Causality-Aware Multi-Graph Convolutional Networks with Critical Node Dynamics for Electric Vehicle Charging Station Load Forecasting

Yaohui Huang⬤, Senzhen Wu⬤, Zhijin Wang⬤, Xiufeng Liu⬤, Chendan Li⬤ and Yue Hu⬤

Abstract—**Accurately forecasting the load of electric vehicle charging stations (EVCSs) is crucial for optimizing grid operations and facilitating EV integration, yet existing methods struggle to capture the intricate spatio-temporal dependencies and the impact of influential EVCSs within charging networks. To address this, we propose a novel framework, Causality-Aware Dynamic Multi-Graph Convolutional Network (CADGN), a multi-graph convolutional network that integrates causal inference and critical node modeling. It consists of two core modules: the Causality-Aware Graph Learning Module (CAGLM) uncovers and represents causal relationships between EVCSs, while the Critical Relationship Graph Learning Module (CRGLM) dynamically models the evolving connections among critical EVCS nodes. Temporal patterns extracted from these modules are then fused to generate accurate load predictions. Extensive experiments using real-world datasets of hourly charging data from multiple cities demonstrate CADGN's superiority over state-of-the-art EVCS load forecasting models, particularly for short-term and mid-term horizons. Notably, our model achieves an average 4.7% reduction in Mean Absolute Error (MAE) compared to Graph WaveNet across all datasets and prediction horizons, highlighting the practical benefits of considering both causal and critical relationships for enhanced grid operations and EV integration. These results emphasize the importance of incorporating causality and the identification of critical relationships in the EVCS load forecast to achieve higher accuracy.**

Index Terms—**Electric Vehicle Charging Stations, Load Forecasting, Graph Convolutional Networks, Causal Inference, Time Series Analysis**

I. INTRODUCTION

A. Background and Motivation

The widespread adoption of electric vehicles (EVs) is rapidly transforming transportation systems, offering promising solutions to reduce carbon emissions and dependency on fossil fuels. However, the increasing demand for electricity from EV charging poses significant challenges to the stability and efficient power management of the electric grid [1]. Electric Vehicle Charging Stations (EVCS), playing a pivotal role in facilitating this transition to electric mobility, necessitate accurate and robust load forecasting techniques to predict future electricity demands and ensure a reliable and sustainable charging infrastructure [2]. Accurate EVCS load forecasts provide invaluable information for various stakeholders in the power system, enabling informed decisions in grid operations,

planning, and market management. For grid operators, precise load predictions are essential for mitigating peak demand stress and optimally scheduling generation and distribution resources to meet EV charging requirements effectively. In planning and investment decisions, reliable forecasts serve as critical inputs for infrastructure expansions and upgrades, ensuring adequate capacity and resilience for supporting the growing EV fleet. Moreover, for EVCS operators, accurate load forecasts support strategic decisions on infrastructure investments, enabling optimal placement and capacity of charging stations to enhance profitability and customer satisfaction. In smart grid systems with dynamic pricing, forecasting plays a critical role in offering consumers personalized charging cost insights, encouraging behavior shifts that flatten demand peaks and reduce operational costs [3]. Additionally, accurate forecasting allows operators to fine-tune pricing strategies, maximizing profits during high-demand periods while keeping costs affordable during off-peak times. These advancements contribute to both business sustainability and grid efficiency, driving broader adoption of EVs and expanding the economic viability of EV charging networks [4].

The inherent challenge in EVCS load forecasting arises from the complex interplay of spatial and temporal factors governing charging patterns [5]. Spatially, EVCS load demands exhibit inherent dependencies, with nearby stations often experiencing correlated fluctuations due to factors such as user mobility patterns, proximity to popular destinations, and regional transportation networks. Temporally, EVCS loads demonstrate distinct periodicities influenced by daily and weekly routines, socioeconomic patterns, and charging behavior variability, encompassing both long-term (seasonal) and short-term (hourly or sub-hourly) variations [6]. While traditional time series forecasting methods have been applied to address EVCS load forecasting [7], they struggle to model the rich spatial interdependencies between charging stations, often neglecting crucial information relevant to capturing the intricate dynamics of network-level EVCS load patterns. To capture the inherent relationships between charging station loads, Graph Neural Networks (GNNs) have gained significant traction due to their ability to model and learn on complex graph structures [8]. By explicitly encoding EVCS spatial relationships into the model's architecture, GNNs provide a powerful framework for capturing complex spatio-temporal interactions and achieving higher prediction accuracy [9].

The above-mentioned are not the only factors that affect the charging load dynamics. While GNN-based models show great potential for capturing these spatio-temporal relationships, most existing approaches treat all EVCS nodes (charging

stations) and their interconnections uniformly during representation learning, overlooking two crucial aspects that limit their forecasting accuracy. First, they do not differentiate between correlations and true causal relationships, potentially leading to less robust and interpretable predictions, particularly as the network dynamics evolve [10]. Fig. 1 highlights this distinction by contrasting the structure of a causal graph against a correlation graph for EVCS loads. The **causal graph** is driven by specific events or behavioral changes, such as short-term station service interruptions or pricing adjustments, which result in user transfers between stations and subsequently impact station load dynamics. This forms a clear directional causal chain. As shown in Fig. 1(a), when station v_2 halts some of its charging services, users are forced to shift to stations v_1 and v_4. However, the excessive load at v_1 and v_4 further leads to an increase in demand at station v_3. In contrast, the **correlation graph** is driven by overarching systemic trends, such as time-based or regional demand patterns. The relationships between variables are predominantly influenced by common external factors. As illustrated in Fig. 1(b), variations in load are attributed to shared factors, such as daytime or nighttime traffic flow and demand patterns, but without a clearly defined causal pathway. Furthermore, beyond direct causal effects, more complex mechanisms also influence load dynamics, including joint interventions [11], where simultaneous changes at multiple stations trigger cascading effects, as well as mediation effects [12] and interaction effects [13], where intermediate variables or interdependencies shape the overall impact. Second, they fail to recognize that certain EVCS nodes can have disproportionately larger impacts on the overall network load behavior. For instance, charging stations located near transportation hubs or popular commercial areas may experience significantly higher demand and could exert a stronger influence on surrounding stations due to shifting traffic patterns and user preferences. By not explicitly identifying and leveraging the unique role of these influential "critical" EVCS nodes during forecasting, existing GNN methods might underutilize crucial information, leading to reduced performance and less-informed insights.

Motivated by considering both spatial and temporal factors from this overlayed transportation network and electrical network, as well as the significant impact of the critical nodes, we introduce CADGN, a novel framework that explicitly models both causality and the dynamics among critical nodes in EVCS load forecasting. CADGN leverages two key modules: the Causality-Aware Graph Learning Module (CAGLM) and the Critical Relationship Graph Learning Module (CRGLM), respectively. The CAGLM infers a causality graph from the historical load data, representing cause-and-effect relationships between EVCS. It then performs dynamic graph convolution on this causality graph to learn representations informed by the flow of causal dependencies. Complementing the CAGLM, the CRGLM focuses on modeling the intricate dynamics among the most influential EVCS in the network. It achieves this by first identifying "critical" nodes – those with substantial influence on other EVCS loads based on a calculated centrality score – and then constructing a dynamic critical relationship graph representing their interconnectedness. By dynamically
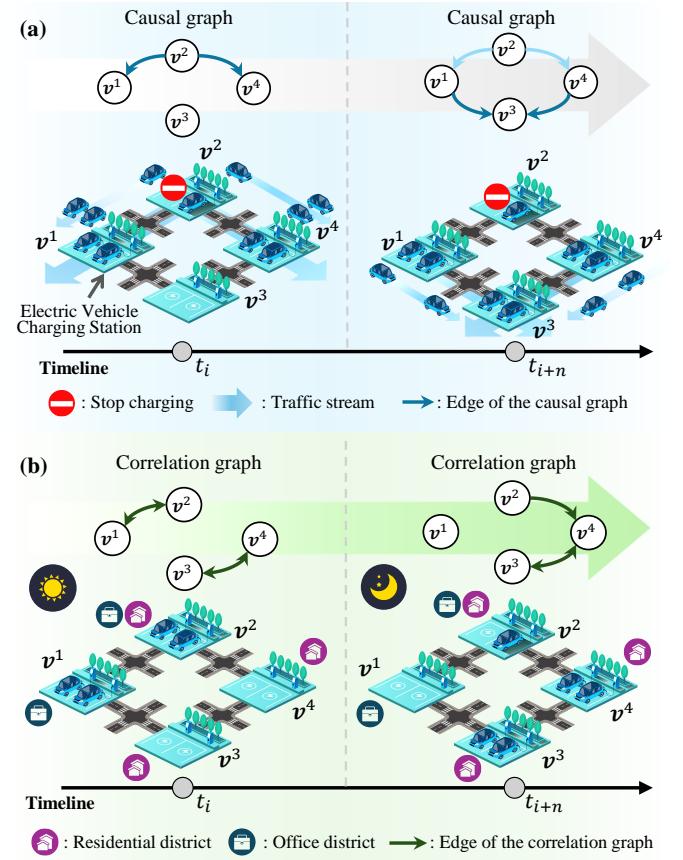

Fig. 1. Conceptual differences between causal and correlation graphs in EVCS networks. (a) Causal graphs show directed relationships where events like service interruptions cause load redistribution among stations. (b) Correlation graphs illustrate statistical associations driven by shared external factors, such as time-of-day demand patterns.

capturing and representing these relationships, the CRGLM extracts critical information often overlooked by conventional graph-based models. The representations learned from these modules are subsequently fused and processed through a Temporal Representation Fusion Module (TRFM) and an output layer, ultimately leading to enhanced forecasting accuracy, particularly for short-term and mid-term horizons.

B. Literature Review

1) EVCS Load Forecasting and the Application of Graph Neural Networks: Accurate prediction of EVCS load demand is paramount for ensuring efficient power grid management and successfully integrating Electric Vehicles into the existing energy infrastructure. Traditional approaches predominantly relied on classical statistical methods [14] and conventional machine learning methods [7] to model temporal dependencies in EVCS load patterns. While these models offered simplicity and interpretability, their limited capacity to capture non-linear dependencies and inter-station interactions rendered them inadequate for complex, large-scale networks [15]. The rise of deep learning, particularly in time series forecasting [16], [17], [18], [19], has paved the way for tackling these more intricate load prediction challenges. Zheng et al. [15] introduced a hierarchical forecasting model using partial input convex neural networks, capturing spatial dependencies across

EVCS and temporal demand dynamics. Huang et al. [20] proposed MetaProbformer, a meta-learning-enhanced Transformer framework, effectively addressing spatial correlations and long-term temporal dependencies in charging load forecasting. Li et al. [21] and Morteza et al. [22] both presented reinforcement learning-assisted deep learning approach to handle spatial interactions and temporal uncertainties in EVCS power forecasting. However, the preassumption of correlations across all EVCS may introduce extraneous complexity, noise, and redundancy, thereby limiting their predictive efficacy and scalability in complex network settings.

The advent of Graph Neural Networks (GNNs) has introduced transformative capabilities in capturing the spatiotemporal correlations inherent in EVCS networks [23]. Their adeptness in representing arbitrary relations and dynamically adapting to evolving graph topologies makes them a compelling choice for addressing the complexities of EVCS load forecasting [24], [25]. Several studies have demonstrated the successful application of various GNN architectures to enhance the accuracy of EVCS load forecasting [14]. Specifically, Shi et al. [26] proposed a novel GNN-based model that leveraged spatial and temporal information, combined with carefully selected features evaluated using the maximal information coefficient, to predict EVCS load demand, achieving superior performance compared to traditional methods. Kim et al. [9] introduced a parallel-structured GNN model that effectively captured multi-level spatial-temporal dependencies using a mutual adjacency matrix, demonstrating strong forecasting accuracy. Qu et al. [27] incorporated meta-learning with graph and temporal attention mechanisms, integrating multimodal prior knowledge to achieve superior accuracy and reliability in regional EVCS demand prediction. Li et al. [28] proposed an EGAT-LSTM model that integrates edge aggregation graph attention networks with LSTMs, further enhancing the predictive performance. However, Ankitha et al. [29] observed that existing GNN-based models largely focus on capturing correlational patterns in load data without explicitly disentangling true cause-and-effect relationships, which can result in capturing spurious correlations and reducing the model's robustness and generalizability [30]. Moreover, the influence of certain "critical" EVCSs, particularly those experiencing high demand and exerting significant influence on surrounding stations, remains largely underexplored [31].

2) Beyond Correlations: The Need for Causality and Critical Node Focus in EVCS Networks: To enhance EVCS load forecasting, it is essential to transition from simply modeling statistical correlations and instead focus on understanding true cause-and-effect relationships between stations. This requires integrating causality principles, as initially proposed by Granger [32]. Recent advancements in deep learning have shown promise in incorporating causal concepts into forecasting models, leading to improved accuracy [33], [34]. Despite this progress, the integration of causal reasoning in EVCS load forecasting remains underexplored, presenting an opportunity to extend these approaches for greater accuracy and understanding [35].

Beyond explicitly representing causal dependencies, the recognition of critical EVCS nodes, particularly those with

TABLE I
COMPARATIVE ANALYSIS OF EXISTING STUDIES AND THE PROPOSED CADGN MODEL.

Ref.	Temporal Modeling	Spatial Modeling	Node Relationships	Critical Node Identification	Causal Awareness
[9]	✔	✔	Dynamic	✗	✗
[15]	✔	✗	✗	✗	✗
[17]	✔	✗	✗	✗	✗
[18]	✔	✔	✗	✗	✗
[19]	✔	✔	✗	✗	✗
[20]	✔	✔	✗	✗	✗
[21]	✔	✗	✗	✗	✗
[22]	✔	✗	✗	✗	✗
[23]	✔	✔	Static	✗	✗
[24]	✔	✔	Dynamic	✗	✗
[25]	✔	✔	Dynamic	✗	✗
[26]	✔	✔	Dynamic	✗	✗
[27]	✔	✔	Static	✗	✗
[28]	✔	✔	Dynamic	✗	✗
[31]	✔	✔	Dynamic	✔	✗
Ours.	✔	✔	Dynamic	✔	✔

outsized influence on the surrounding charging patterns within a network, represents another vital area of advancement. Kong et al. [31] emphasize the importance of identifying key network components, referred to as "influencers" or "hubs", which play pivotal roles in driving information flow and influencing overall system dynamics. In EVCS networks, stations near transportation hubs or busy commuter routes likely exert more influence on load patterns than less strategically located stations. Although research on applying critical node analysis to EVCS load forecasting is limited, related work in traffic prediction [36] underscores the importance of these influential elements in enhancing predictive accuracy and understanding system dynamics [37]. These studies demonstrate the advantages of incorporating critical nodes into the modeling framework. Our proposed Causality-Aware Dynamic Multi-Graph Convolutional Network (CADGN) addresses these gaps by explicitly modeling both causal relationships and the influence dynamics among critical EVCS nodes, aiming to achieve more robust and causality-aware load forecasting. Table I provides a detailed comparison analysis between existing studies in the literature and the proposed approach.

C. Our Contributions

We propose a prediction framework with better performance, and richer causal understanding. Specifically:

- We propose CADGN, the first framework that explicitly integrates causal relationships and a dynamic focus on influential ("critical") EVCS nodes for enhanced load forecasting in EVCS networks.
- We present the novel CAGLM and CRGLM modules within CADGN. CAGLM leverages causal inference to learn informative node representations based on cause-and-effect dependencies between EVCS. CRGLM selectively focuses on dynamically capturing evolving relationships among the most influential EVCSs, offering insights that conventional GNN models miss.
- Through extensive experimental evaluation on real-world EVCS data, we demonstrate that CADGN achieves substantial performance gains compared to various state-of-the-art EVCS load forecasting approaches, showcasing

the practical benefits of incorporating both causality and critical nodes in this crucial task.

The remainder of the paper is structured as follows. Section II outlines the problem definition and key notations. Section III details the proposed CADGN model. Section IV presents the experimental setup and case study. Section V discusses the key findings. Finally, Section VI concludes the paper and discusses future research directions.

II. PRELIMINARY

This section establishes the foundational concepts for EVCS load forecasting and introduces the mathematical frameworks of EVCS load networks, critical relationship graphs, and causality graphs.

Definition 2.1 (EVCS Load Forecasting Problem): The EVCS load forecasting problem is defined as a multivariate time series forecasting task. Given previous N_T time steps historical load data from N_S EVCSs, represented as $\boldsymbol{X}^{(t_e)} = \{X^{(t_b)}, X^{(t_b+1)}, \ldots, X^{(t_e)}\} \in \mathbb{R}^{N_T \times N_S}$, where each $X^t = \{x^{(t,1)}, x^{(t,2)}, \ldots, x^{(t,N_S)}\}$ is a vector of load values at time step t, the goal is to predict future load demands $\{X^{(t_e+1)}, X^{(t_e+2)}, \ldots, X^{(t_e+N_P)}\} \in \mathbb{R}^{N_P \times N_S}$, where N_P represents the predictive step size. Mathematically, we aim to learn a mapping function $\mathbb{F}(\cdot)$ as follows:

$$\{\hat{X}^{(t_e+1)}, \ldots, \hat{X}^{(t_e+N_P)}\} = \mathbb{F}\big((X^{(t_b)}, \ldots, X^{(t_e)}); \Theta\big), \quad (1)$$

where Θ denotes the model's trainable parameters, $\hat{X}^{(t+N_P)}$ represents the predicted load values at $t + N_P$ time step .

Definition 2.2 (EVCS Load Network): An EVCS load network can be modeled as a dynamic directed graph, $\boldsymbol{\mathcal{G}} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{A})$, to represent the interdependencies between EVCSs: $\boldsymbol{V} = \{v^{(1)}, v^{(2)}, \ldots, v^{(N_S)}\}$ is a set of nodes, with each node $v^{(i)}$ corresponding to an EVCS, $\boldsymbol{E} = \{e^{(i,j)}\}$ represents a set of directed edges, where $e^{(i,j)}$ indicates a directional influence from node $v^{(i)}$ to node $v^{(j)}$, and $\boldsymbol{A} \in \mathbb{R}^{N_S \times N_S \times N_T}$ represents a dynamic adjacency tensor, with $a^{(i,j,t)} \in \boldsymbol{A}$ quantifying the strength of direct influence from EVCS $v^{(i)}$ to $v^{(j)}$ at time step t.

Definition 2.3 (Critical Relationship Graph): A critical relationship graph, $\boldsymbol{\mathcal{G}}_{cr} = (\boldsymbol{V}, \boldsymbol{E}_{cr}, \boldsymbol{A}_{cr})$, is a subgraph of the EVCS load network focusing on statistically significant connections. In this graph, $\boldsymbol{V}$ inherits the node set from the load network, and each node represents an EVCS. $\boldsymbol{E}_{cr} \subseteq \boldsymbol{E}$ represents a subset of edges from the load network, where $e_{cr}^{(i,j)} \in \boldsymbol{E}_{cr}$ represents a relationship exceeding a significance threshold. These edges can be bidirectional to reflect mutual influences. $\boldsymbol{A}_{cr} \in \mathbb{R}^{N_S \times N_S \times N_T}$ represents a dynamic adjacency tensor, with each entry $a_{cr}^{(i,j,t)}$ denoting the strength of the relationship at time step t. The specific method for determining the significance and quantifying relationships will be detailed in Section III.

Definition 2.4 (Causality Graph): A causality graph, $\boldsymbol{\mathcal{G}}_{ca} = (\boldsymbol{V}, \boldsymbol{E}_{ca}, \boldsymbol{A}_{ca})$, is a directed acyclic graph (DAG) that represents the cause-and-effect relationships. Similar to EVCS Load Network and Critical Relationship Graph, $\boldsymbol{V}$ utilizes the same node set. $\boldsymbol{E}_{ca} \subseteq \boldsymbol{E}$ represents a subset of directed edges where $e_{ca}^{(i,j)} \in \boldsymbol{E}_{ca}$ implies that EVCS i has a causal effect on EVCS j. The acyclic property enforces unidirectional causal paths. $\boldsymbol{A}_{ca} \in \mathbb{R}^{N_S \times N_S \times N_T}$ represents the dynamic causal influence strength over time.

III. METHODOLOGY

This section introduces the CADGN, consisting of four key modules: the Causality-Aware Graph Learning Module (CAGLM), the Critical Relationship Graph Learning Module (CRGLM), the Temporal Representation Fusion Module (TRFM), and a final output layer, as illustrated in Fig. 2.

A. Causality-Aware Graph Learning Module (CAGLM)

The Causality-Aware Graph Learning Module (CAGLM) aims to capture the inherent cause-and-effect relationships between EVCSs and utilize them to learn informative representations from the load data. This module operates in two stages: (1) construction of a causality graph, and (2) causal-dynamic graph learning using the constructed graph to derive feature representations. The architecture of CAGLM is illustrated in Fig. 3. CAGLM starts by normalizing the input EVCS load data, $\boldsymbol{X} \in \mathbb{R}^{N_T \times N_S}$, where N_T represents the number of time steps and N_S represents the number of EVCS. Normalization ensures that features contribute equally to model training and aids in faster convergence during optimization:

$$\tilde{\boldsymbol{X}} = \frac{\boldsymbol{X} - \mu(\boldsymbol{X})}{\sqrt{\frac{1}{N_T} \sum_{t=1}^{N_T} \left(X^{(t)} - \mu(\boldsymbol{X})\right)^2 + \varepsilon}}, \quad (2)$$

where $\tilde{\boldsymbol{X}}$ represents the normalized data matrix, $\mu(\boldsymbol{X}) = \frac{1}{N_T} \sum_{t=1}^{N_T} X^{(t)}$ denotes the mean values of $\boldsymbol{X}$ along the temporal dimension, and ε is a small constant to prevent division by zero.

1) Causality Graph Construction: Instead of analyzing the entire normalized data $\tilde{\boldsymbol{X}}$, we focus on a data segment $\tilde{\boldsymbol{X}}^{(t_b:t_b+t_m)}$, where $t_m < N_T$, to efficiently capture causal structures. This segment is used to infer a causal directed acyclic graph (DAG), $\boldsymbol{\mathcal{G}}_{ca} = (\boldsymbol{V}_{ca}, \boldsymbol{E}_{ca})$, representing the cause-and-effect relationships between EVCSs, where $\boldsymbol{V}_{ca} = \{v^{(1)}, \ldots, v^{(N_s)}\}$ represents the set of EVCS (nodes) and $\boldsymbol{E}_{ca}$ is the set of causal edges. The goal is to find the DAG, $\boldsymbol{\mathcal{G}}_{ca}$, that maximizes the cross-validated likelihood score $\mathcal{L}_{cv}$, subject to the acyclicity constraint of the graph. This optimization problem is addressed using the Time-Fragment Greedy Equivalence Search (TFGES) algorithm (Algorithm 1), based on a combination of the Generalized Score Function (GSF) [38] and cross-validated likelihood. TFGES is a greedy search approach for solving this optimization problem. It iteratively explores potential edges, adding them to the graph only if they lead to a statistically significant increase in the cross-validated likelihood score, $\mathcal{L}_{cv}$, while ensuring the graph remains acyclic. This is achieved by first constructing a temporary graph $\mathcal{G}'$ for each candidate edge $(v^{(i)}, v^{(j)})$, and then evaluating its fit using $\mathcal{L}_{cv}$. If the addition significantly improves $\mathcal{L}_{cv}$, the change is accepted, and the graph $\boldsymbol{\mathcal{G}}_{ca}$ is updated. The process continues until no further statistically significant improvements are observed, or a maximum of 100 iterations is reached.

To improve generalization and robustness to potential noise and inconsistencies in the data, the input to the SGCN is not the raw data $\tilde{\boldsymbol{X}}$, but rather a temporally masked representation, $\boldsymbol{X}_m$, created as follows:

$$\boldsymbol{X}_m^{(i,t)} = \begin{cases} 0, & \text{if } \mathbb{M}_{ca}(R_d) = 1, \\ \boldsymbol{X}^{(i,t)}, & \text{otherwise,} \end{cases} \quad (6)$$

where $\boldsymbol{X}_m^{(i,t)}$ represents the value for EVCS i at time step t after masking. $\mathbb{M}_{ca}(R_d)$ is a Bernoulli random variable with probability R_d (the masking ratio), determining whether a given input element is set to 0.

By stacking multiple GCN layers with appropriate non-linearities, batch normalization (BN), and dropout regularization, the SGCN can learn hierarchical representations that capture both the global influence of causal dependencies from $\boldsymbol{\mathcal{G}}_{ca}$ and the local temporal patterns within each EVCS's load history. The intermediate representations $\boldsymbol{H}_{ca}^{(g,i)}$ obtained from each SGCN block are then combined through element-wise addition (denoted as $\oplus$). To further capture local temporal dependencies, a separate Multilayer Perceptron (MLP) processes the entire (non-masked) data segment and produces a local representation $\boldsymbol{H}_{ca}^{(local)}$. The two elements are then fused to create $\boldsymbol{H}_{ca}^{(r)}$ as follows:

$$\boldsymbol{H}_{ca}^{(r)} \leftarrow \underset{\forall i \in 1,\cdots,N_{ca}}{\text{Aggregate}}\Big(\boldsymbol{H}_{ca}^{(g,i)}\Big) \oplus \boldsymbol{H}_{ca}^{(local)}. \quad (7)$$

Finally, two consecutive 1-dimensional temporal convolutional layers are applied to refine $\boldsymbol{H}_{ca}^{(r)}$:

$$\boldsymbol{H}_{ca}^{(o)} = \text{Conv}_{ca}^{(2)}\Big(\sigma\big(\text{Conv}_{ca}^{(1)}(\boldsymbol{H}_{ca}^{(r)})\big)\Big), \quad (8)$$

where $\text{Conv}_{ca}^{(1)}$ and $\text{Conv}_{ca}^{(2)}$ represent the two temporal convolutional operations, each parametrized with learnable kernel weights and incorporating non-linear activations to model complex patterns within the temporal dimension of the input features. By stacking these convolutional layers, CAGLM performs hierarchical feature extraction, where higher-level features capture more abstract and complex patterns from the fused causal and temporal information present in $\boldsymbol{H}_{ca}^{(r)}$. This multi-level representation $\boldsymbol{H}_{ca}^{(o)}$ then serves as a refined summary of the load dynamics from the CAGLM's perspective, capturing both the network's causal structure and localized temporal trends within the input data segment.

B. Critical Relationship Graph Learning Module (CRGLM)

The Critical Relationship Graph Learning Module (CRGLM), depicted in Fig. 4, plays a crucial role by capturing significant evolving relationships between particularly influential EVCS within the network. It achieves this through a two-step process: identifying the subset of critical nodes (influential EVCS) and then dynamically learning from the graph representing their relationships.

To identify the critical nodes, a measure of centrality based on historical load dependencies is employed. First, a pairwise relationship score $e^{(i,j)}$, signifying the degree of influence between EVCS $v^{(i)}$ and $v^{(j)}$, is calculated as follows:
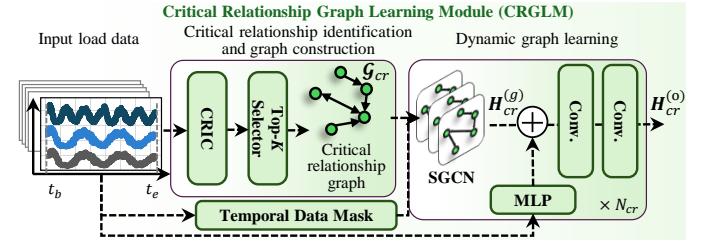
Fig. 4. The overall illustration of the CRGLM.

$$e^{(i,j)} = \frac{\sum_{k=1}^{N_T}\Big(\boldsymbol{X}^{(i,k)} \cdot \boldsymbol{X}^{(j,k-T_d)}\Big)w^{(i,j)}}{\Big(\sqrt{\sum_{k=1}^{N_T}(\boldsymbol{X}^{(i,k)})^2} \cdot \sqrt{\sum_{k=1}^{N_T-T_d}(\boldsymbol{X}^{(j,k)})^2}\Big) + \epsilon}, \quad (9)$$

where $\boldsymbol{X}^{(i,k)}$ is the input feature vector for EVCS i at time step k; T_d represents a temporal shift to account for potential lagged relationships; N_T denotes the historical time window considered for calculating this score, as defined in Eq.(1). The term $w^{(i,j)}$ acts as a learnable weight, allowing the model to dynamically emphasize or de-emphasize specific pairwise EVCS relationships during training. A small constant ϵ is introduced to enhance numerical stability by preventing division by zero.

Using these pairwise relationship scores from Eq. (9), the CRGLM module calculates a centrality score $s^{(i)}$, for each EVCS $v^{(i)}$. This score aggregates the relationship scores of an EVCS with all others in the network, aiming to identify EVCSs with significant overall influence:

$$s^{(i)} = \sum_{j=1}^{N_s}\Big(e_{cr}^{(i,j)} + e_{cr}^{(j,i)}\Big). \quad (10)$$

Critically, by summing both $e_{i,j}$ and $e_{j,i}$, this formulation considers bidirectional influence between EVCS, acknowledging that influential charging stations impact others while also being influenced in return. Following this, the CRGLM employs a Top-K selection mechanism, guided by a threshold parameter R_k, to identify the most influential electric vehicle charging stations (EVCSs) based on their centrality scores. This process formally defines the set of critical nodes $\boldsymbol{C}_{cr}$ as follows:

$$\boldsymbol{C}_{cr} = \text{Top-}K\Big(\boldsymbol{S}, R_k\Big) = \Big\{i \mid \text{rank}\Big(s^{(i)}, \boldsymbol{S}\Big) \le R_k\Big\}, \quad (11)$$

where $\text{rank}(s^i, \boldsymbol{S})$ represents the rank of EVCS i's centrality score s^i within the entire set of centrality scores (denoted by $\boldsymbol{S} = \{s^{(1)}, s^{(2)}, \ldots, s^{(N_s)}\}$). The model deems only those EVCSs as critical whose centrality scores fall within the top R_k proportion, thereby concentrating its attention on the most significant interactions.

These identified critical nodes become the constituent elements of the Critical Relationship Graph. This graph, $\boldsymbol{\mathcal{G}}_{cr} = (\boldsymbol{V}_{cr}, \boldsymbol{E}_{cr}, \boldsymbol{A}_{cr})$ (refer to Definition 2.3), inherits the same set of nodes as the full EVCS network, i.e., $\boldsymbol{V}_{cr} = \boldsymbol{V}$. However, $\boldsymbol{\mathcal{G}}_{cr}$ differentiates itself from the full network graph through its meticulously constructed adjacency matrix $\boldsymbol{A}_{cr}$, specifically designed to exclusively represent interactions between the

critical EVCS. Each entry $\boldsymbol{A}_{cr}^{i,j,t}$ within the time-dependent adjacency tensor $\boldsymbol{A}_{cr}$ at time step t is defined as follows:

$$\boldsymbol{A}_{cr}^{i,j,t} = \begin{cases} 1, & \text{if } (i \in \boldsymbol{C}_{cr} \vee j \in \boldsymbol{C}_{cr}) \wedge (i \neq j), \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

where i and j denotes the different EVCSs. Essentially, $\boldsymbol{A}_{cr}$ encodes a dynamic graph structure where connections are established between critical nodes and other nodes in the network. The time-dependent nature of this adjacency structure enables CRGLM to capture evolving influences between these vital nodes throughout the load forecasting process.

After identifying the critical nodes and forming the initial graph $\mathcal{G}_{cr}$ using Eq. (12), the CRGLM module employs a similar approach to the CAGLM, but on the critical graph $\mathcal{G}_{cr}$. First, CRGLM utilizes a dynamic graph learning structure with SGCN across N_{cr} layers to process the masked data $\boldsymbol{X}_m$. Each layer generates an intermediate representation $\boldsymbol{H}_{cr}^{(g,i)}$, specifically capturing patterns localized around critical nodes. The formula for $\boldsymbol{H}_{cr}^{(g,i)}$ mirrors Eq. (13), but now using the critical graph's adjacency matrix $\boldsymbol{A}_{cr}$ during convolution operations. Analogous to the CAGLM, the representations of $\boldsymbol{H}_{cr}^{(g,i)}$ are aggregated and then concatenated with a separate, locally-focused temporal representation using Eq. (7). The fused feature representation from the CRGLM then undergoes further refinement through a similar two-layer convolutional process as in CAGLM, following the pattern described by Eq. (8), resulting in the CRGLM's final output representation $\boldsymbol{H}_{cr}^{(o)}$:

$$\boldsymbol{H}_{cr}^{(o)} = \text{Conv}_{cr}^{(2)} \left(\sigma \left(\text{Conv}_{cr}^{(1)}(\boldsymbol{H}_{cr}^{(r)}) \right) \right), \tag{13}$$

where $\text{Conv}_{cr}^{(1)}$ and $\text{Conv}_{cr}^{(2)}$ represent the temporal convolutional operations in CRGLM.

C. Temporal Representation Fusion Module and Output Layer

The CAGLM and CRGLM serve complementary roles in capturing the intricate spatiotemporal dependencies of EVCS load dynamics. CAGLM constructs a causality graph to model directional dependencies among EVCSs, leveraging a SGCN to propagate causal influences and refine feature representations. This approach enhances interpretability by explicitly encoding cause-and-effect relationships. In contrast, CRGLM selects a subset of critical nodes based on centrality measures to efficiently model dominant interactions, prioritizing the most influential EVCSs while potentially overlooking finer patterns. The combination of CAGLM and CRGLM provides a more detailed and stable representation of load dynamics. CRGLM improves efficiency by selecting critical nodes but may overlook local load variations, a limitation CAGLM addresses through causal modeling. While CAGLM enhances interpretability with its causality graph, CRGLM captures changing relationships over time, ensuring a balanced and comprehensive understanding of EVCS load dynamics.

The Temporal Representation Fusion Module (TRFM) acts as a bridge, receiving the outputs $\boldsymbol{H}_{ca}^{(o)}$ and $\boldsymbol{H}_{cr}^{(o)}$ from the CAGLM and CRGLM, respectively. Its core responsibility is to effectively integrate the rich, multi-faceted temporal patterns present in these inputs, thus consolidating the insights obtained

by considering both causal relationships and critical EVCS interdependencies. First, $\boldsymbol{H}_{ca}^{(o)}$ and $\boldsymbol{H}_{cr}^{(o)}$ are combined through channel-wise concatenation. Subsequently, two temporal 1D convolutional layers are applied to the combined features, drawing inspiration from their effectiveness in capturing local patterns and dependencies within sequential data, resulting in a temporally rich output representation:

$$\boldsymbol{H}_{tr}^{(o)} = \text{Conv}_{tr}^{(2)} \left(\psi \left(\text{Conv}_{tr}^{(1)}(\boldsymbol{H}_{ca}^{(o)} \oplus \boldsymbol{H}_{cr}^{(o)}) \right) \right), \tag{14}$$

where $\boldsymbol{H}_{tr}^{(o)}$ denotes the TRFM's final output; $\text{Conv}_{tr}^{(1)}$ and $\text{Conv}_{tr}^{(2)}$ represent the consecutive temporal convolution operations within the module, each associated with its own set of learnable parameters, and $\psi(\cdot)$ represents the Rectified Linear Unit (ReLU) activation function, adding complexity and non-linearity crucial for modeling the inherent complexities in EVCS load dynamics. By processing the concatenated representations in this manner, the TRFM aims to discover intricate interplays between the temporal patterns present in $\boldsymbol{H}_{ca}^{(o)}$ (representing long-range causal influences and local temporal variations) and $\boldsymbol{H}_{cr}^{(o)}$ (encapsulating the dynamics among critical EVCSs within the network), ultimately providing the downstream layer with a more informative and robust basis for generating forecasts.

This refined temporal representation $\boldsymbol{H}_{tr}^{(o)}$ from the TRFM then feeds into the output layer to generate the EVCS load forecasts. The input is formed by concatenating $\boldsymbol{H}_{tr}^{(o)}$ with the original, pre-processed (unmasked) EVCS load data spanning from time steps t_b to t_e. Finally, this concatenated input is processed by a fully connected network (FCN), which consists of a series of linear layers that learn to map the fused spatial-temporal representations into the target load forecasts, effectively capturing the diverse relationships learned throughout the various modules of CADGN. Mathematically, this output generation process can be represented as:

$$\hat{\boldsymbol{X}}^{t_e+1:t_e+N_P} = \text{FCN}\left(\boldsymbol{H}_{tr}^{(o)} \oplus \boldsymbol{X}^{t_b:t_e}; \Theta_{out} \right), \tag{15}$$

where $\hat{\boldsymbol{X}}^{t_e+1:t_e+N_P}$ represents the final EVCS load prediction outputted by the CADGN model for time steps from $t_e + 1$ to $t_e + N_P$.

The performance of the CADGN during training is evaluated using the Mean Squared Error (MSE) loss function, calculated by comparing the generated forecasts $\hat{\boldsymbol{X}}^{t_e+1:t_e+N_P}$ with the corresponding ground truth values $\boldsymbol{X}^{t_e+1:t_e+N_P}$:

$$\mathcal{L}_{total}(\Theta) = \frac{1}{N_P N_S} \sum_{i=1}^{N_P} \sum_{j=1}^{N_S} \left(\hat{X}^{i,j} - X^{i,j} \right)^2, \tag{16}$$

where $\mathcal{L}_{total}(\Theta)$ represents the loss for the model parameters Θ, which encompass parameters from all constituent modules like SGCN layers, traditional convolutional layers, the MLP, and other learnable components. By minimizing $\mathcal{L}_{total}(\Theta)$ through an optimization algorithm, the model is driven to generate EVCS load forecasts that align as closely as possible with the historical patterns and relationships it has learned.

TABLE II
THE STATISTICS OF CHARGING LOAD DATASETS.

Area	Time range	Number of stations	Average load (kW)	Peak load (kW)	Standard deviation (kW)
Palo	2016/01/01–2017/08/01	27	1.865	18.645	2.080
Boulder	2022/01/01–2023/06/22	40	0.525	15.892	1.166
Korea	2021/09/30–2022/09/30	14	10.351	147.424	10.437

IV. EVALUATION

A. Dataset Description

To assess the efficacy and performance of the proposed CADGN model for EVCS load forecasting, a comprehensive evaluation is conducted on real-world electric vehicle charging station load datasets gathered from three distinct urban environments: Palo Alto, USA[1], Boulder, USA[2], and a metropolitan area within South Korea [39]. Each dataset intrinsically encompasses records for charging sessions, providing granular information including the start and end timestamps, as well as the total energy consumed (measured in kWh) during each session.

While valuable in its raw form, session-level data must be transformed to become suitable for the time series forecasting task addressed in this work. Accordingly, the raw session data for each EVCS is converted into an hourly aggregated time series, representing the total load demand within each hour at the respective charging station. The transformation process involves three primary steps: (1) The raw dataset, which includes attributes such as start and end times, charging durations, and energy delivered, is cleaned to ensure consistency. Charging sessions with durations exceeding 24 hours, energy delivery over 150 kWh, or invalid timestamps (e.g., start time later than end time) are discarded. (2) The cleaned data is divided into minute-level intervals. Energy delivered during each session is distributed proportionally across its duration, resulting in a detailed temporal representation of the energy usage. (3) The minute-level data is aggregated into hourly intervals. The total energy delivered across all charging stations is calculated, adhering to the temporal resolution commonly used by prior studies [20], [21], [40] and utility providers for effective planning and decision-making.

We use a 7:1:2 data split ratio for training, validation, and testing. Table II summarizes essential descriptive statistics for each processed dataset, highlighting potential inter-dataset differences in EV charging patterns, network sizes, and load characteristics which may affect forecasting difficulty and influence model performance.

B. Experimental Settings

In this study, three commonly used metrics in the time series forecasting domain are adopted: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The experiments are implemented within the PyTorch framework (version 1.13.1). All models were conducted using an NVIDIA RTX A6000 with 48GB of memory. Adam optimizer is employed for training all models. The hyperparameter settings for CADGN and the

[1]https://data.cityofpaloalto.org

[2]https://open-data.bouldercolorado.gov/datasets/

baseline models are optimized using a grid search strategy, which evaluates a range of values for key parameters. The detailed hyperparameter search space and the optimal configurations for CADGN and the baseline models are presented in Appendix B, Table A2. The review window size N_T of all models was set to 72h (3 days) for a fair comparison. The initial learning rate was 0.001, batch size was 32. Model training is conducted for a maximum of 100 epochs. Early stopping mechanisms are implemented, particularly for deep learning baselines prone to overfitting, based on monitoring validation performance. The predictive horizons evaluated are 3, 6, and 12 hours ahead, reflecting the need for both short-term and mid-term EVCS load predictions for grid operational and planning applications. To ensure reliable and statistically significant results, each experiment, including baseline and CADGN evaluations across different parameters and forecasting horizons, is repeated 5 times with varying random model weight initializations. The average performance from these runs is presented in the results.

C. Multi-step Point Estimation for EVCS Loads

This subsection presents a quantitative analysis of the multi-step ahead forecasting performance for the proposed CADGN model and a comprehensive set of baseline models. Dlinear [41] is a novel lightweight forecasting method. GRU [42] is a commonly used technique for sequence data learning. As for multivariate time series forecasting, we selected advanced models including Informer [17] (Inf.), Crossformer [18] (Cross.), PatchTST [19], and NHiTS [43], all of which have been applied to load forecasting tasks previously. Additionally, we incorporated several typical graph neural networks for comparison: GAIN [23], MAGNet [24], AGCRN [8], and Graph WaveNet [25] (GWN). Among these, GAIN and Graph WaveNet are widely used in load forecasting. MAGNet and AGCRN focus on pivotal nodes in graph learning, aligning closely with our approach.

Table III summarizes the results for each model and dataset combination, MAE, RMSE, and MAPE at 3-hour, 6-hour, and 12-hour prediction horizons. Quantitative evaluation reveals that CADGN demonstrates consistently strong performance, frequently exceeding, or at a minimum equaling, the accuracy of the baseline models across the majority of the evaluated forecasting scenarios. This superior performance is demonstrably evident in the often lower MAE and RMSE values achieved by CADGN, indicating its capacity for generating more precise EVCS load predictions, particularly for mid-term forecasting horizons (6 and 12 hours). Among the advanced models for multivariate time series forecasting, Informer, Crossformer, PatchTST, and NHiTS demonstrated varying degrees of success. However, none consistently outperformed CADGN, particularly on longer forecasting horizons. GNN-based models, including GAIN, MAGNet, AGCRN, and GWN, are notable for their application in load forecasting. Despite their strengths, these models were consistently outperformed by CADGN, which leveraged causality-aware components and dynamic graph structures to better capture the underlying data relationships. These observations lend empirical support to our initial hypothesis: integrating both

TABLE III

FORECAST RESULTS WITH 72 REVIEW WINDOW AND PREDICTION LENGTH {3, 6, 12}. THE BEST RESULT IS IN BOLD, FOLLOWED BY UNDERLINING.

Area	Steps	Metrics	GRU	Inf.	Cross.	Dlinear	PatchTST	NHiTS	GAIN	MAGNet	AGCRN	GWN	CADGN
Boulder	3	MAE	0.647	0.672	0.648	0.571	0.535	0.534	0.682	0.546	0.561	0.555	**0.530**
		RMSE	1.391	1.456	1.294	1.217	1.203	1.214	1.394	**1.199**	1.257	1.205	1.222
		MAPE	2.717	2.720	3.225	2.714	2.396	2.287	3.256	2.492	2.411	2.549	**2.109**
	6	MAE	0.696	0.735	0.713	0.655	0.627	0.633	0.762	0.612	0.657	0.665	**0.610**
		RMSE	1.483	1.542	1.382	1.338	1.321	1.327	1.511	**1.315**	1.374	1.331	1.326
		MAPE	2.856	3.095	3.628	3.145	2.955	2.921	3.600	2.726	3.025	3.318	**2.602**
	12	MAE	0.719	0.726	0.705	0.688	0.667	0.666	0.723	0.656	0.671	0.699	**0.644**
		RMSE	1.505	1.532	1.409	1.371	1.376	1.367	1.482	**1.362**	1.404	1.370	1.363
		MAPE	3.052	2.962	3.286	3.424	3.161	3.196	3.251	2.959	3.016	3.630	**2.862**
Palo	3	MAE	1.333	1.343	1.290	1.249	1.252	1.246	1.304	1.214	1.193	1.288	**1.189**
		RMSE	1.926	1.953	1.850	1.794	1.800	1.798	1.910	1.768	1.798	1.799	**1.746**
		MAPE	4.618	4.616	4.761	4.409	4.407	4.320	4.329	4.081	3.873	5.104	**3.869**
	6	MAE	1.355	1.402	1.390	1.356	1.300	1.307	1.503	1.304	1.317	1.356	**1.274**
		RMSE	1.971	2.028	1.925	1.911	**1.850**	1.870	2.193	1.862	1.889	1.895	**1.850**
		MAPE	4.540	4.725	5.566	4.878	4.665	4.735	5.018	4.528	4.591	5.099	**4.257**
	12	MAE	1.377	1.406	1.303	1.357	1.336	1.311	1.497	1.336	1.341	1.356	**1.283**
		RMSE	1.998	2.037	1.887	1.904	1.896	1.885	2.108	1.893	1.928	1.901	**1.880**
		MAPE	4.646	4.749	4.375	4.959	4.794	4.690	5.327	4.714	4.750	5.048	**4.021**
Korea	3	MAE	7.146	7.839	6.745	6.612	6.562	6.601	6.842	**6.378**	6.587	6.438	6.380
		RMSE	11.526	12.680	11.411	11.033	10.965	11.108	11.341	10.768	11.055	10.787	10.710
		MAPE	12.807	13.850	10.348	10.445	10.194	10.246	11.531	9.792	10.413	10.246	9.869
	6	MAE	7.553	7.831	7.555	7.090	7.170	7.294	7.797	7.062	8.031	6.984	**6.891**
		RMSE	12.031	12.643	12.404	11.510	11.652	11.777	12.874	11.465	12.700	11.405	11.349
		MAPE	13.864	13.540	11.694	11.869	11.863	13.133	12.356	12.475	14.558	11.825	10.878
	12	MAE	7.791	7.895	7.859	7.347	7.393	7.857	8.011	7.507	7.914	7.339	**7.334**
		RMSE	12.558	12.780	12.732	11.821	11.867	12.321	12.967	11.958	12.725	**11.798**	11.820
		MAPE	13.469	13.563	12.980	12.823	12.938	15.313	13.314	13.818	13.702	13.165	**12.585**

causality awareness and a dynamic focus on critical EVCS network interactions facilitates the modeling of complex interdependencies, often neglected by traditional methods and conventional GNN models.

Furthermore, as the prediction horizon is extended, all forecasting methods predictably exhibit diminished accuracy, reflected by an overall increase in the evaluated metrics. However, the degradation in performance for CADGN is observed to be consistently less pronounced in comparison to the other baseline approaches. This robustness across various conditions highlights CADGN's effectiveness and adaptability, thus offering valuable advantages in scenarios demanding greater predictive horizons.

D. Ablation Study

To thoroughly investigate the contribution of each individual component within CADGN to the improved overall model performance, a comprehensive ablation study is conducted. Four distinct model variants are meticulously designed, strategically removing specific elements while maintaining the core framework of the network: **1) w/o-TEM**: The temporal masking component implemented within CAGLM and CR-GLM are deactivated, allowing an assessment of the impact of data augmentation through temporal masking on model generalization. **2) w/o-CRG**: This variant entirely removes the CRGLM from the architecture, allowing for the investigation into the performance impact of modeling dynamically changing relationships between the identified critical EVCS nodes. **3) w/o-IR**: Both the local representations $H^{(local)}$ within CAGLM and CRGLM are removed. This study examines the necessity of representing local, non-causal, temporal trends alongside global graph structure information. **4) w/o-CAG**: In this variant, the entire CAGLM is removed, enabling the study of the necessity for causal inference to the model's overall performance.

TABLE IV

PERFORMANCE OF ABLATION STUDY. THE BEST RESULT IS IN BOLD, FOLLOWED BY UNDERLINING.

Area	Metrics	CADGN	w/o-TEM	w/o-CRG	w/o-IR	w/o-CAG
Boulder	MAE	**0.530**	0.535	0.541	0.534	0.541
	RMSE	1.222	1.227	1.226	**1.218**	1.228
	MAPE	**2.109**	2.165	2.269	2.193	2.281
Palo	MAE	**1.189**	1.207	1.229	1.208	1.209
	RMSE	**1.746**	1.762	1.791	1.762	1.763
	MAPE	**3.869**	3.977	3.920	4.268	3.947
Korea	MAE	**6.380**	6.400	6.531	6.504	6.418
	RMSE	**10.710**	10.743	10.817	10.886	10.752
	MAPE	9.869	**9.865**	10.536	10.163	10.066

The performance of each CADGN variant, evaluated on the three datasets using the same setup as in the previous experiment, is meticulously documented and compared to the performance of the full CADGN model (reported in Table III). Table IV summarizes the results for all the variant model configurations, enabling in-depth insights into the role and importance of each module within the overall network design. The ablation study results presented in Table IV reveal that the full CADGN model consistently achieves the highest EVCS load forecasting performance, emphasizing the synergy of its components. The most notable observation is that the exclusion of the Critical Relationship Graph Learning Module (CRGLM) leads to the most significant degradation in performance, with substantial increases in MAE and RMSE across all datasets. This underscores the importance of dynamically focusing on critical EVCS interactions, which CRGLM effectively captures, providing valuable insights for improving load prediction accuracy. This crucial role of CRGLM resonates with our hypothesis regarding the significance of capturing critical EVCS dynamics for improved forecasting.

Additionally, the CAGLM is also essential for capturing causal relationships, though the removal of its temporal masking feature showed minor effects. This suggests that while

CAGLM's causal inference is crucial, the specific implementation details like temporal masking may vary in importance depending on the dataset's characteristics, such as its robustness to noise and variability in historical data patterns. These ablation results clearly demonstrate that incorporating causal inference (through CAGLM) and modeling dynamic critical EVCS relationships (via CRGLM) individually enhances predictive performance, and together, these modules work in conjunction to achieve superior and robust forecasting accuracy, validating the key arguments and motivations for this work.

E. Parameter Sensitivity Analysis

This part evaluates the impact of key CADGN parameters affect the model's forecasting accuracy and efficiency, highlighting sensitivities that can guide model tuning. Specifically, the investigation centers on: 1) temporal delay window size T_d used during critical node identification in Eq. (9). This parameter reflects the duration of past data used to assess EVCS relationships. 2) Data masking ratio R_d used in Eq. (6). This parameter, ranging from 0.1 to 1, determines the probability of masking individual load values within CAGLM's and CRGLM input, assessing the impact of data augmentation. 3) Top-k Node selection ratio R_d for CRGLM used in Eq. (11). This ratio dictates the fraction of all EVCS selected as "critical" nodes within the network.

Fig. 5(a) illustrates the effect of varying T_d on MAE (red curve) and RMSE (blue curve). Both metrics rise significantly when $log_2(T_d)$ exceeds 3, indicating that a longer temporal delay window may capture irrelevant relationships, thereby reducing accuracy. Conversely, employing shorter temporal windows appears to facilitate sharper critical node identification and maintain higher predictive accuracy.

Fig. 5(b) investigates the influence of R_d on MAE and RMSE. We observe relatively stable MAE (red curve) and RMSE (blue curve) across the explored range. This relative robustness indicates that CADGN remains comparatively resilient to missing information or masked inputs during the causality learning phase, while hinting that datasets with higher overall variation in loads might benefit from tuning R_d to attain further performance gains.

Fig. 5(c) shows that both MAE and RMSE improve as the critical node selection ratio R_k increases to around 0.4, but both metrics rise significantly when R_k exceeds 0.6. This suggests that moderately expanding the critical node set enhances accuracy by capturing key inter-station dependencies, while an overly large set leads to diminishing returns by diluting the influence of genuinely critical nodes. Additionally, as shown in Fig. 5(d), inference time increases nearly linearly with larger R_k values, reflecting the higher computational cost of larger graphs. Therefore, balancing R_k is crucial to optimize both accuracy and computational efficiency.

F. Study of Prediction Performance

Fig. 6 compares the EVCS load forecasting performance of CADGN against benchmark models (GWN, MAGNet, PatchTST) across three datasets. A key challenge in EVCS load forecasting lies in accurately predicting volatile peak

(a) Effect of T_d on MAE and MSE (b) Effect of R_d on MAE and MSE

(c) Effect of R_k on MAE and MSE (d) Effect of R_k on efficiency
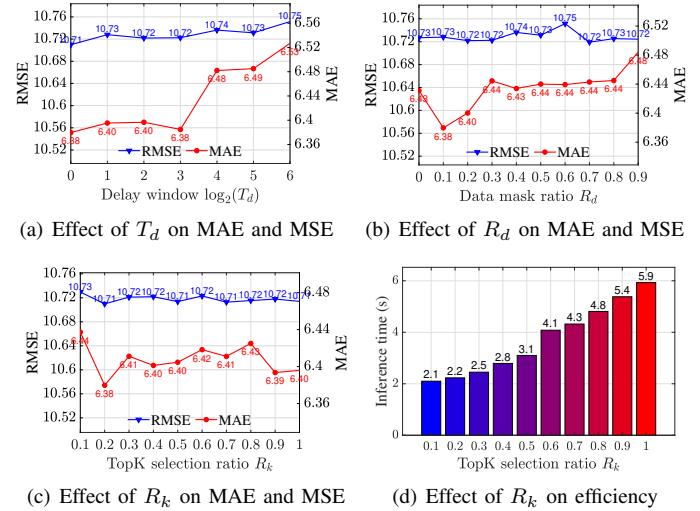
Fig. 5. Performance of sensitivity analysis.

demands, which are influenced by sudden fluctuations in user behavior, grid constraints, and time-dependent usage patterns. CADGN consistently demonstrates superior performance by closely tracking actual load trajectories across both high- and low-demand periods. This is especially evident in the Korea dataset, which exhibits more pronounced peaks and troughs compared to the other datasets. In particular, CADGN outperforms the other models in capturing sharp demand surges during peak hours, as highlighted in yellow in Fig. 6. This enhanced performance is attributed to CADGN's unique ability to model both causal relationships (via CAGLM) and the dynamics between critical nodes (through CRGLM), allowing it to effectively capture the complex interdependencies and load fluctuations within EVCS networks. As a result, CADGN achieves an maximum improvement of 6.9% in MAE over the benchmark models on the Korea dataset, highlighting its superior capacity to provide more accurate predictions across a range of demand scenarios.

However, for some particularly extreme data points, the prediction performance remains relatively less accurate. This is primarily due to the inherent unpredictability of extreme events, which are often driven by rare, non-recurring factors such as sudden surges in demand, unexpected grid constraints, or anomalous user behavior patterns. While CADGN is effective in modeling complex dependencies, the stochastic nature of these extreme variations poses a fundamental challenge for any forecasting model. Nevertheless, CADGN still outperforms comparable methods overall, demonstrating superior adaptability across diverse demand scenarios.

G. Study of Model Efficiency

Table V summarizes the efficiency comparison between CADGN and benchmark models for a typical instance in a batch learning process, covering training time, GPU memory usage, inference time, MAE, and RMSE. In terms of computational efficiency, CADGN demonstrates superior performance with a training time of 0.025 seconds, significantly outperforming Cross. (5.1× slower), MAGNet (4.3× slower), and AGCRN (16× slower). While PatchTST and GWN train
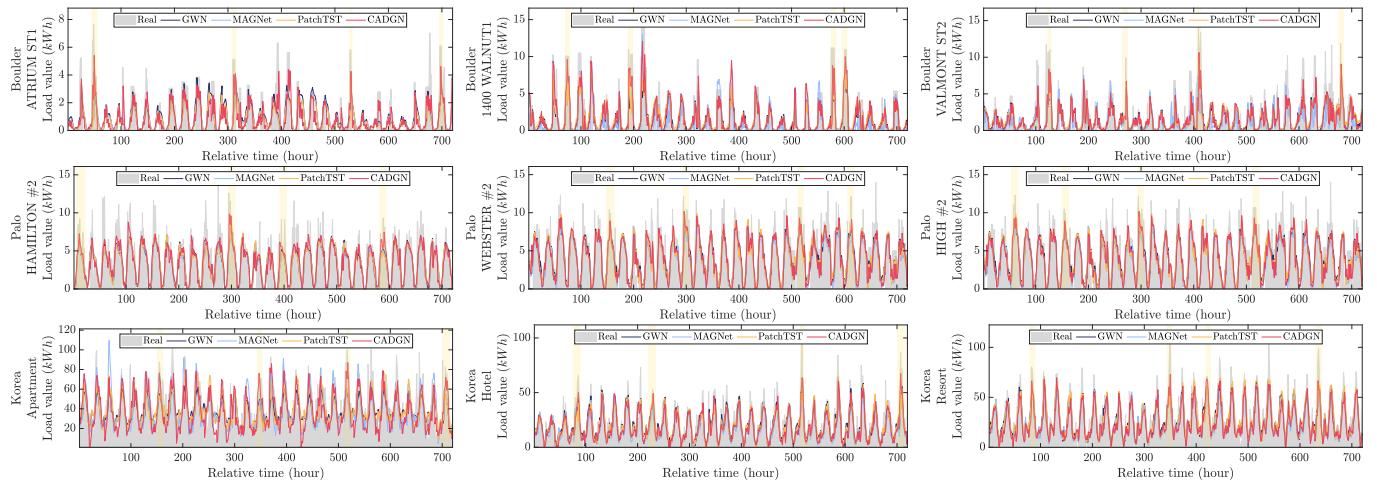
Fig. 6. Predictions from the proposed CADGN and benchmark models for various scenarios with a 3-hour prediction horizon.

slightly faster (1.1× and 0.4× relative speeds, respectively), CADGN requires the least GPU memory due to CRGLM's sparse graph representation, which focuses only on critical EVCS nodes, thereby reducing the memory footprint. This efficiency makes CADGN well-suited for large-scale smart grid applications with hardware constraints. Regarding inference time, CADGN ranks in the mid-range, outperforming AGCRN but slightly trailing behind PatchTST. In terms of forecasting accuracy, CADGN surpasses all benchmarks, achieving a relative improvement in MAE ranging from 1.8% to 8.6%. The RMSE of 1.363 remains competitive with MAGNet's 1.362 while delivering 4.3× higher training efficiency and reducing memory consumption by 25% compared to MAGNet. Overall, CADGN achieves a strong balance between computational efficiency and prediction accuracy, making it well-suited for hourly or finer-grained EVCS load forecasting.

TABLE V
EFFICIENCY COMPARISON OF CADGN AND BENCHMARK MODELS.

Model	Training time (s)	GPU memory usage (MB)	Inference time (s)	MAE	RMSE
Cross.	0.139 (5.1×)	5443 (3.4×)	0.043 (2.1×)	0.705	1.409
PatchTST	0.029 (1.1×)	1795 (1.1×)	0.005 (0.2×)	0.667	1.376
MAGNet	0.108 (4.3×)	2113 (1.3×)	0.030 (1.4×)	0.656	**1.362**
AGCRN	0.409 (16×)	2013 (1.3×)	0.171 (8.4×)	0.671	1.404
GWN	**0.009** (0.4×)	1619 (1.0×)	**0.002** (0.1×)	0.699	1.370
CADGN	0.025	**1585**	0.020	**0.644**	1.363

H. Illustration of Graph Learning Module Differences

Fig. 7 offers a visual comparison between the CAGLM and CRGLM, showcasing the contrasting nature of the learned graphs and their impact on feature representations. Fig. 7(a) and Fig. 7(b) depict the adjacency matrices for CAGLM and CRGLM, respectively. The sparsity of CAGLM's adjacency matrix indicates its focus on representing specific and direct causal relationships. Conversely, the denser structure of CRGLM's adjacency matrix reflects its role in capturing the broader influence of critical nodes on other EVCS within the network. Fig. 7(c) and Fig. 7(d) display the corresponding GCN features learned by each module. The heterogeneous patterns within CAGLM's GCN feature map suggest the
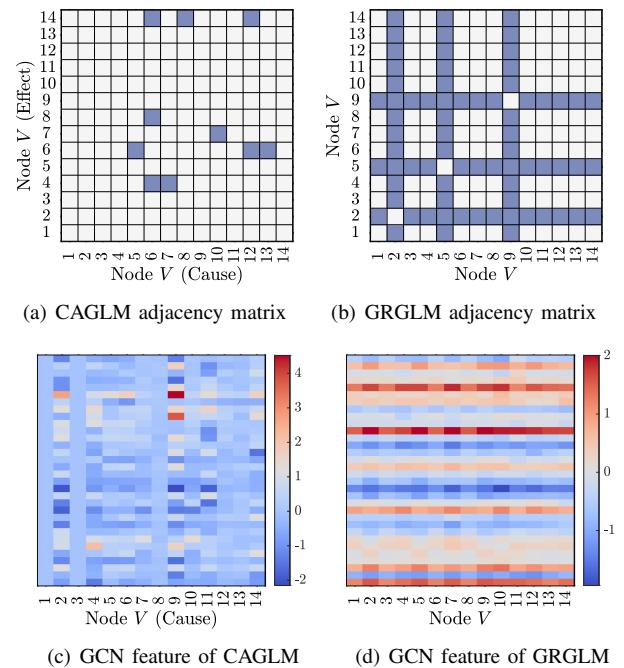
Fig. 7. Examples of adjacency matrices and GCN features for CAGLM and GRGLM.

extraction of detailed and specific information, vital for representing nuanced causal dependencies. In contrast, the relatively uniform values in CRGLM's GCN feature map indicate the learning of general, global trends among the most influential nodes.

These contrasting features highlight the complementary roles of the two modules: CAGLM focuses on precision and localized causal inference, while CRGLM aims for generalization and captures network-wide patterns driven by critical node dynamics. By fusing their respective representations, CADGN leverages a multifaceted understanding of the EVCS network to achieve superior prediction accuracy.

I. Verification of Causality in EVCS Datasets

To establish the presence of causal relationships in EVCS load data, we employ Granger causality tests [32] to learn

causal graph structures. Granger causality determines whether one time series provides statistically significant information for predicting another and is widely used in time-series analysis [44]. Fig. 8(a) illustrates the overall prevalence of causality in three EVCS datasets. Each dataset is partitioned into batches, and pairwise Granger causality tests are conducted on the first sample of each batch. Although the data segments in each batch are relatively small, a notable portion of causal links remain statistically significant. This suggests that meaningful causal dependencies exist within the EVCS load data, rather than being purely random fluctuations.

Due to the large number of stations in real-world EVCS datasets, which introduce intricate and indirect causal relationships that are difficult to present clearly, a synthetic case is used for better interpretability. To further demonstrate the capacity to detect causal effects in multi-station scenarios, Fig. 8(b) illustrates a synthetic case involving two clearly interpretable events. In Event 1, an external social activity near Station A increases its load, which subsequently influences the load at Station B. In Event 2, a local fault or maintenance at Station C causes a sudden load drop, prompting users to shift to Station D, thereby increasing its load.
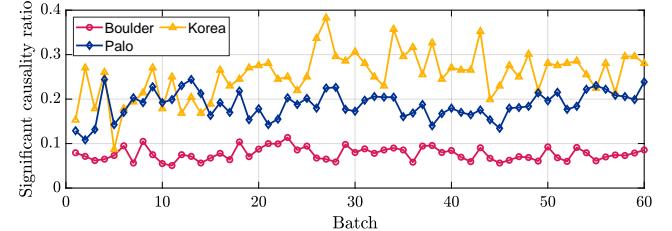
These synthetic events are validated using Granger causality tests. Fig. 8(c) and Fig. 8(d) present the p-values for each station pair during Events 1 and 2, respectively. The highlighted low p-values confirm the hypothesized causal links (e.g., A $\rightarrow$ B for Event 1 and C $\rightarrow$ D for Event 2). Subsequently, we apply the proposed TFGES algorithm to learn the underlying causal graphs. Fig. 8(e) and Fig. 8(f) display the resulting graph structures and their causal scores, which align with the causal directions identified by the Granger tests. Overall, these results provide the evidence of latent causal relationships in EVCS load behavior. The alignment between the Granger causality tests and the learned causal graphs further validates the effectiveness of CADGN in capturing and leveraging causal dependencies for improved forecasting accuracy.

J. Analysis of Adjacency Structures and Learned Feature Representations
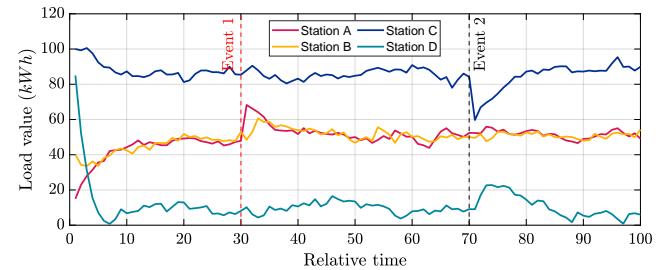
To evaluate the effectiveness of the proposed method in capturing meaningful causal relationships and refining feature representations, this part analyze both the adjacency structures and the learned feature maps. Fig. 9 presents a comparative visualization of adjacency matrices obtained through different methods and the corresponding feature representations learned by the graph-based models.

Fig. 9(a) presents the adjacency matrix derived from statistically significant connections ($p < 0.05$) identified using Granger causality tests, while Fig. 9(b) shows the corresponding matrix of p-values. The Granger-based approach captures a broad set of potential causal links, yet it tends to retain connections of marginal significance, leading to relatively dense adjacency structures. This density may introduce redundancy and increase model complexity, potentially affecting interpretability and computational efficiency.
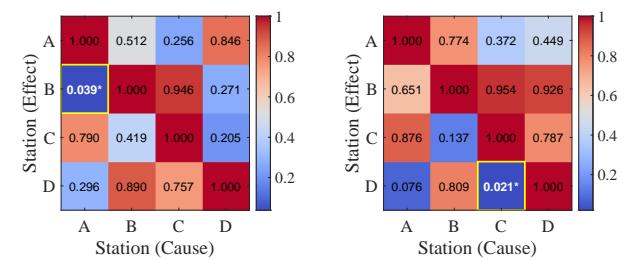
In contrast, Fig. 9(c) illustrates the adjacency matrix obtained through the TFGES method, which employs a causal

(a) Variation of significant causality ratio across different datasets
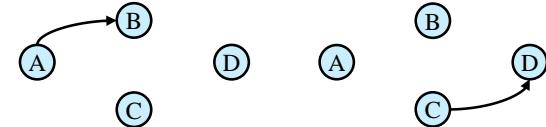
(b) Synthetic case data with two events

(c) p-values from Granger Causality Tests for Event 1

(d) p-values from Granger Causality Tests for Event 2

Causal Score: 1544.19

(e) The causal graph and causal score learned by TFGES for Event 1

Causal Score: 2266.77

(f) The causal graph and causal score learned by TFGES for Event 2

Fig. 8. Causality analysis for three EVCS datasets and a synthetic case.

structure search strategy to construct a more refined and parsimonious graph. This sparser representation filters out weak connections, improving the efficiency of downstream learning tasks.

Fig. 9(d) and Fig. 9(f) depict the first-layer feature maps learned by the SGCN model under the CAGLM and CRGLM frameworks, respectively. In Fig. 9(d), feature activations appear in regions where Granger causality previously assigned low significance, enhancing the effectiveness of the proposed method in refining causal structures. Meanwhile, Fig. 9(f) highlights critical node regions that correspond closely with the adjacency matrix in Fig. 9(e), demonstrating that CRGLM effectively prioritizes critical nodes in the EVCS network.

V. DISCUSSION

The experimental results from Section IV provide comprehensive evidence that the proposed CADGN framework

(a) Significant p-value adjacency matrix from Granger causality

(b) p-value matrix from Granger causality

(c) Adjacency matrix learned by the proposed TFGES method

(d) First-layer feature map from SGCN in CAGLM

(e) Critical node adjacency matrix

(f) First-layer feature map from SGCN in CRGLM
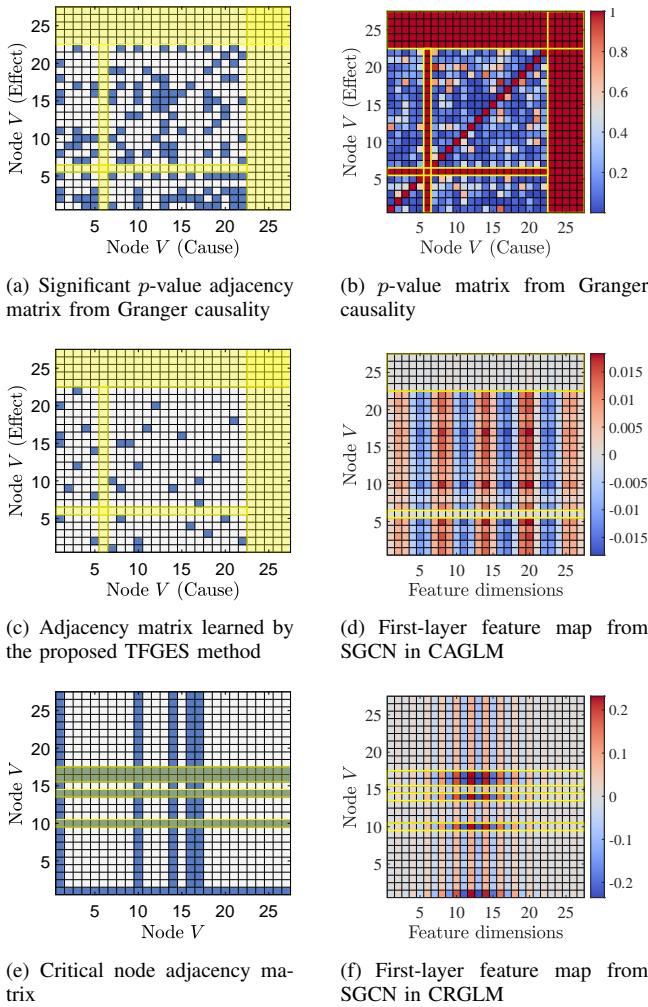
Fig. 9. Visualization of causal adjacency matrices and SGCN feature maps on Palo dataset with 72 review window and prediction length 3.

effectively addresses the challenges of multi-step EVCS load forecasting, offering enhanced accuracy, robustness, and interpretability compared to existing approaches. Several key observations can be highlighted: (1) CADGN consistently outperforms transformer-based time series models and GNN-based approaches across multiple datasets, demonstrating superior forecasting accuracy, particularly for mid-term horizons (6 and 12 hours). Compared to the best-performing baselines, CADGN achieves up to a 8.1% reduction in MAPE, confirming its ability to capture both localized temporal dependencies and global network interactions. As forecasting horizons extend, all models exhibit declining accuracy, but CADGN remains more robust, with a notably smaller performance drop. (2) The ablation study confirms that CADGN's key components contribute significantly to its performance. Removing CRGLM leads to the most substantial degradation, increasing MAE and RMSE by over 6%, underscoring its critical role in dynamically modeling influential EVCS nodes. The exclusion of CAGLM also weakens performance, highlighting the necessity of causal inference in improving prediction accuracy. (3) Parameter sensitivity analysis reveals that excessively large temporal delay windows $T_d > 2^3$ and

critical node selection ratios $R_k > 0.6$ introduce redundant relationships, degrading both accuracy and efficiency. CADGN remains relatively stable across different data masking ratios, suggesting its resilience to missing data. (4) CADGN demonstrates strong forecasting robustness, effectively tracking sharp load fluctuations and peak demand variations. In the Korea dataset, which exhibits significant volatility, CADGN outperforms other models in capturing demand surges. (5) The computational efficiency of CADGN is also competitive, with training times 5.1× faster than Crossformer and 16× faster than AGCRN, while maintaining lower memory consumption due to CRGLM's sparse graph representation. (6) Graph structure analysis confirms that CAGLM captures precise causal relationships, while CRGLM focuses on broader network dynamics. Granger causality tests validate that meaningful causal dependencies exist in EVCS load data, further supporting CADGN's effectiveness in leveraging these relationships. Compared to dense adjacency structures obtained via direct statistical testing, CADGN's refined causal graph representations improve interpretability and learning efficiency. Overall, the results demonstrate that CADGN effectively integrates causality awareness and dynamic node selection to enhance multi-step forecasting accuracy while maintaining a balance between efficiency and scalability. These advantages position CADGN as a robust solution for real-world EVCS load forecasting applications.

VI. CONCLUSIONS

In this paper, we propose the Causality-Aware Dynamic Multi-Graph Convolutional Network (CADGN), a novel deep learning framework designed to improve EVCS load forecasting. CADGN addresses the limitations of traditional time-series and graph-based methods by incorporating two key modules: the Causality-Aware Graph Learning Module (CAGLM) and the Critical Relationship Graph Learning Module (CRGLM). CAGLM identifies cause-and-effect relationships among EVCS loads to extract informative representations, while CRGLM determines and models the dynamic relationships of critical EVCS based on a novel centrality scoring mechanism. A Temporal Representation Fusion Module (TRFM) integrates the spatial and temporal insights from these modules to provide accurate load predictions. Empirical evaluations on real-world datasets from diverse geographical locations demonstrated CADGN's superior performance across multiple forecasting horizons, often surpassing state-of-the-art baselines. The ablation study highlighted the importance of each module, particularly CRGLM, in achieving high accuracy. This work serves as a significant contribution toward developing more effective and insightful forecasting models to address the challenges associated with widespread EV adoption. By integrating deeper causal understanding and prioritizing the role of critical network interactions, approaches such as CADGN pave the way for developing more resilient and adaptable infrastructure ready to cope with the complexity of future electrified transportation systems.

While CADGN demonstrates strong performance in EVCS load forecasting, there are still areas for improvement. First,

the model faces challenges in predicting extreme load variations, particularly sudden spikes or drops in charging demand, which are inherently difficult to forecast due to the stochastic nature of EV charging behavior. Second, the causal graph construction and graph learning processes remain separate, inevitably increasing computational costs. Third, while CADGN captures causal relationships, it does not fully disentangle complex causal dependencies in EVCS load dynamics or trace them to specific influencing factors.

To address these limitations, we will explore extreme regression modeling and post-processing correction to enhance forecasting accuracy for rare but critical EVCS load fluctuations. To optimize causal discovery and graph learning simultaneously during training and reduce computational overhead, we plan to integrate neural causal inference. Additionally, we will investigate counterfactual analysis and attention-based causal discovery to better understand the causal mechanisms behind EV charging patterns and improve model interpretability.

APPENDIX A
ADDITIONAL EXPERIMENTAL RESULTS

TABLE A1
PERFORMANCE COMPARISON ON PERTH DATASET.

Steps	Metrics	Dlinear	PatchTST	MAGNet	AGCRN	GWN	CADGN
3	MAE	2.876	2.853	2.731	2.827	2.713	**2.662**
	RMSE	5.565	5.563	5.558	5.625	5.566	**5.510**
	MAPE	10.305	10.043	8.714	9.689	8.515	**8.082**
6	MAE	2.927	2.902	2.798	3.042	2.841	**2.700**
	RMSE	5.645	5.648	5.631	5.800	5.663	**5.604**
	MAPE	10.539	10.188	9.193	11.490	8.872	**8.110**
12	MAE	2.916	2.970	2.824	3.086	2.812	**2.747**
	RMSE	5.704	5.690	5.687	5.955	5.708	**5.637**
	MAPE	10.116	10.863	9.266	11.227	9.090	**8.587**

To further evaluate the effectiveness of CADGN, additional experiments were conducted using the Perth dataset[3], which contains EVCS load data from 13 stations in Perth, Australia. The dataset spans from January 1, 2017, to August 31, 2019, with an hourly frequency. While the number of stations is similar to that of the Korea dataset, the average load is significantly lower at 2.084 kW. However, the peak load (152.195 kW) is higher than that of the Korea dataset, indicating greater local fluctuations. In comparison, Boulder (USA) has more stations but lower peak loads, whereas Palo (USA) exhibits a more balanced load distribution. These regional variations influence charging behaviors, making Perth an ideal dataset for evaluating forecasting robustness under high-variability conditions. To ensure consistency in experimental comparisons, the dataset was processed using the same data organization procedures as the other three datasets.

The performance comparison on the Perth dataset is summarized in Table A1. CADGN consistently outperforms benchmark models across all forecasting horizons, achieving the lowest MAE, RMSE, and MAPE. Notably, CADGN achieves up to a 9% improvement in MAPE, particularly outperforming both graph-based models and transformer-based models, highlighting its effectiveness in capturing causal dependencies and

[3] https://data.pkc.gov.uk/datasets/

TABLE A2
HYPER-PARAMETER SETTINGS.

Model	Parameter	Option range
GRU	Hidden size	$\{2^4, 2^5, 2^6\}$
Inf.	Encoder layers	1-5 (1 per step)
	Decoder layers	1-5 (1 per step)
Cross.	Label length	1-10 (1 per step)
	Hidden dimension	$\{2^4, 2^5, 2^6, 2^7, 2^8\}$
PatchTST	The numbers of heads	$\{2^2, 2^3, 2^4, 2^5\}$
	Patch length	2-24 (2 per step)
Dlinear	Decomposition kernel size	3-15 (2 per step)
NHiTS	The number of blocks	1-5 (1 per step)
	The layer number	1-3 (1 per step)
	Hidden dimension	$\{2^4, 2^5, 2^6, 2^7, 2^8\}$
	Pooling size	2-8 (2 per step)
GAIN	Hidden dimension	$\{2^4, 2^5, 2^6, 2^7, 2^8\}$
	The number of heads of GAT	$\{2^0, 2^1, 2^2, 2^3, 2^4\}$
	GRU layer number	$\{2^0, 2^1, 2^2, 2^3, 2^4\}$
	CNN kernel size	$\{2^4, 2^5, 2^6, 2^7\}$
	CNN out channels size	$\{2^4, 2^5, 2^6, 2^7, 2^8\}$
MAGNet	Label length	1-10 (1 per step)
	Input channel size	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
	Skip channel size	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
AGCRN	RNN units	$\{2^4, 2^5, 2^6, 2^7, 2^8\}$
	The layer number	1-5 (1 per step)
GWN	Dilation channel size	$\{2^4, 2^5, 2^6, 2^7, 2^8\}$
	Blocks number	2-8 (2 per step)
CADGN	Hidden dimension	$\{2^4, 2^5, 2^6, 2^7, 2^8\}$
	GCN out channels size	$\{2^0, 2^1, 2^2, 2^3, 2^4\}$
	GCN layer number	1-3 (1 per step)
	Delay window	1-6 (1 per step)
	Data mask ratio	0.0-0.9 (0.1 per step)
	TopK selection ratio	0.1-1.0 (0.1 per step)

evolving critical relationships. Unlike baseline models, which exhibit increasing errors over longer forecasting horizons, CADGN maintains stable RMSE values, demonstrating its capability to capture both short-term fluctuations and long-term dependencies. These results further underscore CADGN's robustness and adaptability in handling diverse EVCS load patterns, making it a promising solution for real-world EVCS load forecasting applications.

APPENDIX B
PARAMETER SETTINGS

The search space of hyper-parameters for CADGN and baseline methods is summarized in Table A2. The hyper-parameters are optimized using grid search with a 5-fold cross-validation on the training set. The best hyper-parameters are selected based on the lowest MAE on the validation set.

REFERENCES

[1] S. Powell, G. V. Cezar, L. Min, I. M. Azevedo, and R. Rajagopal, "Charging infrastructure access and operation to reduce the grid impacts of deep electric vehicle adoption," *Nature Energy*, vol. 7, no. 10, pp. 932–945, 2022.

[2] Y. Jiang, T. H. Ortmeyer, M. Fan, and X. Ai, "Data-driven low-rank tensor approximation for fast grid integration of commercial ev charging stations considering demand uncertainties," *IEEE Transactions on Smart Grid*, vol. 14, no. 1, pp. 517–529, 2023.

[3] Y. Zheng, D. R. Keith, S. Wang, M. Diao, and J. Zhao, "Effects of electric vehicle charging stations on the economic vitality of local businesses," *Nature Communications*, vol. 15, no. 1, p. 7437, 2024.

[4] O. N. Nezamuddin, C. L. Nicholas, and E. C. dos Santos, "The problem of electric vehicle charging: State-of-the-art and an innovative solution," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4663–4673, 2021.

[5] C. He, J. Zhu, J. Lan, S. Li, W. Wu, and H. Zhu, "Optimal planning of electric vehicle battery centralized charging station based on ev load forecasting," *IEEE Transactions on Industry Applications*, vol. 58, no. 5, pp. 6557–6575, 2022.

[6] W. Khan, W. Somers, S. Walker, K. de Bont, J. Van der Velden, and W. Zeiler, "Comparison of electric vehicle load forecasting across different spatial levels with incorporated uncertainty estimation," *Energy*, vol. 283, p. 129213, 2023.

[7] Y. Wu, P. Cong, and Y. Wang, "Charging load forecasting of electric vehicles based on vmd–ssa–svr," *IEEE Transactions on Transportation Electrification*, vol. 10, no. 2, pp. 3349–3362, 2024.

[8] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," in *Proceedings of the 2020 Advances in Neural Information Processing Systems*, vol. 33. Virtual: Curran Associates, Inc., Dec 2020, pp. 17 804–17 815.

[9] H. J. Kim and M. K. Kim, "Spatial-temporal graph convolutional-based recurrent network for electric vehicle charging stations demand forecasting in energy market," *IEEE Transactions on Smart Grid*, vol. 15, no. 4, pp. 3979–3993, 2024.

[10] K. Yi, Q. Zhang, W. Fan, H. He, L. Hu, P. Wang, N. An, L. Cao, and Z. Niu, "Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective," *Proceedings of the 2023 Advances in Neural Information Processing Systems*, vol. 36, Dec. 2023.

[11] J. M. Mooij, S. Magliacane, and T. Claassen, "Joint causal inference from multiple contexts," *Journal of machine learning research*, vol. 21, no. 99, pp. 1–108, 2020.

[12] J. Pearl, *Direct and Indirect Effects*, 1st ed. New York, NY, USA: ACM, 2022, p. 373–392.

[13] K. Imai and T. Yamamoto, "Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments," *Political Analysis*, vol. 21, no. 2, pp. 141–171, 2013.

[14] M. Rashid, T. Elfouly, and N. Chen, "A comprehensive survey of electric vehicle charging demand forecasting techniques," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 1348–1373, 2024.

[15] K. Zheng, H. Xu, Z. Long, Y. Wang, and Q. Chen, "Coherent hierarchical probabilistic forecasting of electric vehicle charging demand," *IEEE Transactions on Industry Applications*, 2023.

[16] E. Yaghoubi, E. Yaghoubi, A. Khamees, D. Razmi, and T. Lu, "A systematic review and meta-analysis of machine learning, deep learning, and ensemble learning approaches in predicting ev charging behavior," *Engineering Applications of Artificial Intelligence*, vol. 135, p. 108789, 2024.

[17] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Virtual Event: AAAI Press, Feb 2021, pp. 11 106–11 115.

[18] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *Proceedings of the 11th International Conference on Learning Representations*. Kigali, Rwanda: OpenReview.net, May 2023.

[19] Y. Nie, N. H. Nguyen, mPhanwadee Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *Proceedings of the 11th International Conference on Learning Representations*. Kigali, Rwanda: OpenReview.net, May 2023.

[20] X. Huang, D. Wu, and B. Boulet, "Metaprobformer for charging load probabilistic forecasting of electric vehicle charging stations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 10 445–10 455, 2023.

[21] Y. Li, S. He, Y. Li, L. Ge, S. Lou, and Z. Zeng, "Probabilistic charging power forecast of evcs: Reinforcement learning assisted deep learning approach," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 344–357, 2023.

[22] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement learning-based load forecasting of electric vehicle charging station using q-learning technique," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4229–4237, 2021.

[23] Z. Wang, X. Liu, Y. Huang, P. Zhang, and Y. Fu, "A multivariate time series graph neural network for district heat load forecasting," *Energy*, vol. 278, p. 127911, 2023.

[24] L. Chen, D. Chen, Z. Shang, B. Wu, C. Zheng, B. Wen, and W. Zhang, "Multi-scale adaptive graph neural network for multivariate time series forecasting," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 10, pp. 10 748–10 761, 2023.

[25] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao, China: ijcai.org, Aug 2019, pp. 1907–1913.

[26] J. Shi, W. Zhang, Y. Bao, D. W. Gao, and Z. Wang, "Load forecasting of electric vehicle charging stations: Attention based spatiotemporal multi–graph convolutional networks," *IEEE Transactions on Smart Grid*, vol. 15, no. 3, pp. 3016–3027, 2024.

[27] H. Qu, H. Kuang, Q. Wang, J. Li, and L. You, "A physics-informed and attention-based graph learning approach for regional electric vehicle charging demand prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 10, pp. 14 284–14 297, 2024.

[28] X. Li and Q. Han, "An ev charging station load prediction method considering distribution network upgrade," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 4360–4371, 2024.

[29] A. N. Prasanna, P. Grecov, A. D. Weng, and C. Bergmeir, "Causal effect estimation with global probabilistic forecasting: A case study of the impact of covid-19 lockdowns on energy demand," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 3417–3430, 2024.

[30] H. Lan, H. C. Hou, Z. Gou, and M. S. Wong, "Spatiotemporal analysis and forecasting of pv systems, battery storage, and ev charging diffusion in california: A graph network approach," *Renewable Energy*, vol. 230, p. 120868, 2024.

[31] W. Kong, Z. Guo, and Y. Liu, "Spatio-temporal pivotal graph neural networks for traffic flow forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, 2024, pp. 8627–8635.

[32] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[33] W. Liu, Y. He, J. Guan, and S. Zhou, "Multivariate time series forecasting with causal-temporal attention network," in *Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 5735–5739.

[34] L. Massidda and M. Marrocu, "Total and thermal load forecasting in residential communities through probabilistic methods and causal machine learning," *Applied Energy*, vol. 351, p. 121783, 2023.

[35] J. Liang, Y. Qiu, P. Liu, P. He, and D. L. Mauzerall, "Effects of expanding electric vehicle charging stations in california on the housing market," *Nature Sustainability*, vol. 6, no. 5, pp. 549–558, 2023.

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

[37] E.-Y. Yu, Y.-P. Wang, Y. Fu, D.-B. Chen, and M. Xie, "Identifying critical nodes in complex networks via graph convolutional networks," *Knowledge-Based Systems*, vol. 198, p. 105893, 2020.

[38] B. Huang, K. Zhang, Y. Lin, B. Schölkopf, and C. Glymour, "Generalized score functions for causal discovery," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. London, UK: ACM, August 2018, pp. 1551–1560.

[39] K. Baek, E. Lee, and J. Kim, "A dataset for multi-faceted analysis of electric vehicle charging transactions," *Scientific Data*, vol. 11, no. 1, p. 262, 2024.

[40] T. Wang, C. Ren, D. Z. Yang, and C. Yip, "Domain-adaptive clustered federated transfer learning for ev charging demand forecasting," *IEEE Transactions on Power Systems*, vol. Early Access, 2024.

[41] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI conference on Artificial Intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.

[42] K. Cho, B. van Merrienboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: ACL, Oct 2014, pp. 1724–1734.

[43] C. Challu, K. G. Olivares, B. N. Oreshkin, F. G. Ramírez, M. M. Canseco, and A. Dubrawski, "NHITS: neural hierarchical interpolation for time series forecasting," in *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Washington, DC, USA: AAAI Press, Feb 2023, pp. 6989–6997.

[44] A. Shojaie and E. B. Fox, "Granger causality: A review and recent advances," *Annual Review of Statistics and Its Application*, vol. 9, no. 1, pp. 289–319, 2022.

Yaohui Huang (Graduate Student Member, IEEE) is currently pursuing the Ph.D. degree in Control Science and Engineering with the School of Automation, Central South University, Changsha, China. His research interests include energy forecasting, deep learning, time series analysis, and graph neural networks.

Yu Hue received the M.Sc. degree from the College of Computer Science and Engineering, Macau University of Science and Technology, Macau, China, in 2015. He is currently with the Chengyi College, Jimei University, Xiamen, China. His research interests include time series forecasting, energy load forecasting, and artificial intelligence.

Senzhen Wu is currently pursuing the B.S. degree in Computer Science at Jimei University, Xiamen, China. His research focuses on energy forecasting, deep learning, and time series analysis, with applications in smart grid resilience and renewable energy integration.

Zhijin Wang received the Ph.D. degree from the Department of Computer Science and Technology, East China Normal University, Shanghai, China, in 2016. He is currently an Associate Professor with Jimei University, Xiamen, China.

He is the founder of the forecasting and time series (FAST) community and the leader of the pyFAST library for time series analysis. His research interests include recommendation systems, time series forecasting, and artificial intelligence in health and healthcare.

Xiufeng Liu (Member, IEEE) received the Ph.D. degree in computer science from Aalborg University, Denmark, in 2012. He was a Post-Doctoral Researcher with the University of Waterloo and a Research Scientist with IBM, Canada, from 2013 to 2014. He is currently a Senior Researcher with the Department of Technology, Management and Economics, Technical University of Denmark. His research interests include smart meter data analysis, data warehousing, energy informatics, and big data.

Chendan Li (Member, IEEE) received a B.S. degree in electrical engineering from Nanjing Agricultural University, Nanjing, China, in 2009, an M.S. degree in electrical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, in 2012, and the Ph.D. degree from Aalborg University, Denmark, in 2016. She is currently a Research Fellow with the Electrical Eng. Department at the University of Genova, Italy. Her main research interests include control and operation for microgrids, stability of power electronic based power systems, and data-driven applications to power systems.