Mobile Traffic Prediction in Consumer Applications: A Multimodal Deep Learning Approach

Weiwei Jiang, *Member, IEEE*, Yang Zhang, Haoyu Han, Zhaolong Huang, Qiting Li, and Jianbin Mu

***Abstract*—Mobile traffic prediction is an important yet challenging problem in consumer applications because of the dynamic nature of user behavior, varying application quality of service (QoS) requirements, network congestion, and proliferation of diverse mobile devices. The mobile traffic prediction problem with multiple services, e.g., SMS, call, and Internet, is defined as the mapping from historical traffic data to future traffic prediction. Both grid and graph-based mobile traffic prediction formulations have been extensively considered in the literature. However, an effective multimodal deep learning approach with both grid and graph modals has not yet been fully considered. This study proposes a multimodal convolutional neural network (CNN)-graph neural network (GNN) hybrid framework for single-step mobile traffic prediction, in which the information extracted from SMS, call and Internet services are fused to make a precise prediction for future traffic consumption of consumers in the next hour. The CNN module is built using ConvLSTM, the GNN module is built using adaptive graph convolutional network (AGCN), and a fusion layer is designed to combine the outputs from the CNN and GNN modules. Numerical experiments based on a real-world dataset demonstrate the effectiveness of the proposed framework, which achieves a prediction error lower than ten baselines.**

***Index Terms*—Consumer electronics, deep learning, convolutional neural networks, graph neural networks.**

I. INTRODUCTION

CONSUMER applications refer to software programs and technological tools designed for end-users or consumers with the primary aim of enhancing their daily lives, experiences, and interactions. These applications span a wide range of domains, including but not limited to communication, entertainment, productivity, health, and finance. Enterprise applications are designed for businesses and organizations, which have strict standards, fixed processes and high prices. Consumer applications are designed for individuals and households, with moderate functionalities, optional features and affordable prices. Among modern consumer applications, the mobile phone is an important example, which has revolutionized the way individuals communicate, access information, and engage with a myriad of services [1]. These handheld devices have evolved from mere communication tools to multifaceted platforms, offering a diverse array of applications catering to the needs and preferences of end-users [2].

Consumer applications on mobile phones encompass a broad spectrum, including social media, messaging services, entertainment, productivity tools, health and fitness trackers, and navigation systems [3], [4]. The ubiquity and personal nature of mobile phones have made them integral to daily life, shaping the dynamics of how people connect, work, and entertain themselves. With the integration of artificial intelligence and mobile edge computing, consumer applications on mobile phones have become more intelligent and smart than before [5], [6].

In this study, multimodal fusion for consumer electronics is considered with a specific application, i.e., mobile traffic prediction, which addresses the challenge of forecasting the volume, patterns, and characteristics of the data exchanged over mobile networks [2]. Deep learning-based fusion and mining of multimodality data for this specific problem are considered and a multimodal convolutional neural network (CNN)-graph neural network (GNN) framework for single-step traffic prediction is proposed as an effective solution. Even though multimodal data fusion from GNNs and CNNs has been considered in the literature for other problems, it has not been fully considered for consumer applications, including mobile traffic prediction. With exponential growth in mobile device usage, the demand for mobile data continues to surge. Mobile traffic prediction is a crucial area of study and application because it plays a pivotal role in optimizing network resources, enhancing user experience, and facilitating efficient network management [7], [8].

Efficient mobile traffic prediction is essential for mobile network operators, service providers and policymakers. This enables them to optimize resource allocation, plan network upgrades effectively, and enhance the overall quality of service for mobile users. Additionally, accurate predictions can assist in mitigating congestion issues, improving energy efficiency, and supporting the deployment of emerging technologies such as 5G. Researchers in this field have explored innovative

Manuscript received 1 December 2023; revised 8 January 2024; accepted 29 January 2024. Date of publication 1 February 2024; date of current version 26 April 2024. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2023RC16, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ24F030023. *(Corresponding authors: Qiting Li; Jianbin Mu.)*

Weiwei Jiang, Yang Zhang, and Haoyu Han are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, and the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: jww@bupt.edu.cn; zy2023140030@bupt.edu.cn; 786898479@bupt.edu.cn).

Zhaolong Huang is with the College of Science, Tianjin University of Technology, Tianjin 300384, China (e-mail: zhao_long_h@163.com).

Qiting Li is with the Research and Development Department, China Academy of Launch Vehicle Technology, Beijing 100076, China (e-mail: jeetinlee@gmail.com).

Jianbin Mu is with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: jianbinmu@zjut.edu.cn).

Digital Object Identifier 10.1109/TCE.2024.3361037

methodologies such as machine learning algorithms, statistical models, and data analytics, to predict mobile traffic patterns. Addressing the challenges associated with mobile traffic prediction is crucial for building resilient and adaptive mobile networks that can satisfy the growing demands of an increasingly connected world [9].

Deep learning methods have proven effective for mobile traffic prediction in recent years [10], which can be roughly divided into three categories: recurrent neural networks (RNNs), CNNs [11], and GNNs [12]. Compared with traditional linear prediction models, e.g., autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH) and vector autoregression (VAR), deep learning models have presented a strong learning ability for non-linear patterns in historical traffic data and have achieved state-of-the-art prediction performance. RNNs, e.g., long short-term memory (LSTM) and gated recurrent unit (GRU), are designed for sequence input and are an ideal option for capturing both short-term and long-term temporal dependencies in sequential traffic data. CNNs are suitable for capturing spatial dependencies in a unified grid format, in which each cell has the same size, and the influence of neighbors is extracted with the convolution operation. GNNs take a step further by modelling the cross-region relationship with graph construction, and the convolution operation is extended to the graph format, e.g., graph convolutional network (GCN) [13] and graph attention networks (GAT) [14].

The problem of mobile traffic prediction is still challenging because of the dynamic nature of user behavior, varying application quality of service (QoS) requirements, network congestion, and proliferation of diverse mobile devices. Predicting mobile traffic involves anticipating the amount of data that will traverse the network, the types of services users will access, and the locations and times at which the peak demand is likely to occur. Considering the complex correlation among different services and both spatial and temporal dependencies in mobile traffic prediction, the single-modal deep learning approach is insufficient and multimodal deep learning is used in this study instead, for three different services including call, short message service (SMS), and Internet access.

Multimodal deep learning refers to the integration of information from diverse sources or modalities, such as text, images, and audio, into a unified deep learning model [15]. Traditional deep learning models often focus on a single modality, such as processing only images or text. However, many real-world applications involve data spanning multiple modalities, and leveraging this multimodal information can lead to more robust and comprehensive models. The key idea behind multimodal deep learning is to enable a model to learn and make predictions based on combined information from different modalities. This approach is particularly powerful for tasks in which multiple sources of data provide complementary insights. For example, in natural language processing, combining textual and visual information can enhance the understanding and context. The advantages of multimodal deep learning over single-modal approaches are two-folds. First, the architecture of a multimodal deep learning model typically involves separate pathways for each modality, with shared layers or connections that allow the model to learn relationships and correlations between modalities. Secondly, multimodal deep learning enables the model to capture complex patterns and representations that may not be apparent when each modality is considered in isolation.

In this study, we leverage the advantages of multimodal deep learning and propose a multimodal CNN-GNN hybrid framework for single-step mobile traffic prediction, in which the information extracted from SMS, call and Internet services are fused to make a precise prediction. In the proposed framework, the CNN and GCN modules are designed with a careful selection of basic blocks, namely Convolutional LSTM (ConvLSTM) and Adaptive GCN (AGCN). Numerical experiments are conducted using a real-world mobile dataset to compare the proposed approach with baseline methods and evaluate different model design schemes.

The contributions of this study are summarized as follows:

- The multimodal deep learning is introduced in this study for mobile traffic prediction in consumer applications, based on the combination of grid and graph modals.
- A multimodal CNN-GNN framework for single-step traffic prediction is proposed as an effective solution.
- Numerical experiments with a real-world mobile dataset demonstrate the effectiveness of the proposed solution over ten baselines in terms of prediction errors.

The remainder of this paper is organized as follows. Section II discusses relevant studies categorized into different problem formulations. Section III introduces the dataset and formulates a single-step traffic prediction problem. Section IV presents the proposed multimodal CNN-GNN framework, along with the introduction to different modules. Section V validates the effectiveness of the proposed framework through numerical experiments. Finally, Section VI concludes this paper.

The acronyms for the terminologies used in this paper are listed in Table I.

II. Related Work

A. Overview

Deep learning based mobile traffic prediction is an indispensable part for empowering future consumer applications, considering diverse communication service requirements and limited resources on mobile devices [16]. Extensive research efforts have been put on relevant research topics about mobile traffic prediction. From the methodology perspective, deep learning has been proven the state-of-the-art solutions and the discussion in this section would mainly talk about deep learning approaches. From the problem formulation perspective, three popular modals have been used in the literature, e.g., time series, grid and graph. Since time series can only describe simple cases with limited variables, grid-based and graph-based problem formulations are drawing more attention in recent years. In this section, the relevant studies would be categorized and discussed from the problem formulation perspective. Due to space constraint, only the most recent and relevant studies are reviewed in this section. For a more

TABLE I
THE ACRONYMS USED IN THIS PAPER

Abbreviation	Full Name
AGCN	Adaptive Graph Convolutional Network
ARIMA	AutoRegressive Integrated Moving Average
BLSTM	Bidirectional LSTM
CDF	Cumulative Distribution Function
ConvGRU	Convolutional GRU
ConvLSTM	Convolutional LSTM
eMBB	Enhanced Mobile Broadband
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GASTN	Graph Attention Spatial-Temporal Network
GAT	Graph Attention Network
GCN	Graph Convolutional Network
CDF	Cumulative Distribution Function
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ORAN	Open Radio Access Network
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SMS	Short Message Service
STAGCN	Spatial-temporal Aggregation Graph Convolution Network
STECA-GCN	spatial-temporal-event Cross Attention Graph Convolution Neural Network
STCNet	Spatial-temporal Cross-domain Neural Network
STG-STAN	State Transition Graph-based Spatial-Temporal Attention Network
ST-3DDMCRN	Spatio-temporal 3D DenseNet Multiscale ConvLSTM-ResNet Network
uRLLC	Ultra-reliable and Low-latency
VAR	Vector Autoregression
POT	Point-of-interest

comprehensive literature review, interested readers can refer to recent surveys and reviews [7], [17] for more related work discussion.

B. Time Series-Based Mobile Traffic Prediction

In the time series-based mobile traffic prediction formulation, the traffic captured in different locations with mobile devices is recorded and stored as time series data with both values and timestamps. As a time series prediction problem, the historical traffic sequence is regarded as input and the future traffic demand is the prediction target. The basic principle of time series predict is to recognize the continuous traffic patterns, while fully considering the randomness due to the influence of accidental factors, e.g., weather and holiday factors. The basic idea is that history repeats itself and a constant relationship exists between the historical traffic data and the future traffic demand, for which a time series model is built to learn and fit. In order to eliminate the effects of random fluctuations, historical data can be used for statistical analysis and the data can be properly processed for trend prediction. Traditional linear prediction models including ARIMA, GARCH, and VAR have been used for traffic time prediction for a long time [18]. More recently, deep learning models, such as LSTM and GRU, become dominant in this research area and various deep learning models are proposed for time series-based mobile traffic prediction.

User-Cybertwin Asynchronous Learning (UCAL) [19] is a time series-based traffic prediction model for heterogeneous networks, which is based on a LSTM model for capturing temporal dependencies and a pattern extraction method for mining the relationship between the user and cyber spaces. Curriculum learning is incorporated into the AI-native wireless communication system for traffic prediction [20], which requires fewer parameters and reduced carbon emissions and demonstrates improved prediction performance over standard deep learning models. FedMIC [21] is a federated learning framework for wireless traffic prediction, which utilizes mutual information clustering. FedMIC employs a sliding window scheme, spectral clustering, and hierarchical aggregation architecture to enhance client model learning, address statistical heterogeneity, and improve prediction accuracy. For proactive network resource allocation based on QoS requirements, the integration of RNN and bidirectional LSTM (BLSTM) is utilized as the prediction model for multivariate QoS-aware network traffic, achieving an average accuracy of 97.68% over 13 hours [22]. To optimize resource utilization for enhanced mobile broadband (eMBB) and ultra-reliable and low-latency (uRLLC) services in Open Radio Access Network (ORAN), a joint LSTM-based traffic prediction, flow-split distribution, dynamic user association, and radio resource management framework is proposed [23].

C. Grid-Based Mobile Traffic Prediction

Inspired by the success of CNNs for image classification, grid-based mobile traffic prediction has been considered in a spatial region that can be divided into regular grids and further regarded as 2D matrices [24]. The traffic generated in a grid is aggregated both in spatial and temporal axis and the traffic matrix sequence can be seen as a video. In the grid-based mobile traffic prediction problem formulation, historical traffic matrices are used as inputs and the traffic matrix in the next time slot is the prediction target. This formulation is similar to the video prediction problem, in which a frame has a 2D matrix format. CNNs are dominant for grid-based mobile traffic prediction problems and the attention mechanism can further enhance the prediction performance.

STDenseNet [25] treats traffic matrices as images and employs densely connected CNNs to effectively capture both spatial and temporal dependencies in cell traffic dynamics with a parametric matrix-based fusion scheme for citywide mobile traffic prediction. Spatial-temporal cross-domain neural network (STCNet) [16] utilizes a convolutional LSTM model to capture spatial-temporal dependencies in mobile traffic data and a clustering algorithm, along with a successive inter-cluster transfer learning strategy [26], is proposed to enhance knowledge reuse by segmenting city areas. DMFS-MT [27] is a multi-task deep learning framework for mobile traffic prediction that incorporates a dual modular feature sharing layer and a multi-task learning layer to capture long-term spatio-temporal dependencies and local spatio-temporal fluctuation trends. Spatio-temporal 3D DenseNet multiscale ConvLSTM-ResNet network (ST-3DDMCRN) [28] utilizes a 3D DenseNet network to capture local regional spatio-temporal information, a multiscale ConvLSTM-ResNet network to address long-range spatial correlation, and a region-squeeze-and-excitation unit to quantify spatio-temporal heterogeneity for citywide traffic flow prediction. A joint traffic matrix completion and prediction framework is considered, in which a 3D-UNet architecture for multi-scale spatio-temporal correlation exploitation in traffic matrix sequence completion and an LSTM2D architecture for leveraging spatio-temporal dependencies in traffic matrix prediction [29].

D. Graph-Based Mobile Traffic Prediction

While the grid-based mobile traffic prediction formulation is intuitive and simple to implement, the cross-region and long-range spatial dependency is difficult to model in the grid-based format. To amend this shortcoming, graph-based mobile traffic prediction is further introduced, in which nodes can be either base stations or regions. The edges connecting nodes can be defined in a flexible approach, including both physical distance and virtual correlation relationships. External factors can also be incorporated into the graph construction process to better reflect the hidden spatio-temporal dependencies. GNNs are dominant solutions for graph-based mobile traffic prediction and recent studies focus on developing more advanced GNN variants.

Considering both local geographical dependencies and distant inter-region relationships through a constructed spatial relation graph, graph attention spatial-temporal network (GASTN) [30] utilizes structural RNNs and two attention mechanisms for capturing the dynamic spatial relations, and a collaborative global-local learning strategy that leverages knowledge from both global and local models for individual regions. Spatial-temporal aggregation GCN (STAGCN) [31] using an aggregation GCN for complex spatial-temporal correlation modeling, and a regression module to generate the final prediction when considering external factors. Utilizing state transition graphs to identify semantic context information, state transition graph-based spatial-temporal attention network (STG-STAN) [32] incorporates a spatial attention extraction module with GCNs and a parallel LSTM module to capture dynamic evolution and temporal correlation. An adaptive GCN model is proposed to address limitations in graph structure modeling, capturing latent spatial dependency with self-adaptive matrices and acquiring temporal dependency using RNNs [33]. Spatial-temporal-event cross attention GCN (STECA-GCN) [34] considers event dimension features and incorporating direct cross-fusion among diverse features, which is further combined with a deep reinforcement learning strategy to facilitate dynamic load balancing decisions.

E. Summary

In existing studies, time series prediction has the simplest format and the most wide application scenario, e.g., traffic, energy, and finance. The grid or graph structures are not necessary in a typical time series prediction formulation. However, time series prediction solutions are not enough for modelling mobile traffic prediction in a heterogeneous environment, which involves both temporal and spatial dependencies. Grid-based and graph-based mobile traffic prediction formulations have their advantages and disadvantages, and both have applicable cases for consumer electronics. Compared with graph-based mobile prediction, grid-based mobile prediction is easier to build and solve with CNNs, but the tradeoff is that it is difficult to capture the cross-region spatial dependency in grid-based mobile prediction effectively.

Multimodal data fusion from GNNs and CNNs has been considered in the literature for medical applications [35], e.g., Covid-19 classification [36] and breast cancer classification [37]. However, it has not been fully considered for consumer applications, including mobile traffic prediction. The challenges of mobile traffic prediction in a heterogeneous environment include the dynamic nature of user behavior, varying application QoS requirements, network congestion, and proliferation of diverse mobile devices. To solve these challenges, multimodal deep learning is introduced in this study for mobile traffic prediction with multiple consumer services and a multimodal CNN-GNN framework for single-step traffic prediction is proposed as an effective solution.

Compared with existing models, the proposed framework has several advantages. The architecture of the proposed multimodal deep learning model involves separate pathways for each modality, with shared layers or connections that allow the model to learn relationships and correlations between two modalities including graph and grid modalities. Secondly, the proposed multimodal deep learning model manages to capture complex patterns and representations that may not be apparent when each modality is considered in isolation and achieves better prediction performance than existing methods.

III. Problem Description

In this study, we consider the mobile traffic prediction problem in consumer applications with multiple services, e.g., SMS, call, and Internet, based on a real-world mobile dataset named the Telecom Italia dataset [38]. The Telecom Italia dataset used in this study is publicly available.[1] The dataset is collected in the city of Milan, which is divided into the grid format with a size of $H \times W = 100 \times 100$. Each grid is also regarded as a cell in this study. The mobile traffic is collected from 2023-11-01 00:00 to 2014-01-01 00:00 with a temporal interval of 10 minutes. In this study, both in-flow and out-flow traffic in a cell are added together. There are many cells with rare traffic in a 10-min time interval. To solve the data sparisty problem, the traffic is further aggregated in an hour in this study. The heatmap figures for a randomly chosen time slot are shown in Fig. 1. For a better visualization, only the central 50×50 cells are plotted in Fig. 1, for three services separately. The mobile traffic time series visualization for a single cell (50, 50) in the first 7 days (i.e., 168 hours) is shown in Fig. 2. An obvious daily pattern can be observed in Fig. 2. Besides, different curves have a strong positive correlation and show a similar periodical pattern, which confirms the correlation among SMS, call and Internet services.

In this study, the mobile traffic matrix prediction problem is considered, as shown in Fig. 3. The traffic of a single service in time slot t is the prediction target, which is denoted as $\boldsymbol{X}_t$. The inputs include the recent traffic matrices $\boldsymbol{X}_{t-p}, \ldots, \boldsymbol{X}_{t-2}, \boldsymbol{X}_{t-1}$ with a time length p and the periodical traffic matrices

[1] https://github.com/zctzzy/STCNet

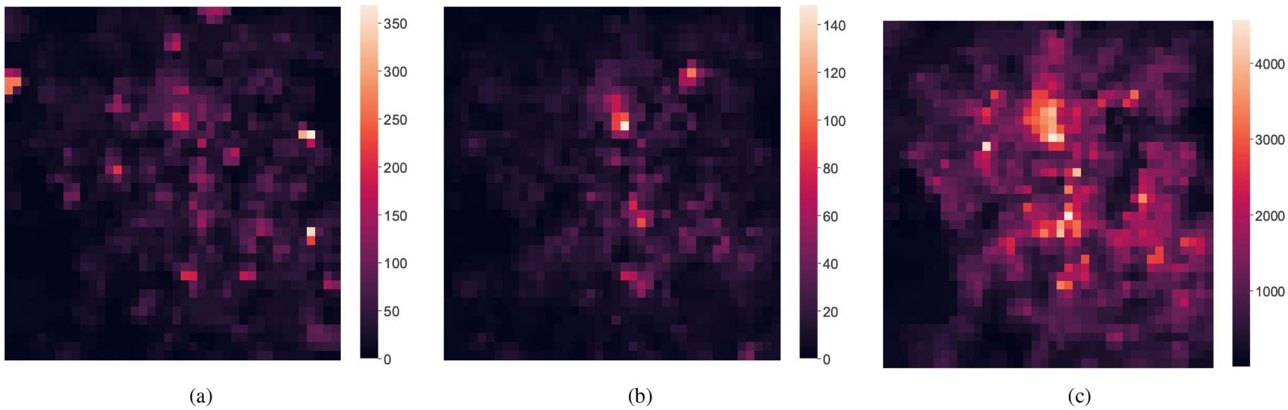


Fig. 1. Heatmap figures for data visualization. (a) SMS; (b) Call; (c) Internet.

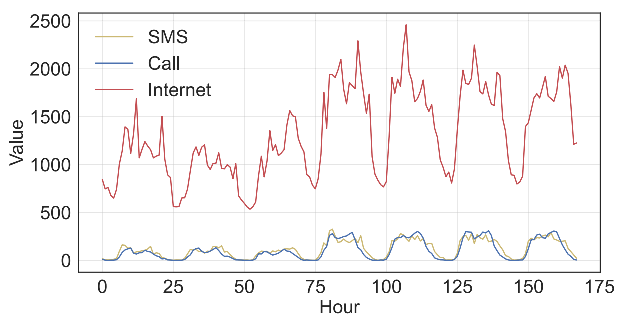


Fig. 2. The mobile traffic time series visualization for a single cell (50, 50) in the first 7 days (i.e., 168 hours).

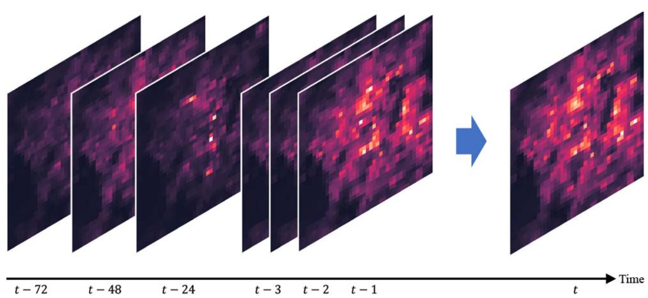


Fig. 3. The mobile traffic prediction problem.

$\boldsymbol{X}_{t-q*24}, \ldots, \boldsymbol{X}_{t-48}, \boldsymbol{X}_{t-24}$ with a number of q, based on the daily pattern observed from Fig. 2.

Compared with existing problem formulations discussed in the related work section, e.g., time series-based, grid-based, and graph-based formats, our problem formulation is more similar to the grid-based format, in which the city of Milan is divided into regular grids. However, the difference is that in the proposed service-fusion multimodal CNN-GNN hybrid framework, the graph modal is constructed and multimodal deep learning method is leveraged. In this study, we only consider the single-step prediction problem, e.g., the prediction for the next hour, and leave the multi-step prediction problem for future consideration.

IV. Methodology

A. Framework

The proposed service-fusion multimodal CNN-GNN hybrid framework for mobile traffic prediction is shown in Fig. 4. The basic idea is to combine three services, namely, SMS, call, and Internet, and two different modals, i.e., grid and graph, to obtain a better prediction performance for mobile traffic. While the relationship of different services has been explored and exploited in previous studies, in which the traffic prediction for a service is based on all three services as we do in this study, the fusion of both grid and graph modals has not been fully considered in the literature for mobile traffic prediction. This section describes different modules used in our proposed framework.

B. Graph Construction

The first step is to construct the graph modal from the grid modal of raw data. A graph is formally defined as $G = (V, E, A)$, where V is the set of vertices or nodes, E is the set of edges between the nodes, and A is the adjacency matrix. Element a_{ij} of A represents the edge weight between nodes i and j. In this study, each cell is regarded as a node in the graph modal. Spatial proximity is used as the criterion for building a graph in this study, which is a simple yet effective graph construction method. Specifically, distance-based adjacency matrix is used, which represents the spatial closeness between nodes. In the binary adjacency matrix, each element value is determined by whether the two cells share a common boundary. The value is set to 1 if connected and 0 otherwise. The binary adjacency matrix can be easily extended to a travel distance based variant. In our future studies, external information would be considered for building more complex graph structures, e.g., the context and point-of-interest (POI) information.

C. CNN Module

In the CNN module, ConvLSTM is chosen as the basic unit because it has been proven effective for traffic matrix prediction problems [39]. Comparative experiments are further conducted in this study to demonstrate its superiority over other options, e.g., Conv3D and Convolutional GRU (ConvGRU).

ConvLSTM combines both CNN and LSTM and is suitable for grid modal sequence modeling. Following a typical LSTM description, t is the time slot index, x_t is the input data sample in a time series, h_t is the output value, and c_t is the cell state.

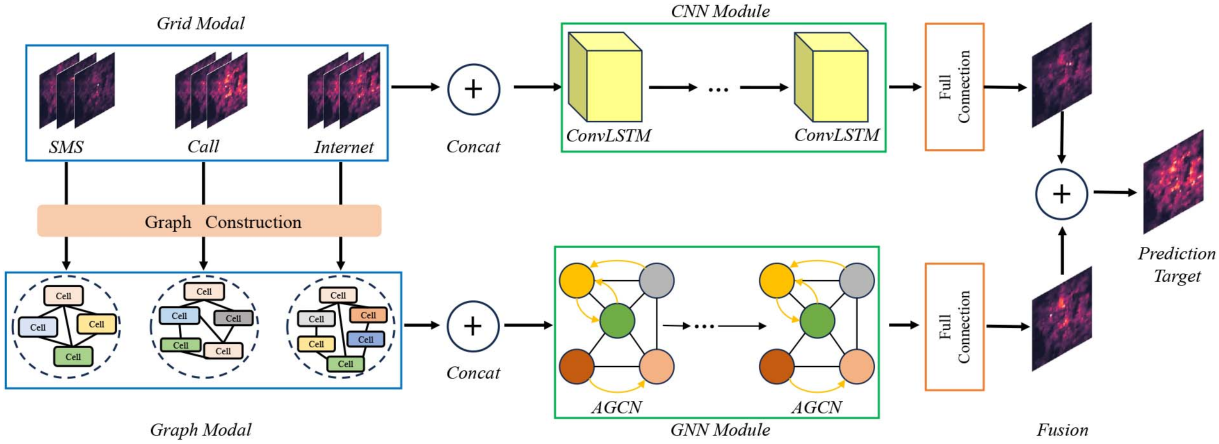


Fig. 4. The proposed framework for mobile traffic prediction.

ConvLSTM extends these variables to fit the grid-format case, in which X_t is the traffic frame with a size of $H \times W$ in time slot t. The simple matrix multiplication in LSTM is also replaced with the two-dimensional convolution operation in ConvLSTM. The three gate mechanisms introduced in LSTM is retained in ConvLSTM and are described as follows.

The forget gate decides what information to throw away from the cell state and is denoted as follows:

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f\right) \quad (1)$$

where $\sigma(*)$ is the activation function, i.e., sigmoid function, $*$ is the convolution operation, $\circ$ is the Hadamard product, W_* and b_* are trainable parameters.

The input gate decides what new information is to be stored in the cell state and is denoted as follows:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \quad (2)$$
$$\tilde{C}_t = tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3)$$
$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \quad (4)$$

where $\tilde{C}_t$ is a matrix of new candidate values that are added to the cell state, and $tanh(*)$ is the tanh activation function.

The output gate decides what information to output from the cell state and is as follows:

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \quad (5)$$
$$H_t = o_t \circ tanh(C_t) \quad (6)$$

Both input and output variables of a ConvLSTM layer are traffic matrix sequences in a 3D matrix format. Multiple ConvLSTM Layers can be stacked to further extract useful information. A fully connected layer is used to map the output of the CNN model into a single matrix output for the fusion module. The input matrix sequences are flattened and then fed into the fully connected layer. The fully connected layer maps the input vector into an output vector with the same number of elements with the expected matrix output. In this study, the expected matrix output is 100×100, which means the output vector size is 10000. Then the output vector of the fully connected layer is reshaped into the expected matrix output format.

D. GNN Module

In the GNN module, AGCN is chosen as the basic unit which outperforms both GCN and GAT in the numerical experiments. A short introduction to GCN, GAT and AGCN is presented in this section. The original GNN [12] aims to update the state of a node with its neighbors, based on the message passing mechanism. Afterwards, the success of CNN in image recognition has inspired the introduction of the convolution operation into GNNs and two families of GCNs are proposed in the literature, including spectral-based and spatial-based GCNs. Based on graph signal processing, spectral-based GCNs define the convolution operation in the spectral domain techniques. Some famous GCN variants belong to this family, including ChebNet [40] and GCN [13].

Following the mathematical formulation introduced in the graph construction step, the degree matrix $D_{ii} = ||\mathcal{N}(v_i)||$ is introduced to measure the number of neighbors, where $\mathcal{N}(v_i)$ is the neighbor node set of node v_i. $\mathbf{X} \in R^{N\times d}$ denotes the node feature matrix, where d is the feature dimension and N is the node number. $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the Laplacian matrix and $\tilde{\mathbf{L}} = \mathbf{I}_N - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ is the normalized Laplacian matrix, where $\mathbf{I}_N$ is the identity matrix of size N. The graph convolution operation $*G$ in GCN is defined as follows:

$$\mathbf{X}_{*G} = \mathbf{W}\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\right)\mathbf{X} \quad (7)$$

where $\mathbf{W}$ is the trainable parameter, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ and $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$.

AGCN extends GCN with a self-adaptive adjacency matrix $\tilde{\mathbf{A}}_{adpt}$, which is based on a trainable distance matrix, and the graph topological structure in AGCN can be updated to minimize the training loss. In AGCN, the self-adaptive adjacency matrix $\tilde{\mathbf{A}}_{adpt}$ is constructed by the distance matrix with a Gaussian kernel,

$$\tilde{\mathbf{A}}_{adpt} = e^{-D(x_i,x_j)/(2\sigma^2)t} \quad (8)$$

where the distance matrix $D(x_i, x_j)$ is parameterized with a parameter matrix M, which is a positive semi-definite matrix,

$$D\left(x_i, x_j\right) = \sqrt{\left(x_i - x_j\right)^T M\left(x_i - x_j\right)} \quad (9)$$

TABLE II
THE EVALUATION RESULTS FOR DIFFERENT PREDICTION MODELS

Model	R^2			MAE			RMSE		
	SMS	Call	Internet	SMS	Call	Internet	SMS	Call	Internet
Last value	0.852	0.905	0.969	9.548	7.605	29.702	27.570	20.222	78.055
HA	0.625	0.672	0.829	14.714	11.249	52.858	43.856	37.587	182.972
STDenseNet [25]	0.855	0.907	0.969	9.452	7.529	29.405	27.295	20.020	77.274
STCNet [16]	0.860	0.911	0.970	9.283	7.352	29.042	26.804	19.548	76.320
GASTN [30]	0.865	0.916	0.971	9.113	7.174	28.679	26.314	19.076	75.366
STAGCN [31]	0.870	0.920	0.972	8.943	6.997	28.316	25.824	18.604	74.412
DMFS-MF [27]	0.875	0.924	0.972	8.773	6.819	27.953	25.334	18.132	73.458
Conv3D [42]	0.880	0.928	0.973	8.604	6.642	27.590	24.844	17.661	72.504
ConvLSTM [28]	0.884	0.931	0.974	8.434	6.465	27.227	24.354	17.189	71.550
ConvGRU [43]	0.889	0.935	0.975	8.264	6.287	26.864	23.864	16.717	70.596
Proposed	**0.898**	**0.942**	**0.976**	**7.925**	**5.932**	**26.138**	**22.883**	**15.773**	**68.688**

The attention mechanism is introduced in GAT [14] to learn the most relevant information from neighbors. While GCN is limited to regular and well-structured graphs, GAT is designed to handle irregularities and complexities in the spatial relationships [41]. The multi-head attention mechanism with K heads is denoted as follows:

$$\mathbf{X}_i^{(t)} = \|_k \sigma \left(\sum_{j \in \mathcal{N}(v_i)} \alpha^k \left(\mathbf{X}_i^{(t-1)}, \mathbf{X}_j^{(t-1)} \right) \mathbf{W}^{(t-1)} \mathbf{X}_j^{(t-1)} \right) \quad (10)$$

where $\|$ is the concatenation operation, σ is the activation method, and $\alpha^k(\cdot)$ is the k-th attention mechanism.

E. Fusion Module

The last step in the proposed framework is the fusion of multimodal features, which is conducted with a parametric matrix based fusion method. Denote the outputs of the CNN and GNN modules as $\boldsymbol{X}_c$ and $\boldsymbol{X}_g$. The final prediction output after fusion is denoted as

$$\hat{\boldsymbol{X}}_t = \sigma\left(\boldsymbol{W}_c \circ \boldsymbol{X}_c + \boldsymbol{W}_g \circ \boldsymbol{X}_g\right) \quad (11)$$

where $\sigma(*)$ is the activation function, $\circ$ is the Hadamard product, $\boldsymbol{W}_c$ and $\boldsymbol{W}_g$ are learnable parameters. With the fusion module, the relative weights of two modules, namely, grid and graph modules, can be learned from the training data, instead of a simple average.

The loss function is defined as the Frobenius norm between the predicted value $\hat{\boldsymbol{X}}$ and the ground truth value $\boldsymbol{X}$, i.e.,

$$\mathcal{L}(\theta) = arg \min_{\theta} ||\hat{\boldsymbol{X}} - \boldsymbol{X}||_2^2 \quad (12)$$

where θ denotes the trainable parameters in the proposed prediction model.

V. EXPERIMENTS

A. Basic Settings

For the considered mobile traffic prediction problem, the input time length p is 6 and the periodical value number q is 3. Minmax normalization is used as the prepocessing to transform the dataset into a range of $[0, 1]$. An inverse transformation is conducted to revert the normalization and to evaluate the prediction performance in the original value range. For the proposed multimodal CNN-GNN hybrid framework, a total of 3 ConvLSTM layers and 3 AGCN layers are used in the numerical experiments.

Ten baselines are used in this study to demonstrate the superiority of the proposed methodology, covering simple methods, CNNs and GNNs. *Last value* uses the last known value, e.g., the value of time slot $t-1$, as the prediction for time slot t. *Historical average* (HA for short) uses the historical average of 3 past periodical values, i.e., the values in the same time slot in the past 3 days, $t-24, t-48$ and $t-72$ more specifically. The remaining baselines all belong to deep learning, including STDenseNet [25], STCNet [16], GASTN [30], STAGCN [31], DMFS-MF [27], Conv3D [42], ConvLSTM [28], ConvGRU [43]. The Python programming language and Tensorflow framework is used to implement both the proposed and baseline deep learning models in this study. The experiments are conducted on a desktop computer with 16 GB RAM and 8GB GPU. The batch size is 32 and the training epoch is set to 100. Adam is chosen as the optimizer, with a learning rate of 10^{-3}.

Three evaluation metrics are used in this study for comparing different approaches, namely, coefficient of determination R^2, root mean square error (RMSE) and mean absolute error (MAE). Denote the true values as $\mathbf{y}$ and the predicted values as $\hat{\mathbf{y}}$, these evaluation metrics are defined as follows:

$$R^2(\mathbf{y}, \hat{\mathbf{y}}) = 1 - \frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N} \left(y_i - \bar{y}\right)^2} \quad (13)$$

$$\text{RMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2} \quad (14)$$

$$\text{MAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (15)$$

where N is the number of data samples in the test set, and $\bar{y}$ is the average value in the test set. Besides RMSE and MAE, the cumulative distribution function (CDF) is further used to describe the overall prediction error quantitatively.

B. Results and Discussion

The evaluation results for different prediction models are summarized in Table II. It is noticed that the last value method is still a simple yet competitive baseline, which outperforms the historical average method. However, the deep learning methods perform better than the last value method and the

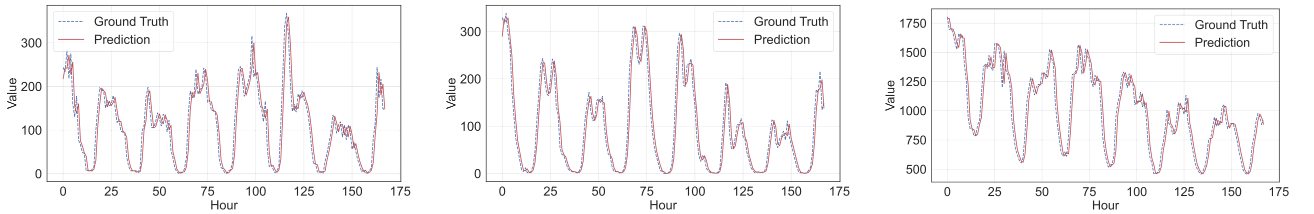


Fig. 5. The comparison between the prediction and ground truth for cell (50, 50) in the first 7 days of the test subset (i.e., 168 hours). (a) SMS; (b) Call; (c) Internet.

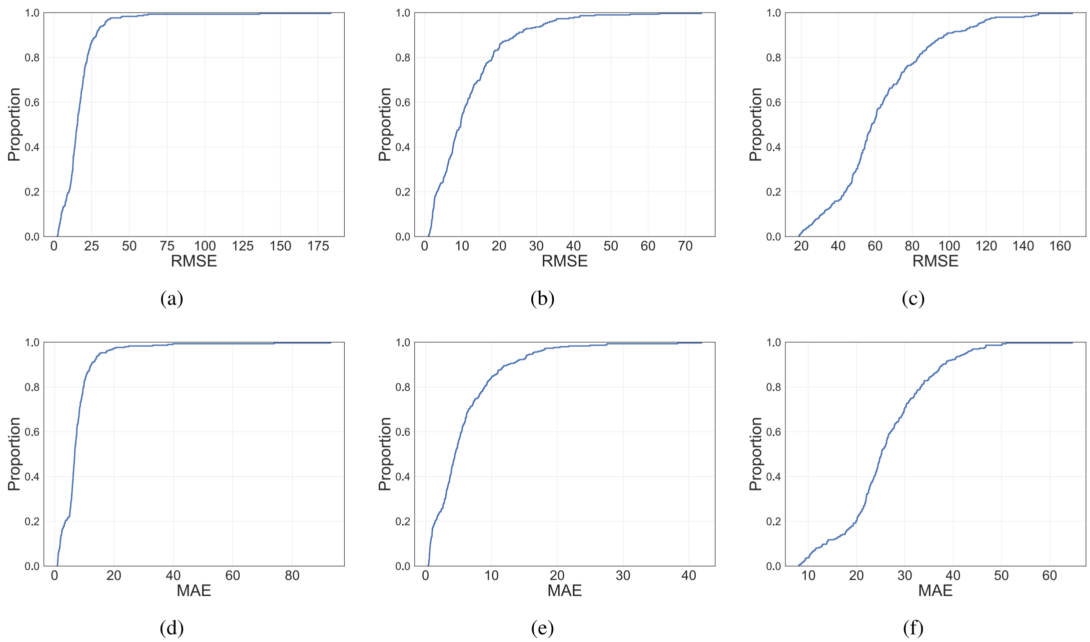


Fig. 6. Empirical cumulative distribution function figures of prediction errors. (a) SMS RMSE; (b) Call RMSE; (c) Internet RMSE; (d) SMS MAE; (e) Call MAE; (f) Internet MAE.

proposed multimodal CNN-GNN approach achieves the lowest error in all cases.

The predictions of the proposed multimodal CNN-GNN approach are compared with ground truth values in Fig. 5, for cell (50, 50) in the first 7 days of the test subset (i.e., 168 hours). The predictions are very close to the ground truth values, demonstrating the effectiveness of the proposed prediction model. To better understand the error distributions, the empirical cumulative distribution function (CDF) figures of prediction errors in different hours are shown in Fig. 6. It can be observed from Fig. 6 that the predictability for different services vary a lot. For example, the Internet service traffic is more difficult to predict.

To prove the rationality of the model design in Section IV, a series of ablation study is conducted by altering the model structure. The first step is to consider the necessity of considering both grid and graph modals. The evaluation of different modals is shown in Fig. 7(a) and Fig. 7(d), which validate the effectiveness of the fusion module. For simplicity, only RMSE and MAE are plotted in the ablation study. It is demonstrated that the idea of multimodal learning applied in mobile traffic prediction is effective and worthy a further investigation with more advanced multimodal deep learning models.

The evaluation results of different CNN and GNN modules are also shown in Fig. 7. From Fig. 7(b) and Fig. 7(e), it can be concluded that ConvLSTM is a better choice than Conv3D and ConvGRU, which exhibits lower RMSE and MAE. Similarly, from Fig. 7(c) and Fig. 7(f), AGCN is a better choice than GCN and GAT. This preliminary study is only a start and more complex model would be considered in our future studies.

The influence of the input time length p and the periodical value number q is shown in Fig. 8. From Fig. 8(a) and Fig. 8(c), a larger p brings a better prediction, and the improvement is not obvious when p is equal to or greater than 6. Similarly, from Fig. 8(b) and Fig. 8(d), a larger q brings a better prediction, and the improvement vanishes after q is equal to or greater than 2. The results in Fig. 8 are reasonable when a longer historical input sequence contains more information for predicting the future situations. However, when the input exceeds some value, the benefits of using more input data do not help and the risk of overfitting increases.

C. Potential Impacts for Consumer Electronics

A better mobile traffic prediction performance helps consumer electronics in several aspects. The tremendous adoption

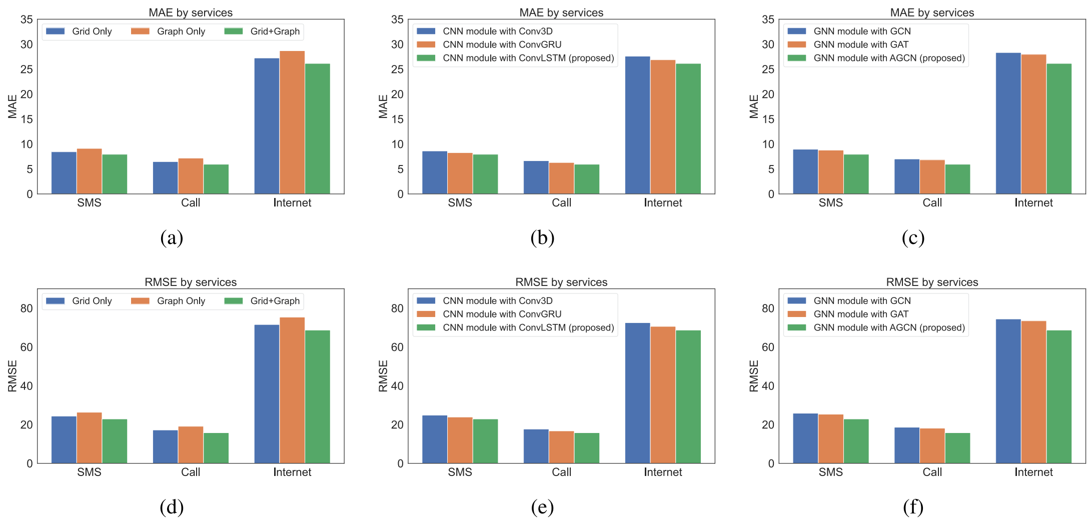


Fig. 7. The ablation study results. (a) MAE of different modals; (b) MAE of different CNN modules; (c) MAE of different GNN modules; (d) RMSE of different modals; (e) RMSE of different CNN modules; (f) RMSE of different GNN modules.

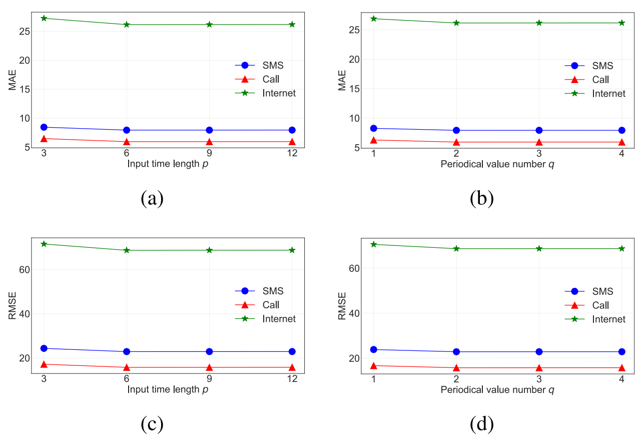


Fig. 8. The parameter influence results. (a) MAE of different input time lengths; (b) MAE of different periodical value numbers; (c) RMSE of different input time lengths; (d) RMSE of different periodical value numbers.

of IoT technologies brings a rapid proliferation of consumer electronics and ubiquitous information exchange is required in the growing interaction between the physical world and the cyber world. To provide a sustainable support for IoT-based consumer electronics, traffic prediction is required to manage network facilities and control network congestion, especially for those latency-sensitive applications, e.g., video streaming and virtual reality services. Another benefit of precise mobile traffic prediction is to facilitate network slicing in smart healthcare services, in which the traffic demand is predicted and the dedicated resources for each network slice are allocated to guarantee QoS. In the last but not the least, large-scale connections of consumer electronic devices lead to spectrum scarcity and huge energy consumption, which would tarnish the user experience and shorten the working life of consumer devices. The proposed traffic prediction scheme can better help device maintenance, when the device can sleep when the traffic demand is not very large to reduce energy consumption and extend the usable life of the device.

VI. Conclusion

Mobile traffic prediction is the basis for providing better consumer applications, e.g., short message, video streaming, and electronic commerce services. While important, this problem has not been fully solved considering the complex temporal and spatial dependencies in traffic data patterns. In this study, the mobile traffic prediction problem for consumer applications is considered using a real-world mobile dataset. The prediction problem is defined using a appropriate data representation method. A multimodal CNN-GNN framework is proposed as the solution, in which the CNN module adopts ConvLSTM, and the GNN module adopts AGCN as the basic model unit. Numerical experiments demonstrate that the proposed approach outperforms ten baselines for single-step mobile traffic prediction.

Some future research directions are considered as follows. This study only considers the short-term prediction for the next hour, which has limited impacts for improving consumer application facilities considering the long construction period. The first future research direction is to consider the long-term prediction, e.g., the next week or month, so that the communication infrastructure can be optimized based on the prediction outcomes, e.g., base station sleeping and scheduled service expansion, so that the consumer experience can be further improved. The second future research direction is to consider multi-step prediction, which is also beneficial for consumer applications, when consumers have some planned tasks that be scheduled based on their potential traffic consumption in the next few years, e.g., caching interested content when traffic load is not so heavy. The third future research direction is to further combine the proposed prediction method with consumer electronics for real-world deployment.

References

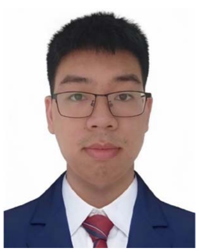
[1] C. K. Wu, C.-T. Cheng, Y. Uwate, G. Chen, S. Mumtaz, and K. F. Tsang, "State-of-the-art and research opportunities for next-generation consumer electronics," *IEEE Trans. Consum. Electron.*, early access, Dec. 27, 2022, doi: 10.1109/TCE.2022.3232478.

[2] K. Abbas, A. Nauman, M. Bilal, J.-H. Yoo, J. W.-K. Hong, and W.-C. Song, "AI-driven data Analytics and intent-based networking for orchestration and control of B5G consumer electronics services," *IEEE Trans. Consum. Electron.*, early access, Oct. 13, 2023, doi: 10.1109/TCE.2023.3324010.
[3] W. Jiang, H. Han, Y. Zhang, and J. Mu, "Federated split learning for sequential data in satellite-terrestrial integrated networks," *Inf. Fusion*, vol. 103, Mar. 2024, Art. no. 102141.
[4] W. Jiang, H. Han, M. He, and W. Gu, "ML-based pre-deployment SDN performance prediction with neural network boosting regression," *Expert Syst. Appl.*, vol. 241, May 2024, Art. no. 122774.
[5] J.-H. Syu, J. C.-W. Lin, G. Srivastava, and K. Yu, "A comprehensive survey on artificial intelligence empowered edge computing on consumer electronics," *IEEE Trans. Consum. Electron.*, early access, Sep. 22, 2023, doi: 10.1109/TCE.2023.3318150.
[6] P. Singh, "Systematic review of data-centric approaches in artificial intelligence and machine learning," *Data Sci. Manage.*, vol. 6, pp. 144–157, Sep. 2023.
[7] W. Jiang, "Cellular traffic prediction with machine learning: A survey," *Expert Syst. Appl.*, vol. 201, Sep. 2022, Art. no. 117163.
[8] P. R. Singh, V. Singh, R. Yadav, and S. Chaurasia, "6G networks for artificial intelligence-enabled smart cities applications: A scoping review," *Telematics Inform. Rep.*, vol. 9, Mar. 2023, Art. no. 100044.
[9] W. Fang, W. Zhang, W. Yang, Z. Li, W. Gao, and Y. Yang, "Trust management-based and energy efficient hierarchical routing protocol in wireless sensor networks," *Digit. Commun. Netw.*, vol. 7, no. 4, pp. 470–478, 2021.
[10] W. Jiang, "Internet traffic prediction with deep neural networks," *Internet Technol. Lett.*, vol. 5, no. 2, p. e314, 2022.
[11] W. Jiang and L. Zhang, "Geospatial data to images: A deep-learning framework for traffic forecasting," *Tsinghua Sci. Technol.*, vol. 24, no. 1, pp. 52–64, Feb. 2019.
[12] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.
[13] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–14.
[14] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–12.
[15] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Memory based fusion for multi-modal deep learning," *Inf. Fusion*, vol. 67, pp. 136–146, Mar. 2021.
[16] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1389–1401, Jun. 2019.
[17] G. O. Ferreira, C. Ravazzi, F. Dabbene, G. C. Calafiore, and M. Fiore, "Forecasting network traffic: A survey and tutorial with open-source comparative evaluation," *IEEE Access*, vol. 11, pp. 6018–6044, 2023.
[18] Y. Gao, K. M. Mosalam, Y. Chen, W. Wang, and Y. Chen, "Auto-regressive integrated moving-average machine learning for damage identification of steel frames," *Appl. Sci.*, vol. 11, no. 13, p. 6084, 2021.
[19] Q. Li, W. Wu, W. Zhang, and X. S. Shen, "Online traffic prediction in multi-RAT heterogeneous network: A user-cybertwin asynchronous learning approach," in *Proc. IEEE 34th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, 2023, pp. 1–7.
[20] S. Bothe, H. Farooq, J. Forgeat, and K. Cyras, "Time-series prediction using nature-inspired small models and curriculum learning," in *Proc. IEEE 34th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, 2023, pp. 1–6.
[21] J. Zhang, X. Hu, Z. Cai, L. Zhu, and Y. Feng, "Federated learning based on mutual information clustering for wireless traffic prediction," *Electronics*, vol. 12, no. 21, p. 4476, 2023.
[22] W. A. Aziz, I. Ioannou, M. Lestas, H. K. Qureshi, A. Iqbal, and V. Vassiliou, "Content-aware network traffic prediction framework for quality of service-aware dynamic network resource management," *IEEE Access*, vol. 11, pp. 99716–99733, 2023.
[23] F. Kavehmadavani, V.-D. Nguyen, T. X. Vu, and S. Chatzinotas, "Intelligent traffic steering in beyond 5G open RAN based on LSTM traffic prediction," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7727–7742, Nov. 2023.
[24] Y. Gao, J. Yang, H. Qian, and K. M. Mosalam, "Multiattribute multitask transformer framework for vision-based structural health monitoring," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 38, no. 17, pp. 2358–2377, 2023.
[25] C. Zhang, H. Zhang, D. Yuan, and M. Zhang, "Citywide cellular traffic prediction based on densely connected convolutional neural networks," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1656–1659, Aug. 2018.
[26] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 9, pp. 748–768, 2018.
[27] X. Sun, B. Wei, J. Gao, D. Cao, Z. Li, and Y. Li, "Spatio-temporal cellular network traffic prediction using multi-task deep learning for AI-enabled 6G," *J. Beijing Inst. Technol.*, vol. 31, no. 5, pp. 441–453, 2022.
[28] R. He, Y. Liu, Y. Xiao, X. Lu, and S. Zhang, "Deep spatio-temporal 3D densenet with multiscale ConvLSTM-Resnet network for citywide traffic flow forecasting," *Knowl.-Based Syst.*, vol. 250, Aug. 2022, Art. no. 109054.
[29] W. Zheng, Y. Li, M. Hong, G. Zhao, and X. Fan, "Network traffic matrix prediction with incomplete data via masked matrix modeling," *Inf. Sci.*, vol. 657, Feb. 2024, Art. no. 119835.
[30] K. He, X. Chen, Q. Wu, S. Yu, and Z. Zhou, "Graph attention spatial-temporal network with collaborative global-local learning for citywide mobile traffic prediction," *IEEE Trans. Mobile Comput.*, vol. 21, no. 4, pp. 1244–1256, Apr. 2022.
[31] N. Zhao, A. Wu, Y. Pei, Y.-C. Liang, and D. Niyato, "Spatial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 587–591, Mar. 2022.
[32] J. Shi, L. Cui, B. Gu, B. Lyu, and S. Gong, "State transition graph-based spatial–temporal attention network for cell-level mobile traffic prediction," *Sensors*, vol. 23, no. 23, p. 9308, 2023.
[33] Y. Zhang et al., "Adaptive graph convolutional recurrent neural networks for system-level mobile traffic forecasting," *China Commun.*, vol. 20, no. 10, pp. 200–211, Oct. 2023.
[34] S. Liu, M. He, Z. Wu, P. Lu, and W. Gu, "Spatial–temporal graph neural network traffic prediction based load balancing with reinforcement learning in cellular networks," *Inf. Fusion*, vol. 103, Mar. 2024, Art. no. 102079.
[35] Y.-D. Zhang et al., "Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation," *Inf. Fusion*, vol. 64, pp. 149–187, Dec. 2020.
[36] S.-H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y.-D. Zhang, "Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network," *Inf. Fusion*, vol. 67, pp. 208–229, Mar. 2021.
[37] Y.-D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, and S.-H. Wang, "Improved breast cancer classification through combining graph convolutional network and convolutional neural network," *Inf. Process. Manage.*, vol. 58, no. 2, 2021, Art. no. 102439.
[38] G. Barlacchi et al., "A multi-source dataset of urban life in the city of milan and the province of Trentino," *Sci. Data*, vol. 2, no. 1, pp. 1–15, 2015.
[39] W. Jiang, "Internet traffic matrix prediction with convolutional LSTM neural network," *Internet Technol. Lett.*, vol. 5, no. 2, p. e322, 2022.
[40] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
[41] S. Mortazavi and E. Sousa, "Efficient mobile cellular traffic forecasting using spatial-temporal graph attention networks," in *Proc. IEEE 34th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, 2023, pp. 1–6.
[42] H. Wu, X. Ma, and Y. Li, "Spatiotemporal multimodal learning with 3D CNNs for video action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1250–1261, Mar. 2022.
[43] T. Yu, Q. Kuang, and R. Yang, "ATMConvGRU for weather forecasting," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Sep. 2022.

Weiwei Jiang (Member, IEEE) received the B.Sc. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He is currently an Assistant Professor with the School of Information and Communication Engineering and the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications. His current research interests include artificial intelligence for networking and communication, satellite communication, and smart grid communication.

Yang Zhang received the B.E. degree in communication engineering from the Beijing University of Chemical Technology, Beijing, China, in 2023. He is currently pursuing the master's degree in communication engineering with the Beijing University of Posts and Telecommunications. His research interests include intersatellite routing and satellite communication networks.

Haoyu Han received the B.E. degree in electronic information engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2021. He is currently pursuing the master's degree in communication engineering with the Beijing University of Posts and Telecommunications. His research interests include satellite communication networks and satellite–ground integrated networks.



Zhaolong Huang received the B.S. degree from the College of Science, Tianjin University of Technology, Tianjin, China, in 2021, where he is currently pursuing the master's degree. His research interests include quantum computing, artificial intelligence, and UAV routing problem.

Qiting Li received the Ph.D. degree from the Department of Mathematical Sciences, Tsinghua University, Beijing, China, in 2017. He is currently a Senior Engineer with the Research and Development Department, China Academy of Launch Vehicle Technology. His current research interests include applied mathematics, quantum computing, and artificial intelligence.

Jianbin Mu received the B.Sc. degree from the Nanjing University of Science and Technology in 2013, and the Ph.D. degree from Shanghai Jiao Tong University in 2021. He is currently a Lecturer with the College of Information Engineering, Zhejiang University of Technology. His research interests include networked control systems and distributed model predictive control.