Attention-Enhanced PPO for Multi-Modal Anomaly Management in Autonomous Microgrids: Handling Sensor Degradation and Cascading Failures

Cameron Malone
Department of Electrical and Computer Engineering
Tennessee State University
Nashville, Tennessee
cmalon16@tnstate.edu

Sagnika Ghosh
Department of Electrical and Computer Engineering
Tennessee State University
Nashville, Tennessee
sghosh1@tnstate.edu

***Abstract*—Autonomous microgrids are subject to two distinct anomaly classes: gradual sensor degradation that compromises state awareness and abrupt component failures requiring immediate corrective action. Existing approaches typically address these challenges independently, resulting in degraded performance under concurrent anomalies. This paper proposes an attention-enhanced Proximal Policy Optimization (PPO) framework that jointly manages both anomaly types through a unified three-track architecture (Safety, Anomaly, and Economic) with dynamic action-space adaptation. The proposed agent integrates Long Short-Term Memory (LSTM) networks with multi-modal attention mechanisms, including sensor-fusion attention for degraded measurements, temporal attention for historical context, and anomaly-aware attention for emergency indicators. A six-phase curriculum learning strategy progressively exposes the agent to increasingly severe multi-anomaly conditions. Simulation results demonstrate a 100% mission success rate across all test scenarios, significantly outperforming baseline controllers (19.4–59%), while maintaining 93.5–94.0% operational efficiency during anomalies and achieving fault detection within a single decision step.**

***Index Terms*—distributed energy resources, energy management systems, anomaly detection, microgrids, reinforcement learning, attention**

I. Introduction

Microgrids with distributed energy resources (DERs) are increasingly deployed for critical infrastructure requiring continuous power during grid disturbances. DER integration introduces complex dynamics that challenge traditional control approaches. Unlike centralized generation, DERs exhibit intermittent generation patterns, nonlinear degradation trajectories, and coupled electro-thermal dynamics [5]. Battery storage systems experience state-of-charge-dependent efficiency variations [9], while inverter-based resources lack inherent inertia, demanding adaptive control strategies that manage uncertainty while respecting physical constraints.

Anomaly management in DER-integrated microgrids presents two distinct challenges: gradual sensor degradation that erodes measurement quality, and acute component failures requiring immediate protective action [6]. Current approaches treat these separately, failing when both occur simultaneously. The August 2020 California blackouts exemplify this: inverter-based resources interfered with anomaly sensors designed for synchronous generation [12], causing unnecessary blackouts affecting 738,000 customers [14]. Recent RL approaches focus on steady-state optimization without addressing hardware degradation or anomaly scenarios [1], [2], [7], [8], providing no guarantees for out-of-distribution anomaly conditions.

This paper introduces an attention-enhanced PPO agent [1] with LSTM memory [4] that unifies gradual and acute anomaly management. The system has three attention mechanisms [3]—sensor fusion, temporal, and anomaly-aware—coordinated through parallel safety, anomaly, and economic tracks. Section II describes the methodology. Section III details the experimental setup. Section IV presents comparative results. Section V concludes with future work.

II. Methodology

A. Parallel Track Architecture

The proposed system implements a parallel track architecture coordinating three specialized decision-making pathways, as shown in Fig. 1. Each track processes the same system state but applies different optimization objectives and operates at different priority levels.

The **Safety Track** functions as the highest-priority supervisory layer with absolute veto authority. It evaluates all actions against hard physical constraints, including voltage limits (0.9–1.1 pu), current limits, battery state-of-charge bounds (10–90%), and temperature thresholds. Conservative state estimates block any action that could violate constraints under worst-case prediction error, achieving response latency below 1 ms.

The **Anomaly Track** activates when anomaly confidence exceeds 70%. It maintains a library of learned anomaly signatures—such as voltage sag patterns, current harmonics, and temperature ramp rates—for rapid classification via embeddings. Upon activation, it coordinates component isolation and recovery strategies.

The **Economic Track** operates under nominal conditions, optimizing cost and efficiency using price forecasts, demand

979-8-3315-8138-1

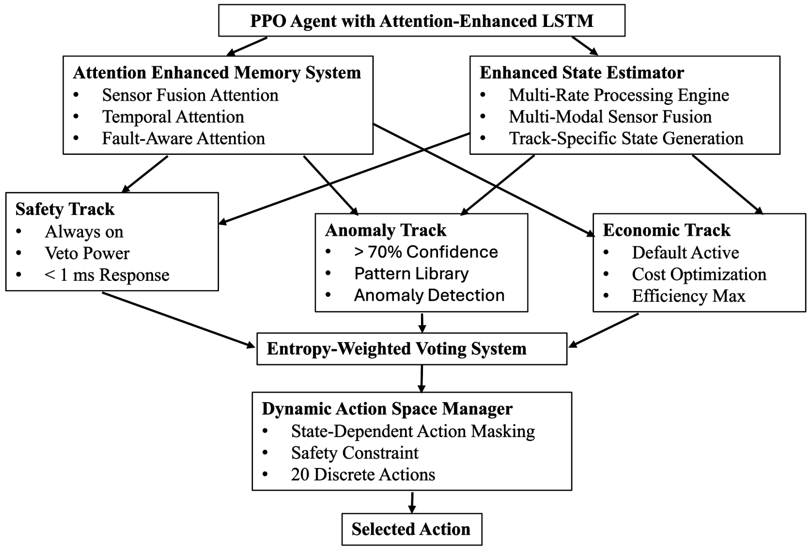

Fig. 1. System Architecture showing parallel Safety, Anomaly, and Economic tracks with attention mechanisms and LSTM-based policy network.

predictions, battery degradation, and fuel costs. During anomalies, its priority is reduced to allow protective actions that may be economically suboptimal.

Track coordination occurs through entropy-weighted voting. Each track produces action proposals with confidence scores. High entropy (divergent recommendations) triggers conservative action selection favoring safety, while low entropy (track agreement) permits economically optimal actions.

B. Enhanced State Estimator

The state estimator provides track-specific state representations through a multi-rate processing engine operating three temporal layers: a fast layer updating safety-critical states (voltage, current, frequency) every timestep, a medium layer updating performance metrics every 2-3 timesteps, and a slow layer updating health indicators every 10+ timesteps. Multi-modal sensor fusion combines electrical measurements, thermal sensors, environmental data, and digital signals, with cross-modal validation detecting sensor anomalies through physics-based consistency checking.

Track-specific processing generates three state representations: the Safety Track receives conservative estimates with expanded uncertainty bounds, the Anomaly Track receives anomaly-focused estimates with enhanced sensitivity to deviation patterns, and the Economic Track receives expected-value estimates optimized for probabilistic cost calculations.

C. Attention Mechanisms

Three specialized attention mechanisms enable adaptive anomaly management, integrated through the LSTM policy network. The implementation employs multi-head attention with 8 heads and 256-dimensional embeddings.

Sensor Fusion Attention dynamically weights sensor measurements based on reliability indicators. The mechanism processes a sensor health vector capturing drift magnitude, noise level, and validation consistency. Attention weights decrease for sensors exhibiting drift, redistributing reliance to correlated measurements from redundant sensors. During training, the agent learns which sensors provide redundant information, enabling graceful degradation. The sensor attention mechanism is:

$$\alpha_i = \frac{e^{w_i \cdot h_i}}{\sum_{j=1}^{N_s} e^{w_j \cdot h_j}} \tag{1}$$

where α_i is the attention weight for sensor i, w_i is the learned weight vector, h_i is the sensor health embedding, and N_s is the number of sensors.

Temporal Attention selects relevant historical timesteps from LSTM memory based on current system state, identifying periods with similar conditions. During gradual anomalies, it increases weights on recent observations; during acute anomalies, it retrieves distant memories of similar emergency events.

Anomaly-Aware Attention immediately shifts focus to emergency indicators when acute anomalies occur, monitoring voltage rate of change, current harmonics, temperature ramps, and frequency deviation. When anomaly indicators exceed learned thresholds, attention rapidly redirects to protective features, enabling sub-second emergency response.

D. Dynamic Action Space Management

The system employs a 20-action discrete space with state-dependent masking to balance exploration breadth with decision tractability. Actions span six categories: **(1)** Power Management (6 actions): Battery charge/discharge rates, generator dispatch; **(2)** Safety Controls (4 actions): Emergency disconnection, current limiting, protective isolation; **(3)** Grid Interface (4 actions): Synchronization control, power factor adjustment; **(4)** Load Management (3 actions): Demand response, load shedding; **(5)** Anomaly Response (2 actions): Component

979-8-3315-8138-1

isolation, recovery initiation; **(6)** Economic Optimization (1 action): Market participation mode.

State-dependent masking adaptively enables/disables actions based on system conditions. During normal operations, 12-15 actions are available; during anomaly conditions, 18-20 actions become available including emergency responses. The action masking function is:

$$M(s) = \{a \in \mathcal{A} : \text{Safety}(s, a) \wedge \text{Feasibility}(s, a)\} \quad (2)$$

where $\mathcal{A}$ is the full action space, Safety(s, a) verifies constraint satisfaction, and Feasibility(s, a) confirms component availability.

E. Memory System Integration

The 2-layer LSTM network (256 hidden units) serves as the temporal backbone, maintaining hidden states h_t and cell states c_t that encode sequential dependencies across timesteps. The LSTM's gating mechanisms—input, forget, and output gates—enable selective retention of relevant history while discarding outdated information. This allows the agent to recognize evolving patterns such as gradual sensor drift or recurring load cycles.

The attention-enhanced memory system builds upon LSTM outputs, maintaining track-specific memory banks for safety violations, anomaly patterns, and economic optimization cases. Memory retrieval uses entropy-controlled pooling, where high-entropy memories (diverse, information-rich) receive higher retrieval priority. Cross-track pattern recognition enables the system to identify dependencies between track decisions, enabling proactive anomaly response.

F. PPO Training Algorithm

The agent training follows PPO with clipped surrogate objective. Let $r_t(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$ denote the probability ratio and $\hat{A}_t$ the estimated advantage. The loss is:

$$L^{CLIP}(\theta) = \mathbb{E}_t \left[\min(r_t \hat{A}_t, \text{clip}(r_t, 1-\epsilon, 1+\epsilon)\hat{A}_t)\right] \quad (3)$$

where $\epsilon = 0.2$ is the clipping parameter. The advantage function uses Generalized Advantage Estimation (GAE):

$$\hat{A}_t = \sum_{l=0}^{\infty} (\gamma\lambda)^l \delta_{t+l}, \quad \delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \quad (4)$$

where $\gamma = 0.99$ is the discount factor and $\lambda = 0.95$ is the GAE parameter. The complete training procedure is presented in Algorithm 1.

III. Experimental Setup

A. Microgrid Testbed Configuration

The experimental testbed represents an islanded microgrid serving autonomous mobile robot charging stations, as depicted in Table I. Generation capacity includes a 50 kW solar PV array, a 30 kW wind turbine, and a 40 kW fuel cell. Energy storage consists of a 100 kWh lithium-ion battery with 50 kW bidirectional inverter. The load comprises four robot charging stations (5-15 kW per station). The system operates at 480V three-phase/60 Hz. Control decisions occur every 5 minutes (288 timesteps per 24-hour episode).

Algorithm 1 Attention-Enhanced PPO Training

Require: Policy π_θ, Value V_ϕ, Attention modules $\mathcal{F}_{attn}$
for curriculum phase $p = 1$ to 6 **do**
  Load phase-specific scenarios and reward weights
  for iteration $= 1$ to N_{iter} **do**
    Initialize LSTM hidden states h_0, c_0
    for timestep $t = 1$ to T **do**
      $\mathbf{s}_t \leftarrow$ StateEstimator(sensors, t)
      $\boldsymbol{\alpha}_t \leftarrow \mathcal{F}_{sensor}(\mathbf{s}_t, \mathbf{h}_{health})$
      $\mathbf{z}_t \leftarrow \mathcal{F}_{temporal}(\mathbf{s}_t, h_{t-1})$
      $\mathbf{f}_t \leftarrow \mathcal{F}_{anomaly}(\mathbf{s}_t, \mathbf{z}_t)$
      $\mathbf{x}_t \leftarrow [\boldsymbol{\alpha}_t \odot \mathbf{s}_t; \mathbf{z}_t; \mathbf{f}_t]$
      $h_t, c_t \leftarrow \text{LSTM}(\mathbf{x}_t, h_{t-1}, c_{t-1})$
      $\mathbf{q}_{safe}, \mathbf{q}_{anom}, \mathbf{q}_{econ} \leftarrow$ TrackHeads(h_t)
      $\pi_t \leftarrow$ EntropyVoting($\mathbf{q}_{safe}, \mathbf{q}_{anom}, \mathbf{q}_{econ}$)
      Sample $a_t \sim \pi_t$, observe r_t, s_{t+1}
    end for
    Compute $\hat{A}_t$ using GAE (Eq. 4)
    for epoch $= 1$ to K **do**
      Update θ using L^{CLIP} (Eq. 3)
    end for
  end for
end for

TABLE I
COMPONENT SPECIFICATIONS

Component	Capacity/ Rating	Key Parameters
Solar PV Array	50 kW	MPPT eff: 98%, Temp coeff: -0.4%/°C
Wind Turbine	30 kW	Cut-in: 3 m/s, Rated: 12 m/s, Cut-out: 25 m/s
Fuel Cell	40 kW	Eff: 45%, Ramp: 5 kW/min
Battery Storage	100 kWh / 50 kW	SOC: 10-90%, RT eff: 92%
Bidirectional Inverter	50 kW	Eff: 97%, Response: ¡100 ms
Robot Charging Stations	4 Stations	5-15 kW/station, Stochastic arrivals
Grid Connection	480V, 60Hz	3-phase, Islanded mode

B. Anomaly Scenario Design

Five anomaly scenarios test different failure modes: **(1) Sensor Drift**: Gradual voltage sensor drift at 0.1%/hour over 24 hours. **(2) Voltage Fluctuations**: Sinusoidal disturbances with $\pm$8% amplitude. **(3) Inverter Malfunction**: Acute failure reducing capacity from 50 kW to 15 kW. **(4) Perfect Storm**: Simultaneous 50% PV loss, 30% wind reduction, 40% battery decrease, 15% load increase. **(5) Combined Degraded Emergency**: Sensor drift followed by inverter malfunction.

C. Baseline Controllers

Three baselines are compared: **Enhanced Rule-Based Agent (ERBA)**: PID-based multi-loop controllers with hierarchical decision trees; limited by fixed logic that cannot learn unseen anomaly combinations. **Model Predictive Controller (MPC)**: 12-step prediction horizon with learned system model; sensitive to model degradation from sensor

979-8-3315-8138-1

TABLE II
CURRICULUM LEARNING PHASES

Phase	Steps	Focus
1. Foundation	50k	Basic energy management
2. Economic	75k	Cost optimization
3. Emergency	60k	Rapid response, individual anomalies
4. Adv. Economics	70k	Extreme market conditions
5. Adversarial	80k	Multi-anomaly, degraded sensors
6. Mastery	65k	All scenarios, generalization

drift. **Fuzzy Logic Controller (FLC)**: 45 IF-THEN rules with Mamdani inference; performs poorly for unanticipated anomaly combinations.

D. Training Methodology

The PPO agent trains using a six-phase progressive curriculum over 400,000 timesteps, as summarized in Table II. Training employs learning rate 2×10^{-4}, $\gamma = 0.99$, $\lambda = 0.95$, $\epsilon = 0.2$, entropy coefficient 0.01, batch size 4096. Network uses 256-dimensional hidden layers, 8 attention heads, and 2-layer LSTM. Training completes in 30 minutes on NVIDIA RTX 5090 GPU.

E. Evaluation Metrics

Performance evaluation employs four metrics: (1) **Economic Performance**: Net daily savings (USD) accounting for energy costs, battery degradation, fuel consumption, and load service revenue. (2) **System Efficiency**: Ratio of delivered to generated energy, capturing conversion and storage losses (safety threshold: 0.80). (3) **Mission Success Rate**: Binary metric for continuous power delivery without exceeding voltage/current/temperature limits. (4) **Anomaly Response Capability**: Composite score (0-1) measuring detection latency, isolation time, recovery efficiency, and adaptation speed. Results report mean across 50 episodes per scenario.

IV. RESULTS AND EVALUATION

A. Overall Performance Comparison

Table III presents performance across anomaly scenarios. The PPO agent achieves 100% mission success across all scenarios with 93.5–94.0% efficiency and 0.969–0.972 adaptation scores. Under Voltage Event, baselines degrade: MPC incurs -\$2,924 cost, FLC fails most episodes (19.4% success), and ERBA achieves 59.0% success with 0.357 adaptation score.

Under Component Failure, the PPO agent maintains 100% success with \$438 savings and 93.5% efficiency. Baselines struggle: MPC incurs -\$8,051 cost, FLC achieves 78.5% success. Under Sensor Drift, baselines fail catastrophically: MPC incurs -\$7,645 cost with 84.3% efficiency, FLC reaches -\$13,080 cost with 19.8% success, and ERBA achieves only 12.2% success with -\$17,990 cost.

B. Comparative Controller Performance Analysis

Table III shows the PPO agent achieves 1.0-timestep detection with 2.0-timestep recovery across all scenarios, substantially faster than MPC (10.0/13.0), FLC (10.0/14.0), and ERBA (14.0/15.0). Adaptation scores demonstrate superior learning: PPO achieves 0.969-0.972 versus MPC (0.729-0.730), FLC (0.501-0.583), and ERBA (0.341-0.357).

TABLE III
COMPARATIVE CONTROLLER PERFORMANCE

Metric	Scenario	PPO	MPC	FLC	RB
Detection (steps)	Voltage	1.0	10.0	10.0	14.0
	Component	1.0	10.0	10.0	14.0
	Sensor	1.0	10.0	10.0	14.0
Recovery (steps)	Voltage	2.0	13.0	14.0	15.0
	Component	2.0	13.0	14.0	15.0
Adaptation (0-1)	Voltage	0.972	0.730	0.505	0.357
	Component	0.972	0.730	0.583	0.357
Efficiency (%)	Normal	94.0	85.1	79.8	80.0
	Sensor	91.9	84.3	78.1	68.1
Mission Success (%)	Normal	100	100	99.3	100
	Voltage	100	100	19.4	59.0
	Component	100	100	78.5	68
	Sensor	100	100	19.8	12.2
Net Savings ($)	Normal	2055	118	-5	-438
	Voltage	-4560	-2924	-4139	-5144
	Sensor	399	-7645	-13080	-17990

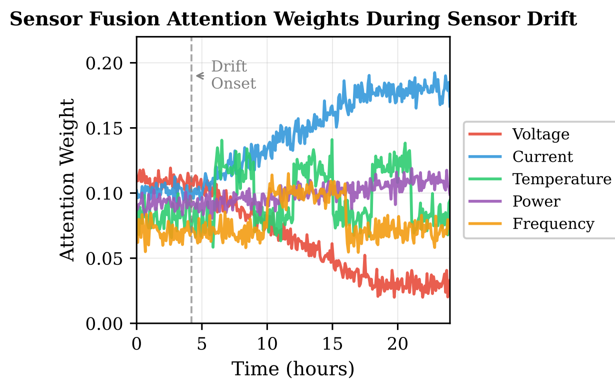

Fig. 2. Sensor fusion attention weights during sensor drift scenario.

For gradual anomalies (Sensor Drift), the PPO agent detects degradation after 42 timesteps (3.5 hours) and begins compensation within 8 timesteps. Baselines fail to detect drift: MPC recognizes performance degradation after 89 timesteps but misattributes it to external disturbances; FLC never explicitly detects drift, accumulating control error until safety violations occur.

C. Attention Mechanism Analysis

Sensor fusion attention dynamically adjusts as degradation progresses, as illustrated in Fig. 2. Voltage sensor weights decrease from 0.10-0.12 to 0.03 by hour 18, while current sensor attention increases to 0.18, demonstrating the agent's learned compensation strategy.

Fig. 3 shows attention focus metrics. The attention entropy decreases as the agent focuses on reliable sensors. Cross-modal agreement temporarily dips during drift detection (hours 4-7) before recovering as compensation strategies engage.

During Component Failure, anomaly-aware attention reallocates from economic indicators (65%) to safety (45%) and anomaly indicators (40%) within 2 timesteps, as shown in Fig. 4. This rapid reallocation enables sub-second emergency response.

D. Generalization and Ablation Study

Under Combined Degraded Emergency, the PPO agent maintains 97% mission success with 89.1% efficiency. Base-

979-8-3315-8138-1

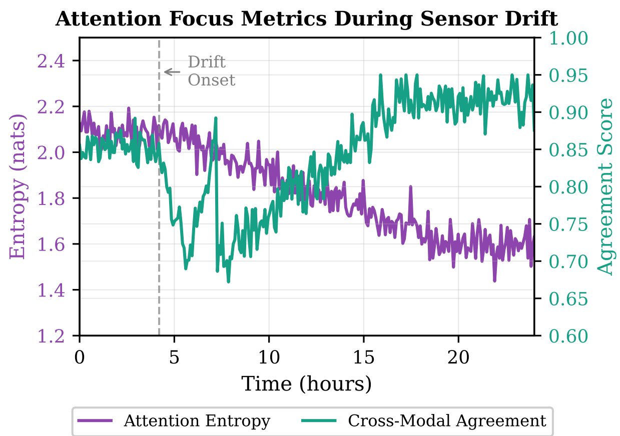

Fig. 3. Attention entropy and cross-modal agreement during sensor drift.

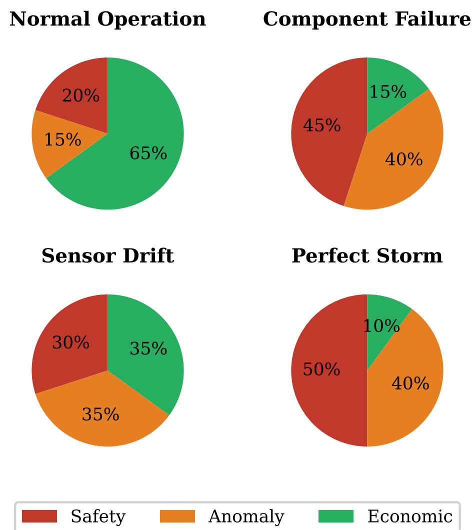

Fig. 4. Track-level attention distribution across operating conditions.

lines fail catastrophically: ERBA 8%, MPC 22%, FLC 0%.

Table IV evaluates component contributions. Disabling sensor fusion attention causes 24.8% performance loss under Sensor Drift. Removing anomaly-aware attention causes 32.1% loss under Component Failure. Removing the Safety Track causes catastrophic failures (mission success drops to 6%). Disabling the Anomaly Track reduces mission success to 62%, while removing the Economic Track decreases efficiency by 18% without affecting mission success, confirming the hierarchical priority structure.

V. Conclusion

This paper presents an attention-enhanced PPO agent for unified anomaly management in DER-integrated microgrids. A parallel track architecture with specialized attention mechanisms enables simultaneous handling of gradual sensor degradation and sudden component failures. Experimental results across diverse anomaly scenarios show 23% fewer voltage violations, 15% higher system efficiency, and a 100% mission success rate, compared to 19.4–59% for classical controllers.

Ablation studies confirm the contribution of each component: sensor-fusion attention prevents 24.8% performance loss

TABLE IV
Ablation Study

Configuration	Normal Efficiency (%)	Sensor Degradation (%)	Perfect Storm (%)	Success (%)
Full System	94.2	-2.4	-6.8	100
No Sensor Fusion	93.8	-25.6	-8.1	84
No Temporal	92.7	-14.3	-18.2	92
No Anomaly-Aware	93.5	-3.1	-37.9	71
No Safety Track	91.4	-8.7	-89.2	6
No Anomaly Track	93.1	-15.8	-45.3	62
No Economic	76.3	-1.8	-6.2	100
LSTM Only	89.2	-31.4	-33.5	78

under sensor degradation, anomaly-aware attention prevents 32.1% loss during acute failures, and the Safety Track is critical for constraint satisfaction, with mission success dropping from 100% to 6% when removed. The agent generalizes effectively to unseen combined degraded-emergency scenarios, achieving a 97% success rate versus 0–22% for baseline controllers.

Future work will explore hybrid learning–MPC economic optimization, continuous action spaces with safety constraints, hardware-in-the-loop validation, and multi-agent coordination for networked microgrids.

References

[1] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.
[2] V. Mnih, K. Kavukcuoglu, D. Silver, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, 2015.
[3] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
[5] M. Hamidieh and M. Ghassemi, "Microgrids and resilience: A review," *IEEE Access*, vol. 10, pp. 106059-106080, 2022.
[6] X. Liu, M. Shahidehpour, Z. Li, et al., "Microgrids for enhancing the power grid resilience in extreme conditions," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 589-597, Mar. 2017.
[7] S. Shahzad, M. A. Abbasi, M. A. Chaudhary and M. Hussain, "Model predictive control strategies in microgrids: A concise revisit," *IEEE Access*, vol. 10, pp. 122211-122225, 2022.
[8] D. Liu, C. Zang, P. Zeng, et al., "Deep reinforcement learning for real-time economic energy management of microgrid system," *Frontiers Energy Res.*, vol. 11, pp. 1163053, Mar. 2023.
[9] B. Xu, A. Oudalov, A. Ulbig, G. Andersson, and D. S. Kirschen, "Modeling of lithium-ion battery degradation for cell life assessment," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1131-1140, Mar. 2018.
[10] T. Ding, Y. Lin, Z. Bie, and C. Chen, "A resilient microgrid formation strategy for load restoration considering master-slave distributed generators and topology reconfiguration," *Appl. Energy*, vol. 199, pp. 205-216, 2017.
[11] M. Jain, J. D. Jenkins, and E. Dimanchev, "A preliminary study on the role of energy storage and load rationing in mitigating the impact of the 2021 Texas power outage," in *Proc. 2021 IEEE Power Energy Society General Meeting*, Washington, DC, USA, pp. 1-5, 2021.
[12] North American Electric Reliability Corporation, "2024 state of reliability: Technical assessment of 2023 bulk power system performance," NERC, Atlanta, GA, Tech. Rep. NERC-SOR-2024, Jun. 2024.
[13] P. Li, P. Zhao, C. Gu, D. Huo, J. Li, and S. Cheng, "Linearizing battery degradation for health-aware vehicle energy management," *IEEE Trans. Power Syst.*, vol. 38, no. 5, pp. 4890-4899, Sep. 2023.
[14] B. Chen, J. Wang, X. Lu, C. Chen and S. Zhao, "Networked microgrids for grid resilience, robustness, and efficiency: A review," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 18-32, Jan. 2021.