KAN-Enhanced Graph Learning for Active Voltage Control in Dynamic Power Systems

Liqian Sun
School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, 518055
Shenzhen, China
25s151073@stu.hit.edu.cn

Hang Xiao
School of Software, Northwestern Polytechnical University, 710072
Xi'an, China
shawh@mail.nwpu.edu.cn

Shuhan Qi*
School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, 518055
Shenzhen, China
shuhanqi@cs.hitsz.edu.cn

Huale Li*
School of Information Science and Engineering, Lanzhou University, 730000
Lanzhou, China
lihuale@lzu.edu.cn

Jiajia Zhang
School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, 518055
Shenzhen, China
zhangjiajia@hit.edu.cn

Xuan Wang
School of Computer Science and Technology, Harbin Institute of Technology Shenzhen, 518055
Shenzhen, China
wangxuan@cs.hitsz.edu.cn

ABSTRACT

The large-scale integration of distributed energy resources has significantly increased the complexity of industrial power dispatch. While existing multi-agent reinforcement learning (MARL) methods leverage graph neural networks for topology-aware voltage control, their ability to capture evolving grid topologies remains limited. Therefore, we propose GKAN-MA, a dual-enhanced MARL framework specifically designed to maintain voltage stability in power systems with highly dynamic topologies and strongly nonlinear voltage-power dynamics. It achieves robust voltage regulation despite frequent grid reconfigurations, while precisely modeling complex relationships between reactive power and nodal voltages. Through persistent topology awareness and accurate nonlinear function approximation, GKAN-MA ensures consistent performance during network changes. Experimental results on IEEE 33-bus and 141-bus systems demonstrate superior controllability and operational efficiency, validating its adaptability to dynamic power system conditions.

KEYWORDS

Active Voltage Control; Multi-agent Reinforcement Learning; Graph Attention Networks; Kolmogorov-Arnold network

ACM Reference Format:
Liqian Sun, Hang Xiao, Shuhan Qi*, Huale Li*, Jiajia Zhang, and Xuan Wang. 2026. KAN-Enhanced Graph Learning for Active Voltage Control in Dynamic Power Systems. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/WDUC5953

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/WDUC5953

1 INTRODUCTION

In the context of global sustainable development, distributed energy resources (DERs) utilizing clean energy, such as rooftop photovoltaics (PV), are increasingly being integrated into power systems, demonstrating significant potential [1]. However, their inherent intermittency and volatility pose substantial challenges to power system operation and dispatch [2]. Effectively integrating rooftop PV into industrial power systems has become a critical issue demanding urgent resolution.

Active voltage control (AVC), as a key technique for ensuring grid stability, plays a crucial role in mitigating the challenges posed by rooftop PV integration [3–6]. Specifically, AVC monitors nodal voltages in real time and dynamically adjusts reactive power injection or absorption through devices such as PV inverters to maintain voltage within acceptable limits and optimize overall system performance. Traditional AVC methods rely on model-based approaches, such as optimal power flow (OPF) [7] and convex relaxation [8], which require accurate real-time mathematical modeling of the grid. However, these methods suffer from high dependence on precise system models and slow response times, rendering them unsuitable for environments characterized by parameter uncertainty and rapid fluctuations in large-scale distributed loads [9, 10].

In recent years, multi-agent reinforcement learning (MARL) has partially alleviated these limitations [11–14]. MARL systems comprise multiple autonomous or semi-autonomous agents, which through decentralized decision-making and information exchange can efficiently explore near-optimal policies in complex, dynamic power dispatch scenarios. Nevertheless, conventional MARL methods typically treat agent states as independent vectors, aggregating features without considering spatial correlations or physical couplings among nodes [15]. This results in collaborative decisions that are often disconnected from the underlying grid topology. Even with frequent agent interaction, communication content frequently lacks explicit modeling of critical factors such as topological

[0] *Corresponding authors

adjacency and electrical distance, thereby undermining the interpretability and generalization capability of control strategies in complex distribution networks.

As a structure-aware learning paradigm, graph neural networks (GNNs) [16] offer a promising data-driven approach by integrating grid topology constraints and electrical characteristics into optimization frameworks [17, 18]. Ma et al. proposed GNN-QL, embedding GNNs into a Deep Q-Network (DQN) [19] architecture to achieve more stable and topology-aware real-time Volt-VAR control [20]. Mu et al. introduced MAGRL, incorporating Graph Convolutional Networks (GCNs) [21] into a multi-agent actor-critic framework to stabilize voltage under high DER penetration, enhanced by an exponential voltage barrier function for safety assurance [22]. Luo et al. developed GAMARL, integrating domain knowledge and Graph Attention Networks (GATs) [23]to enable more accurate and robust voltage regulation by dynamically optimizing inter-agent coupling relationships [24]. However, these existing methods remain largely grounded in static topology assumptions, limiting their adaptability to dynamic topology changes induced by DER disconnections or line faults, thus constraining their generalizability in real-world grids.

To address these challenges, we propose GKAN-MA, a novel MARL framework based on dynamic graph attention and precise nonlinear approximation. Specifically, GKAN-MA first employs the Graph Attention Network variant (GATv2) [25] to dynamically capture evolving grid topology and node dependencies, enabling real-time modeling of voltage fluctuation propagation paths. Subsequently, it leverages the Kolmogorov-Arnold Network (KAN) [26], using learnable B-spline functions to precisely approximate the nonlinear dynamics governing voltage regulation, thereby extracting complex patterns from real-time voltage and power flow data. Through this dual-enhanced design, GKAN-MA effectively mitigates the limitations imposed by static topology assumptions, significantly enhancing its generalization and robustness in dynamic scenarios involving resource integration or line failures.

Our contributions are summarized as follows:

- We introduce GKAN-MA, a MARL framework explicitly designed to address voltage control challenges in dynamic power systems where grid topology continuously evolves due to the integration of DERs and operational reconfigurations, which enables effective voltage stabilization in environments with high renewable energy penetration.
- We propose a dual-enhanced learning paradigm that maintains persistent environmental awareness through continuous adaptation to topology changes while accurately capturing the complex nonlinear dynamics between voltage profiles and power flow, which ensures robust control policy learning and precise voltage regulation in non-steady grid environments with rapidly changing operational conditions.
- Extensive experiments on static and dynamic variants of IEEE 33-bus and 141-bus systems validate that the proposed approach not only achieves superior adaptability and operational efficiency in both stable and dynamically changing grid conditions, but also maintains high voltage controllable ratio with competitive power loss, validating its practical value for real-world power system operations facing unstable topological structures.

2 DEC-POMDP FORMULATION FOR AVC

In this section, we formally define the AVC problem and recast it within a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) framework, which enables its solution via MARL. This formulation provides the theoretical foundation for the proposed GKAN-MA method.

The AVC problem addresses the challenge of mitigating voltage fluctuations in distribution networks induced by the intermittent active power injection from PV units. The primary objective is to stabilize nodal voltages around a reference value through the flexible regulation of PV inverters' reactive power output, while simultaneously minimizing network power losses. The underlying physical system is modeled as an undirected tree-topology graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V} = \{0, 1, ..., N-1\}$ represents N buses and the edge set $\mathcal{E} = \{0, 1, ..., L-1\}$ represents L branches.

To enable a MARL-based solution, we formulate the AVC problem as a Dec-POMDP, defined by the septuple $\langle I, S, A, O, T, R, \gamma \rangle$:

- **I:** The agent set $I = \{1, ..., n\}$, where each agent i controls a PV inverter and is responsible for regulating the reactive power output Q^i_{pv} in its designated area.
- **S:** The global state space $S = L \times P \times \Omega \times V$. $L = \{P^k_L, Q^k_L\}^{N-1}_{k=0}$ represents the load power profile, $P = \{P^i_{pv}\}^{N-1}_{i=0}$ denotes the active power output from all PV units, $\Omega = \{Q^i_{pv}\}^{N-1}_{i=0}$ captures the current reactive power setpoints of all PV inverters, and $V = \{v_k\}^{N-1}_{k=0}$ is the vector of complex nodal voltages, with $v_k = |V_k|e^{j\theta_i}$, where $|V_k|$ and θ_k are the voltage magnitude and phase angle at node k respectively. For any agent $i \in \mathcal{I}$, the net active and reactive power injections are given by $P^i = P^i_{pv} - P^i_L$ and $Q^i = Q^i_{pv} - Q^i_L$, which are fundamental quantities governing the grid's dynamic behavior.
- **A:** The joint action space $A = \prod^n_{i=1} A_i$, where the individual action space for agent i is $A_i = [-Q^{i,\max}_{pv}, Q^{i,\max}_{pv}]$, with $Q^{i,\max}_{pv} = \sqrt{(S_{i,\max})^2 - (P^i_{pv})^2}$. The action $a_i \in A_i$ directly sets the new reactive power output Q^i_{pv} for the i-th inverter. This constraint ensures that the action respects the physical capacity limit of the inverter, defined by its maximum apparent power rating $S_{i,\max}$.
- **O:** The observation space $O = \prod_{i \in I} O_i$. Each agent i observes a local state $O_i = \{v_i, \theta_i, P^i_{pv}, P^i_L, Q^i_{pv}, Q^i_L\}$, which includes the voltage magnitude and phase at its own node, the local PV and load power profiles, and its current reactive power setting.
- **T:** The state transition function $T : S \times A \to S'$ governs the dynamic evolution of the power grid after actions are executed. It is implicitly defined by the AC power flow equations.
- **R:** The reward function R is designed to incentivize both voltage stability and operational efficiency, which can be formally expressed as:

$$R(s, a) = -\lambda_1 \|v - v_{ref}\|^2_2 - \lambda_2 \sum_{i=1}^{N} |Q^i_{pv}| + P_{\text{safety}} \tag{1}$$

where v is the voltage magnitude vector derived from the state $s' = T(s, a)$. The coefficients λ_1 penalizes deviations from the reference and economic efficiency. λ_2 penalizes excessive reactive power consumption, which contributes to network power losses P_{loss}. The term P_{safety} imposes a large penalty if any voltage violates the safety constraint, ensuring the policy learns to maintain all bus voltages within the safe operating range.

- γ: The discount factor $\gamma \in [0, 1)$ balances the importance of immediate versus future rewards.

Notably, the state transition function T and the reward function R in the above Dec-POMDP implicitly encode the complex, nonlinear relationships between grid topology and voltage dynamics. This constitutes the core challenge addressed by the GKAN-MA framework in Section 3. Specifically, when the grid topology undergoes dynamic changes, the transition function T experiences abrupt changes, which traditional static graph neural networks fail to capture effectively. Consequently, Section 3 introduces a GATv2-based dynamic topology perception module to adapt to these real-time changes, cooperating with the KAN's powerful nonlinear approximation capability to precisely estimate Q-values, thereby ensuring the algorithm's robustness in dynamic environments.

3 METHOD

In this section, we introduce the architecture of GKAN-MA, an innovative MARL framework specifically designed for AVC in distribution networks. Built upon the established Multi-Agent Deep Deterministic Policy Gradient (MADDPG) paradigm [27], GKAN-MA leverages its core principle of Centralized Training with Decentralized Execution (CTDE) to stabilize learning in non-stationary multi-agent environments. As illustrated in Figure 1, GKAN-MA comprises two core modules: distributed execution and global value estimation. These correspond to the actor and critic networks within the Actor-Critic framework. The multi-agent distributed execution module coordinates the distributed actions of regional controllers to achieve global voltage optimization while ensuring system safety, by acquiring information from specific locations. The global value estimation module comprises two steps: dynamic topology perception and nonlinear value approximation, detailed in Sections 3.1 and 3.2, respectively. Through coordinated interaction among these components, GKAN-MA effectively addresses the dual challenges of dynamic topology changes and nonlinear voltage-power relationships, providing a robust and efficient solution for voltage control in complex distribution networks.

3.1 GATv2-Based Dynamic Topology Perception

Traditional voltage control algorithms typically assume a static power grid topology. However, modern power systems with high renewable penetration exhibit significant topological dynamics due to factors such as line faults, resource integration, and operational reconfiguration. In this dynamic environment, the spatial dependencies among nodes continuously evolve, making it challenging for fixed-topology methods to accurately capture the real-time propagation paths of voltage fluctuations.

To address the limitations of static topology modeling, we employ GATv2 to dynamically capture changes in grid topology. GATv2 improves upon the classical GAT by introducing a learnable attention mechanism that enables the model to adaptively compute edge weights when aggregating information from neighboring nodes. This allows for more flexible modeling of complex and nonlinear inter-node dependencies. The core update rule is given in Equation (2) and Equation (3)[25], where the new representation h_i' of each node i is computed as a weighted sum of its neighbors' features, with the weights α_{ij} dynamically generated by a learnable attention vector $\mathbf{a}$ and a shared weight matrix W :

$$h_i' = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} W h_j \tag{2}$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [Wh_i \| Wh_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^T [Wh_i \| Wh_k]))} \tag{3}$$

The concatenated multi-agent observations and actions FC are first decoupled into individual node feature vectors x, ensuring independent participation of each agent within the topological interactions. As illustrated in Figure 1, these node features, along with the edge connection matrix E representing the current graph structure at each time step, are fed into the GATv2 network. It employs a multi-head attention mechanism to adaptively learn the relationship weights between nodes, thereby dynamically modeling the evolving topological dependencies within the power grid. As demonstrated in Equation (4), the output of the GATv2 network generates neighborhood-aware embeddings, which are further processed through a ReLU activation function to preserve non-linear interaction information:

$$x_{\text{relu}} = \text{ReLU}(\text{GATv2}(x, E) \in \mathbb{R}^{B \times N \times (h_{\text{out}} \times H)}) \tag{4}$$

where B denotes the batch size, N is the total number of agents, h_{out} represents the output dimension per attention head, and H is the number of attention heads. To abstract local regional information into a global grid state representation, the agent-wise features are first reshaped from a flat sequence of length $B \times N$ back into a structured tensor of shape $[B, N, H \times h_{\text{out}}]$, aligning each agent's feature vector with its corresponding position in the batch. Subsequently, an average pooling operation is applied across the agent dimension, yielding a unified global topological feature vector $x_{\text{global}} \in \mathbb{R}^{H \times h_{\text{out}}}$ for each sample in the batch. This aggregation step effectively summarizes the distributed agent-level information into a compact, fixed-size global state representation, which serves as input to the subsequent KAN module for policy decision-making.

3.2 KAN-based Precise Q-Value Approximation

The strong nonlinearity inherent in power system voltage dynamics poses a significant challenge for accurate Q-value function approximation. Conventional neural networks employ fixed activation functions, whose static nonlinear structures are often insufficient to adequately model the complex mapping relationships between system states and voltage control outcomes. This limitation constrains the accuracy and generalization capability of value estimation, particularly under diverse operating conditions and dynamic topologies.

To overcome the limited expressivity of fixed activation functions, we leverage KAN to explicitly learn complex state-control mappings. As formalized in Equation (5), KAN decomposes the high-dimensional function $f(x)$ into compositions of lower-dimensional

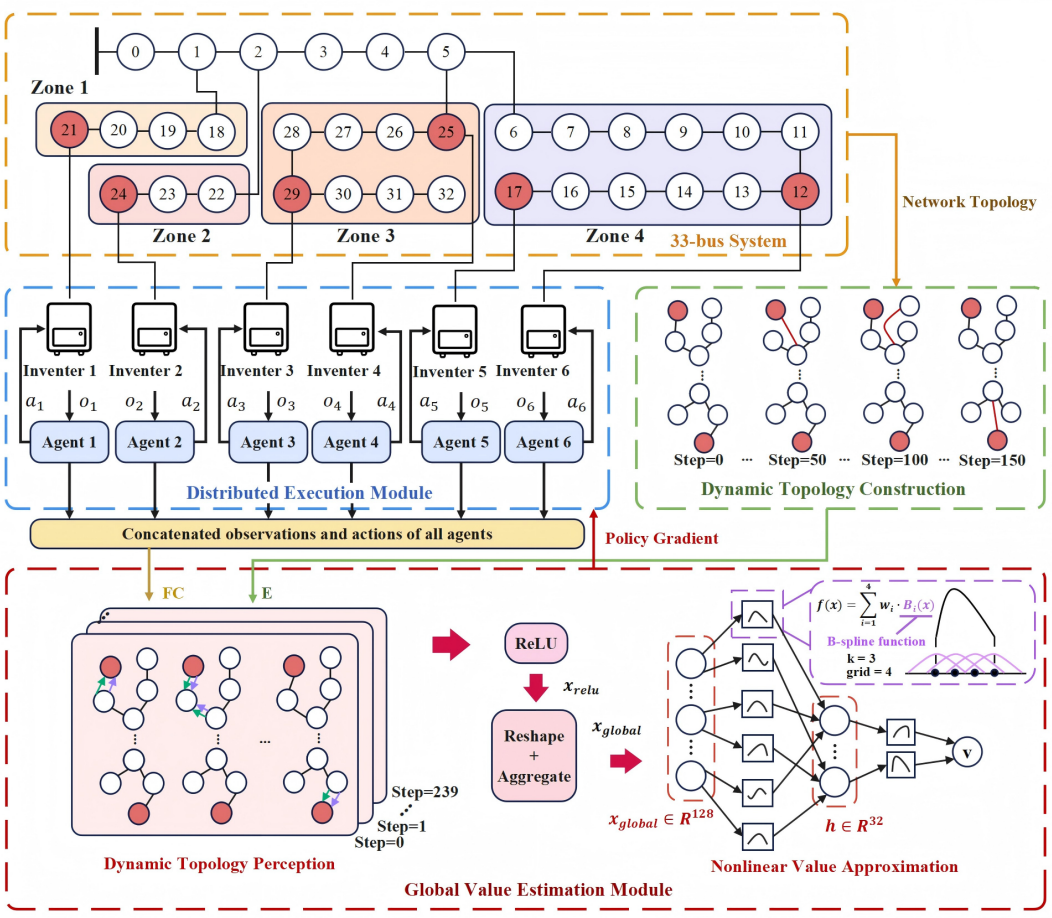


Figure 1: The framework of GKAN-MA on the 33-bus dynamic system. On each step, distributed agents (blue) interact with the environment (yellow), feeding their state-action data and the current topology (green) into the global value estimation module (red). It computes a policy gradient that closes the loop by updating the agents' individual policies.

ones, with each mapping parameterized by learnable basis functions for more flexible and precise approximation [26]:

$$f(x) = \sum_{i=1}^{n} w_i \cdot B_i(x) \tag{5}$$

where w_i represents the learnable weights and $B_i(x)$ denotes the learnable basis functions. This architecture allows KAN to simultaneously optimize both the weights and the shapes of the basis functions during training.

In our method, KAN receives the global topological features x_{global} and processes them through a single hidden layer. As illustrated in Figure 1, learnable B-spline functions, parameterized by grid points and basis functions, serve as activation units that efficiently model complex nonlinear relationships. To prioritize numerical stability, symbolic expression generation is disabled during training. The final Q-values are produced through successive tensor-product spline transformations, which can be formalized as:

$$v = \text{KAN}(x_{\text{global}}) \in \mathbb{R}^{B\times O} \tag{6}$$

furthermore, an intermediate hidden state is derived from the activations within KAN to ensure compatibility with the policy network inputs required by the MADDPG framework.

3.3 Training Stability Optimization

GKAN-MA adopts the paradigm of centralized training and decentralized execution. To ensure the training stability and convergence of GKAN-MA in complex dynamic environments, GKAN-MA employs a replay buffer to store state transition samples (S, A, R, S'), breaking temporal correlations and improving data efficiency. It incorporates delayed-update target networks $\theta^{Q'}$ and $\theta^{\mu'}$ to compute target Q-values $y = R + \gamma Q^{\mu'}(S', A')$, where $A' = \mu'(S')$, mitigating Q-value oscillations caused by frequent parameter updates. A soft update strategy $\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}$ achieves smooth parameter transition, balancing exploration and exploitation. The policy network parameters θ_i^{μ} are optimized via policy gradient, with update direction given by $\nabla_{\theta_i^{\mu}} J = \mathbb{E}_{S,A\sim\mathcal{D}}\left[\nabla_{\theta_i^{\mu}} \mu_i(O_i) \cdot \nabla_{a_i} Q(S, A)\right]$.The critic network takes the concatenated observations and actions

of all agents as node features $X = [O_i; a_i]_i$, and processes them through a GATv2 layer informed by the real-time topology matrix E to produce neighborhood-aware embeddings. These embeddings are then passed through a ReLU activation and average pooling to obtain a global state representation x_{global}, which is finally mapped to Q-values by KAN. The entire critic is trained by minimizing the temporal difference loss: $\mathcal{L}(\theta^Q) = \mathbb{E}_{(S,A,R,S')\sim\mathcal{D}}\left[(Q(S,A)-y)^2\right]$. The complete algorithm is presented in Algorithm 1.

Algorithm 1: GKAN-MA

Input: Power grid topology E, reward function R, discount γ, batch size B, soft update rate τ
Output: Optimal policy μ^*
Initialize:
Policy networks $\{\mu_i\}_{i=1}^N$ with params θ_i^μ
Critic network $\{Q\}$ with params θ^Q
Target networks μ', Q' with $\theta^{\mu'} \leftarrow \theta^\mu$, $\theta^{Q'} \leftarrow \theta^Q$
Experience replay buffer $\mathcal{D}$
for *each training step* **do**
  Observe state $S = \{O_i\}_{i=1}^N$
  foreach *agent* $i \in \{1, ..., N\}$ **do**
    $a_i \leftarrow \mu_i(O_i|\theta_i^\mu)$
  end
  Execute actions $A = \{a_i\}$, observe R, S'
  Store transition (S, A, R, S') in $\mathcal{D}$
  if $|\mathcal{D}| \geq B$ **then**
    Sample mini-batch $\mathcal{B} \sim \mathcal{D}$
    Update critic:
      Compute target Q-values: $y = R + \gamma Q'(S', \mu'(S'))$
      Extract node features: $X = [O_i; a_i]_{i=1}^N$
      Process via GATv2: $H = \text{GATv2}(X, E)$
      Extract global feature: $x_{\text{global}} = \frac{1}{N}\sum_i \text{ReLU}(H_i)$
      Predict Q-values: $\hat{Q} = \text{KAN}(x_{\text{global}})$
      Minimize loss: $\mathcal{L}(\theta^Q) = \frac{1}{|\mathcal{B}|}\sum(\hat{Q}-y)^2$
    Update actor:
      $\nabla_{\theta_i^\mu} J = \frac{1}{|\mathcal{B}|}\sum \nabla_{\theta_i^\mu}\mu_i(O_i)\cdot\nabla_{a_i}Q(S,A)$
      Update θ_i^μ via policy gradient
    Update target networks:
      $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
      $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
  end
end

4 EXPERIMENTS

4.1 Experiment Settings

4.1.1 Environments. To effectively validate the performance of GKAN-MA in dynamic environments, the experimental setup includes both static and dynamic power distribution networks. The static environments are based on modified IEEE 33-bus and 141-bus network configurations [28], which are widely adopted as the primary test cases for contemporary AVC studies [24, 29, 30]. The 33-bus system incorporates 6 PV units distributed across 4 regions, while the 141-bus system contains 22 PV units across 9 regions. The dynamic environments are derived from the static counterparts, with topology changes implemented by removing a non-critical line and adding a new line at specific time steps, thereby simulating real-world grid instability caused by external environmental and human-induced disturbances. The topology reconfiguration operations are randomly distributed across different regions to prevent the model from adapting only to changes in specific areas, ensuring that the experimental results genuinely reflect the algorithm's ability to handle topology changes during actual grid operations. The 33-bus system undergoes topology changes every 50 steps, while the 141-bus system changes every 30 steps.

Data in the distribution network environment are derived from real-world scenarios [22, 28]. The node load profiles are based on three-year electricity consumption records from 232 residential customers in Portugal. The PV generation data were compiled from actual operational records of PV plants across ten regions. Both datasets were converted to a 3-minute resolution, which matches the 3-minute time step used in our experiments.

4.1.2 Baselines. To evaluate the effectiveness of the proposed GKAN-MA method, the experiments compare two categories of methods: MARL algorithms specifically designed for the AVC problem and classical reinforcement learning algorithms.

- **MAGRL [22]:** A pioneering graph-based MARL method that integrates GCNs into the MADDPG framework to explicitly model the topological structure of distribution networks. By aggregating features from neighboring agents, MAGRL effectively captures spatial coupling among inverters, significantly improving voltage controllability and reducing power loss. It is widely recognized as a strong baseline for topology-aware voltage control in active distribution networks.
- **GAMARL [24]:** A recently proposed domain-knowledge-enhanced MARL approach that leverages GATs to adaptively optimize edge weights to generate refined spatial features, enabling precise and robust Volt/Var control under high PV penetration. It achieves leading performance in suppressing voltage violations, which is one of the latest advances in MARL for power systems.
- **MADDPG [27]:** A classical deep MARL method in which each agent acts via an independent actor network, while a centralized critic network evaluates the joint state-action information during training. It has served as a foundational baseline in many MARL applications, including early explorations in power system control [28].
- **MAPPO [31]:** A canonical multi-agent extension of Proximal Policy Optimization (PPO). Building upon PPO's success in single-agent domains, MAPPO maintains independent policy and value networks for each agent while leveraging a shared value function for centralized training. It has become a standard benchmark in complex cooperative tasks.
- **IPPO [32]:** An improved PPO variant developed for robotic manipulation. It enhances exploration-exploitation balance via a Poisson-based action ensemble. While not designed for power systems, its robustness and sample efficiency make it a strong baseline in recent MARL-based grid control [33, 34].

4.1.3 Evaluation Metrics. Consistent with mainstream research methodologies in power dispatch, we adopt classical evaluation

Table 1: Parameter settings of GKAN-MA

Parameters	Values	
	33-bus	141-bus
Number of Attention Heads in GATv2	2	4
Hidden Dimension	64	32
Dropout Rate in GATv2	0.2	0.4
KAN Grid Points per Dimension	4	6
KAN B-spline Degree	3	3

metrics in experimental design to directly reflect the algorithm's capability in enhancing voltage stability and economic efficiency of the system [3, 22, 28]. The specific definitions of these metrics are as follows:

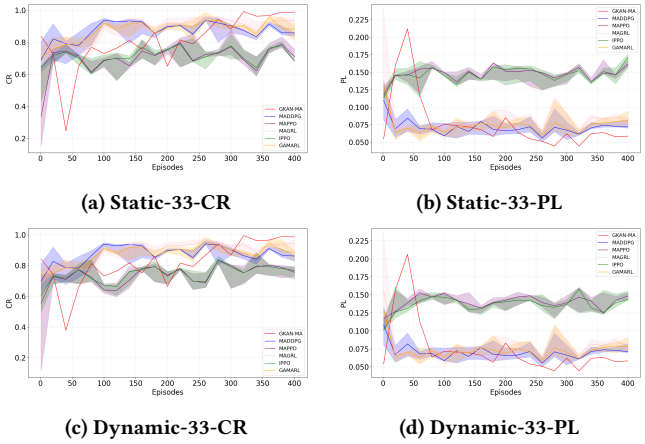
- **Controllable Ratio (CR):** The rate of time steps during which all bus voltages remain within the predefined control range to the total number of time steps per training episode. A higher CR indicates stronger voltage regulation capability.
- **Power Loss (PL):** The average sum of instantaneous power losses across all buses per time step per training episode. A lower PL signifies that the algorithm achieves voltage stability while minimizing unnecessary energy dissipation, thus improving overall operational efficiency.
- **Node Voltage (NV):** The magnitude of voltage in each bus per test episode. Voltages closer to the reference value of 1.0 p.u. and with more uniform distribution indicate a superior suppression of voltage fluctuations.

4.1.4 Other Settings. Algorithms are trained with a policy learning rate and value learning rate, both set to 1.0e-4. Using an experience replay buffer of size 5000 and sample batches of 32 transitions for each update. The voltage control task enforces a safety boundary of [0.95, 1.05] p.u., ensuring all node voltages remain within this operational range during training and evaluation. The specific parameter settings for GKAN-MA are presented in Table 1. All the algorithms are implemented in Python, leveraging the PyTorch deep learning framework in version 1.13.1. The experiment was conducted on a workstation configured with an NVIDIA RTX 3090 GPU.

4.2 Comparison Experiment Results

4.2.1 Performance Analysis on Benchmark Scenarios. For experiments on the benchmark scenarios, all methods were evaluated using five random seeds to ensure statistical robustness and mitigate performance bias introduced by initialization randomness. Furthermore, every 20 training episodes, 10 test episodes were conducted for evaluation, enabling dynamic monitoring of the models' generalization capability and convergence trends during training. The final performance report presents the median value across the five seeds as the representative mean for each metric, supplemented by the 25th and 75th percentiles to form a confidence interval. Figure 2 presents the results of the methods in the 33-bus static and dynamic environments.

In the static environment, GKAN-MA rapidly converges after initial fluctuations and achieves a final CR of 0.9869, which is 5.30% higher than the second-best method, MAGRL, demonstrating exceptional voltage safety compliance. Its PL reaches 0.0588, reducing power loss by 18.22% over MADDPG, confirming high operational efficiency. In the dynamic environment, GKAN-MA achieves optimal performance in both CR and PL metrics. Although GKAN-MA still exhibits noticeable fluctuations in the initial phase, its fluctuation peaks are optimized compared to results in static scenarios. This indicates improved performance under dynamic conditions and stronger robustness against dynamic disturbances.



(a) Static-33-CR (b) Static-33-PL

(c) Dynamic-33-CR (d) Dynamic-33-PL

Figure 2: Training results of algorithms on the 33-bus static and dynamic systems.

Table 2: Algorithms' median values of CR and PL on the 33-bus static and dynamic systems

Methods	33-bus-static		33-bus-dynamic	
	CR	PL	CR	PL
GKAN-MA	**0.9869**$^{\uparrow 5.30\%}$	**0.0588**$^{\downarrow 18.22\%}$	**0.9876**$^{\uparrow 11.92\%}$	**0.0584**$^{\downarrow 17.98\%}$
GAMARL	0.8816	0.0816	0.8824	0.0786
MAGRL	0.9372	0.0836	0.8816	0.0794
MADDPG	0.8582	0.0719	0.8615	0.0712
MAPPO	0.7105	0.1615	0.7561	0.1487
IPPO	0.6820	0.1714	0.7682	0.1435

Across both scenarios, GKAN-MA not only outperforms all baselines but also exhibits narrower confidence intervals, reflecting strong repeatability. This non-monotonic learning dynamic is a known trade-off when introducing richer representational capacity in deep architectures, particularly when two highly expressive components are co-trained. Crucially, once this adaptation phase stabilizes, the policy rapidly converges and surpasses all baselines, confirming that the initial dip reflects learning complexity.In the tabulated experimental results, we annotate the relative improvement percentages with upward (↑) and downward (↓) arrows, indicating the performance gain or reduction relative to the best value among other algorithms for each metric. As shown in Table 2, GKAN-MA not only avoided performance degradation despite increased environmental complexity but also achieved further improvements while maintaining high CR, validating its practicality in small-scale complex power systems.

Table 3: Algorithms' median values of CR and PL on the 141-bus static and dynamic systems

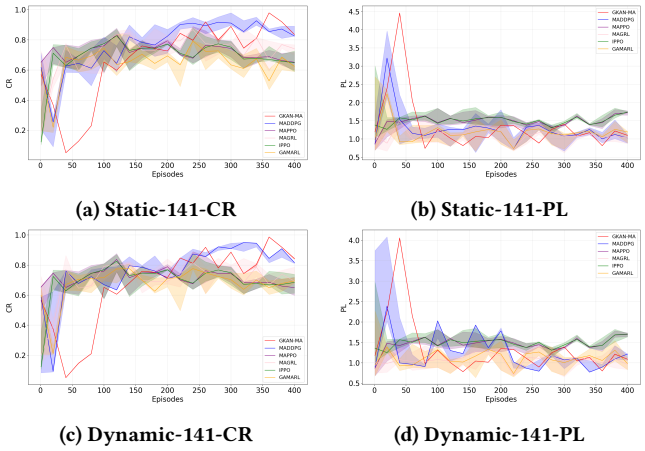
Methods	141-bus-static		141-bus-dynamic	
	CR	PL	CR	PL
GKAN-MA	**0.8356**$^{\uparrow 1.22\%}$	1.0996	**0.8423**$^{\uparrow 3.18\%}$	1.0879
GAMARL	0.6238	1.1929	0.6950	0.9817
MAGRL	0.7410	1.0364	0.6937	**0.9619**
MADDPG	0.8255	**1.0307**	0.8163	1.2159
MAPPO	0.6515	1.7191	0.6498	1.6970
IPPO	0.6477	1.7383	0.6845	1.7041

As the scale of the scenarios expands, all the algorithms show a decline in performance across the two metrics. As illustrated in Figure 3, GKAN-MA maintains the highest CR, outperforming the second-best by 1.22% in static scenarios and 3.19% in dynamic scenarios, highlighting its scalability. This enhanced voltage control performance results in a slight increase in PL. This is attributable to the adoption of more aggressive control measures under frequent dynamic disturbances caused by fluctuations in PV generation, aimed at maintaining voltage stability and reflecting the model's prioritization of safety in complex systems. Although MADDPG and MAGRL respectively achieved the optimal PL in the two scenarios, their low CR undermines overall effectiveness. Furthermore, unlike MADDPG, which becomes more volatile in dynamic settings, GKAN-MA maintains stable performance with tighter confidence intervals, confirming its robustness.

Table 3 further shows that GKAN-MA performs better in dynamic environments, while other methods fail to maintain consistent gains. This strongly emphasizes the effective integration of GATv2's dynamic topology modeling with KAN's nonlinear approximation capabilities within GKAN-MA. This enables GKAN-MA not only to handle complex systems but also to learn and optimize control strategies from dynamic changes, highlighting its potential for practical deployment in increasingly complex power grids with rising renewable energy penetration.

4.2.2 Performance Analysis for Representative Test Days. For experiments on representative test days, we adopt a typical summer PV output pattern under dynamic scenarios. Figure 4 presents the voltage distribution heatmaps of GKAN-MA and various baseline methods on the dynamic 33-bus system during a typical summer day. GKAN-MA demonstrates the most robust performance in voltage control. Its heatmap primarily exhibits uniform dark green to light green colors, indicating that the majority of nodes maintain voltage within the safe operational range throughout all time steps, with only minor fluctuations at a few time points. In contrast, other methods exhibit varying degrees of voltage violations. MADDPG, MAPPO, and IPPO show prominent yellow regions, particularly around bus 15-20 during certain time intervals, indicating their limited capability in suppressing voltage rise. Although MAGRL and GAMARL perform better than conventional RL algorithms overall, they still display sporadic areas of elevated voltage.

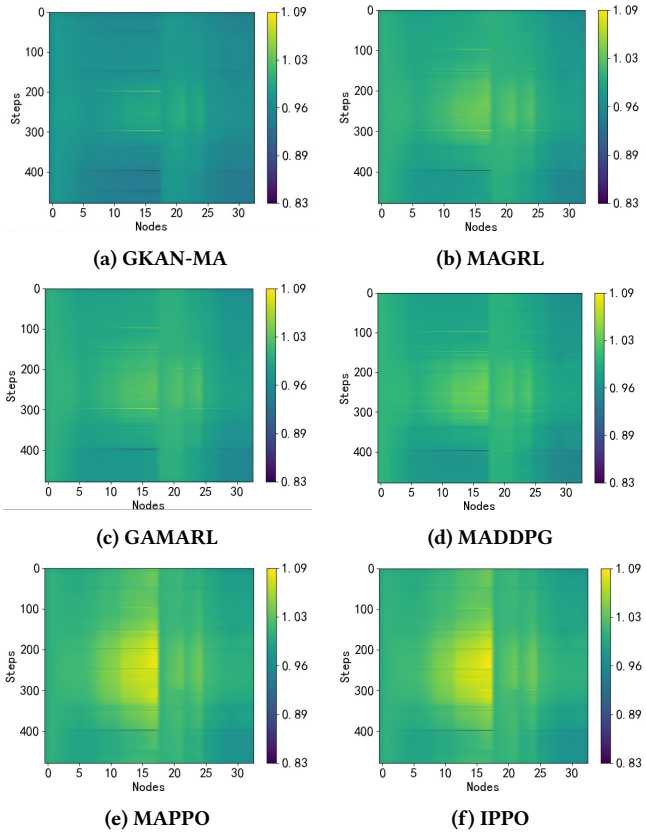
Figure 5 presents the dynamic voltage distribution heatmaps for a typical summer day across the 141-bus system, comparing GKAN-MA with multiple baseline algorithms. The GKAN-MA heatmap



(a) Static-141-CR (b) Static-141-PL

(c) Dynamic-141-CR (d) Dynamic-141-PL

Figure 3: Training results of algorithms on the 141-bus static and dynamic systems.



(a) GKAN-MA (b) MAGRL

(c) GAMARL (d) MADDPG

(e) MAPPO (f) IPPO

Figure 4: Test results of algorithms on the 33-bus dynamic system on a typical summer day.

mainly displays uniform green, indicating that the majority of nodes maintain voltages within safe limits throughout all time steps, with only minor yellow high-voltage areas appearing in a few node regions. In contrast, other methods exhibited more severe voltage

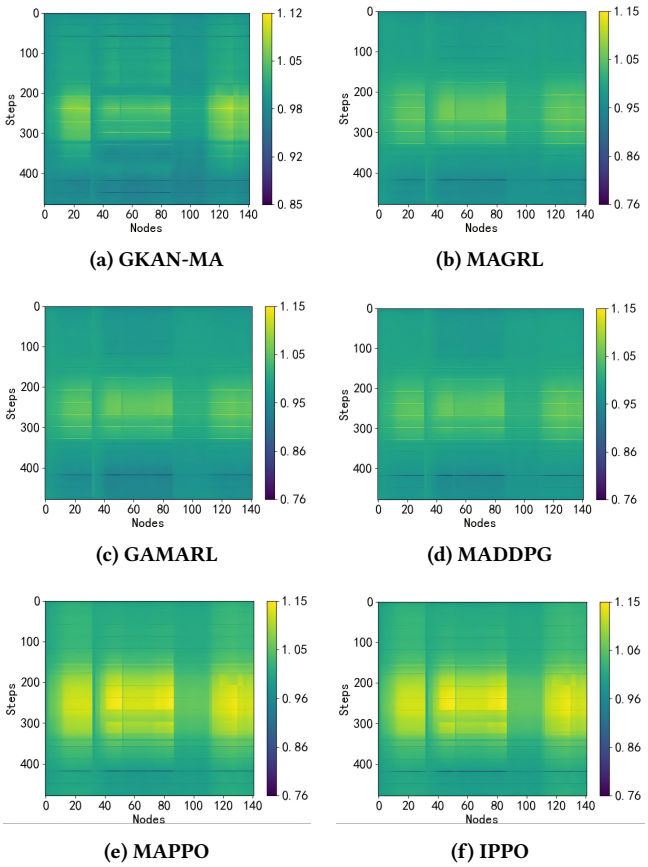


Figure 5: Test results of algorithms on the 141-bus dynamic system on a typical summer day.

out-of-limit issues. MAPPO and IPPO exhibit distinct yellow zones, particularly from bus 40 to bus 80, indicating these algorithms struggle to effectively suppress voltage rises in large-scale systems. While MAGRL and GAMARL generally outperform classical RL algorithms and exhibit voltage variations similar to GKAN-MA, their voltage fluctuation ranges exceed those of GKAN-MA, suggesting slightly inferior voltage control capabilities. Notably, GKAN-MA demonstrates exceptional suppression of inter-node voltage fluctuations under both conditions. Its heatmap exhibits virtually no discernible longitudinal stripe-like oscillations, indicating the method's capability to effectively address dynamic voltage disturbances caused by sharp variations in summer photovoltaic output, which ensures the system operates stably within the safe voltage range throughout the 24 hours.

4.2.3 Ablation Studies. For experiments on the ablation studies, to investigate the impact of each layer on the overall performance of GKAN-MA, we conducted ablation experiments with two configurations: (1) KAN-MA: removing only the GATv2 layer; (2) GATv2-MA: removing only the KAN layer. The experiments involve 10 episodes of testing conducted after training, presenting the average values of the test results CR and PL.

Table 4: The comparison results of ablation studies in 33-bus environments

Methods	33-bus-static		33-bus-dynamic	
	CR	PL	CR	PL
GKAN-MA	**0.9869**$^{\uparrow 13.35\%}$	**0.0588**$^{\downarrow 20.86\%}$	**0.9876**$^{\uparrow 3.85\%}$	**0.0584**$^{\downarrow 24.94\%}$
GATv2-MA	0.8707	0.0772	0.9510	0.0778
KAN-MA	0.8126	0.0743	0.7674	0.0933

In the 33-bus system, GKAN-MA achieves optimal performance in both static and dynamic scenarios. As shown in Table 4, it reaches CR values of 0.9869 and 0.9876, respectively—representing improvements of 13.3% and 3.8% over GATv2-MA, and 21.5% and 28.7% over KAN-MA. Notably, GATv2-MA significantly outperforms KAN-MA, especially in dynamic environments where its CR is 24.0% higher, underscoring the critical role of GATv2 in capturing topological dependencies and modeling voltage fluctuation propagation paths. Conversely, GKAN-MA achieves lower power loss (PL) than GATv2-MA in static environments, highlighting KAN's strength in modeling complex nonlinear control relationships through learnable B-spline functions. This functional complementarity between the two modules explains why GKAN-MA consistently surpasses both ablated variants: GATv2-MA lacks KAN's nonlinear fitting capability, while KAN-MA cannot account for network topology. Moreover, GKAN-MA's slightly higher CR in the dynamic setting compared to the static one demonstrates that the integration of GATv2 enhances the model's adaptability to changing operating conditions, thereby improving both robustness and responsiveness.

5 CONCLUSION

In this paper, we have introduced GKAN-MA, a MARL framework designed to address AVC challenges in power systems with highly dynamic topologies and nonlinear voltage-power dynamics. It maintains continuous topology awareness without static assumptions, ensuring consistent performance during network reconfigurations. Experimental results demonstrate superior controllability and reduced power loss across both IEEE 33-bus and 141-bus systems, validating the critical value of integrating dynamic topology perception with interpretable nonlinear function approximation. This dual-enhanced approach enables reliable adaptive voltage control in modern power grids with high integration of renewable energy.

In the future, we aim to explore the development of responsive techniques to refine GATv2 attention parameters, enabling a more accurate representation of evolving topological configurations in diverse operational contexts. Additionally, considering the prevalence of highly heterogeneous environments in real-world power grids, we will develop representative simulation scenarios and design a specialized algorithm for efficient voltage control.

ACKNOWLEDGMENTS

This Work is supported by National Natural Science Foundation of China (No. 62372139), Natural Science Foundation of Guangdong (No. 2024A1515030024), Research Projects of Shenzhen (No. ZDCY20250901100302003).

REFERENCES

[1] Surender Singh and Saurabh Singh. 2024. Advancements and challenges in integrating renewable energy sources into distribution grid systems: A comprehensive review. *Journal of Energy Resources Technology* 146, 9 (2024), 090801.

[2] Nande Fose, Arvind R Singh, Senthil Krishnamurthy, Mukovhe Ratshitanga, and Prathaban Moodley. 2024. Empowering distribution system operators: A review of distributed energy resource forecasting techniques. *Heliyon* 10, 15 (2024).

[3] Zhaoyang Liu, Liang Ma, Ke Wang, Junnan Zhang, Chenyi Si, Jun Yi, and Chaoxu Mu. 2025. Uncertainty-Aware Model-Based Multi-Agent Deep Reinforcement Learning for Robust Active Voltage Control. *IEEE Transactions on Circuits and Systems I: Regular Papers* (2025).

[4] Pengcheng Chen, Shichao Liu, Xiaozhe Wang, and Innocent Kamwa. 2024. Physics-guided multi-agent adversarial reinforcement learning for robust active voltage control with peer-to-peer (P2P) energy trading. *IEEE Transactions on Power Systems* 39, 6 (2024), 7089–7101.

[5] Wei Wu, Zhiqing Yang, Shan He, Helong Li, Kai Zhang, and Lijian Ding. 2025. Enhanced Operation of Grid-Following VSCs with Alternating-Voltage Control in Ultra-Weak Grids. *IEEE Journal of Emerging and Selected Topics in Power Electronics* (2025).

[6] Hansheng Tang, Ye He, Xiaoming Wang, Hao Zheng, Bin Xu, Wenguang Zhao, and Hongbin Wu. 2024. Two-stage multi-mode voltage control for distribution networks: A deep reinforcement learning approach based on multiple intelligences. *IEEE Transactions on Industry Applications* 60, 4 (2024), 5681–5691.

[7] Yujie Tang, Krishnamurthy Dvijotham, and Steven Low. 2017. Real-time optimal power flow. *IEEE Transactions on Smart Grid* 8, 6 (2017), 2963–2973.

[8] Yonathan Aflalo, Alexander Bronstein, and Ron Kimmel. 2015. On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences* 112, 10 (2015), 2942–2947.

[9] Md Mahmud-Ul-Tarik Chowdhury, Krishna Murari, Md Shamim Hasan, and Sukumar Kamalasadan. 2025. Optimal Power Flow (OPF) Analysis for AC–DC Active Distribution Networks Utilizing Second-Order Cone Programming (SOCP) Approach. *IEEE Transactions on Industrial Informatics* (2025).

[10] Sina Mohammadi, Van-Hai Bui, Wencong Su, and Bin Wang. 2024. Surrogate Modeling for Solving OPF: A Review. *Sustainability* 16, 22 (2024), 9851.

[11] Yang Qu, Jinming Ma, and Feng Wu. 2024. Safety constrained multi-agent reinforcement learning for active voltage control. *arXiv preprint arXiv:2405.08443* (2024).

[12] Bin Zhang, Di Cao, Weihao Hu, Amer MYM Ghias, and Zhe Chen. 2024. Physics-Informed Multi-Agent deep reinforcement learning enabled distributed voltage control for active distribution network using PV inverters. *International Journal of Electrical Power & Energy Systems* 155 (2024), 109641.

[13] Jiapeng Huang, Huifeng Zhang, Ding Tian, Zhen Zhang, Chengqian Yu, and Gerhard P Hancke. 2024. Multi-agent deep reinforcement learning with enhanced collaboration for distribution network voltage control. *Engineering Applications of Artificial Intelligence* 134 (2024), 108677.

[14] Na Xu, Chaoxu Mu, Liang Ma, and Ke Wang. 2025. Real-time Voltage Control in Smart Distribution Network through Multi-agent Cooperative Optimization. *IEEE Transactions on Sustainable Energy* (2025).

[15] Xiangwang Hou, Jingjing Wang, Jun Du, Chunxiao Jiang, and Yong Ren. 2025. Distributed Machine Learning for Autonomous Agent Swarm: A Survey. *IEEE Communications Surveys & Tutorials* (2025).

[16] Gabriele Corso, Hannes Stark, Stefanie Jegelka, Tommi Jaakkola, and Regina Barzilay. 2024. Graph neural networks. *Nature Reviews Methods Primers* 4, 1 (2024), 17.

[17] Ahmed K Khamis and Mohammed Agamy. 2024. Circuit topology aware GNN-based multi-variable model for DC-DC converters dynamics prediction in CCM and DCM. *Neural Computing and Applications* 36, 33 (2024), 20807–20822.

[18] Amir Meydani, Hossein Shahinzadeh, Ali Ramezani, Majid Moazzami, Hamed Nafisi, and Hossein Askarian-Abyaneh. 2024. Comprehensive review of artificial intelligence applications in smart grid operations. In *2024 9th International Conference on Technology and Energy Management (ICTEM)*. IEEE, 1–13.

[19] Yanhua Huang. 2020. Deep Q-networks. In *Deep reinforcement learning: fundamentals, research and applications*. Springer, 135–160.

[20] MA Aoxiang, CAO Jun, and CORTES Pedro RODRIGUEZ. 2024. Graph Neural Network Based Deep Reinforcement Learning for Volt-Var Control in Distribution Grids. In *2024 IEEE 15th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*. IEEE, 1–5.

[21] Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. 2019. Graph convolutional networks: a comprehensive review. *Computational Social Networks* 6, 1 (2019), 1–23.

[22] Chaoxu Mu, Zhaoyang Liu, Jun Yan, Hongjie Jia, and Xiaoyu Zhang. 2023. Graph multi-agent reinforcement learning for inverter-based active voltage control. *IEEE Transactions on Smart Grid* 15, 2 (2023), 1399–1409.

[23] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[24] Fengzhang Luo, Shengyuan Wang, Yunqiang Lv, Ranfeng Mu, Jiacheng Fo, Tianyu Zhang, Jing Xu, and Chengshan Wang. 2025. Domain knowledge-enhanced graph reinforcement learning method for Volt/Var control in distribution networks. *Applied Energy* 398 (2025), 126409.

[25] Shaked Brody, Uri Alon, and Eran Yahav. 2021. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491* (2021).

[26] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. 2024. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756* (2024).

[27] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems* 30 (2017).

[28] Jianhong Wang, Wangkun Xu, Yunjie Gu, Wenbin Song, and Tim C Green. 2021. Multi-agent reinforcement learning for active voltage control on power distribution networks. *Advances in neural information processing systems* 34 (2021), 3271–3284.

[29] Yongdong Chen, Youbo Liu, Hang Yin, Zhiyuan Tang, Gao Qiu, and JunYong Liu. 2024. Multiagent soft actor–critic learning for distributed ess enabled robust voltage regulation of active distribution grids. *IEEE Transactions on Industrial Informatics* 20, 9 (2024), 11069–11080.

[30] Anirban Chowdhury, Ranjit Roy, and Kamal Krishna Mandal. 2024. Enhancement of technical, economic & environmental benefits in multi-point PV & wind-based DG integrated radial distribution network using Aquila optimizer. *Expert Systems with Applications* 252 (2024), 124307.

[31] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems* 35 (2022), 24611–24624.

[32] Yongliang Wang and Hamidreza Kasaei. 2022. IPPO: Obstacle Avoidance for Robotic Manipulators in Joint Space via Improved Proximal Policy Optimization. *arXiv preprint arXiv:2210.00803* (2022).

[33] Leijiao Ge, Jingjing Li, Luyang Hou, and Jingang Lai. 2025. Autonomous Voltage Regulation for Smart Distribution Network with High-Proportion PVs: A Graph Meta-Reinforcement Learning Approach. *IEEE Transactions on Sustainable Energy* (2025).

[34] Yongjiang Zhao, Haoyi Zhong, and Chang Cyoon Lim. 2024. Safety-constrained multi-agent reinforcement learning for power quality control in distributed renewable energy networks. *Comput Mater Contin* 79, 1 (2024), 449–71.