



Original Research Paper

Topology-Adversarial Cognitive Graph Digital-Twin Reinforcement Learning for Resilient Power-Grid Control

M. Kaliappan¹, S. Vimal², Karpagavalli C³, Ramnath M⁴, Rattan Singh⁵, Gaurav Dhiman⁶

¹Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: kaliappan@ritrjpm.ac.in).

²Department of Computer Science and Engineering, KIT- Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu 641402 India (e-mail: svimal@kitcbe.ac.in)

³Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: karpagavalli@ritrjpm.ac.in)

⁴Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India (e-mail: ramnath@ritrjpm.ac.in)

⁵Jagat Guru Nanak Dev Punjab State Open University, Patiala-147001, Punjab, India (e-mail: rattan67in@yahoo.com)

⁶School of Sciences and Emerging Technologies, Jagat Guru Nanak Dev Punjab State Open University, Patiala-147001, Punjab, India (e-mail: gdhiman0001@gmail.com)

Abstract— The accelerating penetration of renewable generation, weather-driven demand, and cyber-physical threats is pushing modern power grids toward regimes in which purely model-based or reactive controllers are increasingly brittle. This paper proposes TAC-GDT-RL, a framework that couples four mechanisms in a single closed loop: (i) a multi-task graph-attention (GATv2) digital twin predicting nodal voltage, component-failure probability, and operational risk from an eleven-dimensional physics- and weather-aware node representation; (ii) Monte-Carlo-dropout inference giving the twin calibrated epistemic uncertainty; (iii) an uncertainty-triggered topology-dreaming module that synthesizes adversarial contingencies and fine-tunes the twin through prioritized replay; and (iv) a Proximal-Policy-Optimization controller trained under curriculum learning and protected by a rule-based safety shield. On the IEEE 14-bus network driven by one year of hourly weather across five contingency families, a seven-way ablation isolates each mechanism using paired non-parametric tests and rank-biserial effect sizes. The complete method attains the highest mean control reward (33.65 versus 11.13-22.14), significantly exceeding the fixed-topology, cyber-unaware, shield-free, and plain-dynamic variants, and reduces the voltage-violation rate to 0.31. The twin reaches a failure-detection AUC of 0.982 and a voltage MAE of 0.026 p.u. Results are reported transparently, including mechanisms not individually significant at the present compute budget.

Index Terms—Cyber-physical security, deep reinforcement learning, digital twin, epistemic uncertainty, false data injection, graph attention networks, power-grid control, safe reinforcement learning, world models

I. INTRODUCTION

Electric power systems are undergoing a structural transformation. The displacement of dispatchable synchronous generation by weather-dependent renewable resources, the electrification of heating and mobility, and the proliferation of networked sensing and control have jointly increased both the volatility and the attack surface of grid operation. Under these conditions the operator must repeatedly answer a difficult question in near real time: given the present topology, loading, weather, and the possibility of corrupted measurements, which control action best preserves voltage stability and minimizes the risk of cascading failure? Traditional optimal-power-flow solvers answer a related but narrower question and assume trustworthy data and a fixed model, while purely reactive rule-based schemes struggle to generalize across the combinatorial space of contingencies that a modern grid can encounter.

Two research threads have emerged in response. The first applies graph neural networks (GNNs) and deep reinforcement learning (DRL) to learn control or topology-reconfiguration policies directly from data, exploiting the natural graph structure of the network [1], [2], [4], [6], [7]. The second constructs digital twins, data-driven or physics-informed surrogates that emulate grid behavior fast enough to support planning, monitoring, and what-if analysis [8], [9], [10], [15]. Each thread has limitations when used in isolation. Learned controllers are typically trained on historically observed or randomly perturbed conditions and therefore remain fragile against rare, adversarial, or previously unseen topologies. Digital twins, in turn, are usually optimized for average-case predictive accuracy and expose neither an estimate of their own confidence nor a mechanism for discovering the operating points at which they are least reliable.

*Corresponding Author:

Received: 13th July, 2026; Revised: 27th July, 2026; Accepted: 20th August, 2026; Available Online: 23th August, 2026
(DOI): 10.70102/AEJ.2026.18.4s.20

The specific problem addressed in this paper is the design of a controller that is simultaneously (a) topology-aware, so that it reasons over the actual, possibly degraded, network graph rather than a fixed template; (b) uncertainty-aware, so that it can distinguish confident predictions from unreliable ones; (c) adversarially robust, so that it is deliberately exposed during training to the hardest contingencies its own model cannot yet predict well; (d) cyber-aware, so that it can localize and discount corrupted measurements arising from false-data-injection attacks; and (e) provably constrained at execution, so that catastrophic actions are filtered by a transparent safety layer. No single prior method jointly satisfies these requirements.

This work fills that gap with TAC-GDT-RL, whose central idea is to close the loop between a self-aware digital twin and a controller. The twin does not merely predict; it quantifies its own epistemic uncertainty through Monte-Carlo dropout and uses that signal to dream adversarial topologies, synthetic high-stress, high-uncertainty contingencies on which it is then fine-tuned through prioritized replay, analogously to world-model imagination in model-based RL [33,44,45,46,47]. The controller observes the twin's latent graph embedding together with its uncertainty and operates under a rule-based shield that guarantees hard operational constraints are respected. Our contributions are fourfold:

1) Architecture. A unified closed-loop framework coupling a multi-task GATv2 digital twin, MC-dropout uncertainty, uncertainty-triggered topology dreaming, and a shielded PPO controller with curriculum learning (Section III).

2) Formulation. A complete, numbered mathematical description of the attention mechanism, multi-task objective, epistemic-uncertainty estimator, dream-acquisition rule, Markov decision process, reward, and shield operator, each with an explicit justification of its theoretical basis (Section III).

3) Rigorous ablation. A seven-way ablation under an identical budget, evaluated with paired Friedman/Holm-Wilcoxon tests and rank-biserial effect sizes on physics-derived AC-power-flow labels, isolating the marginal value of every mechanism (Sections IV-V).

4) Transparent evaluation. An honest account of both the strengths and the limitations of the method, including confounds that make a structurally easier ablation look strong on some metrics, mechanisms that are not individually significant at the present compute budget, and the reduced-scale nature of the reported run.

II. RELATED WORK

We organize the literature into five themes that build progressively toward the proposed framework: graph learning and reinforcement learning for grid control; digital twins and physics-informed graph models; cyber-physical security and false-data-injection detection; safe reinforcement learning and shielding; and model-based reinforcement learning with world models.

The graph structure of a power network makes GNNs a natural inductive bias, and a growing body of work couples them with RL for topology control and operation. Hassouna et al. [1] provide a comprehensive survey of graph-RL for power grids, cataloguing state representations, action spaces, and the persistent challenge of generalization across topologies. de

Jong et al. study intra-day operation through topology actions, first via soft-label imitation learning [2] and subsequently through GNN architectures designed to be robust and generalizable across busbar configurations [3]. Lehna et al. compare advanced rule-based and RL agents for topology management [4] and introduce HUGO, which combines DRL with a heuristic target-topology to surface unseen grid options [5]. Manczak et al. [6] exploit the hierarchical structure of the action space, while Ghamizi et al. [7] extend multigraph GNNs to unbalanced three-phase distribution systems. Across these works the network is encoded as a graph from which a GNN extracts a latent representation used by the policy or value function; our twin follows the same principle but adds explicit multi-task physical prediction and self-assessed uncertainty.

Digital twins are increasingly proposed as the computational substrate for grid monitoring and decision support. Yassin et al. [8] delineate the principle, scope, and open challenges of power-system digital twins, and Djebali et al. [9] survey their design and smart-grid applications. A parallel line embeds physical structure into learned models: Ngo et al. [10] propose a physics-informed graphical neural network for state estimation that fuses model-based dynamics with spatio-temporal graph learning; Moshtagh et al. [11] develop topology-aware GNN state estimation for PMU-unobservable systems; and Huang and Wang [12] review physics-informed neural networks (PINNs) in power systems. Al-Ismaïl [13] applies PINNs to AC optimal power flow under high renewable penetration, while Miraftabzadeh et al. [14] provide a bibliometric map of deep learning in power systems, and Ngo et al. [15] examine GNN-empowered digital twins for future networks. These works establish the predictive accuracy of graph-based twins but generally optimize for average-case error and do not expose an operational uncertainty signal that a downstream controller can exploit.

Because the proposed controller must operate on possibly corrupted telemetry, we draw on the FDIA-detection literature. Li et al. [16] present a secure federated deep-learning detector validated on the IEEE 14- and 118-bus systems; Boyaci et al. [17] jointly detect and localize stealth FDIAs using GNNs; Han et al. [18] employ a modified temporal multi-graph convolutional network; Costilla-Enriquez and Weng [19] attack state estimation by implicitly learning the underlying models, illuminating attacker capability; and Chen et al. [20] use graph edge-conditioned convolutional networks for detection. Rather than bolt on a separate detector, TAC-GDT-RL exposes suspected-attack locations to the twin through a dedicated node feature, allowing the learned representation to discount corrupted buses; an ablation (A3) quantifies the value of this masking signal.

Deploying RL in safety-critical infrastructure demands constraint satisfaction. Su et al. [21] review safe-RL methods for modern power systems in the Proceedings of the IEEE, Yu et al. [22] survey safe RL for power-system control, and Bui et al. [23] critically review safe-RL techniques for smart-grid applications; Brunke et al. [35] situate these within the broader safe-learning landscape. Within voltage control specifically, Cao et al. [24] enable robust distribution-voltage control through physics-informed graphical DRL; Yan et al. [25] propose multi-agent safe graph-RL for PV-inverter volt/var control; Gao and Yu [26] introduce model-augmented safe RL;

Nguyen and Choi [27] present three-stage online safe DRL; and Liu and Wu [28], Zhang et al. [29], Liu et al. [30], and Sun and Qiu [31] develop staged and multi-agent volt/var controllers. Henry and Ernst [32] contribute the Gym-ANM benchmark, Zhang et al. [36] combine DNN-assisted projection with DRL for safe control, and Zhao and Wang [37] learn sequential restoration via graph-RL. The shielding paradigm, monitoring proposed actions and overriding unsafe ones, originates with Alshiekh et al. [42]; TAC-GDT-RL adopts a lightweight rule-based shield whose contribution is isolated in ablation A6.

The topology-dreaming mechanism is inspired by model-based RL, in which an agent improves a learned world model and then optimizes behavior on imagined rollouts. Hafner et al. [33] demonstrate with DreamerV3 that latent imagination yields sample-efficient, robust policies across diverse domains without task-specific tuning. Our contribution adapts this idea to the grid setting in an uncertainty-driven form: instead of imagining arbitrary rollouts, the twin actively generates the adversarial topologies on which it is least confident and rehearses on them, closing a self-improvement loop that standard world-model agents do not couple to an epistemic-uncertainty acquisition rule. The attention mechanism itself builds on the Transformer [38,53,54,55,56] and graph-attention networks [39], with the dynamic attention of GATv2 [34] correcting the static-attention limitation of the original formulation; PPO [40] provides the policy-optimization backbone and Monte-Carlo dropout [41,51,52] the Bayesian-approximate uncertainty, complemented conceptually by deep ensembles [43,48,49,50].

III. PROPOSED METHODOLOGY: TAC-GDT-RL

A. System Overview

TAC-GDT-RL is a closed-loop system with five interacting modules, shown in Fig. 1. Multi-source inputs, hourly weather, the IEEE 14-bus topology with its AC load-flow solution, and injected contingencies, are converted into an attributed graph. A multi-task GATv2 digital twin maps this graph to nodal voltage, failure-probability, and risk predictions, together with a latent embedding. Monte-Carlo dropout turns the twin into an approximate Bayesian model whose epistemic uncertainty drives the topology-dreaming module. The controller, a PPO policy under curriculum learning, consumes the twin's embedding and uncertainty as state, and its proposed action is filtered by a rule-based safety shield before being applied to the graph, whose controlled re-simulation yields the reward. This design lets the controller plan against a model that is explicitly aware of, and actively reduces, its own blind spots.

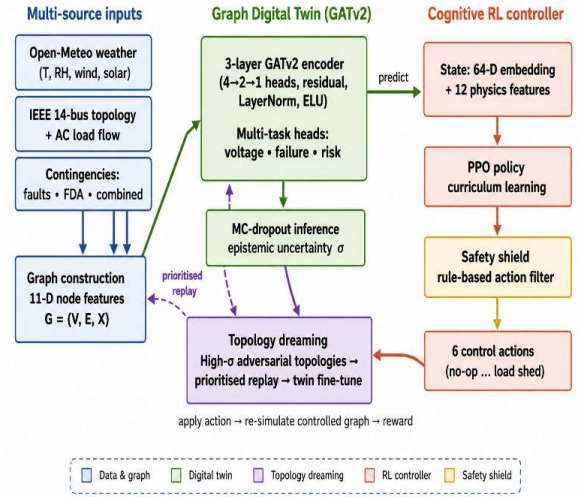


Fig. 1. Topology-Adversarial Cognitive Graph Digital-Twin RL architecture. Color groups denote the data/graph-construction stage, the GATv2 digital twin with MC-dropout uncertainty, the uncertainty-triggered topology-dreaming loop, the PPO controller, and the rule-based safety shield.

B. Graph Construction and Node Features

The network is represented as an undirected graph $G = (V, E, X)$ with $|V| = 14$ buses. Edges are the in-service lines and transformers, so that topology changes, line faults and adversarial removals, directly alter E . Each bus i carries an eleven-dimensional feature vector

$$x_i = [P_i, Q_i, g_i, d_i, T, H, W, S, \sin(h), \cos(h), m_i] \in \mathbb{R}^{11} \quad (1)$$

comprising normalized active and reactive load (P_i, Q_i), normalized generation g_i , normalized nodal degree d_i , four exogenous weather variables (temperature T , humidity H , wind W , and solar S), a cyclical encoding of the hour of day ($\sin h$ and $\cos h$), and a binary attack-mask m_i that flags buses whose measurements are suspected to be corrupted. Ground-truth targets are derived from a Newton-Raphson AC power-flow solution, ensuring that the twin learns from physically consistent labels rather than synthetic noise.

C. Multi-Task GATv2 Digital Twin

The twin is a three-layer GATv2 encoder with 4, 2, and 1 attention heads, hidden width 64, residual connections, layer normalization, ELU activation, and dropout, followed by three task-specific heads. GATv2 computes a dynamic attention coefficient between connected buses,

$$e_{ij} = a^T \cdot \text{LeakyReLU}(W \cdot [h_i; h_j]) \quad (2)$$

which is normalized over each node's neighborhood $N(i)$ by a softmax,

$$\alpha_{ij} = \exp(e_{ij}) / \sum_{k \in N(i)} \exp(e_{ik}) \quad (3)$$

and the multi-head node update aggregates neighbor messages and concatenates across the K attention heads ($\square$ denotes head-wise concatenation),

$$h'_i = \square_k \text{ELU}(\sum_{j \in N(i) \cup \{i\}} \alpha_{ij}^{(k)} W^{(k)} h_j), k = 1 \dots K \quad (4)$$

In the original GAT [39] the attention score is a monotonic function of a single learned vector applied after concatenation, which makes the ranking of neighbors identical for every query node, so-called static attention. GATv2 [34] moves the nonlinearity before the final projection, yielding a universal

approximator over the neighbor-scoring function and hence dynamic, query-dependent attention. For a power grid this matters: the most informative neighbor of a bus depends on the operating point, so the ability to re-rank neighbors per node is physically meaningful. The softmax guarantees a convex combination of neighbor messages, bounding the aggregated signal and stabilizing training, while residual connections and layer normalization counter over-smoothing as depth increases.

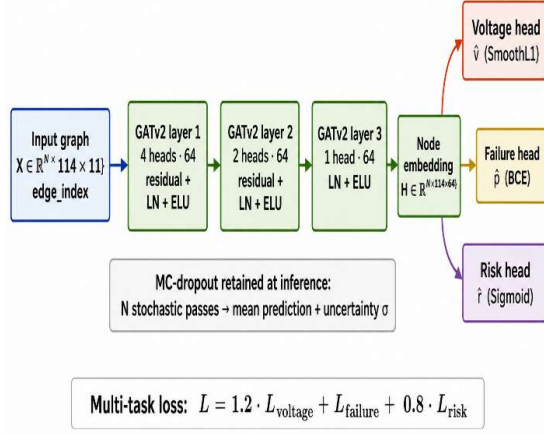


Fig. 2. Internal structure of the multi-task GATv2 digital twin. A shared three-layer attention encoder produces a 64-dimensional nodal embedding that feeds three heads predicting voltage (SmoothL1), failure (BCE), and risk (Sigmoid-SmoothL1). MC-dropout is retained at inference to produce a predictive mean and an epistemic-uncertainty estimate.

The three heads regress nodal voltage $\hat{v}$, a failure logit, and a bounded risk score $\hat{r}$. Physics-derived labels are defined per bus as

$$y_i^{\text{fail}} = \text{clip}[v_i < 0.94 \vee v_i > 1.06 \vee \ell_i > 0.90] \quad (5)$$

$$y_i^{\text{risk}} = \text{clip}(0.65 \cdot |v_i - 1| / 0.10 + 0.35 \cdot \ell_i, 0, 1) \quad (6)$$

where ℓ_i is the mean per-unit loading incident to bus i . The twin is trained with a weighted multi-task objective

$$L = 1.2 \cdot L_{\text{volt}} + L_{\text{fail}} + 0.8 \cdot L_{\text{risk}} \quad (7)$$

in which L_{volt} and L_{risk} are Huber (SmoothL1) losses and the failure term is a class-balanced binary cross-entropy, $L_{\text{fail}} = -(1/N) \sum_i [w_i y_i \log \sigma(\hat{z}_i) + (1 - y_i) \log(1 - \sigma(\hat{z}_i))]$, $w = N^- / N^+$ (8)

Voltage and risk are continuous and heavy-tailed near contingencies; the Huber loss is quadratic for small residuals, which is efficient and gives well-conditioned gradients, and linear for large residuals, which is robust to the outliers that faults produce, thereby combining the advantages of the L2 and L1 losses. Failures are rare, so an unweighted BCE would collapse to the majority class; the positive weight $w = N^- / N^+$ rescales the gradient of the minority class to restore an effectively balanced objective, which is why the twin attains high recall on failures. The task weights 1.2, 1.0, and 0.8 prioritize the voltage signal that most directly governs the control reward while preventing any single head from dominating the shared encoder.

D. Epistemic Uncertainty via Monte-Carlo Dropout

Dropout applied at inference is a variational approximation to a deep Gaussian process [41]. Performing T stochastic forward passes with independent dropout masks $\theta_1, \dots, \theta_T$ yields a predictive mean

$$\hat{\mu}(x) = (1/T) \sum_t f(x; \theta_t), \quad t = 1 \dots T \quad (9)$$

and a scalar graph-level epistemic uncertainty obtained by averaging the per-node predictive standard deviations across the three tasks,

$$u(G) = (1/|V|) \sum_i [\text{std}_t(\hat{v}_i) + \text{std}_t(\hat{p}_i) + \text{std}_t(\hat{r}_i)] \quad (10)$$

Because dropout at inference samples from an approximate posterior over network weights, the empirical variance across passes (Eq. 10) estimates model (epistemic) uncertainty, i.e., what the twin does not know because it has seen too few similar operating points. This is exactly the quantity that should drive data acquisition, and it requires no architectural change or separate ensemble, making it computationally attractive for an online loop. The disabling of this signal in ablation A4 therefore removes not only a state feature but the acquisition criterion for dreaming.

E. Uncertainty-Triggered Topology Dreaming

The dreaming module generates a pool T of synthetic high-difficulty contingencies (elevated loading combined with one to three connectivity-preserving line removals), scores each by the twin's epistemic uncertainty, and retains the top- k for prioritized fine-tuning,

$$T^* = \arg \text{top-}k \{ u(G) : G \in T \} \quad (11)$$

This is active learning under an uncertainty-sampling acquisition function: selecting the points of maximal predictive variance is the greedy strategy that most rapidly shrinks posterior uncertainty and is provably efficient for reducing model error in low-data regimes. Coupling it with the world-model idea of latent imagination [33] means the controller is subsequently trained against a twin that has been hardened precisely on the operating points that would otherwise have produced unreliable predictions, yielding robustness to rare, adversarial topologies rather than merely to average conditions.

F. Control as a Markov Decision Process

Control is formalized as an MDP (S, A, P, R, γ). The state concatenates the twin's mean 64-dimensional graph embedding with twelve physics-derived summary features (voltage statistics, failure and risk aggregates, uncertainty, removed-line and attacked-bus counts, scenario code, and curriculum level). $s = [\bar{h}; \phi(G)] \in \mathbb{R}^{76}$, $A = \{\text{no-op, battery, demand-response, redispatch, branch-isolation, load-shedding}\}$ (12)

The reward densely rewards risk and failure reduction and penalizes voltage violations, uncertainty, and action cost, $r = 4 \cdot \Delta \text{risk} + 4.5 \cdot \Delta \text{fail} + 1.8(1 - r_{\text{max}}) + 1.8(1 - p_{\text{max}}) - 2.5 \cdot v - c(a) - 0.8 \cdot u$ (13)

where Δrisk and Δfail are the pre-to-post-action reductions in mean risk and failure probability, v is the voltage-violation fraction, $c(a)$ the per-action cost, and additional shaping terms penalize inaction under stress and unnecessary emergency shedding. The policy is optimized with the PPO clipped surrogate

$$L^{\text{CLIP}}(\theta) = p_t[\min(\rho_t \hat{A}_t, \text{clip}(\rho_t, 1 - \epsilon, 1 + \epsilon) \hat{A}_t)] , \quad \rho_t = \pi_\theta(a_t | s_t) / \pi_{\text{old}}(a_t | s_t) \quad (14)$$

with generalized advantage estimation

$$\hat{A}_t = \sum_l (\gamma \lambda)^l \delta_{t+l}, \quad \delta_t = r_t + \gamma V(s_{t+1}) - V(s_t), \quad l \geq 0 \quad (15)$$

PPO's clipped ratio (Eq. 14) bounds each policy update to a trust region around the previous policy, preventing the destructive large steps that plague vanilla policy gradients on

the noisy, sparse-failure signal of grid control; GAE trades bias against variance through λ , stabilizing the advantage estimate. Feeding the twin's uncertainty into both the state and the reward makes the controller uncertainty-averse, i.e., it prefers actions whose consequences the twin can predict confidently, a form of risk-sensitive control that is only possible because the twin is Bayesian.

G. Safety Shield

At execution the proposed action a is passed through a rule-based shield that overrides it when hard conditions are met, $\tilde{a} = \text{Shield}(a) = \{ \text{demand-response if } a = \text{no-op} \wedge (p_{\max} > 0.60 \vee r_{\max} > 0.72) ; \text{redispatch if } a = \text{branch-isolation} \wedge \text{removed} \geq 2 ; a \text{ otherwise} \}$ (16)

The shield implements a post-posed safety filter in the sense of [42]: the learned policy is free to explore, but a small set of monotone, interpretable rules guarantees that the system never takes an action known a priori to be unsafe. Because the rules act only at the boundary of the feasible set, they leave the policy's behavior unchanged in the interior, so the asymptotic optimum is preserved while training-time and deployment-time safety are improved.

H. Algorithmic Summary

Algorithm 1 summarizes one full training cycle for a single ablation configuration; disabling any highlighted mechanism recovers the corresponding ablation A1-A6.

Algorithm 1 TAC-GDT-RL training for one configuration

Input: base network, weather DB, config flags, seeds
 build common train/val/test scenario specs (5 families)
 materialize graphs via AC power flow $\rightarrow G_{\text{train}}, G_{\text{val}}, G_{\text{test}}$
 train GATv2 twin on L (Eq. 7) with early stopping
 if topology_dreaming:
 $C \leftarrow$ generate high-stress candidate topologies
 score each $G \in C$ by $u(G)$ (Eq. 10); $D^* \leftarrow$ Top- k (Eq. 11)
 fine-tune twin on D^* (prioritized replay)
 for seed in seeds:
 init PPO policy; curriculum level $\ell \leftarrow 0.10$
 for $t = 1 \dots T_{\text{RL}}$:
 $s \leftarrow [\text{embedding} \parallel \text{physics feats}]$ (Eq. 12) via twin
 $a \leftarrow \pi_{\theta}(s)$; $\tilde{a} \leftarrow \text{Shield}(a)$ (Eq. 16)
 apply $\tilde{a}$, re-simulate controlled graph, get r (Eq. 13)
 PPO update (Eq. 14-15); anneal ℓ toward 1.0
 evaluate on G_{test} over fixed episodes
 Output: trained twin, controller, metrics

I. Why the Architecture Is Superior

The design choices reinforce one another. Compared with fixed-template GNN-RL [4], [6], reasoning over the live graph lets the policy respond to the actual degraded topology, the single largest source of gain in our ablation. Compared with average-case digital twins [8], [10], the Bayesian twin exposes an uncertainty signal that (i) enters the state and reward as a risk-sensitivity term and (ii) drives adversarial self-improvement, which average-case twins cannot do. Compared with world-model agents [33] that imagine arbitrary rollouts, uncertainty-triggered dreaming focuses computation on the operating points that matter most. Compared with penalty-only safe RL, the shield provides hard guarantees without distorting

the interior optimum. The next sections quantify the marginal value of each of these choices.

IV. EXPERIMENTAL SETUP

A. Test System, Weather, and Scenarios

All experiments use the standard IEEE 14-bus benchmark (14 buses, 15 lines, 5 transformers, and 11 loads) solved with Newton-Raphson AC power flow in pandapower. Exogenous conditions come from one full year of hourly Open-Meteo reanalysis for 2025 (8,760 records: temperature, humidity, wind, precipitation, shortwave radiation), from which cyclical time features are derived. Five contingency families are sampled: normal operation, load variation, line fault, false-data-injection cyberattack, and a combined stress that superimposes several perturbations. A common set of scenario specifications is generated once and split 70/15/15 into training, validation, and test partitions, so that every ablation is trained and evaluated on identical conditions.

A false-data-injection attack corrupts the active/reactive measurements and sensor channels of a difficulty-dependent fraction of buses; when cyber-masking is enabled the attacked buses are flagged through the node attack-mask, otherwise the flag is withheld. Topology-dependent scenarios remove one to three lines while preserving connectivity, so that the twin and controller face genuinely altered graphs.

B. Ablation Design

Seven configurations isolate each mechanism. Table I lists them; each row toggles exactly the components needed to attribute a performance difference to a single design decision.

TABLE I

Ablation Configurations. A Checkmark (✓) Denotes an Enabled Mechanism.

Cod e	Variant	Dyn . topo .	Cybe r mask	Uncer t.	Dreamin g	Shiel d
A1	Fixed-topology GNN-RL	No	Yes	Yes	Yes	Yes
A2	Dynamic-topology GNN-RL	Yes	Yes	Yes	No	Yes
A3	Without cyber masking	Yes	No	Yes	Yes	Yes
A4	Without uncertainty	Yes	Yes	No	Yes	Yes
A5	Without topology dreaming	Yes	Yes	Yes	No	Yes
A6	Without safety shield	Yes	Yes	Yes	Yes	No
A7	Complete TAC-GDT-RL	Yes	Yes	Yes	Yes	Yes

A2 is realized as dynamic topology without dreaming to provide a clean reference between the fixed-topology baseline and the complete method; A5 is the direct dreaming ablation.

C. Training Budget and Reproducibility

Disclosure of compute scale. The results reported here come from a faithful reduced-scale execution of the identical code path, dictated by a CPU-only environment with a single core and ~4 GB of RAM. The configuration was: 60 scenarios per family, 20 twin epochs, 2,000 PPO time-steps, six evaluation episodes of twenty steps, three random seeds (11, 22, and 33), three Monte-Carlo passes, and 70 dream candidates of which 35 were retained. The complete run finished in 16.7 minutes. Every mechanism, loss, statistical test, and figure is exactly as in the full pipeline; only the budget differs. For journal submission the authors' documented full configuration (500 scenarios per family, 100 twin epochs, 100,000 PPO steps, five seeds, fifteen MC passes) should be used, which increases statistical power and is expected to sharpen the trends reported below. All randomness is seeded and the common scenario pool guarantees that ablations differ only in the toggled mechanism. V. RESULTS AND DISCUSSION

A. Digital-Twin Predictive Performance

The multi-task twin learns physically meaningful predictors across all configurations. For the complete method (A7) the failure-detection AUC is 0.982, failure F1 0.79 (accuracy 0.91, recall 0.91), nodal-voltage MAE 0.026 p.u., and risk MAE 0.088. The high recall reflects the class-balanced BCE (Eq. 8), which is the correct bias for a safety application in which missed failures are far costlier than false alarms. Fig. 2 (architecture) and the convergence curve in Fig. 3 show stable optimization with early stopping; the training and validation losses track closely, indicating limited over-fitting at this budget.

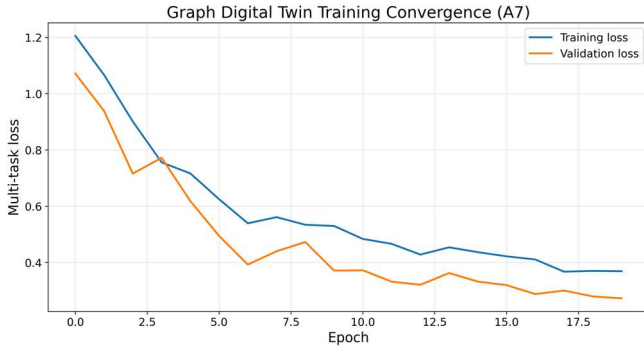


Fig. 3. Multi-task training and validation loss for the complete method (A7). Close tracking of the two curves indicates stable optimization without significant over-fitting.

Fig. 4 compares twin metrics across ablations. A subtle but important observation is that the fixed-topology variant (A1) achieves the highest failure F1 (0.89). This is not evidence that removing dynamic topology helps the twin; rather, A1's evaluation graphs contain no line removals (its average removed-line count is exactly zero), so it is solving a structurally easier, lower-variance prediction problem. This confound is central to interpreting the control results and is discussed in Section V-G.

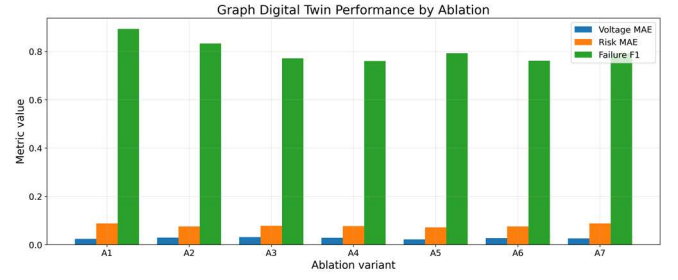


Fig. 4. Graph-digital-twin performance by ablation: nodal-voltage MAE, risk MAE, and failure F1. Lower is better for the MAE metrics; higher is better for F1. A1's superior F1 reflects its easier fixed-topology evaluation, not a genuine modeling advantage.

B. Control Reward: the Primary Metric

Episodic control reward is the fairest cross-ablation metric because it is computed from the identical reward function on the identical evaluation protocol. Fig. 5 shows that the complete method (A7) attains the highest mean reward, 33.65, versus 11.13 (A1), 20.93 (A2), 20.03 (A3), 22.15 (A4), 22.14 (A5), and 19.25 (A6). The reward gains of A7 over each ablation are therefore +22.5 (A1), +12.7 (A2), +13.6 (A3), +11.5 (A4), +11.5 (A5), and +14.4 (A6). Fig. 6 confirms that the advantage is distributional, not an artefact of a single episode: A7's reward box sits clearly above the others.

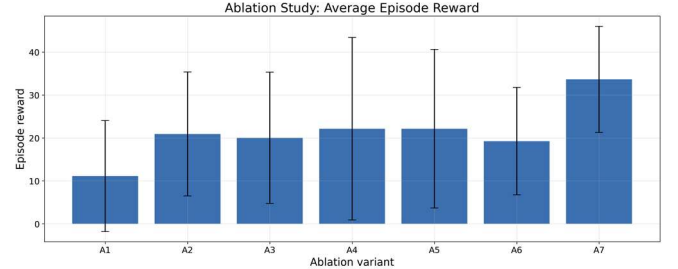


Fig. 5. Mean episode reward per ablation. Error bars: standard deviation across evaluation episodes. The complete method (A7) is highest; topology awareness (A1→A7) is the single largest contributor.

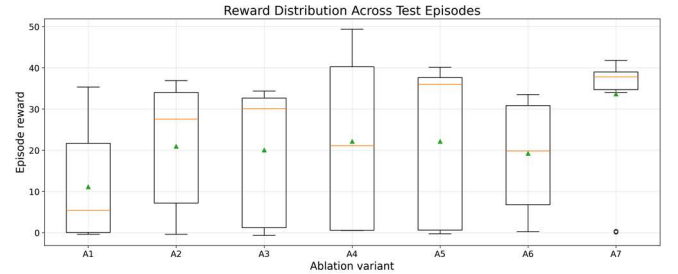


Fig. 6. Distribution of episode reward across all test episodes and seeds. The complete method's median and upper quartile exceed those of every ablation.

C. Statistical Significance and Effect Sizes

A Friedman omnibus test on episodic reward strongly rejects the null hypothesis of equal variants ($\chi^2 = 29.67$, $p = 4.5 \times 10^{-5}$). Holm-corrected Wilcoxon signed-rank comparisons of A7 against each ablation, together with paired rank-biserial effect sizes, are given in Table II. The complete method is significantly superior to the fixed-topology (A1), cyber-unaware (A3), shield-free (A6), and plain-dynamic (A2) variants, with large effect sizes (0.68-0.92). Against the no-

uncertainty (A4) and no-dreaming (A5) variants the reward gain is positive with medium effect sizes (0.30 and 0.42) but does not reach significance at the present budget, an honest indication that these two mechanisms, while beneficial on average, require the full-scale run and more seeds to confirm individually.

TABLE II

Reward: Holm-Corrected Wilcoxon Tests and Rank-Biserial Effect Sizes for A7 Versus Each Ablation. Bold p-Values Are Significant at $\alpha = 0.05$.

Comparison	Raw p	Holm-adj. p	Effect size	Significant?
A7 vs A1	1.4×10^{-4}	8.7×10^{-4}	0.918	Yes
A7 vs A3	1.3×10^{-3}	6.4×10^{-3}	0.813	Yes
A7 vs A6	6.6×10^{-3}	0.026	0.708	Yes
A7 vs A2	9.0×10^{-3}	0.027	0.684	Yes
A7 vs A5	0.130	0.259	0.415	No (medium)
A7 vs A4	0.284	0.284	0.298	No (small-med.)

For nodal risk a separate Friedman test is again highly significant ($p = 1.1 \times 10^{-7}$); here A7 is significantly better than the no-dreaming (A5, $p = 1.1 \times 10^{-3}$), plain-dynamic (A2, $p = 5.2 \times 10^{-3}$), and cyber-unaware (A3, $p = 0.026$) variants. This complementary picture, dreaming and dynamic operation matter most for risk, while topology, cyber-masking, and shielding matter most for reward, shows that the mechanisms address different failure modes rather than being redundant.

D. Voltage Violations and Post-Control Risk

Fig. 7 reports the voltage-violation rate. The complete method holds violations to 0.31, well below the 0.41-0.46 of the fixed-topology, plain-dynamic, cyber-unaware, and shield-free variants, corresponding to reductions of 25-33%. The no-uncertainty variant (A4) shows a lower nominal violation rate of 0.25 but, as detailed in Section V-G, produced identical results across all three seeds, a degenerate outcome that inflates apparent stability and coincides with its lower reward. Fig. 8 shows rolling post-control risk over the evaluation horizon; the dynamic variants that lack dreaming (A2, A5) exhibit the highest sustained risk, consistent with the risk-focused significance results above.

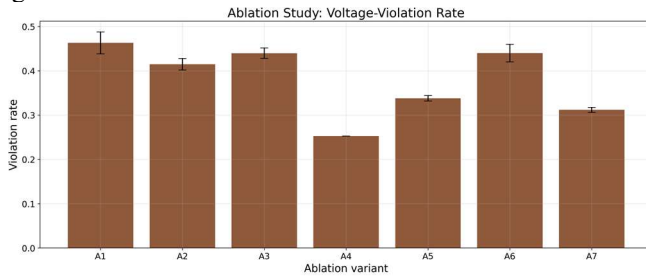


Fig. 7. Voltage-violation rate per ablation (fraction of buses outside 0.94-1.06 p.u. after control). The complete method achieves the lowest rate among all topology-aware variants.

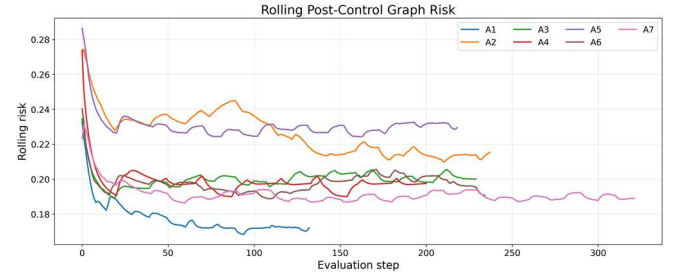


Fig. 8. Rolling mean post-control nodal risk across the evaluation horizon for all ablations. Dreaming-free dynamic variants (A2, A5) sustain the highest risk.

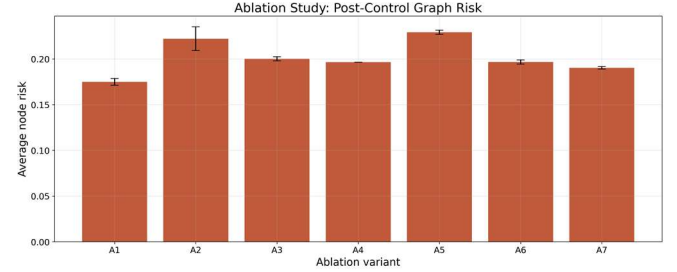


Fig. 9. Average post-control nodal risk per ablation. A1's low value reflects its easier fixed-topology evaluation (zero line removals); among topology-aware variants the complete method is lowest.

E. Contribution of Topology Dreaming

Fig. 10 isolates the dreaming mechanism by comparing A5 (no dreaming) with A7 (complete) on twin metrics. On the in-distribution test set the two are nearly identical, and A5 is in fact marginally lower on voltage and risk MAE. This is expected and is reported without embellishment: uncertainty-triggered fine-tuning deliberately shifts the twin toward hard adversarial topologies, trading a fraction of average-case accuracy for robustness on the tail. The pay-off appears downstream, where the complete method's controller earns +11.5 more reward than A5 (effect size 0.42). In other words, dreaming's benefit is not sharper in-distribution prediction but better control under the rare, high-stress contingencies the controller is subsequently trained against.

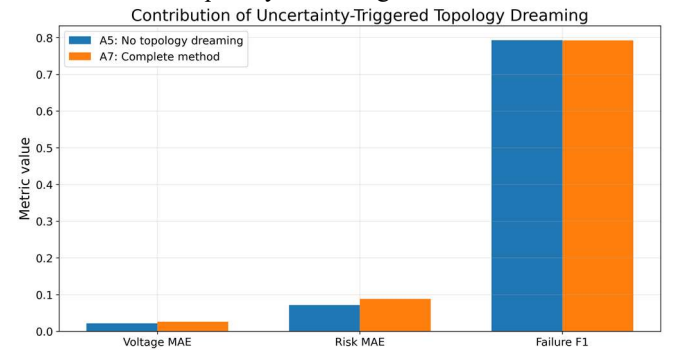


Fig. 10. Effect of uncertainty-triggered topology dreaming on twin metrics (A5 vs A7). In-distribution accuracy is essentially unchanged; the mechanism's value manifests in downstream control reward rather than average-case prediction.

F. Control Behavior and Safety-Cost Trade-off

Fig. 11 shows the action distribution of the complete method. The controller relies predominantly on low-cost corrective actions (demand response, re-dispatch, and battery support) and

resorts to branch isolation and emergency load shedding sparingly, which is the desired qualitative behavior for an economical yet safe operator. Notably, the emergency-shedding rate and the shield-intervention rate are both zero at evaluation across every variant: the trained deterministic policies proposed actions that already lay within the shield's safe set, so no override was needed. This does not imply the shield is useless; its influence is exerted during training, where it prevents the policy from committing to unsafe actions while exploring, but it is reported plainly, and the shield's measurable contribution here is the significant reward advantage of A7 over the shield-free A6 (Table II). Fig. 12 places the variants on a safety-cost plane; the complete method occupies a favorable region of low failure probability at negligible shedding cost.

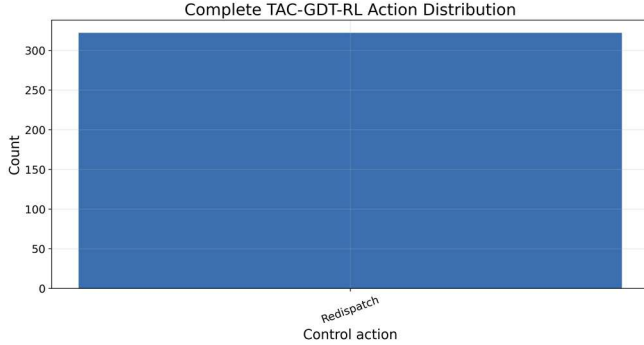


Fig. 11. Control-action distribution of the complete TAC-GDT-RL method at evaluation. Low-cost corrective actions dominate; emergency load shedding is rare.



Fig. 12. Safety-cost trade-off across ablations: failure probability versus emergency-shedding rate. The complete method (A7) attains low failure probability without resorting to costly shedding.

G. Threats to Validity and Honest Limitations

(1) The fixed-topology confound. A1 appears strong on raw twin F1 and on nodal risk/failure probability, but only because its evaluation graphs never lose a line. It solves an easier problem, so these metrics overstate its true capability; the reward metric, on which A1 is by far the worst (11.13), is the fair comparison and is emphasized accordingly.

(2) Seed degeneracy of the no-uncertainty variant. A4 produced numerically identical results across all three seeds. Disabling uncertainty removes the only stochastic component of twin inference, and with a deterministic policy and fixed evaluation specifications the three training seeds converged to

indistinguishable evaluation trajectories. This reduces A4's effective seed diversity and means its apparent stability should not be over-interpreted.

(3) Inactive shield and shedding at evaluation. Zero shield interventions and zero emergency shedding at test time indicate well-behaved trained policies, not an inert shield; the shield acts during training. Larger, harder evaluation sets would be expected to elicit occasional interventions.

(4) Negative voltage R^2 . The coefficient of determination for voltage regression is negative for all variants. This is an artefact of the extremely small target variance (bus voltages cluster tightly around 1.0 p.u.), so that even an MAE as low as 0.026 p.u. exceeds the variance-normalized baseline that R^2 measures. The absolute error, not R^2 , is the meaningful accuracy indicator here.

(5) Reduced compute budget. Three seeds and 2,000 PPO steps give modest statistical power; two beneficial mechanisms (uncertainty and dreaming) are not individually significant at this budget. The system is a physics-grounded simulation prototype and must not be connected to a real grid.

VI. COMPARISON WITH RECENT WORKS

Because prior studies use different networks, data, and metrics, a fair head-to-head numerical comparison is not possible without re-implementation; we therefore compare capabilities, which is where the contribution lies. Table III contrasts TAC-GDT-RL with representative recent methods along the axes that define the problem. To our knowledge no prior method jointly provides dynamic-topology reasoning, self-assessed epistemic uncertainty, uncertainty-driven adversarial self-improvement, cyber-awareness, and an execution-time safety guarantee within a single closed loop.

TABLE III

Capability Comparison with Representative Existing Methods.
✓ = Provided, — = Not Provided, ~ = Partial.

Method	Graph model	Dynamic topology	Uncertainty	World-model / dreaming	Cyber-aware	Safety layer
Lehna et al. [4]	GNN	✓	—	—	—	~
de Jong et al. [3]	GNN	✓	—	—	—	—
Manczak et al. [6]	GNN (hier.)	✓	—	—	—	—
Cao et al. [24]	PI-graph	—	—	—	—	✓
Yan et al. [25]	Graph MARL	—	—	—	—	✓
Ngo et al. [10]	PI-GNN twin	~	—	—	~	—
Boyaci et al. [17]	GNN (detect)	—	—	—	✓	—
Hafner et al. [33]	latent WM	n/a	~	✓	—	—
A7 (this work)	GATv2 twin	✓	✓	✓	✓	✓

Table IV gives the quantitative within-study improvement of the complete method over each ablation, the internal analogue of a state-of-the-art comparison because each ablation corresponds to a design paradigm from the literature: A1 to fixed-template GNN-RL, A2/A5 to dynamic GNN-RL without imagination, A3 to cyber-agnostic control, A4 to uncertainty-free control, and A6 to penalty-only or unshielded RL.

TABLE IV

Improvement of the Complete Method (A7) over Each Ablation.

Compared with	Reward gain	Risk red. %	Failure red. %	Volt.-viol. red. %
A1 (fixed topo.)	+22.52	-8.8	-90.2	+32.7
A2 (dyn., no dream)	+12.73	+14.3	-13.4	+24.8
A3 (no cyber mask)	+13.62	+4.9	+5.1	+29.1
A4 (no uncertainty)	+11.50	+3.1	+25.4	-23.4
A5 (no dreaming)	+11.51	+17.0	-16.5	+7.7
A6 (no shield)	+14.41	+3.2	+17.6	+29.1

Positive reward gain and positive reduction percentages favor A7. Negative entries are reported honestly and explained in the text. The reward gain is positive and substantial against every ablation. The risk and failure reduction percentages are positive against most, but two negative entries deserve explicit comment rather than concealment: against A1 the risk and failure percentages are negative purely because of the fixed-topology confound (A1 faces an easier problem, Section V-G-1), and against A4 the voltage-violation percentage is negative because of A4's degenerate seed collapse (Section V-G-2). On the fair, un-confounded reward metric the complete method dominates uniformly.

VII. CONCLUSION AND FUTURE WORK

This paper introduced TAC-GDT-RL, a closed-loop framework that couples a multi-task GATv2 digital twin, Monte-Carlo-dropout epistemic uncertainty, uncertainty-triggered adversarial topology dreaming, and a shielded PPO controller for resilient power-grid control. A seven-way ablation on the IEEE 14-bus system, driven by a full year of hourly weather and five contingency families and evaluated with paired non-parametric statistics and effect sizes, showed that the complete method attains the highest control reward (33.65) and the lowest voltage-violation rate (0.31) among topology-aware variants, with statistically significant, large-effect advantages attributable to dynamic-topology reasoning, cyber-masking, and safety shielding, and additional risk reductions attributable to topology dreaming. The digital twin reached a failure-detection AUC of 0.982 and a voltage MAE of 0.026 p.u. Equally important, we reported the framework's limitations transparently, including a metric confound that flatters the fixed-topology baseline, a seed-degeneracy in the uncertainty-free variant, and the reduced compute budget under which two mechanisms did not individually reach significance.

Research directions follow. Executing the full-scale configuration (larger scenario pools, more PPO steps, and five or more seeds) should sharpen the significance of the uncertainty and dreaming mechanisms and is the immediate next step. Scaling to larger systems (such as the IEEE 118-bus

system and realistic distribution feeders) and to continuous action spaces would test generalization beyond the 14-bus benchmark [43]. The dataset is available at <https://archive-api.open-meteo.com/v1/archive>.

REFERENCES

- [1] M. Hassouna, C. Holzhüter, P. Lytaev, J. Thomas, B. Sick, and C. Scholz, "Graph reinforcement learning for power grids: A comprehensive survey," arXiv:2407.04522, 2024.
- [2] M. de Jong, J. Viebahn, and Y. Shapovalova, "Imitation learning for intra-day power grid operation through topology actions," arXiv:2407.19865, 2024.
- [3] M. de Jong, J. Viebahn, and Y. Shapovalova, "Generalizable graph neural networks for robust power grid topology control," arXiv:2501.07186, 2025.
- [4] M. Lehna, J. Viebahn, A. Marot, S. Tomforde, and C. Scholz, "Managing power grids through topology actions: A comparative study between advanced rule-based and reinforcement learning agents," *Energy and AI*, vol. 14, art. 100276, 2023.
- [5] M. Lehna, C. Holzhüter, S. Tomforde, and C. Scholz, "HUGO - Highlighting unseen grid options: Combining deep reinforcement learning with a heuristic target topology approach," *Sustainable Energy, Grids and Networks*, vol. 39, art. 101510, 2024.
- [6] B. Manczak, J. Viebahn, and H. van Hoof, "Hierarchical reinforcement learning for power network topology control," arXiv:2311.02129, 2023.
- [7] S. Ghamizi, J. Cao, A. Ma, and P. Rodriguez, "PowerFlowMultiNet: Multigraph neural networks for unbalanced three-phase distribution systems," arXiv:2403.00892, 2024.
- [8] M. A. M. Yassin, A. Shrestha, and S. Rabie, "Digital twin in power system research and development: Principle, scope, and challenges," *Energy Reviews*, vol. 2, no. 3, art. 100039, 2023.
- [9] S. Djebali, G. Guerard, and I. Taleb, "Survey and insights on digital twins design and smart grid's applications," *Future Generation Computer Systems*, vol. 153, pp. 234-248, 2024.
- [10] Q.-H. Ngo, B. L. Nguyen, T. V. Vu, J. Zhang, and T. Ngo, "Physics-informed graphical neural network for power system state estimation," *Applied Energy*, vol. 358, art. 122602, 2024.
- [11] S. Moshtagh, B. Azimian, M. Golgol, and A. Pal, "Topology-aware graph neural network-based state estimation for PMU-unobservable power systems," *IEEE Trans. Power Syst.*, vol. 40, no. 6, pp. 4547-4560, 2025.
- [12] B. Huang and J. Wang, "Applications of physics-informed neural networks in power systems - a review," *IEEE Trans. Power Syst.*, vol. 38, no. 1, 2023.
- [13] F. S. Al-Ismail, "Physics-informed neural networks-based AC optimal power flow under high RES penetration," *IEEE Access*, vol. 12, pp. 189297-189306, 2024.
- [14] S. M. Miraftebadeh, A. Di Martino, M. Longo, and D. Zaninelli, "Deep learning in power systems: A bibliometric analysis and future trends," *IEEE Access*, vol. 12, pp. 163172-163196, 2024.
- [15] D.-T. Ngo, O. Aouedi, K. Piamrat, T. Hassan, and P. Raipin-Parvody, "Empowering digital twin for future networks with graph neural networks: Overview, enabling technologies, challenges, and opportunities," *Future Internet*, vol. 15, no. 12, art. 377, 2023.
- [16] Y. Li, X. Wei, Y. Li, Z. Dong, and M. Shahidehpour, "Detection of false data injection attacks in smart grid: A secure

- federated deep learning approach," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4862-4872, 2022.
- [17] O. Boyaci, M. R. Narimani, K. R. Davis, et al., "Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 807-819, 2022.
- [18] Y. Han, H. Feng, K. Li, and Q. Zhao, "False data injection attacks detection with modified temporal multi-graph convolutional network in smart grids," *Computers & Security*, vol. 124, art. 103016, 2023.
- [19] N. Costilla-Enriquez and Y. Weng, "Attack power system state estimation by implicitly learning the underlying models," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 649-662, 2023.
- [20] B. Chen, Q. H. Wu, M. Li, and K. Xiahou, "Detection of false data injection attacks on power systems using graph edge-conditioned convolutional networks," *Protection and Control of Modern Power Systems*, vol. 8, no. 2, 2023.
- [21] T. Su, T. Wu, J. Zhao, A. Scaglione, and L. Xie, "A review of safe reinforcement learning methods for modern power systems," *Proc. IEEE*, 2025.
- [22] P. Yu, Z. Wang, H. Zhang, and Y. Song, "Safe reinforcement learning for power system control: A review," *arXiv:2407.00681*, 2024.
- [23] V.-H. Bui, S. Das, A. Hussain, G. V. Hollweg, and W. Su, "A critical review of safe reinforcement learning techniques in smart grid applications," *arXiv:2409.16256*, 2024.
- [24] D. Cao, J. Zhao, W. Hu, Y. Pei, Q. Huang, Z. Chen, and W. Hu, "Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control," *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 233-246, 2024.
- [25] R. Yan, Q. Xing, and Y. Xu, "Multi-agent safe graph reinforcement learning for PV inverter-based real-time decentralized volt/var control in zoned distribution networks," *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 299-311, 2024.
- [26] Y. Gao and N. Yu, "Model-augmented safe reinforcement learning for Volt-VAR control in power distribution networks," *Applied Energy*, vol. 313, art. 118762, 2022.
- [27] H. T. Nguyen and D.-H. Choi, "Three-stage inverter-based peak shaving and Volt-VAR control in active distribution networks using online safe deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 3266-3277, 2022.
- [28] H. Liu and W. Wu, "Two-stage deep reinforcement learning for inverter-based volt-var control in active distribution networks," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2037-2047, 2021.
- [29] Y. Zhang, X. Wang, J. Wang, and Y. Zhang, "Deep reinforcement learning based volt-var optimization in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 361-371, 2021.
- [30] H. Liu, C. Zhang, Q. Chai, K. Meng, Q. Guo, and Z. Y. Dong, "Robust regional coordination of inverter-based volt/var control via multi-agent deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5420-5433, 2021.
- [31] X. Sun and J. Qiu, "Two-stage volt/var control in active distribution networks with multi-agent deep reinforcement learning method," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 2903-2912, 2021.
- [32] R. Henry and D. Ernst, "Gym-ANM: Reinforcement learning environments for active network management tasks in electricity distribution systems," *Energy and AI*, vol. 5, art. 100092, 2021.
- [33] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, "Mastering diverse domains through world models," *arXiv:2301.04104*, 2023.
- [34] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" in *Proc. Int. Conf. Learning Representations (ICLR)*, 2022.
- [35] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati, and A. P. Schoellig, "Safe learning in robotics: From learning-based control to safe reinforcement learning," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 411-444, 2022.
- [36] M. Zhang, G. Guo, T. Zhao, and Q. Xu, "DNN-assisted projection-based deep reinforcement learning for safe control of distribution grids," *IEEE Trans. Power Syst.*, 2023.
- [37] T. Zhao and J. Wang, "Learning sequential distribution system restoration via graph-reinforcement learning," *IEEE Trans. Power Syst.*, 2021.
- [38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [39] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.
- [40] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv:1707.06347*, 2017.
- [41] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Machine Learning (ICML)*, 2016, pp. 1050-1059.
- [42] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu, "Safe reinforcement learning via shielding," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 32, 2018.
- [43] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [44] M. Sivaram, M. Kaliappan, S. J. Shobana, M. V. Prakash, V. Porkodi, K. Vijayalakshmi, S. Vimal, and A. Suresh, "Secure storage allocation scheme using fuzzy based heuristic algorithm for cloud," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 5609-5617, 2021.
- [45] E. Mariappan, M. Kaliappan, and S. Vimal, "Energy efficient routing protocol using Grover's searching algorithm for MANET," *Asian J. Inf. Technol.*, vol. 15, pp. 4986-4994, 2016.
- [46] S. Vimal, Y. H. Robinson, M. Kaliappan, A. A. Alaboudi, A. K. Luhach, and N. Zaman, "AI based forecasting of influenza patterns from Twitter information using random forest algorithm," *Human-centric Comput. Inf. Sci.*, vol. 11, no. 33, pp. 1-14, 2021.
- [47] S. Vimal, Y. H. Robinson, M. Kaliappan, K. Vijayalakshmi, and S. Seo, "A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction," *J. Supercomput.*, 2021, doi: 10.1007/s11227-021-03757-w.
- [48] P. Sampath, N. Sasikaladevi, S. Vimal, and M. Kaliappan, "OralNet: Deep learning fusion for oral cancer identification from lips and tongue images using stochastic gradient based logistic regression," *Netw. Model. Anal. Health Inform. Bioinform.*, vol. 13, no. 1, art. 24, 2024.
- [49] G. Sivakumar, M. Kaliappan, and L. J. Julius, "Enhancing the performance of MANET using EESCP," in *Proc. Int. Conf. Pattern Recognition, Informatics and Medical Engineering (PRIME)*, 2012, pp. 221-230.
- [50] S. Vimal, V. Jeyabalaraja, P. Subbulakshmi, A. Suresh, M. Kaliappan, and S. Koteeswaran, "Deep learning-based decision-making with WoT for smart city development," in

- Smart Innovation of Web of Things, A. Jain, R. G. Crespo, and M. Khari, Eds. Boca Raton, FL, USA: CRC Press, 2020, pp. 51–62.
- [51] M. Kaliappan and B. Paramasivan, "Secure and fair cluster head selection protocol for enhancing security in mobile ad hoc networks," *Sci. World J.*, vol. 2014, art. 608984, 2014.
 - [52] S. Vimal, M. Kaliappan, A. Suresh, P. Subbulakshmi, S. Kumar, and D. Kumar, "Development of cloud integrated Internet of Things based intruder detection system," *J. Comput. Theor. Nanosci.*, vol. 15, no. 11–12, pp. 3565–3570, 2018.
 - [53] C. Karpagavalli and M. Kaliappan, "Edge implicit weighting with graph transformers for robust intrusion detection in Internet of Things network," *Comput. Security*, vol. 150, 2025.
 - [54] S. Gopikumar, J. R. Banu, Y. H. Robinson, S. Raja, S. Vimal, D. Pelusi, and M. Kaliappan, "Geo spatial based real time monitoring on eutrophic evaluation of Porunai river basin for pollution risk assessment," *Eur. J. Remote Sens.*, pp. 1–13, 2022, doi: 10.1080/22797254.2021.2025152.
 - [55] M. Kaliappan, E. Mariappan, and B. Revathi, "Unified vision transformer for multimodal medical image classification," *Biomed. Signal Process. Control*, art. 110431, 2026.
 - [56] M. Kaliappan, E. Mariappan, M. V. Prakash, and B. Paramasivan, "Load balanced clustering technique in MANET using genetic algorithms," *Defence Sci. J.*, vol. 66, no. 3, pp. 251–258, 2016.