FMGCN: Federated Meta Learning-augmented Graph Convolutional Network for EV Charging Demand Forecasting

Linlin You, *Senior Member, IEEE,* Qiyang Chen, Haohao Qu, Rui Zhu, Jinyue Yan, Paolo Santi, Carlo Ratti

***Abstract*—Recent booming successes of electric vehicles (EVs) motivate emerging exploration of spatio-temporal EV charging demand forecasting to inform policy making. Recent studies have contributed to remarkable accuracy improvement by developing deep learning methods. However, when they access massive amounts of data and frequently exchange data through the Internet of Things (IoT), data silos and inefficient training emerge as main challenges. To tackle these challenges, this study proposes an integrated approach for regional EV charging demand forecasting, named FMGCN, which consists of two modules, namely 1) Spatio-temporal Learning module, which introduces spatial and temporal attentions to capture the underlying charging patterns between different regions and cities effectively; and 2) Distributed Pretraining module, which incorporates Federated Learning and Meta-Learning to enhance the adaptivity and generalisability of the forecasting model. A comprehensive evaluation based on a real-world dataset of 25,246 public EV charging piles shows that the proposed model outperforms other representative models with 1) an average improvement of 29.9% in forecasting errors; 2) an acceleration of 65% in convergence speed; and 3) a sound adaptability to support varying charging demand.**

***Index Terms*—Charging Demand Forecasting, Graph Convolution Networks, Federated Learning, Meta-Learning.**

I. Introduction

GROWING concerns about climate change are driving emerging exploration of the renewable energy transition, amongst which vehicle electrification plays a major role. This is due to the remarkable potential of electric vehicles (EVs) to reduce carbon emissions and mitigate global warming [1], [2]. Despite these impressive benefits, insufficient battery capacity and limited availability of charging infrastructure still remain significant challenges that hinder the proliferation of EVs. For instance, range anxiety can force EV drivers to charge their cars too frequently, placing additional load on the urban power grid [3]–[5]. This has sparked an upsurge of studies on EV charging demand forecasting to inform EV charging-related smart services that can improve energy and charging space efficiency [6], [7].

Thanks to recent successes of Internet of Things (IoT) techniques, sensing data can be collected and exchanged over interconnected devices, enabling spatio-temporal (ST) EV charging demand forecasting across urban regions and cities. Related to that, several challenges are emerging. First, compared to previous prediction methods, such as Recurrent Neural Network (RNN) and its variants [8], the feature learning capability should be enhanced to capture not just time-series patterns but also spatial correlations, given the increasing connectivity between urban areas. Furthermore, the process of model training requires data security considering the frequent data interactions between clients and servers in IoT environments. Last but not least, the generalisability of the method needs to be enhanced to enable demand prediction across different cities, from where the EV charging demand and related factors (e.g., weather and socio-economic data) exhibit non-IID (independent and identically distributed) patterns deteriorating model aggregation.

Many methods have been proposed to tackle the aforementioned challenges, namely 1) to capture spatio-temporal features by incorporating Recurrent Neural Networks (RNNs) and Graph Neural Networks (GNNs) [9]–[11]; 2) to bridge data silos while ensuring no raw data leakage by applying federated learning (FL) frameworks [12]–[14]; and 3) to transfer shared knowledge between cities by enabling the process of meta-learning [15], [16] for an acceleration of model aggregation and optimization. However, an integrated approach that supports the regional EV charging demand forecasting by building a private-preserving spatio-temporal model with a fast convergence speed and high generalisability is still missing to date.

To fill the gap, a novel approach for real-time demand forecasting, named Federated Meta Learning-based Graph Convolutional Network (FMGCN) is poposed with two modules, i.e., Spatio-temporal Learning and Distributed Pretraining. To be specific, inspired by the attention-based graph convolutional network [17], a graph convolution network (GCN)

L. You and Q. Chen are with the School of Intelligent Systems Engineering, Sun Yat-Sen University, Shenzhen, 510006, China

H. Qu is with Department of Computing, Hong Kong Polytechnic University, Hong Kong, 100872, China

R. Zhu is with Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore 138632, Singapore

J. Yan is with Energy and Buildings, Hong Kong Polytechnic University, Hong Kong, 100872, China

P. Santi and C. Ratti are with the Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139 USA

This paper was supported by the GuangDong Basic and Applied Basic Research Foundation (2023A1515012895), the National Natural Science Foundation of China (62002398), the National Research Foundation Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019), and the RIE 2020 Advanced Manufacturing and Engineering (AME) Programmatic Fund (No. A20G8b0102), Singapore. (*Corresponding author, e-mail: zhur@ihpc.a-star.edu.sg)

model with spatio-temporal attention for regional charging demand prediction (GCNSA) is designed as the backbone. Furthermore, by incorporating federated meta-learning [18], the Distributed Pretraining module is designed and implemented to bridge data silos and enable rapid localization.

Through a comprehensive evaluation based on a real-world dataset of 25,246 EV charging piles in six cities in the Greater Bay Area (GBA) of China, from 11th December 2022 to 14th January 2023 (35 days), the efficiency and effectiveness of the proposed model are tested and compared with other SOTA methods. In particular, the results show that FMGCN outperforms other representative prediction models and distributed training strategies with 1) a reduction of forecasting errors by 37%, 36%, 22%, and 25% in MAE, RMSE, MAPE, and R-square, respectively; 2) an acceleration of convergency speed by 62%; and 3) a sound adaptation to different cities with various changing patterns of EV charging demands.

To sum up, the main contributions of this paper include:

- Adaptation of an effective prediction model with a spatio-temporal attention mechanism for EV charging demand forecasting, which can handle not only underlying time-series patterns but also potential correlations between regions and cities.
- Application of a distributed pretraining step incorporated with Federated Learning and Meta-learning to improve the training performance with data silos bridged, forecasting errors remedied, and convergence speed accelerated.
- An integration of GCN and Federated Meta-learning, which enables knowledge learning and propagation among cities with different feature patterns, thus building a high-quality predictor with high generalisability.

The remainder of this paper is structured as follows. Sections II and III provide overviews of related work and preliminary material, respectively. Then, Section IV introduces the proposed approach FMGCN, which is evaluated in Section V. Finally, Section VI draws conclusions and future works.

II. Literature Review

In this section, the emerging challenges and solutions associated with EV charging demand forecasting and the applied techniques are summarized. Additionally, the main abbreviations used in FMGCN are listed in Table I for the readability of the method of FMGCN.

TABLE I: The main abbreviations used in FMGCN

Notation	Description
GCN	Graph Convolution Network
GNN	Graph Neural Network
ASTGCN	Attention-based Spatio-temporal Graph Convolution Network
STA	Spatio-temporal Attention
SC	Spatial Convolution
TC	Temporal Convoltuion
LD	Linear Decoder
GCNSA	Graph Convolution Network Model with Spatio-temporal Attention
FOMAML	First-order Model-agnostic Meta-learning
FedAvg	Federated Averaging
FMGCN	Federated Meta Learning-based Graph Convolutional Network

A. Emerging Challenges

In general, to establish an efficient and effective model for EV charging demand forecasting, several challenges are emerging, including:

- C.1 Feature extraction: Advanced sensing and IoT techniques can equip prediction methods with more diverse information [19]. However, given not only temporal but also spatial data, EV charging demand forecasting models are required to extract and capture the underlying patterns and correlations properly.
- C.2 Data security: With practical considerations, the model training process has to be optimized to prevent raw data leakage, when exchanging important information (e.g., charging records) through IoT protocols.
- C.3 Knowledge sharing and adaptation: The utilization patterns of EV charging spaces in different regions and cities may vary from each other. Therefore, the distributed and collaborative training step becomes crucial to enable efficient knowledge learning with non-IID data, so that the model can be easily adapted to downstream tasks.

B. Existing Solutions

In recent years, the growing number of EVs has sparked a research surge in charging demand forecasting. Initially, the main focus of most early methods remained on the intrinsic patterns hidden in historical time series [20]. E.g., [21] applied Support Vector Regression (SVR) to model statistical features underlying the power load on EV charging stations, while [22] combined Fuzzy Clustering, Least Squares Support Vector Machine, and Wolf Pack Algorithm for further improvement in predictive accuracy. Although these statistical and traditional machine learning models can be computationally efficient and easy to interpret, there is still a limitation in expressing high-dimensional and non-linear features.

More recently, with the booming successes of deep learning techniques, several studies have proposed to leverage RNNs to extract complex temporal features. Examples include Long Short-Term Memory (LSTM) [23] and Grated Recurrent Unit (GRU) [24], which have contributed to a remarkable improvement of feature learning ability in various forecasting tasks, including charging load prediction. Nevertheless, the neglect of spatial correlations limits the further development of these methods.

To fill the gap, some pioneering work has attempted to employ GNNs, such as GCNs [25]–[27], which can incorporate features hidden in vertexes and edges of a graph structure to model the dependencies of data in the spatial dimension. Specifically, the integration of GNNs and RNNs has gained significant popularity in the field of spatio-temporal forecasting scenarios, such as traffic flow, traffic speed, and

parking availability predictions [28]–[30]. Without exception, it is also used to forecast EV charging demands, e.g., SGCN [31] combines GRU and GCN to model temporal and spatial features of the operating status at EV charging stations, respectively, to better assist the prediction. Furthermore, motivated by model structure optimization, a few studies on traffic flow prediction, such as ASTGCN [17], have facilitated a more flexible and implicit allocation of weights through the use of an advanced technique known as the attention mechanism [32]. Despite these benefits, the potential challenges in data security and training efficiency are overlooked.

Taking into account frequent data exchange, the training process of the spatio-temporal models is required to be not only efficient but also secure in real-world scenarios, where EV drivers are reluctant to disclose their records and city administrators are restricted in data sharing. Various Federated Learning (FL) methods, e.g., FedAvg [33], FedProx [34], FedALA [35], and so on [36]–[39], have been proposed to address the training limitations that arise from data islands and data security concerns. The key idea of these methods is to engage in local data training on each edge device and then upload the updated local models to a central server for global model aggregation. Among these, FedAvg [33] is one of the most widely-used frameworks, due to its simple but effective structure. However, these FL methods still struggle with the problem of non-IID data. In other words, heterogeneous data from different regions and cities have different characteristics that could affect the distributed training process.

For the purpose of facilitating knowledge sharing as well as personalised adaptation, a novel technique known as model-agnostic meta-learning (MAML) [40] starts to be incorporated with the FL frameworks. Unlike previous federated learning, the federated meta-learning framework focuses on training a meta-model with well-optimized initial parameters to enable rapid local adaptation. Given that, the meta-model can adapt to diverse domains and tasks to achieve exceptional predictive performance and high generalisability. However, limited studies have exploited such a framework in the application of EV charging demand forecasting.

As summarized in Table II, which evaluates the reviewed solutions based on their abilities to address the three emerging challenges, the integration of GNNs and RNNs (i.e., STGCN, and ASTGCN) can outperform the typical deep learning models (i.e., SVM, GRU, and LSTM) in spatio-temporal feature modeling. However, the training process in the data-isolated and data-heterogeneous scenarios still remains under-explored for regional EV charging demand forecasting. To tackle the emerging challenges and fill the research gap, this paper proposes a novel approach, named FMGCN, which integrates and enhances ASTGCN, FedAvg, and MAML to more efficiently and effectively support regional EV charging demand forecasting.

TABLE II: The summary of literature review
(● Supported; ○ Not Supported)

Model	CDF[1]	GCN[1]	FL[1]	PA[1]
SVM [20]–[22]	●	○	○	○
LSTM [23]	●	○	○	○
GRU [24]	●	○	○	○
ChebNet [26]	○	●	○	○
STGCN [27]	○	●	○	○
ASTGCN [17]	○	●	○	○
FedAvg [33]	○	○	●	○
FedProx [34]	○	○	●	●
FedAMP [36]	○	○	●	●
FedRep [37]	○	○	●	●
FedALA [35]	○	○	●	●
MAML [40]	○	○	○	●
FMGCN (Ours)	●	●	●	●

[1] CDF, GCN, FL and PA represent charging demand forecasting, graph convolution networks, federated learning and personalised adaptation, respectively.

III. Preliminary

For consistency, this paper uses lowercase letters (e.g., x) to represent scalars, bold lowercase letters (e.g., $\mathbf{x}$) to denote column vectors, bold-face uppercase letters (e.g., $\mathbf{X}$) to denote matrices or high-order tensors and uppercase calligraphic letters (e.g., $\mathcal{X}$) to denote sets. Besides, $\mathbb{R}$ is used to represent the data space, e.g., $\mathbf{X} \in \mathbb{R}^{M \times N}$ means the matrix $\mathbf{X}$ has two dimensions of M and N, while X_{mn} represents an element of m_{th} row and n_{th} column of X. Moreover, the transposition of $\mathbf{X}$ is denoted as $\mathbf{X}^{\mathrm{T}}$.

On this basis, the notations and preliminaries are defined in the order of graph structure, data silos and heterogeneity, and objective formulation.

First, in a city, the studied areas can be structured as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ represent the sets of nodes (e.g., districts or regions) and edges (i.e., neighboring relationships between two nodes). Given N nodes in the set $\mathcal{V}$, the adjacency matrix of the city can be defined as $\mathbf{A} \in \mathbb{R}^{N \times N}$, where $\mathbf{A}_{ij} \neq 0$ if and only if there exists an edge between nodes V_i and V_j in $\mathcal{G}$. Besides, the edges in the set $\mathcal{E}$ are binary scales, determined by whether the two urban regions are adjacent (1) or not (0). By introducing other related factors (e.g., weather and socio-economic conditions) to enhance the forecasting of EV charging demands, the dataset can be denoted as $\mathcal{D} = \mathbf{X}, \mathbf{y}$, where $\mathbf{X} \in \mathbb{R}^{N \times F \times T}$ means the input data of N nodes, F features (including the demand), and T time intervals, while $\mathbf{y} \in \mathbb{R}^N$ are the near-future EV charging demands of the N nodes in the studied city.

Second, when extracting data and knowledge from other cities, in which related data are distributed differently, the problem of data isolation and heterogeneity needs to be considered and defined. Assume that there are M cities as federated clients, the dataset and graph of city m can be denoted by $\mathcal{D}_m$ and $\mathcal{G}_m$, and the same will go for all the variables of this city. Note that, the data of these cities are non-IID, and the graphs are different (i.e., $\mathcal{V}_a \neq \mathcal{V}_b$ and $\mathcal{E}_a \neq \mathcal{E}_b$ for $a \neq b$). Since data is isolated, the raw data of federated

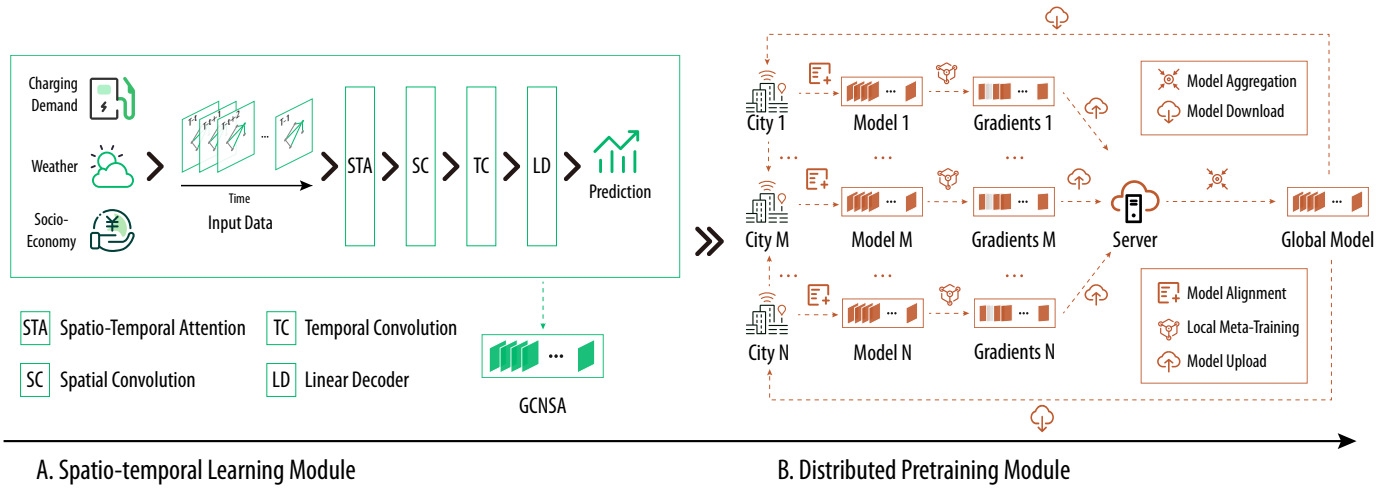


Fig. 1: The overall architecture of the proposed mechanism FMGCN. (A) Spatio-temporal Learning module used to construct the backbone GCN model with temporal and spatial attentions to extract forecasting features from multiple information sources; and (B) Distributed Pretraining module used to train the meta-model through the collaboration between the server and source stations under the IoT scenario that data are isolated and heterogeneous.

clients can only be processed locally, while generated model parameters or gradients can be exchanged between the central server and the local clients.

Finally, in contrast to the objective function in the setting of traditional deep learning formulated as $\min L(\mathbf{y}_m, \mathbf{F}(\mathcal{D}_m, \mathcal{G}_m))$, where only local data is used, the objective function of our task in the setting of federated graph learning to minimize error in predicting the EV charging demand for all collaborated cities can be formulated as:

$$\min \sum_{m=1}^{M} L(\mathbf{y}_m, \mathbf{F}(\mathcal{D}_m, \mathcal{G}_m, \mathbf{F}'(\mathcal{D}, \mathcal{G}))) \tag{1}$$

where $\mathbf{F}$ and $\mathbf{F}'$ represent the personalised and pretrained forecasting models, respectively; L denotes the loss function. To be specific, in such a setting, it is permitted to leverage all data and graphs in a privacy-preserving and collaborative manner to conduct a pretraining step for a globally shareable meta-model and finetune it for a personalised model, so as to achieve higher local forecasting performance. Note that, the process is required to be performed according to the workflow of FL under the constraints of local data protection, i.e., no raw data exchange.

IV. Methodology

As shown in Figure 1, the proposed approach consists of two modules, i.e., 1) a Spatio-temporal Learning module, which utilises both temporal and spatial attention to capture the spatio-temporal dependencies underlying EV charging demands and other factors (e.g., weather and socio-economic information); and 2) a Distributed Pretraining module, which empowers the model with high generalisability and fast convergence speed by incorporating Federated Learning and Meta-learning. In the following sections, the two modules will be described.

A. Spatio-temporal Learning Module

Considering the fluctuations in real-world demand for EV charging, the backbone model is designed as illustrated in Figure 2. In general, it consists of four components, namely spatio-temporal attention, spatial convolution, temporal convolution, and linear decoder. Specificially, spatio-temporal attention is responsible for calculating and enhancing the correlation of charging demand in spatio-temporal dimensions, followed by the spatial and temporal convolutions to extract the charging features across both dimensions, and the final prediction is produced by the linear decoder.

Moreover, it is worth noting that the GCNSA integrates the socio-economic data and weather data as model input to assist the charging demand forecasting. On the one hand, city charging demand is highly correlated with its social economy, and the more prosperous the city, the greater the charging demand for electric vehicles. On the other hand, weather conditions will affect people's desires to travel, and in turn, influnce the urban charging demand. Therefore, socio-economic data, weather data and the charging demand data are embedded as model input to improve the forecasting accuracy of the charging demand, and their impacts are analyzed in Section V-B4. Note that all calculations are equal for each city m, and thus for simplicity and readability, the subscript m is omitted in this module.

1) Spatio-temporal Attention: First, an attention mechanism [41] is used to adaptively capture the dynamic correlations between nodes in the spatial dimension, which can be calculated by (2),

$$\mathbf{S} = \mathbf{W}^s \cdot \sigma\left((\mathbf{X}\mathbf{U}_1)\mathbf{U}_2(\mathbf{U}_3\mathbf{X})^{\mathrm{T}} + \mathbf{B}^s\right) \tag{2}$$

where $\mathbf{S}$ denotes the spatial attention matrix that is dynamically computed according to the input $\mathbf{X} \in \mathbb{R}^{N\times F\times T}$; $\mathbf{W}^s, \mathbf{B}^s \in \mathbb{R}^{N\times N}$, $\mathbf{U}_1 \in \mathbb{R}^T$, $\mathbf{U}_2 \in \mathbb{R}^{F\times T}$, and $\mathbf{U}_3 \in \mathbb{R}^F$ are learnable parameters; σ denotes the sigmoid activation function. Then, a softmax function is utilised to guarantee that

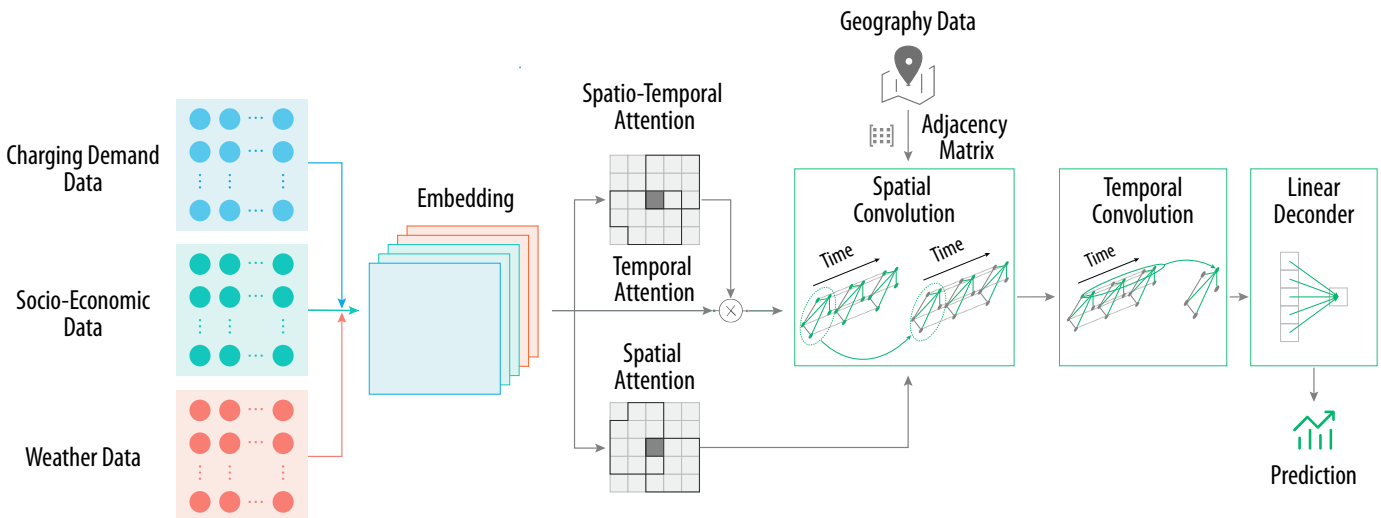


Fig. 2: The structure of the GCN model with spatio-temporal attention for regional charging demand prediction (GCNSA).

the attention weights of a node are summed up to one. To be specific, for node V_i, its spatial attention coefficient for each neighbor node V_j can be normalized by (3).

$$\mathbf{S}'_{ij} = \mathrm{softmax}(\mathbf{S}_{ij}) = \frac{\exp(\mathbf{S}_{ij})}{\sum_{n=1}^{N}\exp(\mathbf{S}_{in})} \tag{3}$$

Accordingly, the temporal attention matrix can be computed with a similar method as described in Formula (4),

$$\mathbf{\Gamma} = \mathbf{W}^{\Gamma}\cdot\sigma\left((\mathbf{X}^{\mathrm{T}}\mathbf{V}_1)\mathbf{V}_2(\mathbf{V}_3\mathbf{X}) + \mathbf{B}^{\Gamma}\right) \tag{4}$$

where $\mathbf{W}^{\Gamma}, \mathbf{V}^{\Gamma} \in \mathbb{R}^{T\times T}$, $\mathbf{V}_1 \in \mathbb{R}^{N}$, $\mathbf{V}_2 \in \mathbb{R}^{F\times N}$ and $\mathbf{V}_3 \in \mathbb{R}^{F}$ are trainable parameters as well. The normalized temporal attention score of time i to time j can be calculated by (5),

$$\mathbf{\Gamma}'_{ij} = \mathrm{softmax}(\mathbf{\Gamma}_{ij}) = \frac{\exp(\mathbf{\Gamma}_{ij})}{\sum_{t=1}^{T}\exp(\mathbf{\Gamma}_{it})} \tag{5}$$

Finally, we directly apply the normalized temporal attention matrix $\mathbf{\Gamma}'$ to the input $\mathbf{X}$ and get $\mathbf{X}' \in \mathbb{R}^{N\times F\times T}$ to dynamically adjust the input by merging relevant information, which can be formulated as (6).

$$\mathbf{X}' = \mathbf{X}\mathbf{\Gamma}' \tag{6}$$

2) Spatial Convolution: In this study, the city connected by urban areas can be considered a graph structure in nature, and the features of each node can be regarded as the signals on the graph. Hence, in order to make full use of the topological relationships among urban areas, at each time slice, we adopt graph convolutions based on spectral graph theory [42] to directly process the signals, exploiting signal correlations on EV charging distribution in the spatial dimension.

First of all, differently from ASTGCN [17], we introduce the distance between neighboring nodes to augment the adjacency matrix, i.e., $\mathbf{A}_{ij}$ is set to the shortest path distance between nodes V_i and V_j. Given that, the Laplacian form of the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ is calculated by (7), where $\mathbf{L} \in \mathbb{R}^{N\times N}$, $\mathbf{D} \in \mathbb{R}^{N\times N}$, and $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ denote the Laplacian, degree, and unit matrices, respectively.

$$\begin{aligned}\mathbf{L} &= \mathbf{D} - \mathbf{A}\\ &= \mathbf{I}_N - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\end{aligned} \tag{7}$$

Thereafter, due to the orthogonality, symmetry, and non-negative eigenvalues of the Laplacian matrix, we can decompose the Laplacian matrix as reported in (8), where $\mathbf{U}$ is Fourier basis and Λ is a diagonal matrix of eigenvalues:

$$\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}}\ . \tag{8}$$

Based on this, the signal of the f-th channel at time t on the graph $\mathcal{G}$, denoted by $\hat{\mathbf{x}} = \mathbf{X}'_{tf} \in \mathbb{R}^{N}$, can be filtered by a kernel g_θ, which can be formulated as (9), where $*_G$ denotes the graph convolution operation.

$$\begin{aligned}g_\theta *_G \hat{\mathbf{x}} &= \mathbf{U}((\mathbf{U}^{\mathrm{T}}g_\theta)(\mathbf{U}^{\mathrm{T}}\mathbf{X}'))\\ &= \mathbf{U}(g_\theta(\Lambda)(\mathbf{U}^{\mathrm{T}}\mathbf{X}'))\\ &= \mathbf{U}g_\theta(\Lambda)\mathbf{U}^{\mathrm{T}}\mathbf{X}'\end{aligned} \tag{9}$$

However, it is computationally expensive to directly perform the eigenvalue decomposition on the Laplacian matrix when the size of the graph is large. Therefore, Chebyshev polynomials are adopted in this paper to solve this problem approximately but efficiently, according to [43], which can be written as (10),

$$g_\theta *_G \hat{\mathbf{x}} = \sum_{k=0}^{K-1}\theta_k C_k(\tilde{\mathbf{L}})\hat{\mathbf{x}} \tag{10}$$

where $\theta \in \mathbb{R}^{K}$ is a vector of polynomial coefficients, while K is the order of graph propagation; $\tilde{\mathbf{L}} = \frac{2}{\lambda_{max}}\mathbf{L} - \mathbf{I}_N$, λ_{max} is the maximum eigenvalue of the Laplacian matrix; and $C_k(x) = 2xC_{k-1}(x) - C_{k-2}(x)$ is the recursive definition of the Chebyshev polynomial, given $C_0 = 0$ and $C_1(x) = x$. Using approximate expansion of Chebyshev polynomial to solve this formulation corresponds to extracting information of the surrounding 0 to $(K-1)$-order neighbors centered on each node in the graph.

To account for spatial dependencies between nodes, for each term of Chebyshev polynomial, $\mathbf{C}_k(\hat{\mathbf{L}})$ is accompanied

with the normalized spatial attention matrix $\mathbf{S}' \in \mathbb{R}^{N\times N}$. Thus, the graph convolution in (10) changes to Formula (11), where $\odot$ denotes the Hadamard product. This definition can be generalised to the graph signal with multiple data channels (i.e., features).

$$g_\theta *_G \hat{\mathbf{x}} = \sum_{k=0}^{K-1} \theta_k C_k(\tilde{\mathbf{L}} \odot \mathbf{S}')\hat{\mathbf{x}} \tag{11}$$

3) Temporal Convolution: After the graph convolution operations that have captured spatial correlations for each node on the graph, a standard convolution layer in the temporal dimension is further stacked to update the signal of a node by merging the knowledge at consecutive time slices, which can be formulated as (12).

$$\mathbf{X}'' = \mathrm{ReLU}(\mathbf{\Phi} * (\mathrm{ReLU}(g_\theta *_G \mathbf{X}'))) \tag{12}$$

where $\mathbf{X}'' \in \mathbb{R}^{N\times T'\times F'}$ is the output of the temporal convolution layer; T' and F' are the changed dimension number of time slices and channels determined by the size of convolution kernels; $*$ denotes a standard convolution operation, while $\mathbf{\Phi}$ is the parameter of the temporal dimension convolution kernel; and the action function is ReLU.

4) Linear Decoder: Finally, after the stacked attention and convolution layers, a decoder layer based on Fully Connected Neural Network is performed to obtain final prediction, which is calculated in (13), where $\mathbf{W}^\Upsilon$ is the learnable parameters.

$$\mathbf{y} = \mathbf{W}^\Upsilon \odot \mathbf{X}'' \tag{13}$$

To sum up, as shown in Figure 2, a GCN model with spatial-temporal attention is designed as the backbone model to extract the complex patterns of different urban areas that vary across space and time, so as to achieve accurate prediction of EV charging demand.

B. Distributed Pretraining module

Although the recent surge in increasing model size has contributed to remarkable improvement in accuracy, deep learning methods require tremendous data for training, which poses new challenges for data security and learning efficiency [44]. In the case of EV charging demand forecasting, as data from different regions and cities are sensitive and non-IID, not only data exchange should be secured to prevent data leakage but also negative transfer needs to be mitigated to facilitate knowledge sharing [45]. To tackle these issues, building upon the designed Spatio-temporal Learning module, a collaborative training mechanism is developed by incorporating Federated Learning [33] and Meta-Learning [40].

Assume that there are M clients, this module works with the following four steps, namely Model Alignment, Local Meta-training, Global Model Updating and Model Personalisation. Specifically, model alignment is initially conducted to ensure consistency in model structures across different cities. Subsequently, local meta-training is designed to perform meta-learning locally instead of centralised training at the server so that the raw data can be better safeguarded without transmitting. After that, different from conventional centralised training methods, global model updating is devised to aggregate individual meta-models from varied cities at the server to generate the global meta-model. Finally, after sufficient rounds of federated meta-training, model personalisation ends up with training and deploying a personalised model for specific forecasting tasks.

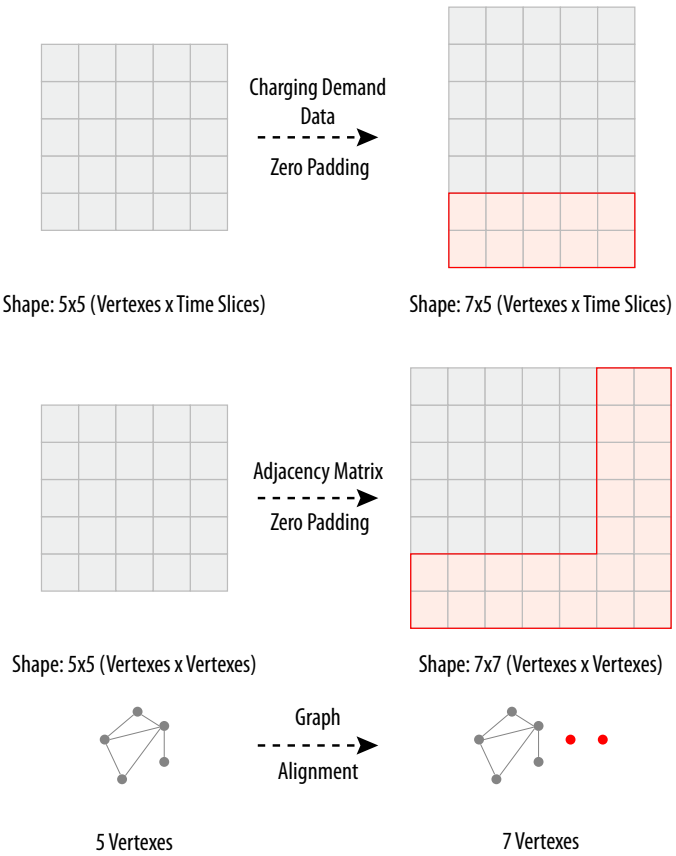


Fig. 3: The process of model alignment by using graph padding. Assume that the predetermined node number $P = 7$, thus the charging demand data and the adjacency matrix need to zero padding until meeting the requirements of P nodes.

1) Model Alignment: The graphs are heterogeneous across cities due to the varying number of nodes. This poses challenges for model aggregation in distributed learning. To enable knowledge sharing, the model of each client is first aligned by graph padding. To be specific, as shown in Figure 3, we zero-pad the graph (e.g., its adjacency matrix) of each client to a predetermined number of nodes P to maintain consistency in the model input dimension.

2) Local Meta-training: Under the distributed training framework of Federated Learning, a simple but effective meta-learning method, i.e., First-order Model-agnostic Meta-learning (FOMAML) [40], is used as a means to enhance the model generalisation capability. Before executing the meta-training, the training set of each city will be divided into support set $\mathcal{D}_m^s$ and query set $\mathcal{D}_m^q$ for local knowledge extraction and global gradient aggregation. In general, the support set $\mathcal{D}_m^s$ is responsible for learning the overall optimization directions across all tasks, while the query set $\mathcal{D}_m^q$ focuses

on learning personalised optimization directions specific to individual tasks. In this study, the support set $\mathcal{D}_m^s$ and query set $\mathcal{D}_m^q$ account for 80% and 20% of the training dataset $\mathcal{D}_m$, respectively.

First, in each epoch, we update the prediction model locally. The calculation can be described in (14):

$$\theta_m = \theta - \alpha \nabla_\theta L(\theta, \mathcal{D}_m^s) \ . \tag{14}$$

where θ_m is the local model of client m trained based on the global model θ and support set $\mathcal{D}_m^s$; L is the loss function; ∇_θ denotes the derivative based on θ; and α is the learning rate of local meta-training.

Given all clients, we then collect their local gradients, instead of model parameters, from the query set $\mathcal{D}_m^q$. The process can be written as (15):

$$\delta_m = \nabla_{\theta_m} L(\theta_m, D_m^q) \ . \tag{15}$$

where δ_m is the gradient obtained from client m, and ∇_{θ_m} denotes the derivative based on θ_m.

3) Global Model Updating: Since there are differences in the amount of data used by each model during the local meta-training process, we assign weights for model aggregation based on the amount of training data used by the clients. Specifically, the more the training data, the higher the weight used for model aggregation. Formally, the weight can be calculated according to (16):

$$\omega_m = \frac{\zeta_m}{\sum_{l=1}^{M} \zeta_l} \ . \tag{16}$$

where ζ_m and ω_m represent the data size and assigned weight of client m, respectively.

After the weight calculation, the central server starts the global model updating by aggregating collected local gradients, which can be formulated as (17):

$$\theta' = \theta - \frac{\beta}{M} \sum_{m=1}^{M} \omega_m \delta_m \ . \tag{17}$$

where θ' is the updated global model, and β is the learning rate of global updating.

4) Model Personalisation: In this step, the server distributes a well-trained global meta-model ϕ (which is θ' after the above-described learning process) and the signal of personalised training to each client. Given the rich knowledge encoded in the global meta-model, when encountering unfamiliar tasks and their associated datasets D_p, local clients can train a personalised model rapidly according to (18):

$$\phi_p = \phi - \gamma \nabla_\phi L(\phi, D_p) \ . \tag{18}$$

where ϕ_p is the personalised model of client p, and γ is the learning rate of personalised training.

C. Algorithms of the proposed approach

For the sake of clarity, the proposed approach FMGCN can be depicted with two pseudocodes for server and clients, respectively, as illustrated in Algorithm 1 and 2.

Algorithm 1 The pseudocode for server in FMGCN

FL Training Mode

Send the predefined graph node number to clients
for m=1,2,...,M **do**
  Transfer θ to clients
  Receive θ_m and ζ_m from clients
  Calculate ω_m according to (16)
  Update the global model θ according to (17)
  if Stop condition is reached **then**
    Switch to personalisation mode
  end if
  Send control signal of continuous learning and updated global model θ' to all clients
end for

Personalised Mode

Deliver the control signal of personalisation to all clients
Deliver the global meta-model ϕ to all clients

Algorithm 2 The pseudocode for clients in FMGCN

FL Training Mode

Conduct model alignment and data alignment based on the predefined graph node number
while True **do**
  Receive control signal
  if The control signal of personalisation is reached **then**
    Break
  end if
  Receive global model of last training epoch θ
  Sample training data of ζ_m size in local data
  Update the model θ_m according to (14) and (15)
  Upload θ_m, and ζ_m to the server
end while
Switch to personalisation mode

Personalised Mode

Receive the global meta-model ϕ
Initialize the local model as ϕ
Update the personalised model ϕ_j as (18)
Apply model ϕ_j to real forecasting

To sum up, a novel approach for EV charging demand forecasting, called FMGCN is proposed with two dedicated modules. On the one hand, a GCN model for regional EV charging demand prediction is designed to capture the underlying patterns that vary across time and space. On the other hand, a collaborative training procedure based on Federated Meta-Learning is implemented to tackle the challenges of data isolation and data heterogeneity. By integrating these two modules, the barriers can be broken down to train a shareable global meta-model in a private-preserving manner, and thus the personalised model for each city can be rapidly built by finetuning the meta-model to make more accurate predictions.

TABLE III: The overall descriptions of evaluation data

City	Node	Edge	Station	Pile	GDP (Billion Yuan)	Population (Million)
Guangzhou	11	21	770	6,152	2,823.20	20.95
Shenzhen	10	16	1,220	14,241	3,059.39	18.95
Foshan	5	6	278	2,048	1,215.65	10.72
Dongguan	32	73	266	2,019	1,120.03	10.44
Zhuhai	3	3	78	432	388.18	2.63
Zhongshan	20	43	61	354	289.05	3.96
Total	81	162	2,673	25,246	8,895.50	67.65

[1] Evaluation data range from December 11th, 2022 to January 14th, 2023, with a 30-minute interval.

V. Performance Evaluation

In this section, the proposed method is tested together with other state-of-the-art forecasting methods under the same evaluation settings. Moreover, the results are analysed to demonstrate the improvements achieved by FMGCN.

A. Evaluation Preparation

1) Data Description: For evaluation purposes, a common dataset is created based on the EV charging demand, geographic information, socio-economic indicators, and weather conditions in six cities of the Great Bay Area (GBA), China, namely Guangzhou, Shenzhen, Foshan, Dongguan, Zhuhai, and Zhongshan. Besides, the studied period is from December 11, 2022, to January 14, 2023. The overall description of the dataset is summarised in Table III. In general, the dataset contains four types of information:

- Charging Demand: The data is organised by the EV charging records of 25,246 piles located in the studied cities. To be specific, the records are aggregated into charging demand with a minimum interval of 30 minutes.
- Geographic Information: The studied regions in each city are defined according to the boundaries of its subordinate administrative districts, in which the region center is calculated by applying a clustering method (k-Nearest Neighbor, KNN) on the distribution of charging station and the edge between neighboring centers is connected based on the shortest path. Accordingly, the region centers and the shortest paths are set as the nodes $\mathcal{V}$ and edges $\mathcal{E}$ in the graph $\mathcal{G}$ (which may vary among cities).
- Socio-economic Indicators: They consist of the population of each sub-region and the gross domestic product (GDP) of each city, which are the best factors reflecting the prosperity of a region. These two indicators are collected from the 2022 government statistical reports.
- Weather Condition: It is represented by the maximum and minimum temperatures in each region.

For evaluation purposes, the dataset is divided into a training set (from 11 December 2022 to 7 January 2023, 28 days) and a test set (from 8 to 14 January 2023, 7 days). Furthermore, according to the setting of meta-learning, the training set is further subdivided into support set and query set, which occupy 80% and 20% of the data, respectively. Note that the dataset used in this paper can be accessed at the link [1].

[1]https://github.com/chenqy87/FMGCN/tree/main/FMGCN_data

2) Metrics: Four metrics are adopted to compare the prediction performance, i.e., Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Coefficient of Determination (R^2), which are defined in Formula (19):

$$\begin{aligned}
\mathrm{MAE} &= \sum_{i=1}^{N} |y_i^{'} - y_i| \\
\mathrm{RMSE} &= \sqrt{\frac{1}{n}\sum_{i=1}^{N}(y_i^{'} - y_i)^2} \\
\mathrm{MAPE} &= \frac{100\%}{n}\sum_{i=1}^{N}\left|\frac{y_i^{'} - y_i}{y_i}\right| \\
\mathrm{R}^2 &= 1 - \frac{\sum_{i=1}^{N}(y_i^{'} - y_i)^2}{\sum_{i=1}^{N}(\bar{y_i} - y_i)^2}
\end{aligned} \tag{19}$$

where $y_i^{'}$ and y_i denote the predicted value and the real value of the demand, respectively. Note that the value domains of MAE, RMSE, MAPE, and R2 are $[0, +\infty)$, $[0, +\infty)$, $[0, +\infty)$ and $(-\infty, 1]$, respectively. In addition, to reduce the random error, each prediction task will run ten times separately, and the averaged value will be used as the final result.

3) Compared Models: The proposed method FMGCN is compared with four neural networks (NNs), two statistical models, and one machine learning model, namely:

- **HA:** Historical Average method. Here, we use the average value of the last 12 time slices to predict the next value;
- **ARIMA [46]:** Auto-Regressive Integrated Moving Average is a well-known time series analysis method for predicting future values;
- **SVR [47]:** Support Vector Regression is a traditional machine learning method for regression tasks, which is derived from Support Vector Machine (SVM);
- **GRU [48]:** Gate Recurrent Unit is an effective recurrent neural network (RNN) simplified from Long Short-term Memory (LSTM);
- **ChebNet [49]:** It is an early graph convolution network (GCN), applying Chebyshev Polynomial as the convolution kernel;
- **STGCN [50]:** It is a recently developed Graph Neural Network (GNN), considering the impacts of spatio-temporal characteristics;
- **GCNSA [51]:** It is the backbone model implemented in FMGCN.

Furthermore, the collaborative training procedure implemented in FMGCN is compared with other ten representative training strategies, namely:

- **Separate:** A simple training strategy, which is to train the model local data only;
- **FedAvg [33]:** A typical decentralized training strategy that aggregates and averages local model parameters or gradients for global model updating;
- **FedProx [34]:** A special federated learning (FL) framework improved based on FedAvg, which resolves data heterogeneity by adding proximal terms;
- **pFedMe [52]:** A popular FL method that aims at personalising model for each client;
- **FedAMP [36]:** An FL method addressing data heterogeneity by maintaining a personalised model for each client at the server;
- **FedDyn [53]:** A dynamic FL method assisting the model to converge to the global optimum in an efficient path;
- **FedRep [37]:** An FL mechanism boosting the model's generalisability by learning shared representation;
- **FedFomo [54]:** An FL method that adapts local models based on different contributions of participated clients to the target tasks;
- **FedALA [35]:** An FL framework that personalises the local model by aggregating the old local model and the global model.

Since the training mechanisms for deep learning models and statistical models are quite different, the evaluation is done in different ways. Specifically, for deep learning models (i.e. GCNSA, STGCN, ChebNet and GRU), each city client is first trained on its own training set, then personalised on the support set of the test set, and finally tested on the query set of the test dataset. However, for statistical methods such as SVR, ARIMA, and HA, the model will be tested directly on the query set of the test dataset without the need to train on the training set.

Finally, several important hyper-parameters and experimental configurations of the compared models are listed in Table IV. Note that to make a fair comparison, the training task is to forecast the future 30-minute charging demand at the county level by using the past 6-hour charging demand (i.e., charging demand of the past 12 time slices) as the input.

4) Running Environment: The evaluation is carried out on a Windows workstation equipped with two NVIDIA GeForce RTX 3090 GPUs, an Intel Gold 5218R Two-Core Processor CPU, and 512G RAM.

B. Evaluation Results

The performance of evaluated methods is analysed in three aspects, namely 1) the forecasting error to illustrate how well the model is to predict the future; 2) the convergence speed to demonstrate how fast the model is to stabilise, and 3) the model generalisability to show how agile the model is to handle contexts with different cities. Moreover, we also discuss the impact of layer number of graph propagation and the introduced factors on the prediction performance. It is worth noting that the bold numbers in the table represent the best performances among all the baseline models.

TABLE IV: The hyperparameter setups of baselines

Model	Notation	Parameter	Value
DL[1]	M	Rounds of global training	1000
	-	Rounds of localization	1
	α	Learning rate of SGD	0.001
	L	Loss function	L1 Loss
	T	Input time slice	12
	-	Batch size	16
	-	Optimizer	Adam
GCN[2]	N	Nodes of alignment	32
	-	CNN filter size	64
SVR	K	Kernel	RBF
	ϵ	Epsilon	0.001
ARIMA	p	Number of time lags	12
	d	Degree of differencing	1
	q	Order of model	12
FMGCN	β	Learning rate of FOMAML update	0.001
FedALA	η	Learning rate of ALA update	0.001
PS[3]	λ	Coefficient of regularization	1.0
	-	Loss function of regularization	L2 Loss

[1] DL denotes deep learning models, including GCNSA, STGCN, ChebNet and GRU.
[2] GCN represents the graph convolution networks, i.e., GCNSA, STGCN and ChebNet.
[3] PS stands for the personalisation strategy of FL methods requiring regularization, namely FedDyn, FedAMP, pFedMe, and FedProx.

1) Forecasting Error: As shown in Table V and Table VI respectively, the assessment of forecasting error is divided into the comparison of varied backbone models and the comparison of different FL frameworks based on the proposed GCNSA model. According to Table V, the proposed method, i.e., FMGCN implementing GCNSA, outperforms other methods in all four metrics with fewer forecasting errors. Specifically, on average, it achieves significant improvements of 16.33% in MAE, 5.48% in RMSE, and 24.17% in MAPE, respectively. Moreover, FMGCN reaches the best fit with less than 10% residuals in R^2. These results illustrate that the regional EV charging demand forecasting model equipped with spatio-temporal attention (i.e., GCNSA) and the model pretraining step designed with Federated Meta-learning can work jointly and smoothly to achieve state-of-the-art performance.

Moreover, the results also show that the models with graph learning ability (i.e., ChebNet, STGCN, and GCNSA) can reduce predictive error significantly, compared to the non-graph ones (i.e., HA, ARIMA, SVR, and GRU), demonstrating the superiority of deploying graph neural networks for spatio-temporal EV charging demand forecasting. Especially, non-GCN models (i.e., HA, ARIMA, SVR, and GRU), exhibit bad performances on R^2 (< -1.00), indicating that the models without adopting GCN are incapable of fitting the correlation between predicted values and data features accurately. It is worth noting that, when the same backbone model is used, i.e., GCNSA, the model with FL (i.e., GCNSA with FMGCN or FedAvg) can outperform the one without, with an improvement of 36.59%, 36.23%, 47.76% and 44.44% in MAE, RMSE, MAPE and R^2, respectively. This demonstrates that federated learning is an effective way to improve model performance by exchanging knowledge among cities.

TABLE V: The performance comparison between different backbone models

Models	MAE	RMSE	MAPE (%)	R^2	Graph-based	FL-based
GCNSA (FMGCN)	**0.26**	**0.44**	**15.95**	**0.91**	✓	✓
GCNSA (FedAvg)	0.44	0.80	21.17	0.73	✓	✓
GCNSA (Separate)	0.41	0.69	30.53	0.63	✓	
STGCN	0.49	0.73	40.26	0.54	✓	
ChebNet	0.84	1.02	76.26	-0.12	✓	
GRU	1.11	1.17	99.97	-1.13		
SVR	0.53	0.87	160.24	-1.15		
ARIMA	0.56	0.94	163.52	-1.09		
HA	0.99	1.17	213.98	-1.81		

[1] Numbers in bold and numbers with underline are the best and second-best performances, respectively.

TABLE VI: The performance comparison between different training strategies on GCNSA

Strategy	MAE							RMSE						
	GZ	SZ	DG	FS	ZS	ZH	Average	GZ	SZ	DG	FS	ZS	ZH	Average
FMGCN	**0.51**	**0.60**	**0.16**	**0.47**	**0.08**	**0.13**	**0.26**	1.15	**0.94**	**0.20**	**0.77**	0.13	**0.28**	**0.44**
FedAvg	1.00	1.01	0.21	0.90	0.12	0.24	0.44	2.39	2.39	**0.20**	1.35	**0.11**	0.34	0.80
FedDyn	0.69	0.62	0.39	0.51	0.37	0.41	0.46	**1.13**	1.02	0.71	0.87	0.69	0.74	0.81
FedAMP	0.91	2.04	0.18	0.66	0.13	0.27	0.52	1.62	1.43	0.70	1.28	0.63	0.79	0.93
FedALA	0.87	0.89	0.45	0.65	0.39	0.43	0.57	1.38	1.45	0.91	1.28	0.80	0.87	1.03
FedProx	1.62	1.50	0.87	1.40	0.85	0.91	1.08	3.39	3.39	1.21	2.35	1.11	1.34	1.81
pFedMe	1.65	1.45	0.95	1.37	0.90	1.02	1.12	3.21	3.45	1.36	1.98	1.25	1.45	1.87
FedFOMO	1.47	1.42	1.19	1.33	1.17	1.21	1.27	2.89	2.66	1.84	2.38	1.75	1.87	2.09
FedRep	2.69	2.61	1.02	1.93	0.94	1.10	1.47	2.91	2.71	2.36	2.56	2.31	2.35	2.48
Separate	1.07	0.85	0.18	0.68	0.13	0.33	0.41	1.38	2.39	0.24	1.09	0.23	0.39	0.69
Strategy	**MAPE**							R^2						
	GZ	SZ	DG	FS	ZS	ZH	Average	GZ	SZ	DG	FS	ZS	ZH	Average
FMGCN	8.26	5.22	**16.20**	**12.84**	**24.66**	**13.51**	**15.95**	**0.90**	**0.96**	**0.90**	**0.91**	**0.89**	**0.92**	**0.91**
FedAvg	8.18	**4.04**	21.67	29.58	31.17	15.05	21.27	0.73	0.81	0.70	0.68	0.74	0.65	0.73
FedDyn	11.66	8.18	21.77	21.42	29.15	19.99	22.00	0.71	0.65	0.73	0.75	0.69	0.74	0.72
FedAMP	11.83	8.22	19.03	20.04	31.41	19.92	20.51	0.72	0.76	0.62	0.64	0.66	0.70	0.68
FedALA	**8.08**	4.15	19.88	17.64	33.84	14.59	20.71	0.69	0.68	0.62	0.68	0.66	0.70	0.66
FedProx	11.68	11.93	49.82	80.07	80.51	44.33	57.08	-0.11	-0.07	-0.05	0.01	-0.11	0.03	-0.07
pFedMe	13.62	9.75	41.17	75.08	89.94	34.56	50.04	-0.22	-0.23	-0.12	-0.06	-0.04	-0.07	-0.12
FedFOMO	24.83	25.08	72.97	89.77	94.66	57.48	77.34	-0.18	-0.59	-0.42	-0.14	-0.94	-0.34	-0.52
FedRep	69.18	65.05	101.87	126.24	134.01	82.35	103.36	-1.07	-0.97	-1.03	-1.00	-1.11	-1.00	-1.05
Separate	20.77	17.36	29.16	26.80	42.51	24.12	30.53	0.61	0.70	0.66	0.67	0.47	0.68	0.63

[1] GZ, SZ, DG, FS, ZS, ZH denote Guangzhou, Shenzhen, Dongguan, Foshan, Zhongshan, Zhuhai, respectively.
[2] Numbers in bold and numbers with underline are the best and second-best performances, respectively.

Finally, as shown in Table VI, FMGCN exhibits superior performance compared to other representative and SOTA training strategies. To be specific, the proposed training strategy can help the backbone model GCNSA to decrease MAE, RMSE and MAPE by 36.59%, 36.23% and 22.23%, respectively, and increase R^2 by 24.66%. This result convincingly demonstrates the merits of FMGCN in facilitating the training of graph convolutional models that can be effectively adapted to a variety of personalised prediction tasks, where the data of each client is isolated and heterogeneous. In addition, to directly visualize the forecasting effects of GCNSA adopting different FL training strategies, the forecasting curves of all the cities are plotted, as shown in Figure 4, which demonstrates that the proposed mechanism FMGCN can assist GCNSA to best fit the ground truth.

2) Convergence Speed: To analyze the convergence speed of different baseline models, we compare the training duration of different models when converging to the target MAE, RMSE, MAPE, and R^2. Based on the second-best values in MAE, RMSE, MAPE, and R^2 and values in multiples of 5 can be better observed in the figure, we set the target values of MAE, RMSE, MAPE, and R^2 as 0.45, 0.80, 0.20, and 0.75, respectively.

As shown in Figure 5, FMGCN has the fastest convergence

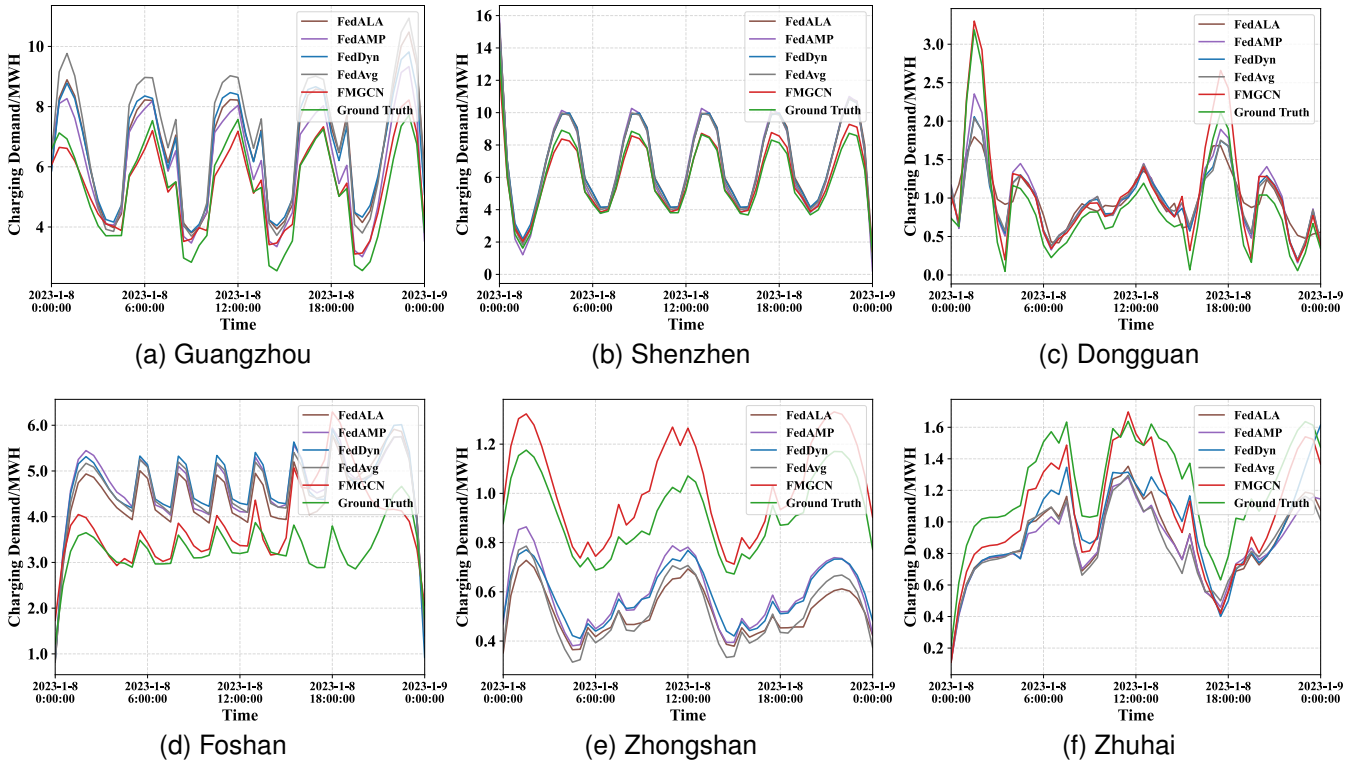


Fig. 4: The fitting curves of GCNSA adopting different FL training strategies in the case of (A) Guangzhou, (B) Shenzhen, (C) Dongguan, (D) Foshan, (E) Zhongshan and (F) Zhuhai.

speed in all four cases. In particular, the MAE curve of FMGCN can reach the target value of 0.45 at the 165th epoch, cutting the number of training rounds by 64.59% compared to the second-best (i.e., the 466th epoch in FedAvg). Similarly, in the case of RMSE, FMGCN only requires 164 epochs to converge to the target value of 0.80, reducing 78.39% compared to the second-fastest FedAvg (759 epochs). Further, FMGCN assists the GCNAS to reach the target MAPE at 743 epochs, decreasing at least 25.70% (even if none of the other training strategies can reach the target value). Lastly, compared to FedAvg which spends 978 epochs to reach the target value 0.75 in R^2, FMGCN (196 epochs) can reach an improvement in convergence speed by 79.96%. To sum up, FMGCN can accelerate the convergence speed by 62.16% on average.

3) Model generalisability: To demonstrate the generalisability of the model trained by FMGCN, we plot a heatmap of prediction accuracy in R^2, which is illustrated in Figure 6. We can see that, with the exception of a few districts in Zhongshan, most of the regions have R^2 higher than 0.85, indicating that the FMGCN can maintain high accuracy for different regions. Especially, in Shenzhen, the proposed approach achieves a remarkable result (> 0.90) for EV charging demands in all the studied regions, which shows its ability to be applied as a basis to support related services, e.g., smart grid. These findings indicate that FMGCN can train a model with high generalisability even if the charging patterns may vary across time and space.

4) Impact analysis: The impact of the layer number of graph propagation and other introduced factors as model inputs are analysed. First, as described in Section (IV-A2), K can be recognized as the range of graph-based information propagation in GNNs. As an important hyperparameter, we deploy different values of K on the proposed model. As shown in Table VII, FMGCN performs the best when K equals 1. Moreover, with the K value growing, the forecasting performance decreases. It shows that the spillover effect of EV charging demand may be limited to neighboring regions.

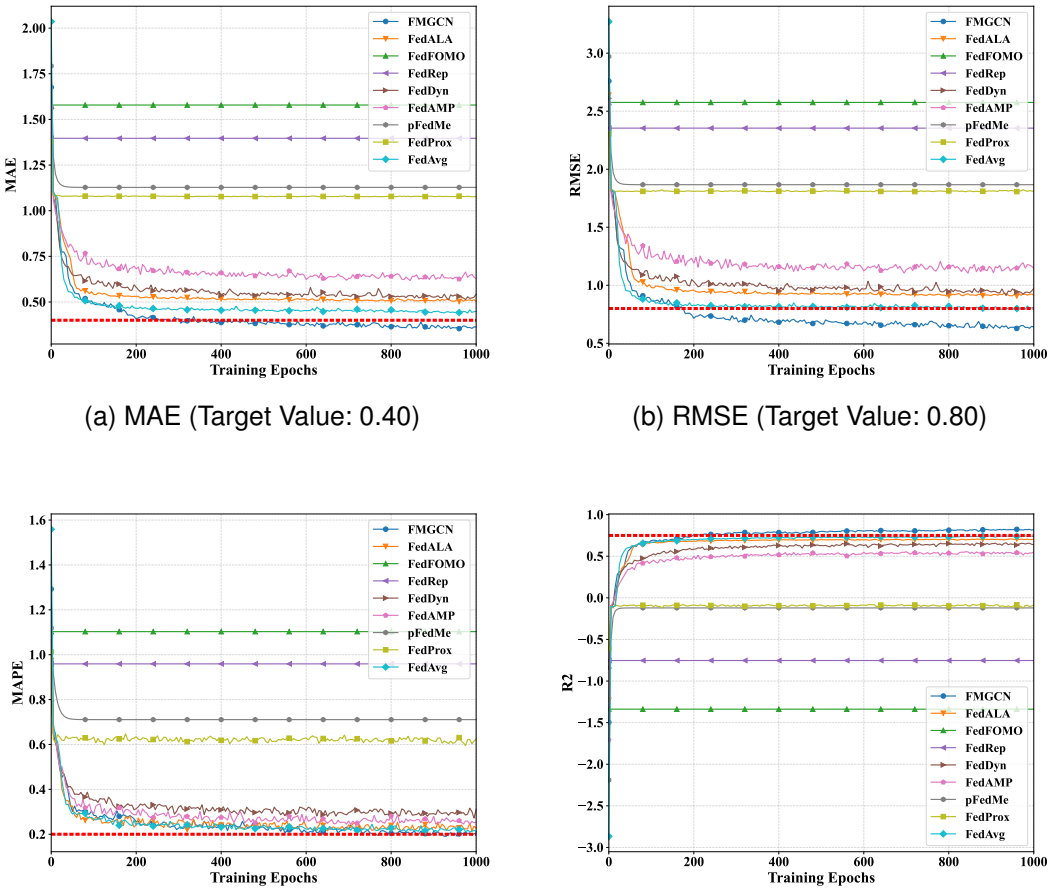
TABLE VII: Performance of FMGCN with different K values

K	MAE	RMSE	MAPE (%)	R^2
1	**0.26**	**0.44**	**15.95**	**0.91**
2	0.35	0.62	17.58	0.83
3	0.43	0.77	19.31	0.76
4	0.44	0.79	20.24	0.75

[1] Numbers in bold and numbers with underline are the best and second-best performances, respectively.

Second, as for the amount of information included in the input, Table VIII shows that 1) the improvement is proportional to the amount of information contained, including charging demand data (CDD), socio-economic data (SED), and weather data (WA); 2) SED (i.e., population and GDP) are more favourable to the model than WD (i.e., temperature) for regional EV charging demand forecasting in short-term (i.e., 30 min).

In summary, the proposed approach is superior in terms



(a) MAE (Target Value: 0.40)

(b) RMSE (Target Value: 0.80)

(c) MAPE (Target Value: 0.20)

(d) R^2 (Target Value: 0.75)

Fig. 5: The evaluation curves for different metrics: (A) MAE, (B) RMSE, (C) MAPE and (D) R^2, and the red dashed lines represent the convergence target value.

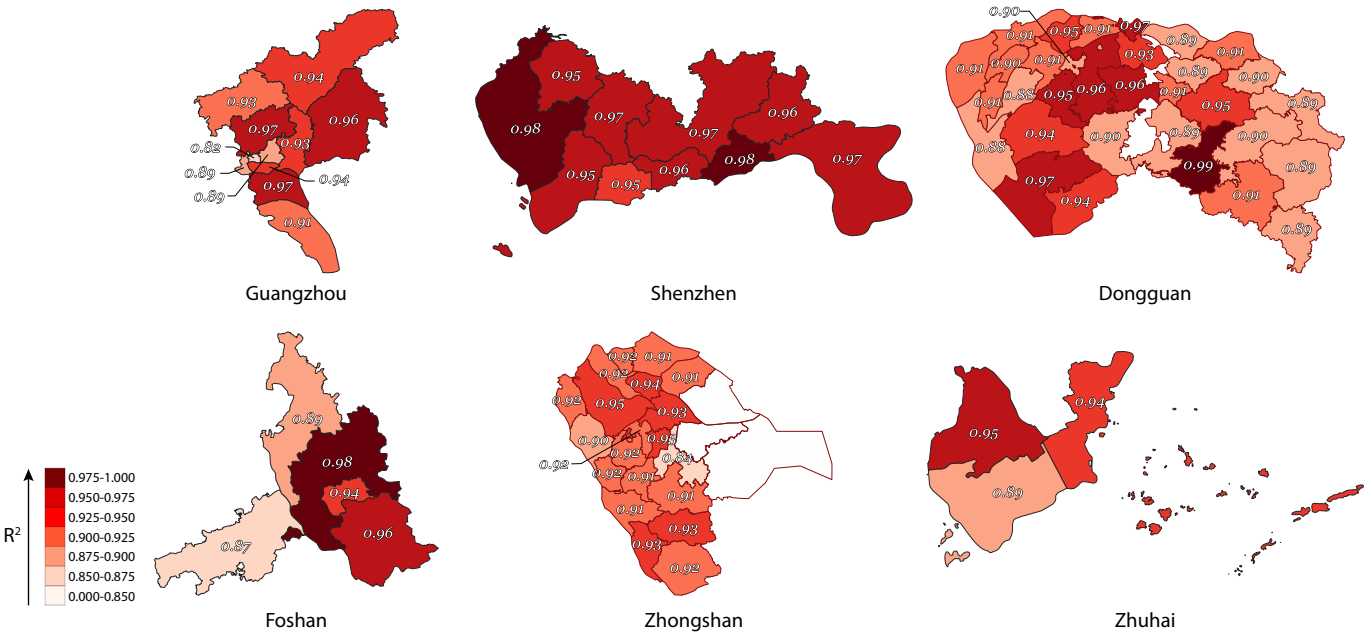


Fig. 6: The R^2 heatmap of charging demand forecasting for different regions. Blank area without numbers means that this area is not used for prediction due to lack of data.

TABLE VIII: The comparison between FMGCN adopting different information as the input vector

Input Data	MAE	RMSE	MAPE	R^2
Only CDD	0.39	0.71	22.20	0.74
CDD+WD	0.37	0.68	19.47	0.77
CDD+SED	**0.26**	**0.44**	17.19	0.88
All the Data	**0.26**	**0.44**	**15.95**	**0.91**

[1] Numbers in bold and numbers with underline are the best and second-best performances, respectively.
[2] CDD, SED and WD represent charging demand data, socio-economic data and weather data, respectively.

of prediction performance, convergence speed, and model generalisability compared to other baselines. In particular, first, FMGCN can outperform the second-best model with an improvement of 36.59%, 36.23%, and 22.23% in MAE, RMSE, and MAPE respectively. Second, FMGCN accelerates the model training speed, resulting in an average reduction of 62.16% in epoch spending to reach the target value. Third, the model can be well personalised for each studied city, showing high model generalisability. Finally, the graph propagation length $K = 1$ used in FMGCN is more suitable for regional EV charging demand prediction, and the extra information about population, GDP, and temperature is beneficial for the forecasting task when it is used as model input.

VI. Conclusions

To promote environmentally friendly and low-carbon lifestyles, regional EV charging demand predictions have been investigated as a way to alleviate the problem of power scarcity caused by spatial and temporal differences in urban charging demand. However, spatio-temporal forecasting and distributed training remain under-explored in the field. To fill the gap, we propose a federated-meta-learning-based graph convolution network for regional charging demand forecasting, called FMGCN. It comprises two main modules, namely 1) Spatio-temporal Learning module, which designs a dedicated GCN model with spatio-temporal attention to discover the dynamic characteristics among cities; and 2) Distributed Pretraining module, which incorporates FedAvg and FOMAML to train a global model with strong personalisation ability to address issues of data isolation and heterogeneity.

Compared to other SOTA models, FMGCN can achieve notable improvements in MAE, RMSE, and MAPE by 36.59%, 36.23%, and 22.23%, respectively, and also a significant enhancement in R^2 by 24.66%. Second, the results reveal that FMGCN can accelerate model convergence by approximately 62.16%. Third, FMGCN can provide a model with high generalisability to support personalisation for different cities to better support forecasting tasks in their own contexts. Finally, the impact analysis shows that most charging demands propagate among neighboring regions (i.e., $K = 1$, one-hop information propagation in GCN), and adding information about population, GDP, and temperature into the model input can be beneficial to further improve the prediction result.

In the future, first, multi-source data fusion will be further explored to further enhance the forecasting capability of the GCN model for EV charging demand. Moreover, an asynchronous update strategy will be studied to resolve the straggler issues caused by the lagging clients during the global model training. Last but not least, an adaptive client selection mechanism will be designed to enable the global model to obtain more beneficial knowledge from high quality clients.

References

[1] S. Zhang and J. J. Q. Yu, "Electric vehicle dynamic wireless charging system: Optimal placement and vehicle-to-grid scheduling," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 6047–6057, 2022.
[2] C. Liu and Y. L. Murphey, "Optimal power management based on q-learning and neuro-dynamic programming for plug-in hybrid electric vehicles," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 1942–1954, 2020.
[3] X. Chen, H. Wang, F. Wu, Y. Wu, M. C. González, and J. Zhang, "Multimicrogrid load balancing through ev charging networks," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5019–5026, 2022.
[4] Y. Cao, D. Li, Y. Zhang, and X. Chen, "Joint optimization of delay-tolerant autonomous electric vehicles charge scheduling and station battery degradation," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8590–8599, 2020.
[5] Y. Yuan, Y. Yuan, P. Memarmoshrefi, T. Baker, and D. Hogrefe, "Lbsp: Load-balanced secure and private autonomous electric vehicle charging framework with online price optimization," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 15 685–15 696, 2022.
[6] C. Napoli, G. Pappalardo, G. M. Tina, and E. Tramontana, "Cooperative strategy for optimal management of smart grids by wavelet rnns and cloud computing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1672–1685, 2016.
[7] S. Xie, W. Zhong, K. Xie, R. Yu, and Y. Zhang, "Fair energy scheduling for vehicle-to-grid networks using adaptive dynamic programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1697–1707, 2016.
[8] Z. Xu, Y. Kang, Y. Cao, and Z. Li, "Spatiotemporal graph convolution multifusion network for urban vehicle emission prediction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3342–3354, 2021.
[9] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans, and M. Messagie, "Day-ahead forecast of electric vehicle charging demand with deep neural networks," *World Electric Vehicle Journal*, vol. 12, no. 4, 2021.
[10] N. Kumar, D. Kumar, and P. Dwivedi, "Load forecasting for ev charging stations based on artificial neural network and long short term memory," in *Advanced Network Technologies and Intelligent Computing*. Cham: Springer International Publishing, 2022, pp. 473–485.
[11] R. Zhang, "Gcn-trn: Efficient transformer based electric vehicle charging demand forecasting system," in *Proceedings of the 5th International Conference on Computer Science and Software Engineering*, ser. CSSE '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 527–533.
[12] L. You, S. Liu, Y. Chang, and C. Yuen, "A triple-step asynchronous federated learning mechanism for client activation, interaction optimization, and aggregation enhancement," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 24 199–24 211, 2022.
[13] Z. Guo, L. You, S. Liu, J. He, and B. Zuo, "Icmfed: An incremental and cost-efficient mechanism of federated meta-learning for driver distraction detection," *Mathematics*, vol. 11, no. 8, 2023.
[14] S. Liu, H. Qu, Q. Chen, W. Jian, R. Liu, and L. You, "Afmeta: Asynchronous federated meta-learning with temporally weighted aggregation," in *2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta)*, 2022, pp. 641–648.
[15] H. Qu, S. Liu, Z. Guo, L. You, and J. Li, "Improving parking occupancy prediction in poor data conditions through customization and learning to learn," in *Knowledge Science, Engineering and Management*. Cham: Springer International Publishing, 2022, pp. 159–172.
[16] S. Liu, L. You, R. Zhu, B. Liu, R. Liu, H. Yu, and C. Yuen, "Afm3d: An asynchronous federated meta-learning framework for driver distraction detection," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–16, 2024.

[17] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 922–929, Jul. 2019.

[18] Q. Chen, S. Liu, H. Qu, R. Zhu, and L. You, "Twafr-gru: An integrated model for real-time charging station occupancy prediction," in *2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta)*, 2022, pp. 1611–1618.

[19] F. Jin and L.-l. Xu, "Research on multi-source heterogeneous sensor information fusion method under internet of things technology," in *Multimedia Technology and Enhanced Learning*. Cham: Springer International Publishing, 2020, pp. 66–74.

[20] X. Liao, X. Kang, M. Li, and N. Cao, "Short term load forecasting and early warning of charging station based on pso-svm," in *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, 2019, pp. 305–308.

[21] Q. Sun, J. Liu, X. Rong, M. Zhang, X. Song, Z. Bie, and Z. Ni, "Charging load forecasting of electric vehicle charging station based on support vector regression," in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2016, pp. 1777–1781.

[22] X. Zhang, "Short-term load forecasting for electric bus charging stations based on fuzzy clustering and least squares support vector machine optimized by wolf pack algorithm," *Energies*, vol. 11, no. 6, 2018.

[23] T. Unterluggauer, K. Rauma, P. Järventausta, and C. Rehtanz, "Short-term load forecasting at electric vehicle charging sites using a multivariate multi-step long short-term memory: A case study from finland," *IET Electrical Systems in Transportation*, vol. 11, no. 4, pp. 405–419, 2021.

[24] Y. Cheng, H. Chang, K. Tang, J. Zou, J. Zhuo, and Y. Cai, "Multistep electricity load forecasting method based on the hybrid gru neural network," in *2022 International Applied Computational Electromagnetics Society Symposium (ACES-China)*, 2022, pp. 1–3.

[25] W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey," *Expert Systems with Applications*, vol. 207, p. 117921, 2022.

[26] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016.

[27] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, ser. IJCAI'18. AAAI Press, 2018, p. 3634–3640.

[28] Q. Zhang, J. Chang, G. Meng, S. Xiang, and C. Pan, "Spatio-temporal graph structure learning for traffic forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 1177–1185, Apr. 2020.

[29] K. Guo, Y. Hu, Z. Qian, H. Liu, K. Zhang, Y. Sun, J. Gao, and B. Yin, "Optimized graph convolution recurrent neural network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1138–1149, 2021.

[30] W. Zhang, H. Liu, Y. Liu, J. Zhou, and H. Xiong, "Semi-supervised hierarchical recurrent graph neural network for city-wide parking availability prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 1186–1193, Apr. 2020.

[31] S. Su, Y. Li, Q. Chen, M. Xia, K. Yamashita, and J. Jurasz, "Operating status prediction model at ev charging stations with fusing spatiotemporal graph convolutional network," *IEEE Transactions on Transportation Electrification*, vol. 9, no. 1, pp. 114–129, 2023.

[32] X. Feng, J. Guo, B. Qin, T. Liu, and Y. Liu, "Effective deep memory networks for distant supervised relation extraction," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 2017, pp. 4002–4008.

[33] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54, 20–22 Apr 2017, pp. 1273–1282.

[34] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proceedings of Machine Learning and Systems*, vol. 2, 2020, pp. 429–450.

[35] J. Zhang, Y. Hua, H. Wang, T. Song, Z. Xue, R. Ma, and H. Guan, "Fedala: Adaptive local aggregation for personalized federated learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, pp. 11 237–11 244, Jun. 2023.

[36] Y. Huang, L. Chu, Z. Zhou, L. Wang, J. Liu, J. Pei, and Y. Zhang, "Personalized cross-silo federated learning on non-iid data," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 9, pp. 7865–7873, May 2021.

[37] L. Collins, H. Hassani, A. Mokhtari, and S. Shakkottai, "Exploiting shared representations for personalized federated learning," in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 2089–2099.

[38] S. Liu, Q. Chen, and L. You, "Fed2a: Federated learning mechanism in asynchronous and adaptive modes," *Electronics*, vol. 11, no. 9, 2022.

[39] L. You, S. Liu, T. Wang, B. Zuo, Y. Chang, and C. Yuen, "Aifed: An adaptive and integrated mechanism for asynchronous federated data mining," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–17, 2023.

[40] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 06–11 Aug 2017, pp. 1126–1135.

[41] X. Feng, J. Guo, B. Qin, T. Liu, and Y. Liu, "Effective deep memory networks for distant supervised relation extraction." in *IJCAI*, vol. 17, 2017, pp. 1–7.

[42] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE signal processing magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[43] M. Simonovsky and N. Komodakis, "Dynamic edge-conditioned filters in convolutional neural networks on graphs," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3693–3702.

[44] H. Qu, S. Liu, J. Li, Y. Zhou, and R. Liu, "Adaptation and learning to learn (all): An integrated approach for small-sample parking occupancy prediction," *Mathematics*, vol. 10, no. 12, p. 2039, 2022.

[45] J. Li, H. Qu, and L. You, "An integrated approach for the near real-time parking occupancy prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 3769–3778, 2022.

[46] D. J. Bartholomew, "Time series analysis forecasting and control." 1971.

[47] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in neural information processing systems*, vol. 9, 1996.

[48] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[49] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems*, vol. 29. Curran Associates, Inc., 2016.

[50] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. AAAI Press, 2018, p. 3634–3640.

[51] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 922–929, Jul. 2019.

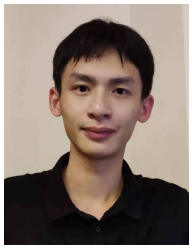
[52] C. T. Dinh, N. Tran, and J. Nguyen, "Personalized federated learning with moreau envelopes," in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp. 21 394–21 405.

[53] D. A. E. Acar, Y. Zhao, R. Matas, M. Mattina, P. Whatmough, and V. Saligrama, "Federated learning based on dynamic regularization," in *International Conference on Learning Representations*, 2021.

[54] M. Zhang, K. Sapra, S. Fidler, S. Yeung, and J. M. Alvarez, "Personalized federated learning with first order model optimization," in *International Conference on Learning Representations*, 2021.

Linlin You is an associate professor at the School of Intelligent Systems Engineering, Sun Yat-sen University, and also a research affiliate at the Intelligent Transportation System Lab, Massachusetts Institute of Technology. He was a senior postdoc at the Singapore-MIT Alliance for Research and Technology and a research fellow at the Architecture and Sustainable Design Pillar of Singapore University of Technology and Design. He received his Ph.D. in Computer Science from the University of Pavia in 2015. He published more than 60 journal and conference papers in the research fields of Smart Cities, Service Orchestration, Multi-source Data Fusion, Machine Learning, and Federated Learning. He is an Associate Editor of Springer Nature Computer Science and Young Editor of The Innovation.

Qiyang Chen received the B.E. degree from the School of Intelligent Systems Engineering, Sun Yat-sen University, China in 2022, where he is currently pursuing the M.E. degree. His research interests include meta-learning, federated learning and their applications in intelligent transportation systems.



Haohao Qu received the B.E. and M.E. degree from the School of Intelligent Systems Engineering, Sun Yat-Sen University, People's Republic of China, in 2019 and 2022, respectively. He is pursuing the PhD degree from the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China. His research interests include Meta-learning, Graph Neural Networks, and their applications in Intelligent Transportation Systems and Recommendation Systems.

Rui Zhu received the Ph.D. degree in Geo-Informatics from The Hong Kong Polytechnic University, MSc degree in Geodesy and Geo-Informatics from KTH - Royal Institute of Technology, and BSc degree in Geographical Information Science from Nanjing Normal University. Zhu is a Senior Scientist at the Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore. Zhu was a Research Assistant Professor at The Hong Kong Polytechnic University and a Postdoctoral Associate at MIT Senseable City Laboratory. His study focused on GIScience, Urban Informatics, and Solar Energy with a publication of more than 60 SCI papers in journals such as Nature Communications, Science Bulletin, and Applied Energy. Zhu is an Associate Editor of Springer Nature Computer Science, Editor of Big Earth Data, and Young Editor of The Innovation and Advances in Applied Energy. He is also the PI/Co-I for several research grants, and Board of Director member of Chinese Professional in Geographic Information Systems. Zhu's study has been reported by international media such as Singapore TV, Lianhe Zaobao, and MIT News.

Jinyue Yan received the Ph.D. degree from the Royal Institute of Technology KTH, Stockholm, Sweden, in 1991. He is currently a Chair Professor of Energy and Buildings at The Hong Kong Polytechnic University. Prof. Yan is an Academician of European Academy of Sciences and Arts, and serves as the advisory expert to the UN, EU, & ADB. His research interests include advanced energy systems, renewable energy, advanced power generation, climate change mitigation technologies and related environment and policy etc. Prof. Yan published about 400 papers including papers in Science, Nature Energy, Nature Climate Change, and Nature Communications and hold 10+ patents with about 25000+ citations and H-index 80. Prof. Yan is the Editor-in-Chief of Advances in Applied Energy, Nexus (Cell), and Handbook of Clean Energy Systems. He has led research platform (Future Energy Profile) with funding of over 80 million Euro by Swedish Knowledge Foundation and industrial partners.

Paolo Santi received the Laurea and Ph.D. degrees in computer science from the University of Pisa, Italy. He is currently a Research Scientist at the MIT Senseable City Laboratory and a Senior Researcher at the Istituto di Informatica e Telematica, CNR, Pisa. His research interests are in the modeling and analysis of complex systems ranging from wireless multi hop networks to sensor and vehicular networks. More recently, his research interests are in smart mobility and intelligent transportation systems. In these fields, he has contributed more than 120 scientific papers and two books. He has been involved in the technical and organizing committee of several conferences in the field. He is a member of the IEEE Computer Society. He has recently been recognized as a Distinguished Scientist by the Association for Computing Machinery. He is an Associate Editor of the IEEE Transactions on Mobile Computing, the IEEE Transactions on Parallel and Distributed Systems, and Computer Networks. He was a Guest Editor of the Proceedings of the IEEE Special Issue on Vehicular Communications: Ubiquitous Networks for Sustainable Mobility in 2011 to which he also contributed a paper.

Carlo Ratti received the joint Graduate degree from the Politecnico di Torino and the École Nationale des Ponts et Chaussées and the M.Phil. and Ph.D. degrees from the University of Cambridge, U.K. He is currently the Founder and the Director of the MIT Senseable City Laboratory, an Architect, and an Engineer by training. He practices in Italy and teaches at the Massachusetts Institute of Technology. He has co-authored over 200 publications and holds several patents. His work has been exhibited worldwide at venues, such as the Venice Biennale; the Design Museum Barcelona; the Science Museum, London; GAFTA, San Francisco; and The Museum of Modern Art, New York. His Digital Water Pavilion at the 2008 World Expo was hailed by Time Magazine as one of the "Best Inventions of the Year." He has been included in Esquire Magazine's "Best and Brightest" list, Blueprint Magazine's "25 People Who Will Change the World of Design," and Forbes Magazine's "Names You Need To Know" in 2011. He was a Presenter at TED 2011. He is serving as a member for the World Economic Forum Global Agenda Council for Urban Management.