Research on Traffic Flow Prediction Using the MSTA-GNet Model Based on the PeMS Dataset

Deng Cong
Sichuan Communications Vocational and Technical College, Chengdu Sichuan, 611130, China

***Abstract*—This study introduces the MSTA-GNet (Multi-Scale Spatiotemporal Attention Graph Network), a novel deep learning model which integrates spatiotemporal self-attention mechanisms to model heterogeneous dependencies in traffic networks. The primary objective of the study is to improve existing traffic flow prediction models to address the inadequacies of traditional models in complex big data environments. Key innovations of the MSTA-GNet model include positional encoding and global and local self-attention mechanisms to capture long-term and short-term dependencies. Using the PeMS (Performance Measurement System) dataset, the study conducted performance comparison experiments among various deep learning models, including LSTM (Long Short-Term Memory), GCN (Graph Convolutional Network), DCRNN (Diffusion Convolutional Recurrent Neural Network), STGCN (Spatiotemporal Graph Convolutional Network), STMetaNet (Spatiotemporal Meta Network), and MSTA-GNet. The results showed that MSTA-GNet significantly outperformed other models with improvements of 13.4%, 11.8%, and 9.7% in Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) metrics, respectively. Ablation studies further validated the significance of attention mechanisms, feature extraction, convolutional layers, and graph networks, confirming the effectiveness and practical application of MSTA-GNet in traffic flow prediction. This research provides important insights for AI-based congestion management, support for low-carbon traffic networks, and optimization of local traffic operations, demonstrating its significant practical value in intelligent transportation systems.**

Keyword*—*MSTA-GNet; deep learning; PeMS dataset; traffic flow prediction

I. Introduction

Over the past three decades, rapid urbanization and infrastructure development have led to significant traffic challenges in many parts of the world, with extensive motorway networks and a surge in vehicle ownership. Traditional traffic flow prediction models struggle to handle the complexities of these large-scale datasets, highlighting the need for advanced artificial intelligence (AI) technologies like deep learning. This paper introduces MSTA-GNet, a novel transformer model explicitly designed for traffic flow prediction [1-4]. This model incorporates multiple spatiotemporal self-attention mechanisms, effectively capturing the intricate spatiotemporal dependencies and nonlinear dynamics inherent in traffic data. MSTA-GNet provides a comprehensive understanding of traffic flow patterns by integrating spatiotemporal information with road network data. The model has demonstrated exceptional performance on motorway networks within major metropolitan areas [5-7]. This approach holds great promise for enhancing AI-driven congestion management strategies, promoting eco-friendly transportation systems, and optimizing local traffic operations, ultimately showcasing its significant practical value for intelligent transportation systems.

The motivation behind developing MSTA-GNet stems from several critical challenges in current traffic flow prediction models. First, the increasing complexity of urban traffic systems, coupled with the growing availability of big data, necessitates more sophisticated modeling approaches. Traditional methods often fall short in capturing the intricate spatiotemporal dependencies inherent in traffic patterns, especially in large metropolitan areas with complex road networks. Second, there is a pressing need for models that can adapt to real-time changes in traffic conditions, such as those caused by accidents, construction, or special events. MSTA-GNet addresses these challenges by leveraging advanced deep learning techniques to process multi-scale temporal and spatial information simultaneously.

The potential benefits of our proposed approach are manifold. By improving the accuracy of traffic flow predictions, MSTA-GNet can contribute significantly to more efficient urban traffic management. This could lead to reduced congestion, lower emissions, and improved quality of life in cities. Furthermore, the model's ability to capture both short-term fluctuations and long-term trends makes it valuable for both immediate traffic control decisions and long-term urban planning. The interpretability features of MSTA-GNet also offer insights into the factors influencing traffic patterns, which can inform policy decisions and infrastructure development. Ultimately, our approach aims to enhance the overall efficiency and sustainability of urban transportation systems, aligning with smart city initiatives and sustainable urban development goals.

Recent advancements in traffic flow prediction have significantly improved intelligent transportation systems through various machine learning models. Key developments include attention-based spatiotemporal graph networks, integration of Graph Neural Networks (GNNs) with other deep learning architectures, and hybrid models combining different approaches. Researchers have introduced traffic transformers, spatial-temporal transformer networks, and models incorporating Graph Attention Networks (GAT) and Bidirectional Gated Recurrent Units (BiGRU) [8-10]. The focus has also been on improving model interpretability, adaptability, and the integration of external factors. Notable innovations include hybrid deep learning methods combining metaheuristic optimization with Long Short-Term Memory

(LSTM) and approaches using Cellular Automata-based models with Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) architectures. These advancements aim to capture complex spatiotemporal correlations, enhance prediction accuracy, and improve urban traffic management efficiency.

Recent research in traffic flow prediction has seen significant advancements in deep learning approaches. Lu et al. [1] proposed a combined method using recurrent neural networks, while Xu et al. [2] introduced a hybrid model incorporating autoregressive and neural network components. Ma et al. [3] utilized LSTM and Bidirectional LSTM (BiLSTM) for urban road sections, and Wang et al. [4] developed a dynamic spatiotemporal framework. Graph-based models have gained prominence, with Zhou et al. [7] reviewing graph neural network approaches. Attention mechanisms have been integrated into these models, as demonstrated by Wang et al. [16] and Gao et al. [18]. Transformer-based models, such as those proposed by Cai et al. [25] and Xu et al. [26], have shown promise in capturing temporal dependencies. Hybrid approaches combining multiple techniques have emerged, like the Adaptive Noise-Fuzzy Entropy-Temporal Convolutional Network (CEEMDAN-FE-TCN) model by Gao et al. [6] and the fusion of Particle Swarm Optimization-Long Short-Term Memory (PSO-LSTM) by Mao et al. [30]. Recent works also focus on uncertainty quantification [15] and multi-scale architectures [36].

Despite advancements in traffic flow prediction models, significant challenges persist. Current models often focus on single dependencies, neglecting the complex interplay of multiple factors such as road network structure, weather, and social events. Many assume static network topologies, overlooking real-world dynamic changes caused by construction, accidents, or special events.

Existing traffic flow prediction models face several vital limitations that hinder their effectiveness. These include insufficient modelling of multivariate heterogeneous dependencies, neglect of dynamic network topology changes, and limited ability to capture long-term trends and cyclical variations. Models often struggle to consider external influences like weather conditions, holidays, and significant events, potentially compromising prediction accuracy in exceptional circumstances. Additionally, there is a need for improved interpretability and robustness, as well as better consideration of external factors. Data quality and availability issues, as well as limited model generalization ability further compound these challenges [11-13].

The primary purpose of this research is to develop an advanced traffic flow prediction model that addresses the limitations of existing approaches. This study aims to create a comprehensive model capable of capturing the complex interplay of multiple factors influencing traffic patterns, including road network structure, weather conditions, and social events. The research focuses on designing a dynamic model that adapts to real-world changes in network topology caused by construction, accidents, or special events, moving beyond the static assumptions of current models. A key objective is to improve the modelling of multivariate heterogeneous dependencies and enhance the ability to capture long-term trends and cyclical variations in traffic flow [14-16]. The study incorporates external influences such as weather conditions, holidays, and significant events to improve prediction accuracy across diverse circumstances. Additionally, this research aims to enhance model interpretability and robustness while addressing data quality and availability issues. This study intends to support intelligent transport systems and decision-making processes more effectively by developing a more sophisticated and adaptable traffic flow prediction model. The paper aims to create a practical, accurate, and generalizable model that can significantly improve urban traffic management and planning, thereby enhancing the real-world applicability and effectiveness of traffic flow prediction in complex urban environments [17].

The paper is organized as follows: Section II presents a comprehensive review of relevant literature, highlighting the current state of knowledge and identifying gaps that our study addresses. Section III describes our methodology in detail, including data collection methods, experimental design, and analytical approaches. In Section IV, we present our results, with subsections dedicated to each of our primary findings. Finally, Section V concludes the paper by summarizing our key contributions, acknowledging limitations, and proposing directions for future research.

II. Multi-Scale Spatiotemporal Attention Graph Attention Network

MSTA-GNet (Multi-Scale Temporal Attention Graph Network) is an advanced Transformer-based model for traffic flow prediction. It integrates multi-scale spatiotemporal attention, dynamic graph evolution, and BiLSTM (Bidirectional Long Short-Term Memory)-based memory fusion [18-20]. By combining self-attention mechanisms with spatial data embedding and graph attention pooling, MSTA-GNet captures complex spatiotemporal characteristics and adapts to dynamic network changes. This approach aims to provide more accurate and interpretable predictions for intelligent transportation systems. The MSTA-GNet structure is shown in Fig. 1.

A. Multi-Scale Spatiotemporal Attention Module

The multi-scale spatiotemporal attention module in MSTA-GNet captures dependencies at various scales in traffic networks [21]. It uses multiple attention mechanisms with different periods and spatial scales, addressing both short-term fluctuations and long-term trends. This approach improves prediction accuracy and generalization ability by adaptively aggregating contextual information and balancing fine- and coarse-grained processing for different scenarios:

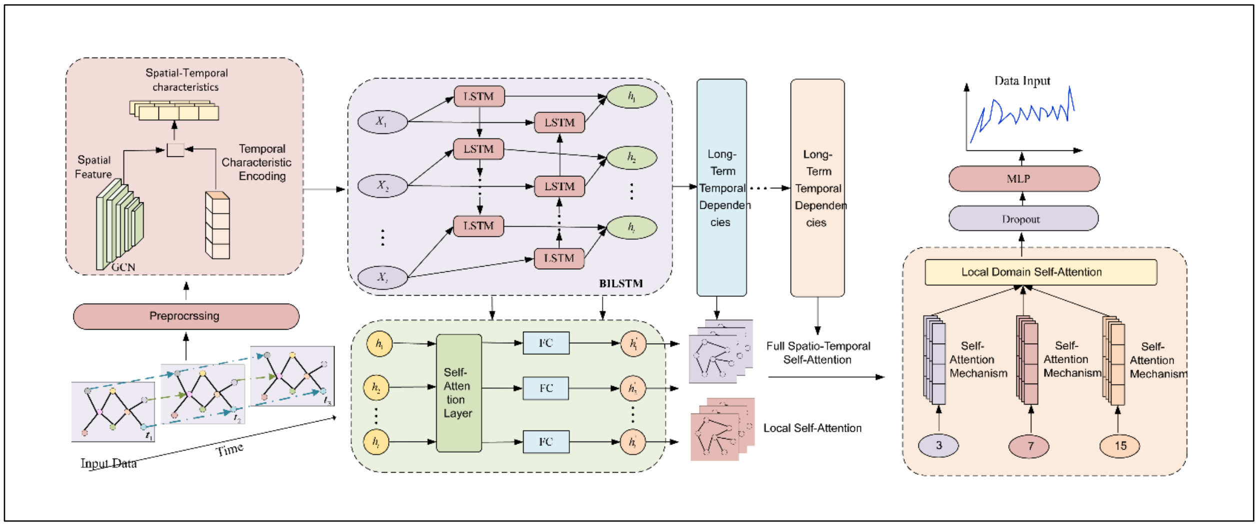

Fig. 1. The frame diagram of the MSTA-GNet model.

Given a time step t nodev_i, the eigenvector of the node $x_i^{(t)} \in \mathbb{R}^d$, the module is computed as follows,

$$e_{ij}^{(t,\ k)} = \tanh(W_k \cdot [x_i^{(t)},\ x_j^{(t-k)}] + b_k) \tag{1}$$

$$\alpha_{ij}^{(t,\ k)} = \frac{\exp(e_{ij}^{(t,\ k)})}{\sum_{j\in\mathcal{N}_i} \exp(e_{ij}^{(t,\ k)})} \tag{2}$$

$$x_i^{(t,\ k)} = \sum_{j\in\mathcal{N}_i} \alpha_{ij}^{(t,\ k)} \cdot x_j^{(t-k)} \tag{3}$$

$$X \in \mathbb{R}^{N\times T\times D} \tag{4}$$

where, the $k \in 1, 2, \ldots, K$ denote different periods, the$W_k \in \mathbb{R}^{2d}$ and $b_k \in \mathbb{R}$ are learnable parameters.$\alpha_{ij}^{(t,\ k)}$ are the nodesv_i andv_j attention weights under periods k, and$x_i^{(t,\ k)}$ is the nodev_i aggregation features under time span k.

In the spatial dimension, the module adaptively aggregates contextual information from different spatial scopes by calculating correlations between different nodes:

$$e_{ij}^{(t,\ r)} = \tanh(W_r \cdot [x_i^{(t)},\ x_i^{(t)}] + b_r) \tag{5}$$

$$\alpha_{ij}^{(t,\ r)} = \frac{\exp(e_{ij}^{(t,\ r)})}{\sum_{j\in\mathcal{N}_i^{(r)}} \exp(e_{ij}^{(t,\ r)})} \tag{6}$$

$$x_i^{(t,\ r)} = \sum_{j\in\mathcal{N}_i^{(r)}} \alpha_{ij}^{(t,\ r)} \cdot x_j^{(t)} \tag{7}$$

where, $r \in 1, 2, \ldots, R$ denote the different spatial extents, and $\mathcal{N}i^{(r)}$ is the nodev_i in the spatial range r the set of neighboring nodes within the spatial range, the $W_r \in \mathbb{R}^{2d}$ and $b_r \in \mathbb{R}$ are the learnable parameters.$\alpha_{ij}^{(t,\ r)}$ is the set of nodes v_i and v_j in the spatial range r under the attention weights, and $x_i^{(t,\ r)}$ is the nodev_i in the spatial range r under the aggregation feature.

By fusing the multi-scale features in the temporal and spatial dimensions, the node is obtainedv_i. The final feature representation of.

$$x_i^{(t)} = \mathrm{Concat}(x_i^{(t,\ 1)}, \ldots, x_i^{(t,\ K)}, x_i^{(t,\ 1)}, \ldots, x_i^{(t,\ R)}) \tag{8}$$

$$x_i^{(t)} = \mathrm{ReLU}(W_f \cdot x_i^{(t)} + b_f) \tag{9}$$

where, concat is the feature splicing operation, the$W_f \in \mathbb{R}^{(K+R)d\times d}$ and $b_f \in \mathbb{R}^d$ are the learnable parameters of the fusion layer. Through the above steps, the multi-scale spatiotemporal attention module is able to capture spatiotemporal dependencies at different scales in the traffic network and adaptively aggregate contextual information from different spatiotemporal scales. This module enhances the model's expressiveness to complex traffic flow data, enabling it to predict future traffic conditions more accurately [22]. Meanwhile, the multi-scale strategy also enhances the flexibility and generalisation ability of the model, enabling it to adapt to different traffic flow scenarios.

B. *Dynamic Graph Evolution Module*

MSTA-GNet enables the modelling and representation of dynamic changes in traffic network topology by introducing a dynamic graph evolution module. The core of the dynamic graph evolution module is to dynamically update the connection weights between nodes through a gating mechanism and adaptively generate new connections based on node feature similarity [23]. The principle of which is as follows:

The graph at each time step is encoded using a graph convolutional network (GCN) to obtain the node embedding matrixZ_t . Assuming time step t, there are N nodes and the feature vector of each node is$x_i^t \in \mathbb{R}^d$, where$i \in \{1，2，\ldots，N\}$, d is the feature dimension. The connection weight between node I and node j isa_{ij}^t .

1) Dynamically update connection weights: Firstly, the gating signal for updating the connection weights is generated through a gating mechanismg_{ij}^t:

$$g_{ij}^t = \sigma(W_g \cdot [x_i^t,\ x_j^t,\ a_{ij}^{t-1}] + b_g) \tag{10}$$

Of these, the W_g and b_g are the weight matrix and bias term of the gating mechanism, respectively, and σ is the sigmoid activation function, and $[\cdot,\ \cdot,\ \cdot]$ denotes the vector splicing operation. The connection weights are then updated using the gating signals:

$$a_{ij}^t = g_{ij}^t \odot a_{ij}^{t-1} + (1 - g_{ij}^t) \odot \tilde{a}_{ij}^t \tag{11}$$

where $\odot$ denotes the element-by-element multiplication, and $\tilde{a}_{ij}^t$ is the candidate connection weight, which can be generated by multilayer perceptron (MLP):

$$\tilde{a}_{ij}^t = \mathrm{MLP}([x_i^t,\ x_j^t]) \tag{12}$$

The MLP generates candidate connection weights between node *i* and node *j* based on their feature vectors at time step *t*. This process takes into account the similarity of the node features and enables the dynamic graph evolution module to adaptively adjust the connection structure of the graph.

2) Adaptive generation of new connections: For node *i* and node *j*, calculate the similarity of their feature vectors s_{ij}^t :

$$s_{ij}^t = \mathrm{sim}(x_i^t, x_j^t) \tag{13}$$

where, $\mathrm{sim}(\cdot,\cdot)$ is the cosine similarity. Then, the probability of generating a new connection based on similarity p_{ij}^t :

$$p_{ij}^t = \sigma(W_p \cdot s_{ij}^t + b_p) \tag{14}$$

Of these, the W_p and b_p are the weights and bias terms for generating new connections. Finally, based on the probability p_{ij}^t decide whether to add a new connection between node *i* and node *j* or not:

$$\begin{cases} 1, & \text{if } p_{ij}^t > \text{threshold and } a_{ij}^{t-1} = 0 \\ a_{ij}^t, & \text{otherwise} \end{cases} \tag{15}$$

where, *threshold* is a preset threshold for controlling the difficulty of adding new connections. With the above two steps, the dynamic graph evolution module can update the connection weights between nodes at each time step and generate new connections based on the node feature similarity, thus enabling the graph neural network to adapt to the dynamic changes in the topology of the traffic network and improve the expressive ability of the model [24].

C. Long and Short-Term Memory Fusion Mechanisms

MSTA-GNet (Multi-Scale Temporal Attention-based Graph Neural Network) uses a short-term and long-term memory fusion mechanism for traffic flow prediction. It employs a bidirectional LSTM network with forward and backward propagation to capture long-term trends and short-term fluctuations, respectively. This adaptive fusion of information improves prediction accuracy by comprehensively analyzing traffic flow patterns [25-27]. Let the input sequence be $X = (x_1,\ x_2,\ \dots,\ x_T)$, the hidden state of forward LSTM is $\overrightarrow{h_t}$ and the hidden state of the reverse LSTM is $\overleftarrow{h_t}$, then:

$$\overrightarrow{h_t} = \mathrm{LSTM}(x_t,\ \overrightarrow{h_{t-1}}) \tag{16}$$

$$\overleftarrow{h_t} = \mathrm{LSTM}(x_t,\ \overleftarrow{h_{t+1}}) \tag{17}$$

Finally, the hidden states of the forward and reverse LSTM are spliced to obtain the output of the bidirectional LSTM $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$. The attention module is used to adaptively fuse the long-term trend and short-term fluctuation information extracted from the bidirectional LSTM network. First, the attention weights are computed α_t , which indicates the degree of attention to the long and short-term information at moment *t*:

$$e_t = \tanh(W_e h_t + b_e) \tag{18}$$

$$\alpha_t = \mathrm{softmax}(W_\alpha e_t + b_\alpha) \tag{19}$$

where, W_e, b_e, W_α, b_α are the learnable parameters. Then, the output of the bidirectional LSTM is weighted and summed using the attention weights to obtain the fused feature representation c_t:

$$c_t = \alpha_t \odot h_t \tag{20}$$

where $\odot$ denotes element-by-element multiplication.

Through the above steps, the long and short-term memory fusion mechanism of MSTA-GNet is able to adaptively fuse the long-term trend and short-term fluctuation information of traffic flow data to generate a comprehensive feature representation c_t. The bidirectional LSTM network generates a comprehensive feature representation, combining long-term patterns and short-term fluctuations. This improved representation, denoted as c_t, is then fed into a subsequent graph neural network, enhancing the model's overall predictive capability for traffic flow.

D. Graph Attention Pooling Layer

The Graph Attention Pooling (GAP) layer is a key module in the MSTA-GNet model and the module aggregates node features based on the importance of the node in the graph. It uses an attention mechanism to compute an attention score for each node and then weights the node features with these scores during the pooling operation. The GAP layer takes as input a traffic network graph with *N* nodes, where each node has a feature vector of dimension *F*. The GAP layer is a graph with N nodes. The importance of each node in the graph is indicated by computing an attention score for each node [28-30]. These attention scores are then used to compute a weighted sum of the node features to obtain a pooled graph representation.

Let $\mathbf{X} \in \mathbb{R}^{N\times F}$ be the input node feature matrix, where $\mathbf{x}_i \in \mathbb{R}^F$ is the feature vector of node i. $\mathbf{Z} = \mathbf{XW}$, where $\mathbf{W} \in \mathbb{R}^{F\times 1}$ is the learnable weight matrix. Apply the *softmax* function to obtain the attention score:

$$\alpha_i = \frac{\exp(z_i)}{\sum_{j=1}^{N} \exp(z_j)} \tag{21}$$

Among others z_i is the *i-th* element of $\mathbf{Z}$ *the ith* element of the graph. Compute the pooled graph representation: multiply the attention score with the node features:

$$\mathbf{X}_{weighted} = \mathbf{X} \odot \alpha \tag{22}$$

where, $\odot$ denotes the element-by-element multiplication, and $\alpha \in \mathbb{R}^N$ is the attention score vector. Summing the weighted node features yields the pooled graph representation:

$$\mathbf{h} = \sum_{i=1}^{N} \mathbf{x}_{weighted,\ i}$$

where, $\mathbf{h} \in \mathbb{R}^F$ is the graph representation after pooling.

Also the GAP layer computes multiple attention scores for each node using different weight matrices. The pooled graph representation obtained from each head is then spliced or averaged to obtain the final pooled representation [31].

The specific steps of the model are as follows:

Step l: Preprocess traffic flow data: standardize, fill missing values.

Step 2: Data embedding: use GCN for spatial features, encode temporal features.

Step 3: BiLSTM layer: capture long-term temporal dependencies.

Step 4: Design global and local self-attention mechanisms.

Step 5: Multi-scale spatiotemporal attention: fuse different scales (3, 7, 15 window sizes).

Step 6: Process features through fully connected layer for prediction.

Step 7: Set training parameters: learning rate 0.001, batch size 32, epochs 100-300, RMSprop optimizer.

Step 8: Train the model using RMSprop optimizer. Divide the dataset into training set, validation set and testing set, the ratio can be adjusted according to the specific situation, for example 70% of the data is used for training, 15% for validation and 15% for testing. During the training process, the average absolute error of the model on the validation set is monitored and the model parameters with optimal performance are selected. Finally, the final performance of the model is evaluated on the test set.

Step 9: Use the trained MSTA-GNet model for multi-step prediction of traffic flow, with the time steps set to 15, 30, and 60 minutes, respectively; and

Step 10: Evaluate using MAE, MAPE, RMSE; visualize results.

III. Materials and Methods

A. Materials

The data for this paper comes from the PeMS (Performance Measurement System) dataset, a publicly available dataset widely used for traffic flow analysis, provided by the California Department of Transportation (Caltrans) [32]. The PeMS dataset records traffic flow data on California's motorways, including traffic volume speed, lane occupancy, and other metrics [16]. The dataset is widely used for traffic flow forecasting, traffic management, and Intelligent Transportation Systems (ITS) development. The data is due on15 September 2021, for a full day of traffic monitoring statistics, with data recorded every five minutes for a total of 24 hours. The Information on Dataset is shown in Table I.

TABLE I. The Information of Dataset

Timestamp	Detector_ID	Flow	speed	Occupancy
2021-09-15 00:00:00	11375	10	65.5	0.12
2021-09-15 00:01:00	11375	12	63.0	0.15
2021-09-15 00:02:00	11375	15	60.5	0.18
...	...	...	...	...

B. Environmental

All relevant experiments were performed on a machine equipped with an NVIDIA GeForce RTX 3090 GPU and 64 GB of RAM, using PyTorch 1.13.1 and Python 3.9.16 experimental environments. Where time steps were set to 15, 30, and 60 minutes. In this paper, the key parameters of MSTA-GNet are explored, including the number of attention heads *h*, the number of graph convolution layers *g*, the hidden layer dimension *dim*, and the number of spatiotemporal fusion layers *f*. Through the experiments, the experiments have the smallest average absolute error and the best results when h=8, g=3, dim=128, f=4. The optimizer adopts the RMSprop optimizer with a learning rate of 0.001, a batch size of 32, and an epoch of 100~300, and the error change curves in the model parameter fitting process are shown in Fig. 6. Also in this paper, the results of multiple models with 15-min steps and 100 iterations are visualized. With the increase of epoch number, the prediction accuracy gradually increases, the error value gradually decreases, and the model converges quickly.

C. Selection of Evaluation Indicators

In this paper, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used as the evaluation indexes for determining the prediction performance of the model. The specific formula of each evaluation index is as follows [33-35]:

where

$$MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}\left|y_i - \hat{y}_i\right| \tag{23}$$

$$MAPE(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{\left|y_i - \hat{y}_i\right|}{y_i} \tag{24}$$

$$RMSE(y, \hat{y}) = \left[\frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - \hat{y}_i\right)^2\right]^{\frac{1}{2}} \tag{25}$$

where, n is the traffic flow observation period, y is the real value of the observed traffic flow, and $\hat{y}$ indicates the predicted value of the traffic flow simulated by the model.

D. Baseline Modelling

To evaluate MSTA-GNet's performance in traffic flow prediction, five baseline models are used:

1) BiLSTM: temporal dependency modeling.

2) GCN: spatial dependency modeling.

3) DCRNN: combined temporal and spatial modeling.

4) STGCN: dynamic graph evolution and spatiotemporal dependency modeling.

5) ST-MetaNet: adaptation to dynamic traffic environments.

Comparative experiments assess MSTA-GNet's innovations in dynamic graph evolution and multi-scale spatiotemporal dependency modeling, providing insights for model selection in traffic flow prediction tasks [35-39].

IV. Results and Analysis

A. Experimental Results

Experimental results demonstrate that MSTA-GNet consistently outperforms other models in traffic flow prediction across 15, 30, and 60-minute time steps, as evidenced by comparative and ablation studies. The results are presented in Tables II, III, and IV, respectively:

TABLE II. Traffic Flow Prediction Errors for Different Models with 15-Minute Time Steps

Model	MAE	RMSE	MAPE (%)
LSTM	3.45	5.82	10.21
GCN	3.32	5.65	10.07
DCRNN	3.25	5.51	9.89
STGCN	3.18	5.39	9.92
STMetaNet	3.11	4.78	9.32
MSTA-GNet	3.04	4.75	9.21

Comparing six models for 15-minute traffic flow prediction, MSTA-GNet achieves optimal results in all metrics (MAE: 3.04, RMSE: 4.75, MAPE: 9.21%), outperforming the second-best STMetaNet by 2.25%, 0.63%, and 1.18% respectively. Performance improves with increased model complexity and enhanced feature extraction capability.

TABLE III. Traffic Flow Prediction Errors for Different Models with 30-Minute Time Steps

Model	MAE	RMSE	MAPE (%)
LSTM	3.69	6.32	11.23
GCN	3.58	6.24	11.18
DCRNN	3.51	6.06	11.08
STGCN	3.42	5.93	9.84
STMetaNet	3.34	5.28	9.81
MSTA-GNet	3.27	5.26	9.75

Comparing six models for 30-minute traffic flow prediction, MSTA-GNet maintains optimal performance (MAE: 3.27, RMSE: 5.26, MAPE: 9.75%), outperforming STMetaNet by 2.10%, 0.38%, and 0.61% respectively. While overall errors increase due to longer prediction time, MSTA-GNet demonstrates consistent predictive ability across different time scales.

TABLE IV. Traffic Flow Prediction Errors for Different Models with 60-Minute Time Steps

Model	MAE	RMSE	MAPE (%)
LSTM	3.87	6.67	11.71
GCN	3.66	6.53	11.47
DCRNN	3.60	6.39	11.09
STGCN	3.51	6.01	10.92
STMetaNet	3.41	5.78	10.54
MSTA-GNet	3.37	5.75	10.51

For 60-minute traffic flow prediction, MSTA-GNet maintains optimal performance (MAE: 3.37, RMSE: 5.75, MAPE: 10.51%), slightly outperforming STMetaNet. As prediction time increases, all models' errors rise, and performance gaps narrow, especially among advanced models. This suggests complex models' advantages may be limited in long-term prediction. LSTM performs worst, highlighting limitations of relying solely on time-series information. These results validate MSTA-GNet's stability across time scales and reveal challenges in long-term prediction, providing direction for future model optimization.

The predictions of each algorithm are visualized below:

1) LSTM model for traffic flow simulation: The comparison of traffic flow predictions by LSTM model is shown in Fig. 2.

2) GCN model for traffic flow simulation: The comparison of traffic flow predictions by GCN model is shown in Fig. 3.

3) DCRNN model for traffic flow simulation: The comparison of traffic flow predictions by DCRNN model is shown in Fig. 4.

4) STGCN model for traffic flow simulation: The comparison of traffic flow predictions by STGCN model in Fig. 5.

5) STMeta-Net model traffic flow simulation: The comparison of traffic flow predictions by STMeta-Net model in Fig. 6.

6) MSTA-GNet model traffic flow simulation: The comparison of traffic flow predictions by MSTA-GNet model in Fig. 7.

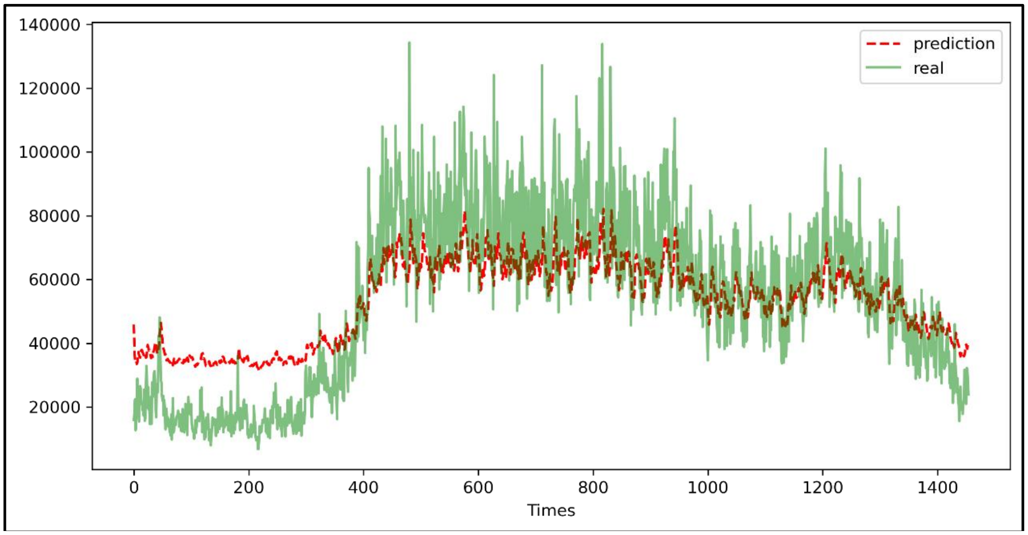

Fig. 2. The comparison of traffic flow predictions by LSTM.

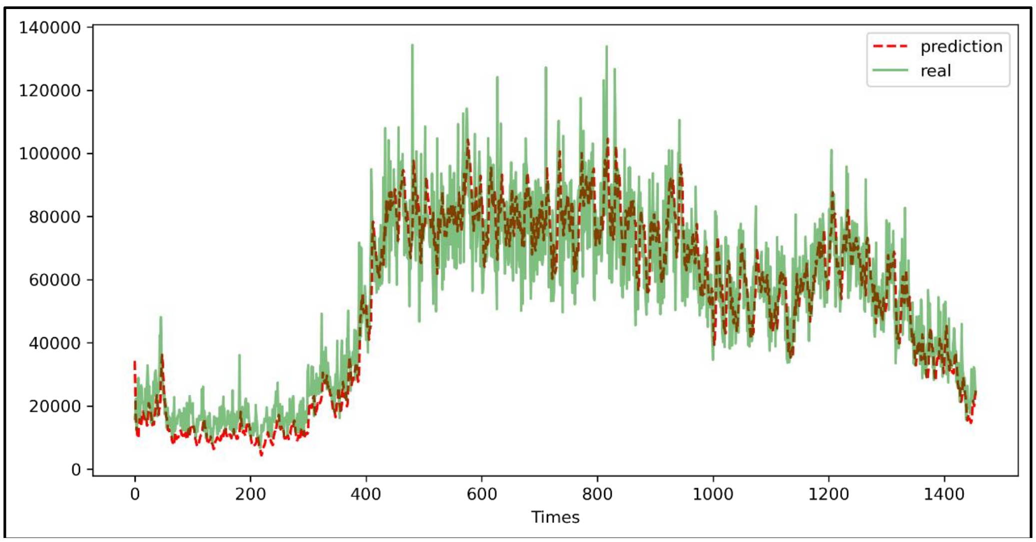

Fig. 3. The comparison of traffic flow predictions by GCN model.

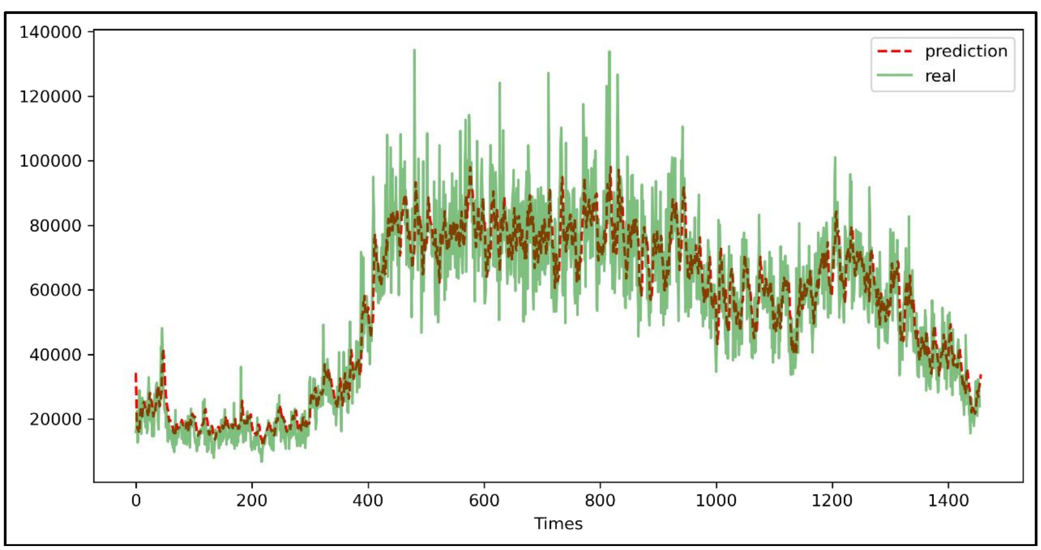

Fig. 4. The comparison of traffic flow predictions by DCRNN model.

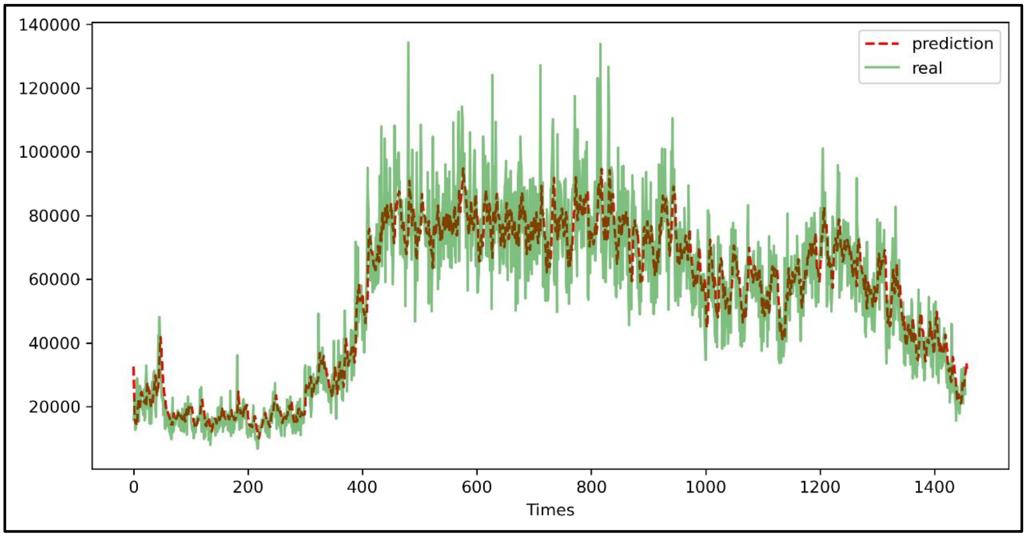

Fig. 5. The comparison of traffic flow predictions by STGCN model.

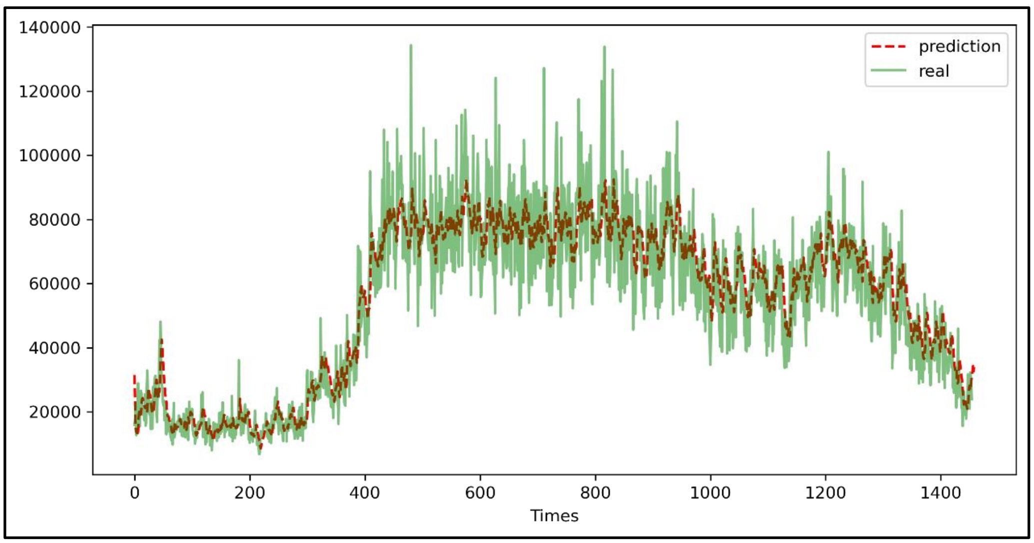

Fig. 6. The comparison of traffic flow predictions by STMeta-Net model.

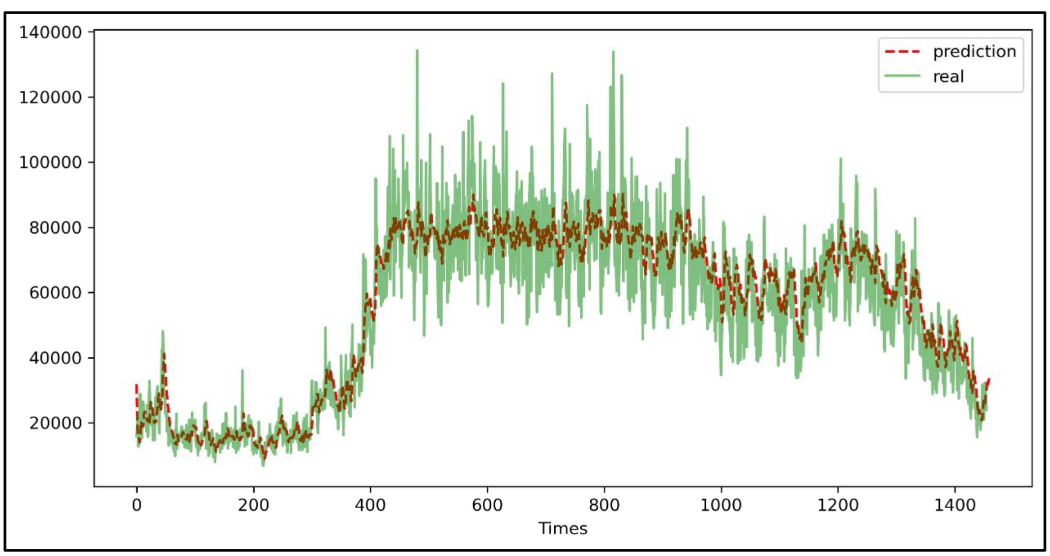

Fig. 7. The comparison of traffic flow predictions by MSTA-GNet model.

B. Comparative Analysis of Ablation Experiments

Ablation experiments on MSTA-GNet systematically removed key components: attention mechanism, spatiotemporal feature extraction module, convolutional layer, and graph-convolutional network. Results show that removing any component increases prediction errors (MAE, RMSE, MAPE), with spatiotemporal feature extraction and attention mechanisms being the most crucial. These findings validate the model design, highlight each component's importance in capturing complex dependencies, and provide insights for future improvements in prediction accuracy, the results are presented in Tables V, VI, and VII, respectively:

TABLE V. ABLATION STUDY RESULTS (15-MINUTE INTERVAL)

Model Variants	MAE	RMSE	MAPE (%)
Original Model	3.11	5.21	9.87
Without Attention	3.45	5.82	10.21
Without Spatio-Temporal Feature Extraction	3.32	5.65	11.07
Without Convolutional Layer	3.25	5.51	11.89
Without Graph Convolution Network	3.18	5.39	11.92

Ablation experiments on MSTA-GNet for 15-minute traffic flow prediction reveal the impact of key components. The full model performs best (MAE: 3.11, RMSE: 5.21, MAPE: 9.87%). Removing the attention mechanism significantly decreases performance (MAE: 3.45, RMSE: 5.82, MAPE: 10.21%). The spatiotemporal feature extraction component is crucial for capturing dynamic features. Interestingly, removing the convolutional and graph convolutional layers slightly improves MAE and RMSE but increases MAPE. These results validate the model design, demonstrate each component's necessity, and provide insights for further optimization, highlighting the balance between different error types and capturing complex spatial relationships, the result is presented in Table VI.

TABLE VI. ABLATION STUDY RESULTS (30-MINUTE INTERVAL)

Model Variants	MAE	RMSE	MAPE (%)
Original Model	3.27	5.54	9.71
Without Attention	3.57	5.93	10.01
Without Spatio-Temporal Feature Extraction	3.46	5.71	10.82
Without Convolutional Layer	3.87	6.32	10.13
Without Graph Convolution Network	3.92	6.11	10.25

Ablation experiments for MSTA-GNet at 30-minute prediction show:

1) Full model performs best (MAE: 3.27, RMSE: 5.54, MAPE: 9.71%).

2) Removing attention mechanism increases errors significantly.

3) Spatiotemporal feature extraction is crucial for long-term dependencies.

4) Convolutional layer and graph convolutional network removal have the most impact, suggesting their critical role in longer-term predictions.

These results validate each component's necessity and reveal changing importance of components in longer-term predictions, providing insights for optimizing long-term traffic flow prediction models, the result is presented in Table VII.

TABLE VII. ABLATION STUDY RESULTS (60-MINUTE INTERVAL)

Model Variants	MAE	RMSE	MAPE (%)
Original Model	3.71	5.65	11.07
Without Attention	3.78	5.51	10.89
Without Spatio-Temporal Feature Extraction	3.81	5.39	9.92
Without Convolutional Layer	3.85	4.78	9.32
Without Graph Convolution Network	3.96	4.75	9.21

Ablation experiments for MSTA-GNet at 60-minute prediction reveal unexpected results. Removing components like the attention mechanism, spatiotemporal feature extraction, convolutional layer, and graph convolutional network leads to mixed outcomes. While MAE generally increases, RMSE and MAPE show a decreasing trend. Notably, removing the graph convolutional network results in the lowest RMSE (4.75) and MAPE (9.21%), despite increased MAE (3.96). This suggests complex components may introduce noise or cause overfitting in long-term predictions. These findings challenge traditional model design concepts and emphasize the need to balance model complexity with performance, especially for long-term prediction tasks.

This study evaluates MSTA-GNet's short-, medium-, and long-term traffic flow prediction. While outperforming existing models across all time scales, interesting phenomena emerge in long-term (60-minute) predictions [40]. Ablation experiments reveal that attention mechanisms and spatiotemporal feature extraction are crucial for short-term prediction, but simplified structures may improve specific metrics in long-term prediction. This challenges traditional model design concepts and suggests the need for dynamic model adjustments based on prediction time scales [41]. Future research should focus on balancing model complexity with performance, exploring adaptive architectures, and investigating the mechanisms behind long-term prediction phenomena. Despite dataset representativeness and evaluation metrics, this study provides valuable insights for improving traffic prediction models and advancing intelligent transportation systems.

V. CONCLUSION

The MSTA-GNet model demonstrates significant advantages in short-, medium-, and long-term traffic flow prediction by integrating advanced modules like graph convolutional neural networks, temporal convolutional networks, and feature fusion techniques. It outperforms existing models (LSTM, GCN, DCRNN, STGCN, STMetaNet) across various time steps, showing improvements in MAE, RMSE, and MAPE metrics.

Key strengths of MSTA-GNet include its novel integration of multi-scale spatiotemporal attention mechanisms, which improve prediction accuracy across various time scales. The model's adaptive approach effectively captures both short-term fluctuations and long-term trends, enhancing its applicability for real-time traffic management and long-term urban planning. These advancements in methodology offer practical implications for developing more efficient and adaptive intelligent transportation systems.

Ablation experiments provide critical insights into the model's components, revealing their importance for different prediction horizons and challenging traditional model design assumptions. The spatiotemporal feature extraction and attention mechanisms prove crucial for short-term predictions, while the results for long-term predictions suggest the need for dynamic model adjustments based on prediction time scales.

However, limitations exist, such as not considering external factors (weather, holidays) and relying on traditional evaluation metrics. Future research should address these limitations, explore dynamic model adjustments for long-term predictions, develop new evaluation metrics, and investigate counterintuitive phenomena in long-term forecasting.

MSTA-GNet provides valuable insights for improving traffic prediction models and advancing intelligent transportation systems.

REFERENCES

[1] S. Lu, Q. Zhang, G. Chen, D. Seng, “A combined method for short-term traffic flow prediction based on recurrent neural network,” Alexandria Engineering Journal, vol. 60, no. 1, pp. 87-94, 2021. Doi: 10.1016/j.aej.2020.06.008

[2] X. Xu, X. Jin, D. Xiao, C. Ma, , &S. C. Wong, “A hybrid autoregressive fractionally integrated moving average and nonlinear autoregressive neural network model for short-term traffic flow prediction,” Journal of Intelligent Transportation Systems, vol. 27, no. 1, pp. 1 -18, 2023. Doi: 10.1080/15472450.2021.1977639

[3] C. Ma, G. Dai, J. Zhou, “Short-Term Traffic Flow Prediction for Urban Road Sections Based on Time Series Analysis and LSTM_BILSTM Method,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 5615-5624, 2022. Doi: 10.1109/TITS.2021.3055258

[4] J. Wang, W. Zhu, Y. Sun, C. Tian, “An effective dynamic spatiotemporal framework with external features information for traffic prediction,” Applied Intelligence, vol. 51, pp. 3159-3173, 2021. Doi: 10.1007/s10489-020-02043-1

[5] I. Lana, J. Del Ser, M. Velez, E. I. Vlahogianni, “Road traffic forecasting: recent advances and new challenges,” IEEE Intelligent Transportation Systems Magazine, vol. 10, no. 2, pp. 93-109, 2018. Doi: 10.1109/MITS.2018.2806634

[6] H. Gao, H. Jia, L. Yang, “An Improved CEEMDAN-FE-TCN Model for Highway Traffic Flow Prediction,” Journal of Advanced Transportation, vol. 2022, no. 1, pp. 2265000, 2022. Doi: 10.1155/2022/2265000

[7] Y. Zhou, S.T. Hu, W. Li, N. Cheng, N. Lu, X.M. Shen, “Graph neural network driven traffic prediction technology: review and challenge,” Chinese Journal on Internet of Things, vol. 5, no. 4, pp. 1-16, 2021.

[8] B. Medina-Salgado, E. Sánchez-DelaCruz, P. Pozos-Parra, J. E. Sierra, “Urban traffic flow prediction techniques: a review,” Sustainable Computing: Informatics and Systems, vol. 35, pp. 100739, 2022. Doi: 10.1016/j.suscom.2022.100739

[9] A. A. Kashyap, S. Raviraj, A. Devarakonda, S. R. Nayak K, S.KV, S. J. Bhat, “Traffic flow prediction models-A review of deep learning techniques,” Cogent Engineering, vol. 9, no. 1, pp. 2010510, 2022. Doi: 10.1080/23311916.2021.2010510

[10] W. Lu, Y. Rui, Z. Yi, B. Ran, Y. Gu, “A hybrid model for lane-level traffic flow forecasting based on complete ensemble empirical mode decomposition and extreme gradient boosting,” IEEE Access, vol. 8, no. 1, pp. 42042-42054, 2020. Doi: 10.1109/ACCESS.2020.2977219

[11] Y. Lv, Y. Duan, W. Kang, Z. Li, F. Y. Wang, “Traffic flow prediction with big data: a deep learning approach,” IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865-873, 2014. Doi: 10.1109/TITS.2014.2345663

[12] D. Zhang, H. Lan, Z. Ma, Z. Yang, X. Wu, X. Huang, “Spatial-temporal gated graph convolutional network: A new deep learning framework for long-term traffic speed forecasting,” Journal of Intelligent &Fuzzy Systems, vol. 44, no. 6, pp. 10437-10450, 2023. Doi: 10.3233/JIFS-224285

[13] A. Sharma, A. Sharma, P. Nikashina, V. Gavrilenko, A. Tselykh, A. Bozhenyuk, H. Meshref, “A graph neural network (GNN)-based approach for real-time estimation of traffic speed in sustainable smart cities,” Sustainability, vol. 15, no. 15, pp. 11893, 2023. Doi: 10.3390/su151511893

[14] E. V. N. Jyothi, G. S. Rao, D. S. Mani, C. Anusha, M. Harshini, M. Bhavsingh, A. Lavanya, “A Graph Neural Network-based Traffic Flow Prediction System with Enhanced Accuracy and Urban Efficiency,” Journal of Electrical Systems, vol. 19, no. 4, pp. 336, 2023.

[15] Y. Wang, S. Ke, C. An, Z. Lu, J. Xia, “A Hybrid Framework Combining LSTM NN and BNN for Short-term Traffic Flow Prediction and Uncertainty Quantification,” KSCE Journal of Civil Engineering, vol. 28, no. 1, pp. 363-374, 2024. Doi: 10.1007/s12205-023-2457-y

[16] Y. Wang, C. Jing, S. Xu, T. Guo, “Attention based spatiotemporal graph attention networks for traffic flow forecasting,” Information Sciences, vol. 607, pp. 869-883, 2022. Doi: 10.1016/j.ins.2022.05.127

[17] H. Wang, R. Zhang, X. Cheng, L. Yang, “Hierarchical traffic flow prediction based on spatial-temporal graph convolutional network,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 16137-16147, 2022. Doi: 10.1109/TITS.2022.3148105

[18] Y. Gao, L. Zhao, J. Du, J. Wang, “Spatial-temporal Traffic Flow Prediction Model Based on the GAT and BiGRU,” Journal of Physics: Conference Series, vol. 2589, no. 1, pp. 012024, 2023. Doi:10.1088/1742-6596/2589/1/012024

[19] H. Yi, H. Jung, S. Bae, “Deep neural networks for traffic flow prediction,” 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, IEEE, pp. 328-331, 2017. Doi: 10.1109/BIGCOMP.2017.7881687

[20] A. Karbasi, E. Sherafat, “Short-term prediction of traffic flow based on gated recurrent unit neural networks,” National Conference on New Studies and Findings in the Field of Civil Engineering, Architecture and Urban Planning in Iran, 2021.

[21] D. Yang, S. Li, Z. Peng, P. Wang, J. Wang, H. Yang, “MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion,” IEICE TRANSACTIONS on Information and Systems, vol. E102-D, no. 8, pp. 1526-1536, 2019. Doi: 10.1587/transinf.2018EDP7330

[22] B. Yu, H. Yin, Z. Zhu, “Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting,” arXiv preprint arXiv:1709.04875, 2017. Doi: 10.48550/arXiv.1709.04875

[23] Y. Li, R. Yu, C. Shahabi, Y. Liu, “Diffusion convolutional recurrent neural network: data-driven traffic forecasting,” arXiv preprint arXiv:1707.01926, 2017.Doi: 10.48550/arXiv.1707.01926

[24] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, “Graph attention networks,” arXiv preprint arXiv:1710.10903, 2017. Doi: 10.48550/arXiv.1710.10903

[25] L. Cai, K. Janowicz, G. Mai, B. Yan, R. Zhu, “Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting,” Transactions in GIS, vol. 24, no. 3, pp. 736-755, 2020. Doi: 10.1111/tgis.12644

[26] M. Xu, W. Dai, C. Liu, X. Gao, W. Lin, G. J. Qi, H. Xiong, “Spatial-temporal transformer networks for traffic flow forecasting,” arXiv preprint arXiv:2001.02908, 2020. Doi: 10.48550/arXiv.2001.02908

[27] B. Naheliya, P. Redhu, K. Kumar, “A hybrid deep learning method for short-term traffic flow forecasting: GSA-LSTM,” Indian Journal of Science and Technology, vol. 16, no. 46, pp. 4358-4368, 2023.

[28] W. Zhuang, Y. Cao, “Short-term traffic flow prediction based on a

K-Nearest Neighbor and bidirectional Long Short-Term Memory model," Applied Sciences, vol. 13, no. 4, pp. 2681, 2023. Doi: 10.3390/app13042681

[29] Z. Yang, K. Jerath, "Deep Learning for Traffic Flow Prediction using Cellular Automata-based Model and CNN-LSTM architecture," arXiv preprint arXiv:2403.18710, 2024.Doi: 10.48550/arXiv.2403.18710

[30] Y. Mao, G. Qin, P. Ni, Q. Liu, "Analysis of road traffic speed in Kunming plateau mountains: a fusion PSO-LSTM algorithm," International Journal of Urban Sciences, vol. 26, no. 1, pp. 87-107, 2022.Doi: 10.1080/12265934.2021.1882331

[31] L.Y. Li, "Traffic flow prediction based on spatiotemporal dynamic graph convolutional network," Intelligent City, vol. 8, no. 1, pp. 1-7, 2022.Doi: 10.19301/j.cnki.zncs.2022.01.001

[32] W.Z. Zhao, G. Yuan, Y.M. Zhang, S.J. Qiao, S.Z. Wang, L. Zhang, "Multi-view Fused Spatial-temporal Dynamic GCN for Urban Traffic Flow Prediction," Journal of Software, vol. 35, no. 4, pp. 1751-1773, 2024.Doi: 10.13328/j.cnki.jos.007018

[33] W.D. Li, L. Yu, "Construction of urban traffic flow prediction model under neural network model," Cyberspace Security, no. Z3, pp. 56-59+77, 2021.

[34] X. T. Zhang, J. Y. Zheng, Q. Shen, D. F. Sun, Y. L. Jiang, "Multi-channel spatial-temporal traffic flow prediction based on hybrid static-dynamic graph convolution," Telecommunications Science, vol. 39, no. 9, pp. 97-110, 2023.

[35] P. Qiao, "Research on traffic flow detection based on deep learning and edge task offloading [D]," Xi'an Electronic Science and Technology University, 2020.Doi: 10.27389/d.cnki.gxadu.2019.001182

[36] R. Tian, C. Wang, J. Hu, Z. Ma, "MFSTGN: a multi-scale spatial-temporal fusion graph network for traffic prediction," Applied Intelligence, vol. 53, no. 19, pp. 22582-22601, 2023.Doi: 10.1007/s10489-023-04703-4

[37] Y. Qin, H. Luo, F. Zhao, Y. Fang, X. Tao, C. Wang, "Spatio-temporal hierarchical MLP network for traffic forecasting," Information Sciences, vol. 632, pp. 543-554, 2023.Doi: 10.1016/j.ins.2023.03.063

[38] Y.P. Liang, Z.Y. Mao, W.B. Zou, R. Xu, "Short-term Traffic Flow Prediction Based on Similar Data Aggregation and KNN with Varying K-value," Journal of Geo-Information Science, vol. 20, no. 10, pp. 1403-1411, 2018.

[39] X.L. Zhang, "A study on short-time traffic volume forecasting based on non-parametric regression [D]," Tianjin University, 2007.

[40] S. Liu, H. Chen, X.Y. Chen, J.J. He, "Double Branch Spatial-temporal Graph Convolutional Neural Network for Traffic Flow Prediction," Information and Control, no. 3, pp. 391-404+416, 2023. Doi: 10.13976/j.cnki.xk.2023.2092

[41] L.L. Wu, L.L. Yin, Q.L. Ren, "A Short-Term Traffic Flow Forecasting Method Based on EMD and DE-BPNN Combined Optimization," Journal of Chongqing University of Technology (Natural Science), vol. 35, no. 12, pp. 155-163, 2021.