Transp. Res.* 24 (5), 533–557.
- Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we're going. *Transp. Res. C* 43, 3–19.
- Wagner-Muns, I.M., Guardiola, I.G., Samaranyake, V., Kayani, W.L., 2017. A functional data analysis approach to traffic volume forecasting. *IEEE Trans. Intell. Transp. Syst.* 19 (3), 878–888.
- Wang, W., Bai, Y., Yu, C., Gu, Y., Feng, P., Wang, X., Wang, R., 2018a. A network traffic flow prediction with deep learning approach for large-scale metropolitan area network. In: *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium. IEEE*, pp. 1–9.
- Wang, J., Boukerche, A., 2020. The scalability analysis of machine learning based models in road traffic flow prediction. In: *ICC 2020-2020 IEEE International Conference on Communications. ICC, IEEE*, pp. 1–6.
- Wang, D., Cao, W., Li, J., Ye, J., 2017. DeepSD: Supply-demand prediction for online car-hailing services using deep neural networks. In: *2017 IEEE 33rd International Conference on Data Engineering. ICDE, IEEE*, pp. 243–254.
- Wang, S., Cao, J., Yu, P., 2020a. Deep learning for spatio-temporal data mining: A survey. *IEEE Trans. Knowl. Data Eng.*
- Wang, J., Chen, R., He, Z., 2019a. Traffic speed prediction for urban transportation network: A path based deep learning approach. *Transp. Res. C* 100, 372–385.
- Wang, C., Cheng, X., Li, J., He, Y., Xiao, K., 2021a. A survey: Applications of blockchain in the internet of vehicles. *EURASIP J. Wireless Commun. Networking* 2021 (1), 1–16.
- Wang, J., Deng, W., Guo, Y., 2014a. New Bayesian combination method for short-term traffic flow forecasting. *Transp. Res. C* 43, 79–94.
- Wang, Z., Fu, K., Ye, J., 2018b. Learning to estimate the travel time. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 858–866.
- Wang, L., Geng, X., Ma, X., Liu, F., Yang, Q., 2019b. Cross-city transfer learning for deep spatio-temporal prediction. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. pp. 1893–1899.
- Wang, J., Gu, Q., Wu, J., Liu, G., Xiong, Z., 2016. Traffic speed prediction and congestion source exploration: A deep learning method. In: *2016 IEEE 16th International Conference on Data Mining. ICDM, IEEE*, pp. 499–508.
- Wang, X., Guan, X., Cao, J., Zhang, N., Wu, H., 2020b. Forecast network-wide traffic states for multiple steps ahead: A deep learning approach considering dynamic non-local spatial correlation and non-stationary temporal dependency. *Transp. Res. C* 119, 102763.
- Wang, J., Jiang, J., Jiang, W., Li, C., Zhao, W.X., 2021b. LibCity: An open library for traffic prediction. In: *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. pp. 145–148.
- Wang, Y., Jing, C., Xu, S., Guo, T., 2022. Attention based spatiotemporal graph attention networks for traffic flow forecasting. *Inform. Sci.* 607, 869–883.

- Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C., Yu, J., 2020c. Traffic flow prediction via spatial temporal graph neural network. In: Proceedings of the Web Conference 2020. pp. 1082–1092.
- Wang, S., Mao, G., 2018. Missing data estimation for traffic volume by searching an optimum closed cut in urban networks. *IEEE Trans. Intell. Transp. Syst.* 20 (1), 75–86.
- Wang, Y., Qin, K., Chen, Y., Zhao, P., 2018c. Detecting anomalous trajectories and behavior patterns using hierarchical clustering from taxi GPS data. *ISPRS Int. J. Geo-Inf.* 7 (1), 25.
- Wang, S., Sun, Y., Musco, C., Bao, Z., 2021c. Public transport planning: When transit network connectivity meets commuting demand. In: Proceedings of the 2021 International Conference on Management of Data. pp. 1906–1919.
- Wang, H., Tang, X., Kuo, Y.-H., Kifer, D., Li, Z., 2019c. A simple baseline for travel time estimation using large-scale trip data. *ACM Trans. Intell. Syst. Technol.* 10 (2), 1–22.
- Wang, Q., Xu, C., Zhang, W., Li, J., 2021d. GraphTTE: Travel time estimation based on attention-spatiotemporal graphs. *IEEE Signal Process. Lett.* 28, 239–243.
- Wang, Y., Yin, H., Chen, H., Wo, T., Xu, J., Zheng, K., 2019d. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1227–1235.
- Wang, D., Zhang, J., Cao, W., Li, J., Zheng, Y., 2018d. When will you arrive? Estimating travel time based on deep neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, (no. 1).
- Wang, W., Zhang, H., Li, T., Guo, J., Huang, W., Wei, Y., Cao, J., 2020d. An interpretable model for short term traffic flow prediction. *Math. Comput. Simulation* 171, 264–278.
- Wang, Y., Zhang, D., Liu, Y., Dai, B., Lee, L.H., 2019e. Enhancing transportation systems via deep learning: A survey. *Transp. Res. C* 99, 144–163.
- Wang, Y., Zheng, Y., Xue, Y., 2014b. Travel time estimation of a path using sparse trajectories. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 25–34.
- Wang, J., Zhu, W., Sun, Y., Tian, C., 2020e. An effective dynamic spatiotemporal framework with external features information for traffic prediction. *Appl. Intell.* 1–15.
- Wei, H., Wang, Y., Forman, G., Zhu, Y., 2013. Map matching: Comparison of approaches using sparse and noisy data. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 444–447.
- Werbos, P.J., 1990. Backpropagation through time: What it does and how to do it. *Proc. IEEE* 78 (10), 1550–1560.
- Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *J. Transp. Eng.* 129 (6), 664–672.
- Wu, Q., Fu, Q., Nie, M., 2020a. Graph wavelet long short-term memory neural network: A novel spatial-temporal network for traffic prediction. *J. Phys.: Conf. Ser.* 1549 (4), 042070.
- Wu, P., Huang, Z., Pian, Y., Xu, L., Li, J., Chen, K., 2020b. A combined deep learning method with attention-based LSTM model for short-term traffic speed forecasting. *J. Adv. Transp.* 2020.
- Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C., 2020c. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 753–763.
- Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C., 2019. Graph WaveNet for deep spatial-temporal graph modeling. In: The 28th International Joint Conference on Artificial Intelligence. IJCAI, International Joint Conferences on Artificial Intelligence Organization.
- Wu, Y., Tan, H., Qin, L., Ran, B., Jiang, Z., 2018. A hybrid deep learning based traffic flow prediction method and its understanding. *Transp. Res. C* 90, 166–180.
- Wu, S., Yang, Z., Zhu, X., Yu, B., 2014. Improved k-nn for short-term traffic forecasting using temporal and spatial information. *J. Transp. Eng.* 140 (7), 04014026.
- Xie, P., Li, T., Liu, J., Du, S., Yang, X., Zhang, J., 2020. Urban flow prediction from spatiotemporal data using machine learning: A survey. *Inf. Fusion* 59, 1–12.
- Xu, T., Li, X., Claramunt, C., 2018. Trip-oriented travel time prediction (TOTTP) with historical vehicle trajectories. *Front. Earth Sci.* 12 (2), 253–263.
- Xu, S., Zhang, R., Cheng, W., Xu, J., 2020. MTLM: A multi-task learning model for travel time estimation. *Geoinformatica* 1–17.
- Yang, F., Chen, L., Zhou, F., Gao, Y., Cao, W., 2019. Relational state-space model for stochastic multi-object systems. In: International Conference on Learning Representations.
- Yang, B., Guo, C., Jensen, C.S., 2013. Travel cost inference from sparse, spatio temporally correlated time series using Markov models. *Proc. VLDB Endow.* 6 (9), 769–780.
- Yang, X., Wang, B., Yang, K., Liu, C., Zheng, B., 2017. A novel representation and compression for queries on trajectories in road networks. *IEEE Trans. Knowl. Data Eng.* 30 (4), 613–629.
- Yang, X., Zou, Y., Tang, J., Liang, J., Ijaz, M., 2020. Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models. *J. Adv. Transp.* 2020.
- Yao, W., Qian, S., 2021. From Twitter to traffic predictor: Next-day morning traffic prediction using social media data. *Transp. Res. C* 124, 102938.
- Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., Li, Z., 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, (no. 1).
- Ye, J., Sun, L., Du, B., Fu, Y., Tong, X., Xiong, H., 2019. Co-prediction of multiple transportation demands based on deep spatio-temporal neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 305–313.
- Ye, J., Zhao, J., Ye, K., Xu, C., 2020. How to build a graph-based deep learning architecture in traffic domain: A survey. *IEEE Trans. Intell. Transp. Syst.*
- Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., Yin, B., 2021. Deep learning on traffic prediction: Methods, analysis and future directions. *IEEE Trans. Intell. Transp. Syst.*
- Yu, B., Lee, Y., Sohn, K., 2020. Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). *Transp. Res. C* 114, 189–204.
- Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors* 17 (7), 1501.
- Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. pp. 3634–3640.
- Yuan, H., Li, G., 2021. A survey of traffic prediction: From spatio-temporal data to intelligent transportation. *Data Sci. Eng.* 6 (1), 63–85.
- Yuan, H., Li, G., Bao, Z., Feng, L., 2020. Effective travel time estimation: When historical trajectories over road networks matter. In: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. pp. 2135–2149.
- Zafar, N., Haq, I.U., Chughtai, J.u.R., Shafiq, O., 2022. Applying hybrid Lstm-Gru model based on heterogeneous data sources for traffic speed prediction in urban areas. *Sensors* 22 (9), 3348.
- Zhang, Q., Benveniste, A., 1992. Wavelet networks. *IEEE Trans. Neural Netw.* 3 (6), 889–898.
- Zhang, Q., Chang, J., Meng, G., Xiang, S., Pan, C., 2020a. Spatio-temporal graph structure learning for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 01), pp. 1177–1185.
- Zhang, D., Ding, M., Yang, D., Liu, Y., Fan, J., Shen, H.T., 2018a. Trajectory simplification: An experimental study and quality analysis. *Proc. VLDB Endow.* 11 (9), 934–946.
- Zhang, Z., Li, M., Lin, X., Wang, Y., He, F., 2019a. Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. *Transp. Res. C* 105, 297–322.

- Zhang, L., Liu, Q., Yang, W., Wei, N., Dong, D., 2013. An improved k-nearest neighbor model for short-term traffic flow prediction. *Procedia-Soc. Behav. Sci.* 96, 653–662.
- Zhang, P., Qian, Z.S., 2018. User-centric interdependent urban systems: Using time-of-day electricity usage data to predict morning roadway congestion. *Transp. Res. C* 92, 392–411.
- Zhang, Y., Shi, W.h., Liu, Y.c., 2010. Comparison of several traffic forecasting methods based on travel time index data on weekends. *J. Shanghai Jiaotong Univ. (Science)* 15 (2), 188–193.
- Zhang, J., Shi, X., Xie, J., Ma, H., King, I., Yeung, D.Y., 2018b. GaAN: Gated attention networks for learning on large and spatiotemporal graphs. In: 34th Conference on Uncertainty in Artificial Intelligence 2018. UAI 2018.
- Zhang, Y., Wang, S., Chen, B., Cao, J., Huang, Z., 2019b. Trafficgan: Network-scale deep traffic prediction with generative adversarial nets. *IEEE Trans. Intell. Transp. Syst.*
- Zhang, D., Xiao, F., Shen, M., Zhong, S., 2021a. DNEAT: A novel dynamic node-edge attention network for origin-destination demand prediction. *Transp. Res. C* 122, 102851.
- Zhang, Y., Zhang, Y., Haghani, A., 2014. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transp. Res. C* 43, 65–78.
- Zhang, C., Zhang, S., James, J., Yu, S., 2021b. FASTGNN: A topological information protected federated learning approach for traffic speed forecasting. *IEEE Trans. Ind. Inform.*
- Zhang, X., Zhang, Z., Jin, X., 2020b. Spatial-temporal graph attention model on traffic forecasting. In: 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. CISP-BMEI, IEEE, pp. 999–1003.
- Zhang, S., Zhang, J., Qi, L., 2021c. Trajectory data compression algorithm based on motion state changing. *Math. Probl. Eng.* 2021.
- Zhang, J., Zheng, Y., Qi, D., 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Thirty-First AAAI Conference on Artificial Intelligence.
- Zhang, C., Zhu, F., Lv, Y., Ye, P., Wang, F.Y., 2021d. MLRNN: Taxi demand prediction based on multi-level deep learning and regional heterogeneity analysis. *IEEE Trans. Intell. Transp. Syst.*
- Zhao, P., Cai, D., Zhang, S., Chen, F., Zhang, Z., Wang, C., Li, J., 2018. Layerwise recurrent autoencoder for general real-world traffic flow forecasting.
- Zhao, J., Chen, C., Huang, H., Xiang, C., 2020. Unifying uber and taxi data via deep models for taxi passenger demand prediction. *Pers. Ubiquitous Comput.* 1–13.
- Zhao, J., Liu, Z., Sun, Q., Li, Q., Jia, X., Zhang, R., 2022. Attention-based dynamic spatial-temporal graph convolutional networks for traffic speed forecasting. *Expert Syst. Appl.* 204, 117511.
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., Li, H., 2019a. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 21 (9), 3848–3858.
- Zhao, F., Zeng, G.Q., Lu, K.D., 2019b. EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization. *IEEE Trans. Veh. Technol.* 69 (1), 101–113.
- Zhao, B., Zhang, X., 2018. A parallel-res GRU architecture and its application to road network traffic flow forecasting. In: Proceedings of 2018 International Conference on Big Data Technologies. pp. 79–83.
- Zhao, J., Zhu, T., Zhao, R., Zhao, P., 2019c. Layerwise recurrent autoencoder for real-world traffic flow forecasting. In: International Conference on Intelligent Science and Big Data Engineering. Springer, pp. 78–88.
- Zheng, Y., 2015. Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol.* 6 (3), 1–41.
- Zheng, C., Fan, X., Wang, C., Qi, J., 2020a. Gman: A graph multi-attention network for traffic prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (no. 01), pp. 1234–1241.
- Zheng, H., Lin, F., Feng, X., Chen, Y., 2020b. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.*
- Zhong, Z., Lee, E.E., Nejad, M., Lee, J., 2020a. Influence of CAV clustering strategies on mixed traffic flow characteristics: An analysis of vehicle trajectory data. *Transp. Res. C* 115, 102611.
- Zhong, Z., Nejad, M., Lee, E.E., 2020b. Autonomous and semiautonomous intersection management: A survey. *IEEE Intell. Transp. Syst. Mag.* 13 (2), 53–70.
- Zhou, X., Mahmassani, H.S., Zhang, K., 2008. Dynamic micro-assignment modeling approach for integrated multimodal urban corridor management. *Transp. Res. C* 16 (2), 167–186.
- Zhou, Z., Yang, Z., Zhang, Y., Huang, Y., Chen, H., Yu, Z., 2022. A comprehensive study of speed prediction in transportation system: From vehicle to traffic. *Iscience* 103909.
- Zhou, F., Yang, Q., Zhang, K., Trajcevski, G., Zhong, T., Khokhar, A., 2020. Reinforced spatiotemporal attentive graph neural networks for traffic forecasting. *IEEE Internet Things J.* 7 (7), 6414–6428.
- Ziat, A., Delasalles, E., Denoyer, L., Gallinari, P., 2017. Spatio-temporal neural networks for space-time series forecasting and relations discovery. In: 2017 IEEE International Conference on Data Mining. ICDM, IEEE, pp. 705–714.
- Zou, Z., Yang, H., Zhu, A.X., 2020. Estimation of travel time based on ensemble method with multi-modality perspective urban big data. *IEEE Access* 8, 24819–24828.