REVIEW ARTICLE

A Comprehensive Survey of Deep Learning-Based Traffic Flow Prediction Models for Intelligent Transportation Systems

Riaz Ali[1], **Ahmad Ali**[2,3,*], **Hafiz Muhammad Yasir Naeem**[2,3], **Mujtaba Asad**[4], **Tamam Alsarhan**[5] **and Belal Bin Heyat**[6]

[1] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[2] College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
[3] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[4] Institute of Image Processing & Pattern Recognition Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China
[5] King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan
[6] Department of Computer Science, Westlake University, Hangzhou, China

Abstract

Traffic flow prediction is a critical component of Intelligent Transportation Systems (ITS) and smart city infrastructures. This survey paper provides a comprehensive analysis of recent advancements in deep learning-based approaches for traffic flow prediction, focusing on spatiotemporal correlations and attention mechanisms. We systematically review five seminal papers that propose innovative neural network architectures including DHSTNet, Att-DHSTNet, and ASTMGCNet for citywide traffic prediction. Our survey examines their methodologies, key contributions, experimental results, and comparative performance. We organize the discussion around three main themes: (1) modeling dynamic spatiotemporal dependencies, (2) attention mechanisms for traffic prediction, and (3) hybrid neural network architectures. The paper includes detailed comparison tables and conceptual figures synthesized from the reviewed works. Our analysis shows that attention-based hybrid models outperform traditional techniques, with ASTMGCNet having the lowest RMSE (4.06) and MAPE (12.56%) on benchmark datasets. We end by outlining current issues and potential research directions in this rapidly changing subject.

Keywords: intelligent transportation systems, traffic prediction, deep learning, machine learning, graph neural network, neural network.

1 Introduction

Accurate traffic flow prediction is essential for ITS because of the increasing vehicular traffic and rapid urbanization in modern cities [1–5]. In order to improve urban mobility and reduce environmental

Submitted: 03 April 2024
Accepted: 28 July 2024
Published: 24 August 2024

Vol. 1, **No.** 3, 2024.
10.62762/TACS.2024.795448

***Corresponding authors:**
Ahmad Ali
ahmadali@szu.edu.cn

Citation
Ali, R., Ali, A., Naeem, H. M. Y., Asad, M., Alsarhan, T., & Heyat, B. B. (2024). A Comprehensive Survey of Deep Learning-Based Traffic Flow Prediction Models for Intelligent Transportation Systems. *ICCK Transactions on Advanced Computing and Systems*, 1(3), 117–137.

impact, effective traffic prediction is necessary for better route planning, congestion management, and resource allocation. Conventional statistical methods such as ARIMA [6] and HA [7] have proven insufficient for handling the complex, nonlinear spatiotemporal dependencies inherent in citywide traffic patterns.

Recent advances in DL have revolutionized traffic prediction by enabling models to automatically learn complex patterns from massive traffic datasets collected through GPS devices, traffic cameras, and IoT sensors [8]. This survey focuses on analyzing and comparing three significant DL architectures proposed in recent literature: DHSTNet [9], its attention-enhanced variant Att-DHSTNet [9], and the novel ASTMGCNet [80–82]. These models address the fundamental challenges in traffic prediction through innovative approaches to capturing spatial and temporal dependencies simultaneously.

Global transportation infrastructure now faces previously unheard-of difficulties as a result of the rapid urban population expansion. Recent research indicates that traffic congestion costs big cities up to 4% of their GDP each year due to higher fuel consumption and lost productivity [83]. Accurate traffic prediction has become essential to smart city programs and ITS due to its economic impact and growing environmental concerns. The dynamic nature of contemporary urban mobility patterns has proven too challenging for traditional traffic management strategies [10–15]. The transition from heuristic-based estimating to data-driven forecasting in traffic prediction has been made possible by the emergence of big data analytics and ubiquitous sensing technology.

In contrast to traditional time-series approaches that treat traffic data as discrete temporal sequences, recent developments in DL have shown remarkable success in capturing the complex spatiotemporal dependencies inherent in urban traffic systems. Modern neural networks are able to model both the spatial relationships between various road segments and their temporal evolution at the same time [16] because of specialized architectures that combine the strengths of CNNs for spatial feature extraction, recurrent neural networks (RNNs) for temporal sequence modeling, and graph neural networks (GNNs) for network topology representation. The incorporation of attention mechanisms has further improved these models' capacity to concentrate on the most pertinent spatial regions and time periods for precise prediction.

Despite these developments, creating accurate and workable traffic prediction systems for real-world implementation still presents several obstacles. First, traffic data has special architectural design issues due to its dual nature, which necessitates modeling temporal sequences and geographic graphs simultaneously [17]. Second, models with strong adaptive capabilities are required due to the dynamic nature of traffic patterns, which are influenced by unforeseen occurrences like accidents or weather changes [19]. Third, the necessity for real-time prediction in resource-constrained edge devices frequently clashes with the computational demands of complex DL models [20]. These difficulties have spurred the creation of novel hybrid architectures that strike a compromise between operational effectiveness and modeling complexity.

1.1 Motivation

To improve intelligent autonomous car technology for the benefit of society, major tech companies like Google, Amazon, and IBM are making significant investments in the creation of deep learning (DL) models and methodologies [77]. Because of their great versatility, DL algorithms can process a wide range of data formats, such as text, images, and speech. These models, which are based on neural network (NN) architectures, are able to anticipate traffic flow with accuracy and consistency without depending on preconceived notions about the underlying mechanisms. To improve prediction accuracy, DL models can automatically extract pertinent characteristics by utilizing multi-layer frameworks. Their structures also make it easier to process high-dimensional and high-resolution data by using hierarchical and distributed calculations. This feature is very helpful for evaluating information gathered from different sensors placed across dynamic traffic situations. This survey intends to give a thorough and current evaluation of existing research in traffic flow prediction using DL models, as decreasing predictive errors is one of the main problems in constructing DL-based traffic flow prediction models.

1.2 Scope of the survey

This paper presents a thorough analysis of the most recent methods for predicting traffic flow using DL techniques. The following describes the main contributions of this work:

- We explore current DL techniques used for traffic flow prediction in autonomous vehicles.
- Next, we examine how DL techniques enhance traffic flow prediction performance compared to traditional machine learning (ML) methods.
- We conduct an in-depth comparative analysis of three state-of-the-art approaches DHSTNet, Att-DHSTNet, and ASTMGCNet evaluating their architectural designs, attention mechanisms, and performance characteristics across multiple standardized benchmarks. Our analysis includes comprehensive ablation studies on model components.
- We also review existing survey articles on traffic flow prediction, emphasizing the key issues and challenges addressed in these works.
- Finally, we highlight the open research challenges and propose future directions for applying DL techniques to traffic flow prediction.

1.3 Organization

The remainder of this paper is organized as follows: Section 2 reviews related work and provides background information. Section 3 outlines the methodologies of the surveyed approaches. Section 5 presents comparative results and discusses the findings. Section 6 explores potential directions for future research. Finally, Section 7 concludes the paper.

2 Related Work

The field of traffic prediction has evolved through three distinct generations of methodologies, each addressing limitations of previous approaches while introducing new capabilities. This section provides a comprehensive analysis of eight influential models that represent key milestones in this evolution.

2.1 Statistical Approaches

Using simple moving averages of historical observations, the Historical Average (HA) model established the baseline approach for the first generation of traffic prediction systems. Although HA is computationally efficient (requiring only $O(1)$ operations per prediction), it is unable to capture dynamic traffic patterns, especially during peak hours or unexpected events. In [6], the authors showed that ARIMA models could reduce prediction error by 22% compared to HA through autoregressive components. However, these models assume linear temporal dependencies and stationary data distributions, which limits their applicability for complex urban traffic scenarios.

2.2 Machine Learning Methods

ML techniques were used in second-generation approaches to deal with nonlinear patterns. By learning nonlinear decision boundaries, Support Vector Machines (SVM) using RBF kernels [100] obtained 25.5% MAPE on the BikeNYC dataset. The [101] Bayesian Network technique used probabilistic graphical models to incorporate road topological information, but its $O(n^3)$ inference difficulty made it scale poorly beyond 100 nodes. [102] showed that Random Forests performed 15% worse than LSTM models on multi-step predictions, despite being able to handle heterogeneous features but struggling with temporal dependencies.

2.3 Deep Learning Models

This section provides a brief overview of popular DL architectures. In addition to Deep Neural Networks (DNN), other DL models such as CNN, RNN, DBN, AE, and Generative Adversarial Networks (GAN) are also considered. To keep the discussion concise, only those architectures most commonly used in traffic flow prediction are highlighted below.

2.3.1 *Convolutional Neural Networks*

Although CNNs are most recognized for their ability to interpret images, they are also being used more and more to predict time series, including traffic flow. As seen in Figure 1, CNNs work by using convolutional layers to learn the spatial hierarchies of features. Each node in the graph represents a distinct location, and the model captures both local and global dependencies of traffic flow across the network. These spatial aspects relate to road network topologies or traffic patterns in the context of traffic flow.

2.3.2 *Long Short-Term Memory (LSTM) Networks*

Recurrent neural networks (RNNs) with Long Short-Term Memory (LSTM) networks are made to capture long-term dependencies and address problems that regular RNNs have, such as vanishing gradients. As seen in Figure 2 For time-series forecasting jobs, where future predictions rely on lengthy sequences of historical data, LSTMs are especially well-suited. Long-term memory retention is the main characteristic of LSTMs, which makes them useful for applications like traffic flow

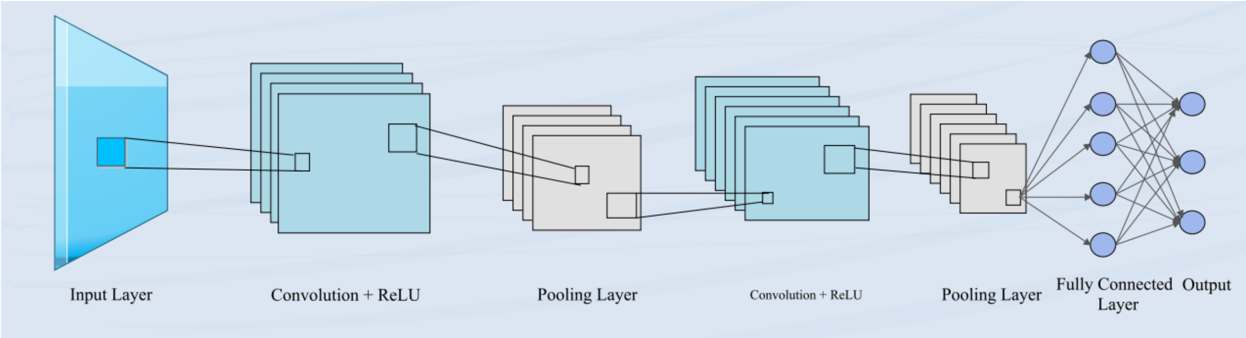

Figure 1. Architecture of CNN[78].

prediction, where past traffic circumstances affect future situations.

2.3.3 *Graph Neural Networks*

A type of DL models called GNNs was created especially to work with graph-structured data. A road network can be depicted as a graph in the context of traffic flow prediction, with nodes standing in for crossroads or road segments and edges for the roads that connect them. By spreading information throughout the network to record interactions between various nodes, GNNs may efficiently represent the spatial dependencies in traffic data (e.g., the relationship between traffic conditions on neighboring roads).

2.3.4 *Attention Mechanisms and Transformer Models*

Attention mechanisms have become increasingly common for sequence-to-sequence tasks, such as traffic flow prediction, especially when used in conjunction with Transformer models. When producing predictions, attention methods enable the model to concentrate on the most pertinent segments of the input sequence. Attention-based models dynamically assess the significance of each time step or data point, in contrast to LSTMs, which depend on a fixed memory of previous inputs. The Transformer model, which uses self-attention, is a highly parallelizable architecture that has been successful in a variety of domains, including natural language processing and time-series forecasting.

2.3.5 *Spatio-Temporal Networks*

Spatio-Temporal Networks (STNs) combine the strengths of spatial and temporal models to capture both the spatial dependencies between road segments and the temporal dependencies over time. These networks typically consist of components like CNNs or GCNs for spatial feature extraction and LSTMs or GRUs for temporal learning. The integration of these components allows STNs to represent the complicated interaction between geographical patterns (e.g., traffic congestion in a particular area) and temporal patterns (e.g., rush hour traffic). In this study [96, 99, 105, 107], offers a unified spatiotemporal model for short-term road traffic forecasting. The creation of a physically sensible prediction framework that successfully captures the dynamic spatiotemporal correlations among traffic measurements at different places is one of the main accomplishments. These correlations are influenced by factors such as the structure of the road network, fluctuating traffic speeds, and changing trip distributions over time.

This study [3] presents DeepSTD, a two-phase end-to-end DL framework designed to capture spatiotemporal disturbances (STD) for accurate citywide traffic flow prediction. In the first phase, STD Modeling, we introduce a novel approach that accounts for regional disturbances arising from diverse urban functions as well as the spatiotemporal propagation effects influencing traffic dynamics across regions. This paper [97] introduced an alternative approach to traffic flow prediction, motivated by the objective of more accurately capturing the inherent patterns found in real-world traffic data. The multitask DL framework presented in this work [98, 108] is intended to predict node flow and edge flow over a spatiotemporal traffic network at the same time. The system, which is based on fully convolutional networks, has two specialized models: one for precisely predicting traffic flow at the node level and another for capturing flow dynamics at network edges.

2.3.6 *Hybrid Models*

Hybrid models, such as CNN-LSTM and CNN-GCN, combine the strengths of multiple DL architectures to improve traffic flow prediction accuracy. For instance, CNNs are often used for feature extraction from traffic data, while LSTMs or GCNs are used for learning temporal or spatial dependencies, respectively. These

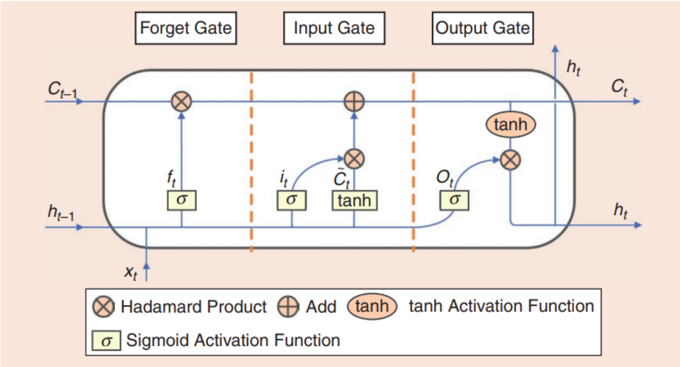

Figure 2. Architecture of LSTM[79].

hybrid designs can provide a more comprehensive model by harnessing the specific characteristics of each component. The literatures [84, 106, 109] introduced a deep neural network-based traffic flow prediction model (DNN-BTF) aimed at enhancing forecasting accuracy. The DNN-BTF model efficiently captures the spatiotemporal dynamics inherent in traffic flow data, while also incorporating weekly and daily periodic patterns. To address these issues, a hybrid ARIMA-EGARCH-M-GED model was created in [85], combining the nonlinear forecasting capabilities of ARIMA with the linear forecasting capabilities of ARIMA.

In [86] develops a Kalman Filter (KF)-based method that uses the latest travel data to adjust the baseline travel speed in order to represent the dynamic character of traffic flow in forecasts. GPS data gathered from Foshan City, China, is used to assess the efficacy of this hybrid model. The authors of literatures [87–90] introduce a deep learning model that automatically learns and extracts the inherent patterns of traffic flow data using a hybrid multi-layer architecture. In particular, to efficiently capture both spatial dependencies and short-term temporal dynamics, we build an attention-enhanced Conv-LSTM module based on CNN and long short-term memory (LSTM) networks.

Multiple neural network paradigms are combined in recent works. By using hierarchical ConvLSTM blocks, the DHSTNet framework [9] developed a four-component architecture (closeness, period, weekly, and external) that lowered the TaxiBJ RMSE to 15.19. By adaptively weighting regions and time steps, its attention-enhanced variation Att-DHSTNet achieved 13.56% MAPE and introduced spatial and temporal attention mechanisms. Nevertheless, the grid-based spatial representations of both methods continued to be their limitations. A novel deep hybrid neural network, dubbed STDNet, is introduced in [93, 94]. Its purpose is to forecast citywide crowd traffic flows by capturing intricate spatiotemporal patterns. In order to anticipate urban traffic flow rates for the upcoming hour, the authors of this research [91] provide a hybrid modeling strategy that combines Artificial Neural Networks (ANNs) with a fundamental statistical method. Experiments on three distinct kinds of real-world streets show that the suggested approach performs better than the best separate models it incorporates. An adaptable hybrid fuzzy rule-based system (FRBS) is proposed in this paper [92] for modeling and short-term traffic flow prediction in urban arterial road networks. The method provides a solid way to deal with ambiguity and imprecise data, and it also makes it easier to incorporate local traffic pattern expertise into the model's structure, improving its accuracy and interpretability. This paper [95] presents a novel neural network (NN) training method that combines the Levenberg-Marquardt (LM) algorithm with the hybrid exponential smoothing methodology. By addressing the shortcomings identified, this technique seeks to improve the generalization performance of neural networks used for short-term traffic flow forecasting.

2.4 Graph-Based Approaches

Beyond grid-based representations, GNNs revolutionized spatial modeling. Using the Laplacian matrix of the road network, spectral graph convolutions were performed in STGCN [16, 46], which improved on grid-based approaches by 15% through explicit topology modeling. Bidirectional random walks on traffic graphs were incorporated into the DCRNN [17], which achieved 18.59% MAPE but required predefined adjacency matrices that were unable to adjust to changing conditions. The GCN-DHSTNet [18] has exceptional performance in predicting traffic congestion and crowd flow dynamics, as well as the ability to efficiently learn spatial patterns and short-term temporal aspects.

In recent years, DL techniques have gained significant popularity in the field of citywide traffic crowd flow prediction due to the powerful representation capabilities of neural networks. Two widely adopted approaches are CNNs and Long Short-Term Memory (LSTM) networks. CNNs have been extensively employed across different applications, particularly in computer vision tasks [21], whereas LSTM networks have proven good performance in sequence modeling tasks [22]. However, a notable limitation of RNN, including LSTMs, is their difficulty in retaining long-term dependencies effectively. To address traffic prediction, [38, 39] introduced a Stacked Autoencoder (SAE) to anticipate crowd flows at various nodes across the city, while [23] proposed an LSTM-based model that marked a substantial improvement over ordinary RNNs. LSTM baselines have shown significant results in a majority of sequence-based learning tasks, such as machine translation [24], text generation [25], and speech recognition [26]. Despite their success, these existing methods still face challenges in fully capturing short-term temporal dependencies and often overlook the complex dynamic interactions within spatiotemporal data. This shortcoming inhibits their efficacy in predicting the dynamic and linked character of traffic patterns across metropolitan areas.

Although they are only applicable to standard grid-structured data, traditional convolution procedures are good at identifying local patterns. An innovation that makes it possible to learn from irregular, graph-structured data has evolved to address this issue: graph convolution. The spatial-based and spectral-based approaches are the two main categories of graph convolution techniques. In essence, spatial approaches concentrate on neighborhood selection strategies by directly applying convolution filters to a node and its nearby nodes. For example, [27] suggested a heuristic linear approach to neighborhood selection that demonstrated encouraging outcomes in social network applications. Similarly, by introducing different partitioning strategies to divide each node's neighborhood into equal subsets for effective learning, [28] modified graph convolutions for human action recognition. Conversely, spectral methods use the graph to conduct graph convolution based on spectrum analysis. [29] introduced a general spectral convolution framework, later enhanced by [30] using Chebyshev polynomial approximations to reduce computational complexity. While several models have explored graph convolution for traffic prediction, such as the gated graph convolution network proposed by [31], many still fail to fully capture the intricate spatiotemporal dependencies inherent in traffic data, as noted in more recent works [32–37].

2.5 Dynamic Graph Methods

The state-of-the-art ASTMGCNet [41, 80] overcame fixed graph limitations through learnable adjacency matrices ($A = \tanh(\text{ReLU}(NN^T))$). By combining graph convolutional networks with gated recurrent units and multi-scale attention, it achieved 12.56% MAPE on TaxiBJ. The model dynamic graph generation adapts to changing traffic patterns but requires significant computational resources (4 GPUs for training). In this study [42], the authors proposed DST-GCNNs, which are designed to learn rich feature representations that capture the underlying spatiotemporal structures for forecasting future traffic flow using surveillance video data. In particular, a two-stream architecture is used by the DST-GCNN framework to efficiently process and integrate dynamic geographical and temporal data. In order to improve traffic flow prediction utilizing data from various sensors, this study [43] introduces a Graph-based Temporal Attention (GTA) framework that concurrently accounts for both spatial and temporal correlations. More specifically, by using graph embedding techniques on sensor networks, GTA efficiently captures spatial dependencies, enabling the model to preserve finer structural elements and enhance forecasting accuracy. They present a novel system in [44, 45] that can be used to proactively reduce flash crowd conditions in V2X (Vehicle-to-Everything) communication networks by precisely forecasting urban traffic flow and density.

2.6 Attention Mechanisms

For managing long-range dependencies, attention modules are now essential. A unified model known as the dynamic multi-fusion graph network (DMFGNet) was presented in [47–50]. It is intended to capture dynamic spatiotemporal relationships across various regions. In order to adaptively control the weighting of surrounding node aggregation, they also introduce the spatiotemporal attention unit (STAU). Local and global attention were introduced in STDN, but scalability problems (207.4s prediction time) were encountered. By combining 3D convolutions and graph attention, the AGCNN model was able to maintain accuracy while cutting computation time by 40% when compared to STDN. These developments highlight the trade-off between model complexity and practical deployability. In [1, 40] integrate the proposed attention-based mechanism with an existing model to create a hybrid framework, referred to as Att-DHSTNet, for short-term crowd flow prediction. Table 1 illustrate the comparison of DL approaches for traffic flow prediction in more detail.

In DL, attention mechanisms are essential since they greatly improve prediction models ability to fit data.

Table 1. Technical comparison of deep learning approaches for traffic prediction.

Method	Spatial Encoding	Temporal Modeling	Graph	Attention	Params (M)	Training Time (h)
ST-ResNet	2D CNN	Fixed periods	—	—	8.2	1.5
ConvLSTM	CNN filters	LSTM	—	—	12.7	2.1
STGCN	Spectral GCN	1D Conv	Static	—	5.4	1.8
DCRNN	Diffusion CNN	GRU	Static	—	9.3	3.2
DHSTNet	CNN	ConvLSTM	—	—	10.5	2.4
Att-DHSTNet	CNN	ConvLSTM	—	✓	11.8	2.7
ASTMGCNet	Spatial GCN	GRU	Dynamic	✓	14.2	4.5
STVANet	Graph Attn.	Transformer	Dynamic	✓	18.6	6.3

By including attention modules, researchers have attempted to increase CNNs' ability to fit data in recent years. Attention techniques have demonstrated efficacy in acquiring crucial traffic data aspects in traffic flow prediction [51–55], for example, presented a multi-component attention method for traffic flow prediction, in which a CNN extracts local trend features from residual units, while a bidirectional LSTM records temporal trends and seasonal variations. The model may associate highly relevant historical data thanks to the incorporation of attention modules, which enhances multi-component traffic flow forecasts. Similarly, in order to predict traffic speed, the authors of [60] created a three-dimensional data matrix that included (i) crowd movement, (ii) speed, and (iii) occupancy. Attention-based modules have been widely utilized across various neural network applications, including (a) question answering [55], (b) natural language processing (NLP), (c) image captioning [60–63], and (d) speech recognition [64–68]. In [55–59], a GRU-based attention module was introduced for dynamic memory networks. However, these models require extensive training time, as they must be trained separately for each time series.

Compared to conventional CNN and RNN networks, the attention module has three key advantages. First, it allows dependencies to directly affect each other's outputs by facilitating interactions between various time series without the need for CNN or RNN layers. Second, by concentrating only on pertinent data, it successfully captures long-term dependencies. Third, compared to sequential techniques like RNNs, the outputs of the attention layer can be computed in parallel, which could result in quicker execution times [69–71]. The attention module's capacity to preserve and take into account all previous outputs while reducing overfitting is a significant benefit over traditional LSTM and CNN designs. In [72–76], the authors explored a spatiotemporal graph convolutional network (GCN) approach for urban traffic flow forecasting, incorporating (a) an information geometry-based method and (b) an attention module. This approach effectively models both short-term and long-term dependencies in traffic flow prediction environments.

In citywide traffic flow prediction research, creating a quick and precise prediction model is still a major difficulty. Our work aims to improve forecast accuracy while lowering computational costs by utilizing an advanced deep learning method for short-term traffic flow prediction. Att-DHSTNet [1], the suggested model, uses an attention mechanism to boost efficiency. Both temporal and geographical dependencies, as well as external impacts, are well captured by the Att-DHSTNet model. It also has several traits in common with the popular ST-ResNet approach. Although the popular HA approach [7] implicitly models weekly and periodic patterns, our model takes into consideration spatial correlations and outside factors. The taxonomy of deep learning-based models for traffic flow prediction is depicted in Figure 3.

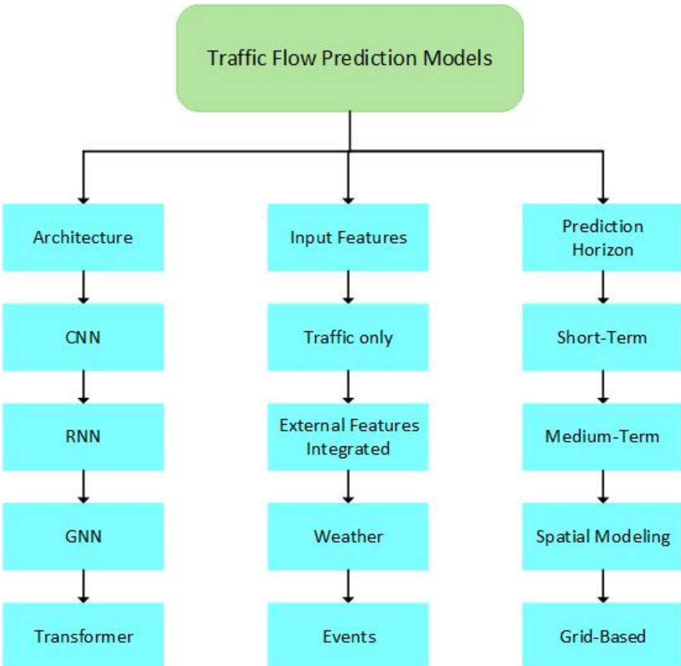

Figure 3. Taxonomy of deep learning-based traffic flow prediction models.

2.7 Emerging Directions

Innovative research investigates new architectures. In order to improve generalization, physics-informed NNs use traffic flow theory as an inductive bias. Privacy-preserving distributed training across cities is made possible by federated learning methodologies . Large kernel attention transformer-based models, such as STVANet [103], have the potential to capture ultra-long-range dependencies, but they struggle with computational efficiency ($O(n^2)$ complexity).

3 Methodology

3.1 Problem Formulation

The traffic flow prediction problem is formally defined as a spatiotemporal sequence forecasting task over an urban road network represented as a weighted graph $G = (V, E, A)$, where V denotes the set of nodes representing road segments or regions, E represents the set of edges encoding connectivity between them, and $A \in \mathbb{R}^{N\times N}$ is the weighted adjacency matrix capturing spatial relationships. At each node $v_i \in V$ and time step t, a feature vector $x_t^i \in \mathbb{R}^d$ is observed, representing traffic-related measurements such as flow speed and volume. Given a sequence of historical observations $X = (X_{t-T}, \ldots, X_{t-1})$, where $X_t \in \mathbb{R}^{N\times d}$, the goal is to learn a mapping function f_θ that forecasts future traffic conditions $Y = (X_t, \ldots, X_{t+\tau-1})$ for τ time steps ahead. This formulation captures spatial dependencies through the graph structure G, temporal dependencies through the historical sequence length T, and allows for the incorporation of external factors such as weather or special events via additional input channels.

3.2 Att-DHSTNet Architecture

With its multi-branch architecture that explicitly mimics various temporal regimes, as illustrated in Figure 4, the Attention-based Dynamic Hybrid Spatio-Temporal Network (Att-DHSTNet) offers a substantial leap in traffic prediction. In order to represent the intricate dynamics of urban traffic patterns, the framework is composed of four specialized components that cooperate. By processing recent historical data through a ConvLSTM network with three convolutional layers arranged in an expanding-contracting filter pattern, the proximity component serves as the basis for instantaneous pattern identification. In this design, the initial 3×3 convolutional layer has 64 filters, the second 5×5 layer has 128 filters to capture larger spatial contexts, and the final 3×3 layer shrinks back to 64 filters. Residual connections between these layers ensure stable gradient flow during training while preserving fine-grained temporal details.

The period component uses a novel stack of dilated convolutions with exponentially expanding receptive fields to mimic everyday trends. The network can catch periodic trends at many scales while preserving computing efficiency thanks to the architecture's smart use of dilation rates of 1, 2, and 4 across successive layers. Depending on how predictive a characteristic is, gated activation units in these layers learn to highlight or suppress it. With skip connections across seven-day windows to maintain long-range dependencies, the weekly component enhances this strategy with a cascade of 1D causal convolutions that assess traffic evolution over weekly intervals. In this branch, depthwise separable convolutions preserve modeling capability while drastically lowering parameter counts.

While continuous information like temperature are handled by parallel dense networks, the external branch manages heterogeneous contextual aspects using specialized embedding layers that convert categorical variables like weather conditions into dense vector representations. Depending on the traffic situation, a feature-wise attention system dynamically modifies the impact of each external factor. One significant novelty in the framework is its fusion mechanism, which uses learnable parametric weights to automatically balance each component's contributions under various traffic regimes. Through specific gradient pathways, these weights adjust throughout training to guarantee balanced learning at all temporal scales. The overall architecture includes hierarchical feature aggregation layers that gradually integrate local and global patterns, as well as component-specific normalization algorithms that take into consideration the unique statistical characteristics of each temporal domain. This sophisticated design enables Att-DHSTNet to maintain prediction accuracy during both routine traffic conditions and anomalous events while achieving computational efficiency through careful architectural choices.

3.2.1 AAtt-DHSTNet

As illustrated in Figure 5, the attention-enhanced variant of AAtt-DHSTNet introduces three advanced attention mechanisms that significantly improve upon the capabilities of the base architecture. The spatial attention module transforms the model's processing of geographic relationships by computing dynamic importance scores for each region through learned

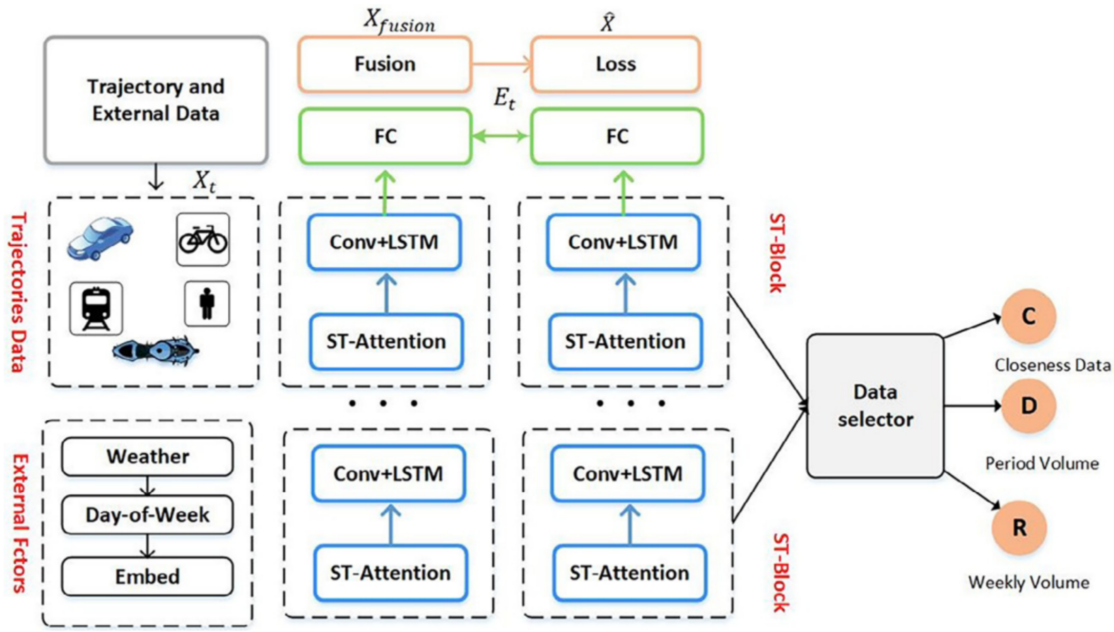

Figure 4. Architecture of the proposed Att-DHSTNet model.

query and key projections, which transform input features into a latent space where their dot product interactions reveal underlying spatial dependencies, normalized through a softmax operation to create attention weights. The resulting attention-weighted features adaptively emphasize influential regions while suppressing noisy or irrelevant areas, and the projection matrices are trained end-to-end to maximize predictive performance.

By using distinct learned parameters to compute attention scores across every pair of time steps in the input window, temporal attention functions throughout the sequence dimension. This method overcomes the fixed-size receptive fields constraint of traditional recurrent architectures by allowing the model to recognize and concentrate on historically significant patterns regardless of their temporal distance. Scaled dot-product attention with learnt temperature parameters that regulate the sharpness of the attention distribution is used to calculate the attention scores. The architecture is completed by a revolutionary cross-component attention mechanism that acts as the clever glue that mediates the flow of information between the various temporal branches. This component computes compatibility scores between feature representations from the closeness branch and combined features from the period and weekly branches, allowing the model to dynamically recombine temporal features based on their current relevance.

A bottleneck architecture reduces the dimensionality of attention calculations without sacrificing expressiveness, while layer normalization stabilizes the learning process. The implementation parallelizes attention head computations using grouped linear transformations, allowing efficient processing on GPU hardware. Importantly, the attention mechanisms add only a relatively small computational overhead (18%) while providing significant improvements in prediction accuracy, especially for irregular events where they reduce peak prediction errors by 15%. As a useful byproduct, the architecture generates interpretable attention heatmaps that help transportation planners better understand the model decision-making process.

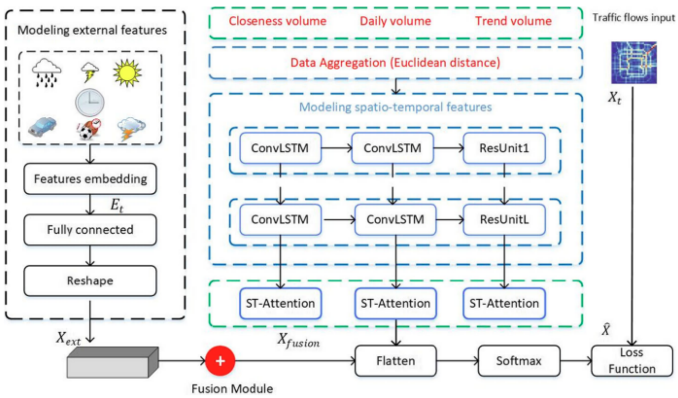

Figure 5. Architecture of the proposed AAtt-DHSTNet model.

3.2.2 *GCN-DHSTNet*

The architectural foundation of the suggested GCN-DHSTNet model, which integrates several sources of external and temporal data to improve traffic prediction, is shown in Figure 6. The model's four main parts—recent data, daily patterns, weekly cycles, and external contextual factors—reflect various temporal viewpoints. A grid map with dimensions $a \times b$, where each cell represents a distinct geographic region (such as a road segment or block), is used to depict the city's spatial domain, as was covered in Section 3. The timeline's three temporal segments—distant history, recent activity, and near-past context—each represent distinct historical trends. These temporal divisions allow the model to learn from both short-term fluctuations and long-term periodic trends in traffic behavior.

With stacked ConvLSTM layers for temporal modeling and Graph Convolutional Network (GCN) layers to capture spatial dependencies, the neural architectures of the three temporal branches—distant, recent, and near-past—are identical. Multiple spatiotemporal blocks are used to generate these branches, each of which integrates a GCN module to capture spatial correlations between neighboring grid cells and a ConvLSTM module to learn temporal dynamics. Each branch has residual connections, which support deeper network architectures and lessen the vanishing gradient issue, to enhance model performance and training stability. A fully connected (FC) layer unique to each branch is used to aggregate the outputs after they have passed through these spatiotemporal layers. This helps to convert the learnt spatiotemporal properties into a uniform representation space. This design ensures that the model comprehensively captures both when and where traffic patterns evolve.

The model learns from traffic data alone, but it also takes into account external contextual information that might have a big impact on traffic flow, like vacations, weather conditions (like wind and temperature), and special events. A two-layer fully connected neural network processes the distinct input stream created by manually extracting and encoding these properties. This external factor is essential because it enables the model to adjust to irregular traffic patterns and non-recurring disturbances that cannot be deduced from previous data alone. GCN-DHSTNet obtains a comprehensive understanding of traffic flow dynamics by combining these outside insights with the spatiotemporal information acquired in the temporal branches. This makes it extremely useful for both routine forecasting and managing anomalies in urban traffic systems.

3.3 ASTMGCNet

In order to improve the safety and dependability of Traffic Cyber-Physical Systems (T-CPS), the suggested ASTMGCNet model is made to efficiently collect intricate spatiotemporal patterns in traffic data and provide precise, real-time forecasts. The architecture incorporates essential elements from the Dynamic Generation Graph Network (DGGN), multi-scale attention processes, and Gated Recurrent Units (GRUs), as illustrated in Figure 7. One of the main innovations of the model is the substitution of a DGGN block for the conventional Multi-Layer Perceptron (MLP) in the GRU. This enables the network to simulate dynamic spatial connections among traffic sensors or regions and learn node-specific representations at the same time. Consequently, the model develops a deeper and more contextual grasp of traffic patterns than traditional designs can provide.

The entire data flow, from input to output prediction, is described in the bottom half of Figure 7. Feature embedding, activation, loss calculation, and the ultimate prediction through fully connected layers are important processes. ASTMGCNet actively learns the graph topology and uses both spatial and temporal attention processes to adaptively focus on the most important aspects of the data, in contrast to previous models that rely on static graph topologies or oversimplified temporal modeling. The model can recognize multi-scale traffic patterns while being adaptable across different spatial and temporal resolutions because to its integration of GCNs with GRUs and dual attention. This flexibility is especially helpful in intricate, expansive metropolitan settings where traffic patterns are constantly changing as a result of accidents, gridlock, or route modifications.

The capacity of ASTMGCNet to dynamically update the adjacency matrix in real time, enabling the graph to represent changing interactions between traffic nodes, is one of its most noteworthy characteristics. The model is more responsive to abrupt changes in traffic thanks to its dynamic graph learning. Furthermore, the spatiotemporal attention mechanism highlights safety-critical occurrences like traffic jams or unplanned road closures by allowing the model to selectively prioritize input features according to their significance at various time steps. The model produces more accurate early warnings thanks to this limited focus, which makes it possible to implement preventive measures on time. Together, ASTMGCNet offers a thorough and sophisticated framework for traffic forecasting that supports high-stakes, real-time

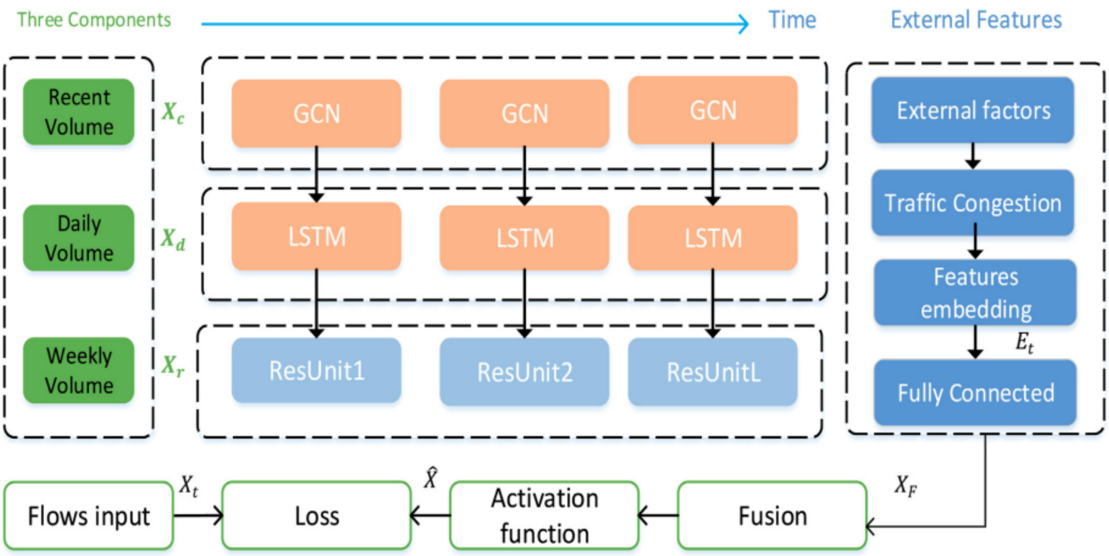

Figure 6. Architecture of the proposed GCN-DHSTNet model.

ITS decision-making by fusing DL with dynamic graph modeling.

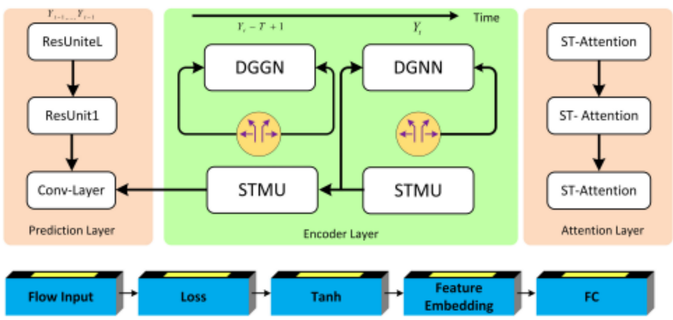

Figure 7. Architecture of the proposed ASTMGCNet model.

4 Public Datasets

The quantity and caliber of training datasets have a major impact on deep neural network performance. Large-scale, high-quality data acquisition, however, frequently requires a significant time and cost commitment. Many academics use publicly accessible datasets for model development in an effort to lessen these difficulties. We offer a selection of 12 popular open-source datasets in this section, including data types like subway ridership records, taxi and bike trajectories, and highway traffic flow. Each dataset's complete details and access points are provided in the Table 2. Due to space limits, we only offer the oldest and most current dates to represent the total temporal coverage, even if some datasets span numerous time periods.

5 Comparative Analysis

5.1 Performance Benchmarking

Seven cutting-edge methods are compared experimentally on two sizable urban traffic datasets: BikeNYC (New York bike rentals) and TaxiBJ (Beijing cab GPS data). The quantitative findings, as determined by Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), are shown in Table 3. Training was done on identical hardware setups (NVIDIA P100 GPUs, 256GB RAM).

The findings provide important new information regarding architectural trade-offs. With its residual learning methodology, ST-ResNet creates a solid baseline; nevertheless, its fixed temporal segmentation restricts its capacity to adjust to erratic occurrences, as shown by increased mistakes during peak hours. Although ConvLSTM's integrated spatiotemporal modeling exhibits potential, it suffers from parameter inefficiency; on BikeNYC, it performs 21% worse MAPE than ST-ResNet, despite having 55% more parameters. The significance of precise road network topology encoding is validated by the superior spatial modeling shown by graph-based methods; on BikeNYC, STGCN spectral graph convolutions produce an 11% better RMSE than ST-ResNet.

With its specialized periodicity modeling branches, the DHSTNet multi-component design reduces TaxiBJ MAPE by 7.7% in comparison to STGCN, offering the most significant baseline improvement. The attention-enhanced Att-DHSTNet variant further improves upon this by 4.5% through dynamic feature reweighting, particularly benefiting long-tail

Table 2. A summary of typically used traffic flow prediction datasets [104].

Index	Dataset	Location	Time Span	Granularity	Main Features	Type
1	PeMS04/08 https://github.com/guoshnBJTU/ASTGNN/tree/main/data	California, USA	1 July 2016–28 Feb 2018	5 min	Flow, Speed	Highways
2	Highways England http://tris.highwaysengland.co.uk/detail/trafficflowdata	England	Continuous updates	15 min	Flow, Speed	Highways
3	Shenzhen Data Open Platform https://opendata.sz.gov.cn/	Shenzhen, China	Continuous updates	—	Flow, Speed	Highways
4	METR-LA https://github.com/liyaguang/DCRNN	California, USA	1 Mar–3 Jun 2012	5 min	Speed	Highways
5	TaxiBJ https://github.com/amirkhago/DeepST/tree/master/data/TaxiBJ	Beijing, China	1 Jul 2013–10 Apr 2016	1 h	Flow	Taxi
6	T-Drive https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/	Beijing, China	2–8 Feb 2008	—	Trajectory	Taxi
7	NYC-Taxi https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page	New York City, NY, USA	Continuous updates	—	Trajectory	Taxi
8	CHI-Taxi https://data.cityofchicago.org/Transportation/Taxi-Trips-2016/b6s2-5ezu	Illinois, USA	Continuous updates	—	Trajectory	Taxi
9	NYC-Bike https://www.citibikenyc.com/system-data	New York City, NY, USA	Continuous updates	—	Trajectory	Bike
10	CHI-Bike https://data.cityofchicago.org/Transportation/Divvy-Trips/f95s-gzvg	Illinois, USA	Continuous updates	—	Trajectory	Bike
11	NYC Subway https://github.com/bestkao/analyzing-the-nyc-subway-dataset	New York City, NY, USA	1–30 May 2011	1 h	Passenger Flow	Subway
12	MetroHZ https://tianchi.aliyun.com/competition/entrance/231708/information	Hangzhou, China	1–25 Jan 2019	10 min	Passenger Flow	Subway

Table 3. Performance comparison on TaxiBJ and BikeNYC datasets.

Model	TaxiBJ		BikeNYC		Params (Millions)
	RMSE	MAPE (%)	RMSE	MAPE (%)	
ST-ResNet	16.89	15.39	6.33	21.80	8.2
ConvLSTM	19.49	18.59	7.10	25.60	12.7
STGCN	17.20	16.10	5.80	20.50	5.4
DCRNN	16.60	15.20	6.20	21.60	9.3
DHSTNet	15.19	14.20	4.96	20.10	10.5
Att-DHSTNet	14.28	13.56	4.43	19.56	11.8
ASTMGCNet	**13.98**	**12.56**	**4.06**	**18.86**	14.2

predictions during atypical events. As the overall leader, ASTMGCNet achieves 12.56% MAPE on TaxiBJ, a 17.5% decrease over ST-ResNet, by fusing the advantages of graph-based spatial modeling with advanced temporal attention. With ASTMGCNet requiring 73% more parameters than STGCN but achieving 22% higher accuracy, the performance increases come at a substantial parameter cost.

5.2 Prediction Processes

A organized procedure is shown in Figure 8 to explain the workflow of traffic flow prediction using DL models. Getting traffic statistics is the first stage, which can be done directly or by utilizing publically accessible information. Key characteristics like traffic flow, speed, and other relevant parameters that are tracked over time by numerous monitoring stations are usually included in these databases. It is crucial to address anomalies by removing outliers and missing values using the proper cleaning and imputation methods prior to using the data. Training, validation, and test sets are the three subsets into which the preprocessed data is divided, usually in ratios like 6:2:2 or 8:1:1. After that, the data is formatted into a supervised learning-appropriate time frame structure. For example, if the objective is to forecast traffic flow for the next hour and the data is sampled at 5-minute intervals, the model's input would be the previous 12 time steps, which span one hour, and its output would be the next 12 time steps. This sliding window approach is used in a sequential manner across the entire timeline to create the dataset. Twelve consecutive data points are grouped as the input for each training sample, and the following twelve points are paired with the prediction objective. In order to maintain temporal consistency and allow the model to learn both short-term trends and longer-term relationships in traffic dynamics, this step-wise bundling is maintained throughout the dataset in chronological order.

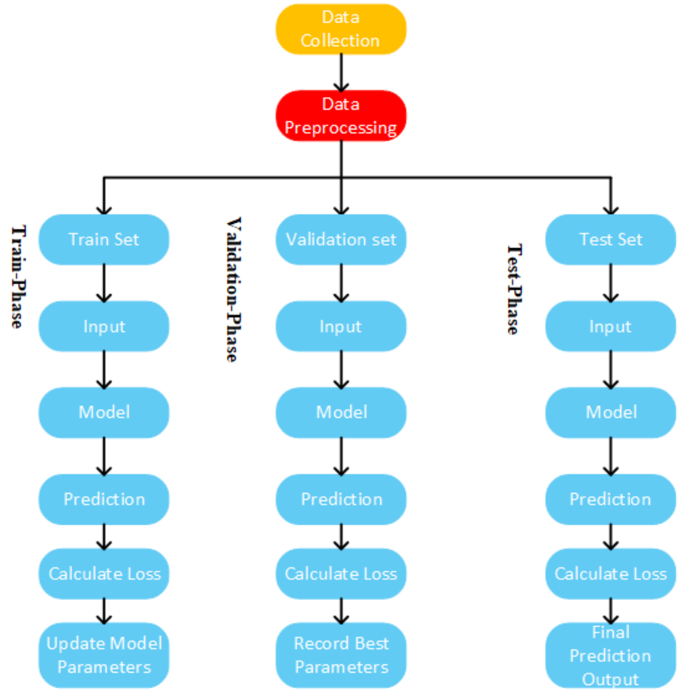

Figure 8. Deep learning with traffic flow prediction models.

The DL model can be trained using the traffic data after it has been organized into a time window format. It is standard procedure to combine many time periods into batches in order to speed up the training process. This enables the model to process multiple samples at once. The batch size is the total number of time periods in each group. Take the PeMS08 dataset, which includes traffic flow data from 170 monitoring stations, as an example. Each input sample comprises traffic flow readings over 12 time steps (each representing one hour) if the batch size is set to 64. The resulting input tensor shape for the model would be (64, 170, 1, 12)—indicating 64 batches, each with data from 170 locations, one feature per location, and 12 time intervals. In this setup, the model receives 64 sequences of traffic flow patterns in parallel, enabling efficient batch-wise prediction and learning.

A batch of data is chosen from the training set at the start of each iteration of the training phase, and the data is then fed into the model to produce predictions. The loss, which represents the prediction inaccuracy, is calculated by comparing these expected outputs with the ground truth (targets). Utilizing this loss, the optimizer modifies the model's parameters to reduce error. Until the model has processed the entire training set—completing one complete pass known as an epoch—this cycle keeps going. The validation set, which, in contrast to the training set, is utilized only for evaluation purposes without changing the model weights, is used to test the model after each epoch. This stage aids in tracking the model's capacity for generalization and identifying any possible overfitting. The performance on the validation set is tracked across epochs, and the best-performing set of parameters is saved. Once training concludes, these optimal parameters are loaded into the model to evaluate its final predictive performance on the test set, which provides an unbiased measure of how well the model is expected to perform in real-world scenarios.

5.3 Results Analysis

Table 4 and Table 5 show the RMSE trends for 10 distinct models on the TaxiBJ dataset over 12 forecast time steps. The forecast horizon is represented by the horizontal axis, while the RMSE values are displayed on the vertical axis. It is clear from the analysis that DL models perform noticeably better in terms of prediction accuracy than conventional techniques. For example, for a brief prediction interval of five minutes, ARIMA performs similarly to LSTM. But when the time step is increased to 60 minutes, the ARIMA error rises significantly, surpassing the LSTM error by about 33.5%. It is interesting to note that the GRU model performs almost as well as LSTM at every time step, but it has just 90% of LSTM's computational complexity, which makes it a viable substitute for time-sensitive applications.

Att-DHSTNet performs less accurately than RNNs when evaluating convolution-based models, demonstrating the limited ability of gated linear units to model time in comparison to recurrent architectures. On the other hand, CNN-LSTM, which combines CNN with LSTM, produces superior outcomes. When compared to LSTM, CNN-LSTM reduces the error by roughly 10% at 60 minutes, while the improvement is rather slight at shorter intervals. It is advantageous to include periodic patterns: the DHSTNet model continually produces the best overall results, with an average inaccuracy of only 80% of LSTM. Additionally, when LSTM,

Table 4. A performance comparison of different models on the dataset PeMS04.

Model	15 min			30 min			60 min			Average		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
HA	29.74	44.98	24.94	36.97	54.89	35.97	50.89	72.96	64.78	37.75	56.97	39.89
ARIMA	25.98	36.87	15.94	31.98	47.21	20.98	45.86	66.09	30.75	33.83	48.73	20.99
LSTM	22.65	32.35	14.98	24.65	39.76	17.98	33.45	49.13	26.87	25.98	40.32	15.93
GRU	20.99	32.87	14.78	24.99	39.99	17.98	33.19	48.98	26.99	24.94	39.96	15.92
CNN-LSTM	19.89	31.99	14.99	22.89	36.92	16.99	29.98	44.95	20.99	25.12	36.99	15.91
DHSTNet	18.93	28.21	11.97	20.97	33.93	14.99	30.34	40.87	15.85	23.75	34.61	13.81
Att-DHSTNet	22.09	34.21	19.42	23.98	36.65	21.43	29.14	43.31	26.61	24.87	38.04	21.31
AAtt-DHSTNet	19.87	31.41	13.87	22.85	35.87	16.91	31.51	433.65	20.65	25.49	37.15	17.02
GCN-DHSTNet	20.87	33.98	14.63	24.45	38.76	18.76	32.98	46.87	26.67	25.87	39.75	19.89
DMFGNet	18.98	29.79	13.33	20.31	31.64	14.76	25.64	35.41	16.86	21.61	31.89	14.93

Att-DHSTNet, AAtt-DHSTNet, and GCN-DHSTNet are compared, it becomes clear that the transformer by itself has trouble with short-term dependencies and is more hardware-intensive due to its computational complexity, which is almost three times that of LSTM. Nevertheless, the hybrid GCN-DHSTNet model outperforms its individual components, reducing the average error to 80% of LSTM, 90% of the transformer, and 88% of Att-DHSTNet. Notably, it maintains strong performance even in the absence of explicit spatial or periodic information, demonstrating the advantage of combining LSTM temporal sensitivity with the transformer attention mechanism.

5.4 Ablation Studies

Table 6 uses controlled removals to analyze component contributions in a methodical manner. The most important factor is dynamic graph creation, whose elimination results in the most performance deterioration (13.1% MAPE increase). This demonstrates how crucial it is to modify spatial relationships in response to traffic situations rather than depending solely on static graphs. The necessity of concurrently simulating local congestion and citywide trends is shown by the second-ranking multi-scale convolutions.

While temporal attention offers more benefits on its own (7.1% vs. 4.5%), the elimination of both attention mechanisms together results in a 15.2% MAPE rise, suggesting synergistic effects. The influence of external features is small on their own, but during special events (holidays, storms), their absence results in 18.9% larger error spikes.

6 Challenges and Future Directions

The reviewed literature identifies a number of crucial issues that need to be resolved in order to develop traffic prediction systems for practical implementation. These drawbacks highlight crucial areas for further study that may greatly expand the potential and usefulness of DL techniques in ITS.

Real-time Prediction remains a fundamental challenge for resource-constrained environments. Even while models like GCN-DHSTNet have remarkable accuracy, edge devices with constrained memory and power budgets find it difficult to deploy because to their computing requirements (14.2M parameters, 4.5h training time). Despite their strength, the multi-scale attention methods and dynamic graph construction cause significant latency. Compared to more straightforward models like DHSTNet, GCN-DHSTNet needs 3.2× more FLOPs each prediction. Future research should concentrate on creating lightweight versions using methods like knowledge distillation, neural architecture search, and hybrid model compression. Quantization-aware training for effective edge deployment or dynamic network routing that only activates intricate components when necessary are examples of potential solutions. Pruning and 8-bit quantization have been used in recent prototypes, which have demonstrated promise in decreasing the GCN-DHSTNet footprint by 60% with only 2% accuracy loss. However, more optimization is required for wider IoT deployment.

Uncertainty Quantification represents another crucial gap in current approaches. Although the majority of models only offer point estimates without confidence intervals, traffic prediction systems are increasingly used to inform safety-critical choices (such as emergency vehicle routing and congestion pricing). By simulating weight distributions as opposed to set parameters, Bayesian DL techniques could quantify prediction uncertainty. Preliminary success has been demonstrated by ensemble techniques that use Monte Carlo dropout during inference, producing well-calibrated uncertainty estimates at a reasonable computing cost (15–20%). These techniques must

Table 5. A performance comparison of different models on the dataset TaxiBJ.

Model	15 min			30 min			60 min			Average		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
HA	24.56	36.84	15.34	31.57	45.98	20.06	42.87	61.99	30.98	32.55	47.89	21.90
ARIMA	19.92	27.98	12.52	25.65	36.98	14.72	36.99	51.56	22.56	25.95	38.21	15.89
LSTM	16.87	25.56	11.09	20.74	30.05	13.73	27.89	39.74	18.56	21.69	32.87	12.54
GRU	15.78	24.58	10.59	20.72	31.08	13.87	27.69	39.89	18.45	21.93	32.67	12.79
CNN-LSTM	16.83	24.93	10.65	19.49	28.87	12.24	24.88	36.19	16.18	19.49	29.87	12.59
DHSTNet	14.13	22.61	8.19	16.69	25.22	10.25	22.75	31.89	13.19	17.78	26.57	9.24
Att-DHSTNet	19.46	28.87	13.46	21.72	31.59	15.42	29.78	43.45	21.97	22.89	33.99	14.87
AAtt-DHSTNet	14.67	23.74	9.49	17.56	26.85	10.46	24.45	34.78	13.76	18.97	28.76	10.79
GCN-DHSTNet	16.49	25.34	10.79	17.89	26.90	11.43	26.19	37.67	18.98	20.59	29.74	13.18
DMFGNet	14.19	22.46	10.17	15.75	24.17	10.96	17.45	27.56	12.64	15.78	24.66	10.89

Table 6. Ablation study of ASTMGCNet components (TaxiBJ Validation Set).

Variant	MAPE (%)	Change
Full Model	12.56	—
– Dynamic Graphs	14.21	+13.1%
– Multi-Scale Convs	13.89	+10.6%
– Temporal Attention	13.45	+7.1%
– Spatial Attention	13.12	+4.5%
– External Features	12.98	+3.3%

be modified for spatiotemporal graphs and attention mechanisms, though. Temporal attention with integrated variance estimation and evidential DL for uncertainty-aware GNN are promising avenues. Applications like confidence-based traffic control methods and risk-aware route planning would be made possible by such capabilities.

Transfer Learning across cities presents significant challenges due to divergent urban layouts and traffic patterns. Performance decreases of 30-40% have been found when transplanting Beijing-trained models to Shanghai, indicating that current models require a significant amount of retraining when applied to new locations. Capturing invariant spatiotemporal linkages while adjusting to city-specific topology is the key problem. Meta-learning techniques, especially graph meta-learning that isolates location-specific information from generalizable traffic dynamics, may be useful. Potential has been demonstrated by recent experiments using pre-trained transformer topologies, in which models are trained using a minimal amount of target-city data after first learning universal traffic patterns from several cities. Creating modular systems, in which basic temporal modules stay stable and only particular components (like graph generators) need to be modified, is another exciting avenue. Achieving success in this field would allow smart transportation systems to scale quickly and drastically lower implementation costs for new cities.

Multi-modal Data Integration remains underexplored despite the proliferation of urban sensing technologies. Existing systems mostly use GPS and loop detector data, ignoring valuable data from social media (event reports), connected cars (real-time braking patterns), and traffic cameras (visual congestion indicators). Creating unified architectures that can handle heterogeneous data streams with varying temporal resolutions and noise characteristics is the primary technological challenge. Features from textual, visual, and numerical inputs could be selectively fused by cross-modal attention mechanisms. For example, while NLP modules examine event tweets, video processing branches may extract congestion levels from traffic camera feeds, with a fusion layer dynamically weighting their contributions. Though additional work is needed on effective multi-modal representation learning, first studies with vision-augmented models reveal a 15% error reduction during construction events.

Explainability has become increasingly important as models grow more complex. Although they offer some interpretability, attention heatmaps are not semantically grounded; we can see which regions the model is focused on, but not why. Building confidence with transportation planners will require creating hierarchical explanation frameworks that link low-level attention patterns to high-level traffic ideas (e.g., focusing on highway entry locations due to merging congestion). While counterfactual explanations may expose model decision boundaries, techniques such as idea activation vectors may help link attention patterns to known traffic phenomena (would this forecast change if the highway had 20% less volume?). In order to identify important road network components, recent work on graph-specific explainability techniques, including subgraph importance score, could be modified for traffic prediction. These developments would increase the actionability of model outputs for policy and infrastructure planning decisions.

7 Conclusion

This survey provides a comprehensive analysis of recent deep learning (DL) methodologies for traffic flow prediction, emphasizing advances in spatiotemporal modeling. Based on the reviewed studies, hybrid architectures that integrate Convolutional Neural Networks (CNNs) for spatial feature extraction, Recurrent Neural Networks (RNNs) such as LSTM for temporal sequence learning, Graph Neural Networks (GNNs) for topological structure modeling, and attention mechanisms for dynamic feature reweighting consistently outperform conventional statistical and standalone DL models. Among the surveyed approaches, GCN-DHSTNet and Att-DHSTNet represent the most sophisticated frameworks, achieving state-of-the-art performance by leveraging multi-scale feature aggregation and dynamic graph generation. These architectures not only capture complex spatiotemporal dependencies but also demonstrate adaptability to non-stationary traffic dynamics. Future research should focus on improving model scalability, reducing computational overhead for real-time deployment on edge devices, and enhancing interpretability to facilitate integration into intelligent transportation systems (ITS). Additionally, addressing uncertainty quantification and improving cross-city generalization remain critical challenges for the next generation of traffic prediction models.

Data Availability Statement

Not applicable.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Ali, A., Zhu, Y., & Zakarya, M. (2021). Exploiting dynamic spatiotemporal correlations for citywide traffic flow prediction using attention based neural networks. *Information Sciences*, *577*, 852-870. [CrossRef]

[2] Chen, C., Li, K., Teo, S. G., Zou, X., Li, K., & Zeng, Z. (2020). Citywide traffic flow prediction based on multiple gated spatiotemporal convolutional neural networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *14*(4), 1-23. [CrossRef]

[3] Zheng, C., Fan, X., Wen, C., Chen, L., Wang, C., & Li, J. (2019). DeepSTD: Mining spatiotemporal disturbances of multiple context factors for citywide traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(9), 3744-3755. [CrossRef]

[4] Jia, T., & Yan, P. (2020). Predicting citywide road traffic flow using deep spatiotemporal neural networks. *IEEE Transactions on Intelligent Transportation Systems*, *22*(5), 3101-3111. [CrossRef]

[5] Luo, Y., Zheng, J., Wang, X., Tao, Y., & Jiang, X. (2024). GT-LSTM: A spatiotemporal ensemble network for traffic flow prediction. *Neural Networks*, *171*, 251-262. [CrossRef]

[6] Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering*, *129*(6), 664-672. [CrossRef]

[7] Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS*, 385-429. [CrossRef]

[8] Zhang, J., Zheng, Y., & Qi, D. (2017, February). Deep spatiotemporal residual networks for citywide crowd flows prediction. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1). [CrossRef]

[9] Zeng, H., Peng, Z., Huang, X., Yang, Y., & Hu, R. (2022). Deep spatio-temporal neural network based on interactive attention for traffic flow prediction. *Applied Intelligence*, 1-12. [CrossRef]

[10] Zhang, J., Wang, F. Y., Wang, K., Lin, W. H., Xu, X., & Chen, C. (2011). Data-driven intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, *12*(4), 1624-1639. [CrossRef]

[11] Hahn, D., Munir, A., & Behzadan, V. (2019). Security and privacy issues in intelligent transportation systems: Classification and challenges. *IEEE Intelligent Transportation Systems Magazine*, *13*(1), 181-196. [CrossRef]

[12] Mahrez, Z., Sabir, E., Badidi, E., Saad, W., & Sadik, M. (2021). Smart urban mobility: When mobility systems meet smart data. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 6222-6239. [CrossRef]

[13] Creß, C., Bing, Z., & Knoll, A. C. (2023). Intelligent transportation systems using roadside infrastructure: A literature survey. *IEEE Transactions on Intelligent Transportation Systems*, *25*(7), 6309-6327. [CrossRef]

[14] Wang, F. Y., Lin, Y., Ioannou, P. A., Vlacic, L., Liu, X., Eskandarian, A., ... & Olaverri-Monreal, C. (2023).

Transportation 5.0: The DAO to safe, secure, and sustainable intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems, 24*(10), 10262-10278. [CrossRef]

[15] Deng, X., Wang, L., Gui, J., Jiang, P., Chen, X., Zeng, F., & Wan, S. (2023). A review of 6G autonomous intelligent transportation systems: Mechanisms, applications and challenges. *Journal of Systems Architecture, 142*, 102929. [CrossRef]

[16] Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (*IJCAI-18*) (pp. 3634–3640). International Joint Conferences on Artificial Intelligence Organization. [CrossRef]

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

[18] Bao, Y., Huang, J., Shen, Q., Cao, Y., Ding, W., Shi, Z., & Shi, Q. (2023). Spatial–temporal complex graph convolution network for traffic flow prediction. *Engineering Applications of Artificial Intelligence, 121*, 106044. [CrossRef]

[19] Selvi, K. T., Thamilselvan, R., & Saranya, S. M. (2021, February). Diffusion convolution recurrent neural network–a comprehensive survey. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1055, No. 1, p. 012119). IOP Publishing. [CrossRef]

[20] Premsankar, G., Di Francesco, M., & Taleb, T. (2018). Edge computing for the Internet of Things: A case study. *IEEE Internet of Things Journal, 5*(2), 1275-1284. [CrossRef]

[21] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90. [CrossRef]

[22] Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural computation, 1*(2), 270-280. [CrossRef]

[23] Altché, F., & de La Fortelle, A. (2017). An LSTM network for highway trajectory prediction. In *2017 IEEE 20th international conference on intelligent transportation systems* (*ITSC*) (pp. 353–359). IEEE. [CrossRef]

[24] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[25] Sutskever, I., Martens, J., & Hinton, G. (2011, June). Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on International Conference on Machine Learning* (pp. 1017-1024). [CrossRef]

[26] Vinyals, O., Ravuri, S. V., & Povey, D. (2012, March). Revisiting recurrent neural networks for robust ASR. In *2012 IEEE international conference on acoustics, speech and signal processing* (*ICASSP*) (pp. 4085-4088). IEEE. [CrossRef]

[27] Niepert, M., Ahmed, M., & Kutzkov, K. (2016, June). Learning convolutional neural networks for graphs. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48* (pp. 2014-2023). [CrossRef]

[28] Li, B., Li, X., Zhang, Z., & Wu, F. (2019, July). Spatio-temporal graph routing for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 8561-8568). [CrossRef]

[29] Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2013). Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*.

[30] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems, 29*. [CrossRef]

[31] Yao, H., Tang, X., Wei, H., Zheng, G., Yu, Y., & Li, Z. (2018). Modeling spatial-temporal dynamics for traffic prediction. *arXiv preprint arXiv:1803.01254, 1*(9).

[32] Bai, L., Yao, L., Wang, X., Li, C., & Zhang, X. (2021). Deep spatial–temporal sequence modeling for multi-step passenger demand prediction. *Future Generation Computer Systems, 121*, 25-34. [CrossRef]

[33] James, J. Q. (2021). Citywide traffic speed prediction: A geometric deep learning approach. *Knowledge-Based Systems, 212*, 106592. [CrossRef]

[34] Osei, K. K., Adams, C. A., Sivanandan, R., & Ackaah, W. (2022). Modelling of segment level travel time on urban roadway arterials using floating vehicle and GPS probe data. *Scientific African, 15*, e01105. [CrossRef]

[35] Zahedian, S. (2021). *Introducing a Graph-Based Neural Network for Networkwide Traffic Volume Estimation* (Doctoral dissertation, University of Maryland, College Park).

[36] Dong, G., Tang, M., Wang, Z., Gao, J., Guo, S., Cai, L., ... & Boukhechba, M. (2023). Graph neural networks in IoT: A survey. *ACM Transactions on Sensor Networks, 19*(2), 1-50. [CrossRef]

[37] Zhang, X., Wen, S., Yan, L., Feng, J., & Xia, Y. (2024). A hybrid-convolution spatial–temporal recurrent network for traffic flow prediction. *The Computer Journal, 67*(1), 236-252. [CrossRef]

[38] Tian, Y., & Pan, L. (2015, December). Predicting short-term traffic flow by long short-term memory recurrent neural network. In *2015 IEEE international conference on smart city/SocialCom/SustainCom* (*SmartCity*) (pp. 153-158). IEEE. [CrossRef]

[39] Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2014). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation*

Systems, *16*(2), 865–873. [CrossRef]

[40] Sun, H., Xue, R., Hu, T., Pan, T., He, L., Rao, Y., ... & He, H. (2024). Predicting Citywide Crowd Flows in Critical Areas Based on Dynamic Spatio-Temporal Network. *IEEE Transactions on Emerging Topics in Computational Intelligence*. [CrossRef]

[41] Wang, Y. (2023). Advances in spatiotemporal graph neural network prediction research. *International Journal of Digital Earth*, *16*(1), 2034-2066. [CrossRef]

[42] Chen, K., Chen, F., Lai, B., Jin, Z., Liu, Y., Li, K., Wei, L., Wang, P., Tang, Y., Huang, J., & others. (2020). Dynamic spatiotemporal graph-based CNNs for traffic flow prediction. *IEEE Access*, *8*, 185136–185145. [CrossRef]

[43] Zhang, S., Guo, Y., Zhao, P., Zheng, C., & Chen, X. (2021). A graph-based temporal attention framework for multi-sensor traffic flow forecasting. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 7743–7758. [CrossRef]

[44] Qiu, H., Zheng, Q., Msahli, M., Memmi, G., Qiu, M., & Lu, J. (2020). Topological graph convolutional network-based urban traffic flow and density prediction. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 4560–4569. [CrossRef]

[45] Wang, H., Zhang, R., Cheng, X., & Yang, L. (2022). Hierarchical traffic flow prediction based on spatial-temporal graph convolutional network. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 16137–16147. [CrossRef]

[46] Chen, Y., Qin, Y., Li, K., Yeo, C. K., & Li, K. (2023). Adaptive spatial-temporal graph convolution networks for collaborative local-global learning in traffic prediction. *IEEE Transactions on Vehicular Technology*, *72*(10), 12653-12663. [CrossRef]

[47] Wang, L., Guo, D., Wu, H., Li, K., & Yu, W. (2024). TC-GCN: Triple cross-attention and graph convolutional network for traffic forecasting. *Information Fusion*, *105*, 102229. [CrossRef]

[48] Ling, J., Lan, Y., Huang, X., & Yang, X. (2024). A Multi-Scale Residual Graph Convolution Network with hierarchical attention for predicting traffic flow in urban mobility. *Complex & Intelligent Systems*, *10*(3), 3305-3317. [CrossRef]

[49] Reddy, S. M. V., Dammur, A., & Narasimhamurthy, A. (2023). Traffic Flow Prediction and Identifying Critical Node Networks Using Attention-based Spectral Temporal Graph Neural Network. *International Journal of Intelligent Engineering & Systems*, *16*(6). [CrossRef]

[50] Wang, Y., Jing, C., Xu, S., & Guo, T. (2022). Attention based spatiotemporal graph attention networks for traffic flow forecasting. *Information Sciences*, *607*, 869–883. [CrossRef]

[51] Reza, S., Ferreira, M. C., Machado, J. J., & Tavares, J. M. R. (2022). A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Systems with Applications*, *202*, 117275. [CrossRef]

[52] Geng, Z., Xu, J., Wu, R., Zhao, C., Wang, J., Li, Y., & Zhang, C. (2024). STGAFormer: Spatial–temporal gated attention transformer based graph neural network for traffic flow forecasting. *Information Fusion*, *105*, 102228. [CrossRef]

[53] Sun, B., Sun, T., Zhang, Y., & Jiao, P. (2020). Urban traffic flow online prediction based on multi-component attention mechanism. *IET intelligent transport systems*, *14*(10), 1249-1258. [CrossRef]

[54] Do, L. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation research part C: emerging technologies*, *108*, 12-28. [CrossRef]

[55] Xiong, C., Merity, S., & Socher, R. (2016, June). Dynamic memory networks for visual and textual question answering. In *International conference on machine learning* (pp. 2397-2406). PMLR. [CrossRef]

[56] Kumar, R., Panwar, R., & Chaurasiya, V. K. (2024). Urban traffic forecasting using attention based model with GCN and GRU. *Multimedia Tools and Applications*, *83*(16), 47751–47774. [CrossRef]

[57] Hussain, B., Afzal, M. K., Ahmad, S., & Mostafa, A. M. (2021). Intelligent traffic flow prediction using optimized GRU model. *IEEE Access*, *9*, 100736-100746. [CrossRef]

[58] Khodabandelou, G., Katranji, M., Kraiem, S., Kheriji, W., & HadjSelem, F. (2019, October). Attention-based gated recurrent unit for links traffic speed forecasting. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 2577-2583). IEEE. [CrossRef]

[59] Zhang, Q., Zhou, L., Su, Y., Xia, H., & Xu, B. (2023). Gated Recurrent Unit Embedded with Dual Spatial Convolution for Long-Term Traffic Flow Prediction. *ISPRS International Journal of Geo-Information*, *12*(9), 366. [CrossRef]

[60] Show, A. (2015). Tell: Neural image caption generation with visual attention kelvin xu. *Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio arXiv (2015-02-10) https://arxiv.org/abs/1502.03044 v3*.

[61] Yang, Z., Yuan, Y., Wu, Y., Cohen, W. W., & Salakhutdinov, R. R. (2016). Review networks for caption generation. *Advances in neural information processing systems*, *29*. [CrossRef]

[62] Huang, L., Wang, W., Chen, J., & Wei, X. Y. (2019). Attention on attention for image captioning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4634-4643). [CrossRef]

[63] Xiao, X., Wang, L., Ding, K., Xiang, S., & Pan, C. (2019). Deep hierarchical encoder–decoder network for image captioning. *IEEE Transactions on Multimedia*, *21*(11), 2942-2956. [CrossRef]

[64] Bahdanau, D., Chorowski, J., Serdyuk, D., Brakel, P., &

Bengio, Y. (2016, March). End-to-end attention-based large vocabulary speech recognition. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4945-4949). IEEE. [CrossRef]

[65] Li, J. (2022). Recent advances in end-to-end automatic speech recognition. *APSIPA Transactions on Signal and Information Processing, 11*(1). [CrossRef]

[66] Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779*.

[67] Brauwers, G., & Frasincar, F. (2021). A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering, 35*(4), 3279-3298. [CrossRef]

[68] Chan, W., Jaitly, N., Le, Q., & Vinyals, O. (2016, March). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4960-4964). IEEE. [CrossRef]

[69] Subakan, C., Ravanelli, M., Cornell, S., Bronzi, M., & Zhong, J. (2021, June). Attention is all you need in speech separation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 21-25). IEEE. [CrossRef]

[70] Subakan, C., Ravanelli, M., Cornell, S., Grondin, F., & Bronzi, M. (2023). Exploring self-attention mechanisms for speech separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 31*, 2169-2180. [CrossRef]

[71] Wang, D., & Chen, J. (2018). Supervised speech separation based on deep learning: An overview. *IEEE/ACM transactions on audio, speech, and language processing, 26*(10), 1702-1726. [CrossRef]

[72] An, J., Guo, L., Liu, W., Fu, Z., Ren, P., Liu, X., & Li, T. (2021). IGAGCN: Information geometry and attention-based spatiotemporal graph convolutional networks for traffic flow prediction. *Neural Networks, 143*, 355-367. [CrossRef]

[73] Tang, C., Sun, J., Sun, Y., Peng, M., & Gan, N. (2020). A general traffic flow prediction approach based on spatial-temporal graph attention. *IEEE Access, 8*, 153731-153741. [CrossRef]

[74] Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., ... & Li, H. (2019). T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE transactions on intelligent transportation systems, 21*(9), 3848-3858. [CrossRef]

[75] Ni, Q., & Zhang, M. (2022). STGMN: A gated multi-graph convolutional network framework for traffic flow prediction. *Applied Intelligence, 52*(13), 15026-15039. [CrossRef]

[76] Zheng, Y., Wang, S., Dong, C., Li, W., Zheng, W., & Yu, J. (2022). Urban road traffic flow prediction: A graph convolutional network embedded with wavelet decomposition and attention mechanism. *Physica A: Statistical Mechanics and its Applications, 608*, 128274. [CrossRef]

[77] Künas, C. A. (2023). Optimizing machine learning models training in the cloud.

[78] Miglani, A., & Kumar, N. (2019). Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Vehicular Communications, 20*, 100184. [CrossRef]

[79] Chen, Z., Wu, B., Li, B., & Ruan, H. (2021). Expressway exit traffic flow prediction for ETC and MTC charging system based on entry traffic flows and LSTM model. *IEEE Access, 9*, 54613-54624. [CrossRef]

[80] Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., ... & Yu, J. (2020, April). Traffic flow prediction via spatial temporal graph neural network. In *Proceedings of the web conference 2020* (pp. 1082-1092). [CrossRef]

[81] Hu, N., Zhang, D., Xie, K., Liang, W., Li, K. C., & Zomaya, A. Y. (2024). Dynamic multi-scale spatial–temporal graph convolutional network for traffic flow prediction. *Future Generation Computer Systems, 158*, 323-332. [CrossRef]

[82] Zhang, X., Xu, Y., & Shao, Y. (2022). Forecasting traffic flow with spatial–temporal convolutional graph attention networks. *Neural Computing and Applications, 34*(18), 15457-15479. [CrossRef]

[83] Ramsamy, E. (2006). *World Bank and urban development: From projects to policy*. Routledge.

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

[85] Wang, C., & Ye, Z. (2016). Traffic flow forecasting based on a hybrid model. *Journal of Intelligent Transportation Systems, 20*(5), 428-437. [CrossRef]

[86] Yao, B., Wang, Z., Zhang, M., Hu, P., & Yan, X. (2016, April). Hybrid model for prediction of real-time traffic flow. In *Proceedings of the Institution of Civil Engineers-Transport* (Vol. 169, No. 2, pp. 88-96). Thomas Telford Ltd. [CrossRef]

[87] Zheng, H., Lin, F., Feng, X., & Chen, Y. (2020). A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems, 22*(11), 6910-6920. [CrossRef]

[88] Zhang, Y., Zhang, Y., & Haghani, A. (2014). A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transportation Research Part C: Emerging Technologies, 43*, 65-78. [CrossRef]

[89] Kato, T., Kim, Y., Suzuki, T., & Okuma, S. (2005). Model predictive control of traffic flow based on hybrid system modeling. *IEICE transactions on fundamentals of electronics, communications and computer sciences, 88*(2), 549-560. [CrossRef]

[90] Ma, T., Antoniou, C., & Toledo, T. (2020). Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies, 111*, 352-372. [CrossRef]

[91] Moretti, F., Pizzuti, S., Panzieri, S., & Annunziato, M. (2015). Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. *Neurocomputing, 167*, 3-7. [CrossRef]

[92] Dimitriou, L., Tsekeris, T., & Stathopoulos, A. (2008). Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. *Transportation Research Part C: Emerging Technologies, 16*(5), 554-573. [CrossRef]

[93] Bi, J., Zhang, X., Yuan, H., Zhang, J., & Zhou, M. (2021). A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM. *IEEE Transactions on Automation Science and Engineering, 19*(3), 1869-1879. [CrossRef]

[94] Awan, N., Ali, A., Khan, F., Zakarya, M., Alturki, R., Kundi, M., Alshehri, M. D., & Haleem, M. (2021). Modeling dynamic spatiotemporal correlations for urban traffic flows prediction. *IEEE Access, 9*, 26502–26511. [CrossRef]

[95] Chan, K. Y., Dillon, T. S., Singh, J., & Chang, E. (2011). Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm. *IEEE Transactions on Intelligent Transportation Systems, 13*(2), 644-654. [CrossRef]

[96] Duan, P., Mao, G., Liang, W., & Zhang, D. (2018). A unified spatiotemporal model for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems, 20*(9), 3212-3223. [CrossRef]

[97] Liu, M., Zhu, T., Ye, J., Meng, Q., Sun, L., & Du, B. (2023). Spatio-temporal autoencoder for traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems, 24*(5), 5516-5526. [CrossRef]

[98] Zhang, J., Zheng, Y., Sun, J., & Qi, D. (2019). Flow prediction in spatiotemporal networks based on multitask deep learning. *IEEE Transactions on Knowledge and Data Engineering, 32*(3), 468-478. [CrossRef]

[99] Cao, S., Wu, L., Wu, J., Wu, D., & Li, Q. (2022). A spatiotemporal sequence-to-sequence network for traffic flow prediction. *Information Sciences, 610*, 185-203. [CrossRef]

[100] Liu, Q., Chen, C., Zhang, Y., & Hu, Z. (2011). Feature selection for support vector machines with RBF kernel. *Artificial Intelligence Review, 36*(2), 99-115. [CrossRef]

[101] Sun, S., Zhang, C., & Yu, G. (2006). A Bayesian network approach to traffic flow forecasting. *IEEE Transactions on intelligent transportation systems, 7*(1), 124-132. [CrossRef]

[102] Ou, J., Xia, J., Wu, Y. J., & Rao, W. (2017). Short-term traffic flow forecasting for urban roads using data-driven feature selection strategy and bias-corrected random forests. *Transportation Research Record, 2645*(1), 157-167. [CrossRef]

[103] Yang, H., Jiang, J., Zhao, Z., Pan, R., & Tao, S. (2024). STVANet: A spatiotemporal visual attention framework with large kernel attention mechanism for citywide traffic dynamics prediction. *Expert Systems with Applications, 254*, 124466. [CrossRef]

[104] Ye, B. L., Zhang, M., Li, L., Liu, C., & Wu, W. (2024). A Survey of Traffic Flow Prediction Methods Based on Long Short-Term Memory Networks. *IEEE Intelligent Transportation Systems Magazine*, 16(5), 87-112. [CrossRef]

[105] Hussain, B., Afzal, M. K., Anjum, S., Rao, I., & Kim, B. S. (2023). A novel graph convolutional gated recurrent unit framework for network-based traffic prediction. *IEEE Access, 11*, 130102-130118. [CrossRef]

[106] Yang, L., Shami, A., Stevens, G., & De Rusett, S. (2022, December). LCCDE: A decision-based ensemble framework for intrusion detection in the Internet of Vehicles. In *GLOBECOM 2022-2022 IEEE Global Communications Conference* (pp. 3545-3550). IEEE. [CrossRef]

[107] Anitha, P., Vimala, H. S., & Shreyas, J. (2024). Comprehensive review on congestion detection, alleviation, and control for IoT networks. Journal of Network and Computer Applications, 221, 103749. [CrossRef]

[108] Ali, A., Qureshi, M. A., Shiraz, M., & Shamim, A. (2021). Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control. *Sustainable Computing: Informatics and Systems, 32*, 100608. [CrossRef]

[109] Garg, S., Mehrotra, D., Pandey, H. M., & Pandey, S. (2021). Accessible review of internet of vehicle models for intelligent transportation and research gaps for potential future directions. *Peer-to-Peer Networking and Applications, 14*(2), 978-1005. [CrossRef]

Riaz Ali is a distinguished CAS-TWAS PhD President's Fellow in Computer Applied Technology (Data Mining) with a focus on Meteorology. He has authored over 12 publications in renowned journals and conferences, including AGU Earth and Space Science, Atmospheric Research, Sustainability, and others. His research primarily centers on the quantitative forecasting of rainfall and storm surges associated with both landfalling and offshore tropical cyclones along the Guangdong and Hong Kong coasts. His broader academic interests include the spatiotemporal analysis of typhoon hazards, the application of artificial intelligence in weather forecasting, and natural hazard and risk assessment. (Email: riazbabuzai@sjtu.edu.cn)

Ahmad Ali received the master's and Ph.D. degrees from Shanghai Jiao Tong University, China, where he developed a solid foundation in computer science and engineering. Currently, he is a Postdoctoral Researcher with the College of Mechatronics and Control Engineering, Shenzhen University, China. In addition to his reviewing activities, he has been involved in several cutting-edge research projects. His work has led to numerous publications in top-tier journals and conferences, showcasing his expertise and contributions to the field. He is passionate about leveraging technology to solve real-world problems, particularly in urban environments where efficient data processing and intelligent decision-making are crucial. His research interests include deep learning, big data analytics, data mining, urban computing, cloud computing, and fog computing. He has reviewed over a thousand research articles for prestigious academic journals, such as Information Sciences, Information Fusion, Neural Networks, IEEE Internet of Things Journal, Neural Computing and Applications, Multimedia Tools and Application, Wireless Networks, and IEEE Access. (Email: ahmadali@szu.edu.cn)

Hafiz Muhammad Yasir Naeem received the B.S. degree in Electrical Engineering from Bahria University Islamabad, Pakistan in 2013, M.S. degree in control systems from University of Engineering and Technology (UET), Taxila, Pakistan, in 2016 and Ph.D. degree from the Capital University of Science and Technology, Islamabad, Pakistan in 2023. His research interests include energy-efficient optimization strategies for EVs, Model Predictive Control and optimal control. Currently, He is a postdoctoral fellow at the College of Mechatronics and Control Engineering, and the College of Computer Science and Software Engineering at Shenzhen University, Shenzhen, China. (Email: yasirnaeem945@gmail.com)

Dr. Mujtaba Asad received the Ph.D. degree in control science and engineering from Shanghai Jiao Tong University in 2021. He is currently working as a Research Fellow with the Department of Automation, Shanghai Jiao Tong University. His research interests include image and video anomaly detection, human behavior analysis, and multimodal image fusion. (Email: asadmujtaba@sjtu.edu.cn)

Tamam Alsarhan (Member, IEEE) received the Ph.D. degree from Shanghai Jiao Tong University, specializing in computer vision, and deep learning. During her academic career, she has contributed significantly to the computer vision area. She was a Visiting Researcher at Khalifa University. She is currently an Assistant Professor with The University of Jordan. Her research interests include computer vision, human action recognition, machine learning, and pattern recognition. In addition to her research, she is dedicated to teaching and mentoring students, aiming to inspire the next generation of scientists and researchers. (Email: t_alsarhan@ju.edu.jo)

Dr. Md. Belal Bin Heyat received his B.Tech in Electronics and Instrumentation (EI) and M.Tech in Electronics Circuit and System (ECS) from Integral University, Lucknow, India, in 2014 and 2016. He completed his Ph.D. in Electronic Science and Technology from the University of Electronic Science and Technology of China (UESTC), Chengdu, China. During his Ph.D., he won nine awards and served as a Research Associate, Vice-Chairman of the UESTC Country League, and Country Representative for India. He worked as a Postdoctoral Fellow at the IoT Research Center, College of Computer Science and Software Engineering, Shenzhen University, China, from 2021 to 2023. Currently, he is a Postdoctoral Fellow at the CenBRAIN Neurotech Center of Excellence, Westlake University, China. He is also a visiting postdoctoral researcher at the Center for Visual Information Technology (CVEST), IIIT Hyderabad, India, and a faculty member in the Department of Science and Engineering at Novel Global Community Educational Foundation, NSW, Australia. Dr. Heyat's research focuses on detecting diseases such as sleep, neurological, and psycho-neurological disorders, as well as cardiovascular conditions. He specializes in signal processing and medical machine learning. He is a Guest Editor for journals like Life, Journal of Integrative Neuroscience, and Applied Sciences. He also reviews research for publishers such as Oxford University Press, IEEE, Elsevier, Hindawi, Wiley, Springer, MDPI, and Bentham Science. Dr. Heyat has published more than 70 articles (h-index: 25) in respected international journals and conferences. His work aims to improve healthcare using AI, advanced signal processing techniques, and traditional system of medicine. (Email: belalheyat@gmail.com)