S J S T
sjst.scst.edu.ly
Surman Journal for Science and Technology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol **8**, No.**2.** Jun – Dec. 2026
Pages: 018 ~ 040

Multi-Agent Reinforcement Learning for Self-Healing Distribution Networks under Power Quality Constraints

Abdulbari Ali Mohamed Frei [1], LTFEI AHMED .A. ABDALSMD [2]

[1] Higher Institute of Sciences and Technology Msallatah, Libya
[2] Faculty of Arts and Sciences, Elmergib University, Msallata, Libya

freiabd@gmail.com[1] , Laabdalsmd@elmergib.edu.ly[2]

الملخص

يجب على شبكات التوزيع استعادة الأحمال السليمة بعد الأعطال دون انتهاك قيود الجهد الحراري أو الطوبولوجيا الشعاعية. تقترح هذه الورقة إطارًا محسنًا لتعلم التعزيز متعدد الوكلاء (MARL) لشبكات التوزيع ذاتية الشفاء تحت قيود جودة الطاقة. تم نمذجة نظام التوزيع الشعاعي IEEE 33-bus في Python باستخدام pandapower، بينما تم تنفيذ نموذج التعلم في PyTorch. تمت صياغة عزل الأعطال واستعادة الخدمة كقرارات تبديل منسقة عبر خمسة مفاتيح ربط مفتوحة عادةً. تم تقديم آلية إصلاح تحافظ على الشعاعية لمنع تكوينات ما بعد الاستعادة الحلقية، وتعمل دالة المكافأة الواعية بجودة الطاقة على إعطاء الأولوية للحمل المستعاد مع معاقبة انتهاكات الجهد، والحمل الزائد على الخطوط، وفقدان الطاقة، والتبديل المفرط، وعدم الشعاعية، وفشل تدفق الطاقة. تم حساب جميع مجموعات استعادة الأعطال الممكنة مسبقًا باستخدام pandapower واستخدامها كذاكرة تخزين مؤقت أثناء التدريب. عبر خمسة سيناريوهات أعطال، استعاد تعلم التعزيز متعدد الوكلاء المحسن جميع الأحمال القابلة للاسترداد، وحافظ على الطوبولوجيا الشعاعية، ولم ينتج عنه أي انتهاكات للجهد أو أحمال زائدة على الخطوط، وحقق متوسط حمل مستعاد بنسبة 82.48%، متطابقًا مع معيار البحث الشعاعي الشامل. اختلف متوسط المكافأة بشكل طفيف فقط عن البحث الشامل (404.31 مقابل 404.53)، بينما كانت الخسائر أقل من خط الأساس DQN أحادي الوكيل.

الكلمات المفتاحية: تعلم التعزيز متعدد الوكلاء؛ شبكات التوزيع ذاتية الشفاء؛ استعادة الخدمة؛ جودة الطاقة؛ إعادة تشكيل شبكة التوزيع؛ pandapower؛ PyTorch.

Abstract

Distribution networks must restore healthy loads after faults without violating voltage, thermal, or radial topology constraints. This paper proposes an enhanced multi-agent reinforcement learning (MARL) framework for self-healing distribution networks under power quality constraints. The IEEE 33-bus radial distribution system is modeled in Python using pandapower, while the learning model is implemented in PyTorch. Fault isolation and service restoration are formulated as coordinated switching decisions over five normally open tie switches. A radiality-preserving repair mechanism is introduced to prevent looped post-restoration configurations, and a power-quality-aware reward function prioritizes restored load while penalizing voltage violations, line overloads, power losses, excessive switching, non-radiality, and power-flow failure. All feasible fault-restoration combinations are precomputed using pandapower and used as a cache during training. Across five fault scenarios, the enhanced MARL restored all recoverable loads, maintained radial topology, produced zero voltage violations and zero-line overloads, and achieved an average restored load of 82.48%, matching the exhaustive radial search benchmark. The average reward differed only slightly from exhaustive search (404.31 versus 404.53), while average losses were lower than the single-agent DQN baseline.

Keywords: *multi-agent reinforcement learning; self-healing distribution networks; service restoration; power quality; distribution network reconfiguration; pandapower; PyTorch.*

Submitted: *10/05/2026* ***Accepted****: 14/06/2026*

1. Introduction

Electrical distribution networks are being pushed toward more automated and flexible operation because they are no longer passive radial feeders serving predictable loads. Active distribution systems increasingly include distributed energy resources, controllable switches, variable demand, and monitoring devices, all of which make post-fault operation more data-driven and time-critical. Recent work on self-healing distribution grids emphasizes that modern outage management must combine fault isolation, network reconfiguration, and service restoration rather than treating restoration as a manual switching exercise (Jo et al., 2024; Jacob et al., 2024; Meteab et al., 2025). The practical challenge is that restoring loads after a fault is not equivalent to simply reconnecting as many buses as possible. A switching plan can energize disconnected customers while simultaneously creating low-voltage buses, overloaded lines, excessive losses, or a non-radial topology. Distribution network reconfiguration studies therefore treat restoration as a constrained optimization problem in which load recovery, voltage quality, loss reduction, and network topology are jointly considered (Lotfi et al., 2024; Jia et al., 2024; Puma et al., 2024). This is especially relevant in radial distribution systems, where looped configurations may be unacceptable for protection coordination and operational practice.

Self-healing operation is normally described as a sequence of fault detection, isolation, and restoration actions. The literature has proposed several optimization-based solutions for this task, including particle swarm optimization, genetic algorithms, arithmetic optimization, robust reconfiguration, and other metaheuristic or mathematical formulations (Choobdari et al., 2024; Huang et al., 2025; Muñoz et al., 2025; Wu et al., 2024). These approaches are valuable, but they often require repeated optimization whenever the system state changes. In near-real-time restoration, a trained decision-making model may offer a faster alternative once it has learned the relationship between fault location, switching options, and power-flow feasibility.

Reinforcement learning (RL) has consequently become a promising direction for distribution network reconfiguration and service restoration. Deep RL has been used to learn switching policies, manage large action spaces, and incorporate power-flow feedback into sequential decision-making (Gholizadeh & Musilek, 2024; Guo et al., 2025; Jo et al., 2024). Graph-based RL has also been explored for outage management because network topology is central to the restoration problem (Jacob et al., 2024; Liu et al., 2025). These developments suggest that learning-based restoration can reduce dependence on repeated online optimization, provided that the learning process respects engineering constraints.

Multi-agent reinforcement learning (MARL) is particularly relevant because distribution restoration naturally involves several switching devices or network zones. Multi-agent methods have been applied to voltage control, converter coordination, distributed Volt-VAR control, and smart-grid control problems where local devices must coordinate through shared operational objectives (Su et al., 2024; Zhao et al., 2023; Zhang et al., 2024; Wu et al., 2025). However, in self-healing restoration, a naive multi-agent design can select tie-switch actions independently and create looped configurations. Thus, a MARL restoration framework must enforce radiality and power quality while still preserving coordinated learning.

The present paper addresses this issue by proposing an enhanced MARL framework for self-healing distribution networks under power quality constraints. The model is developed around the

IEEE 33-bus distribution system, implemented in Python, simulated with pandapower, and trained using PyTorch. The core contribution is not only the use of MARL, but also the integration of a radiality-preserving restoration repair mechanism and a power-quality-aware reward function. The framework is evaluated against fault isolation only, a single-agent DQN baseline, and an exhaustive radial search benchmark.

The main contributions of the paper are as follows. First, it formulates post-fault restoration as a coordinated multi-agent switching problem under voltage, line loading, loss, radiality, and power-flow feasibility constraints. Second, it introduces a radiality-preserving repair mechanism that converts candidate tie-switch actions into valid radial restoration plans. Third, it uses a centralized-training and decoded multi-agent execution structure so that coordinated joint actions are learned and then decoded into individual tie-switch decisions. Fourth, it validates the framework on five fault scenarios and demonstrates near-optimal restoration relative to exhaustive radial search while maintaining power quality constraints.

The remainder of the paper is organized as follows. Section 2 reviews related studies on self-healing networks, distribution reconfiguration, power quality, RL, and MARL. Section 3 presents the system model and problem formulation. Section 4 describes the proposed enhanced MARL methodology. Section 5 explains the simulation setup. Section 6 reports and discusses the results. Section 7 concludes the paper and outlines future work.

2. Literature Review

2.1 Self-Healing and Service Restoration in Distribution Networks

Self-healing distribution networks aim to reduce outage impact by automatically identifying the faulted area, isolating it, and restoring service to healthy sections. Jo et al. (2024) framed self-healing radial distribution network reconfiguration as a deep reinforcement learning problem and emphasized service restoration as a post-fault switching-control task. Jacob et al. (2024) presented graph reinforcement learning for real-time outage management in active distribution networks, explicitly considering switching control and emergency load shedding. Meteab et al. (2025) proposed a self-healing framework integrating fault detection, isolation, restoration, and network reconfiguration, with validation on an IEEE 33-bus radial distribution system.

Recent self-healing studies also show that restoration is increasingly integrated with flexible resources and renewable generation. Jiang et al. (2025) considered wind and solar accommodation in a distribution network fault self-healing strategy, while Pang et al. (2025) combined network partitioning with flexible resource aggregation for fault self-healing. Ni et al. (2025) examined self-healing strategies with AC flexible interconnection devices, and Li et al. (2025) used an E-SOP-based recovery strategy for flexible interconnected distribution networks. These works support the view that restoration should be treated as an integrated operational problem rather than a simple switching sequence.

Although the exact control architectures differ, these studies converge on three central requirements: restoration decisions should be fast, technically feasible, and consistent with distribution-network operating constraints. The present paper follows this direction by focusing on a self-healing framework that restores healthy loads while explicitly evaluating radiality, voltage limits, line loading, and losses after each restoration action.

2.2 Distribution Network Reconfiguration

Distribution network reconfiguration (DNR) provides the main operational mechanism through which service restoration is implemented after fault isolation. DNR changes the status of sectionalizing and tie switches to modify network topology, reduce losses, improve voltage profiles, and restore loads. Lotfi et al. (2024) reviewed DNR techniques and highlighted its importance for efficiency, reliability, and flexibility in modern distribution systems. Jia et al. (2024) proposed an improved arithmetic optimization algorithm for DNR, while Puma et al. (2024) used a hyperbolic tangent particle swarm optimization method to optimize reconfiguration decisions.

Optimization-based DNR methods can produce high-quality solutions, but their online computational cost can increase with network size and number of switching devices. Choobdari et al. (2024) addressed robust reconfiguration under renewable energy uncertainty, and Wu et al. (2024) considered dynamic multiobjective DNR under load and distributed generation variation. These examples show the increasing complexity of realistic reconfiguration problems. For this reason, learning-based decision models are attractive when a large number of reconfiguration cases must be evaluated or when fast response is required.

A key observation from the DNR literature is that restoration must preserve radial topology. Closing a tie switch can create a loop, so another branch must typically be opened to restore a radial configuration. This paper incorporates that principle through a radiality-preserving repair mechanism. The mechanism is consistent with the operational logic of radial distribution feeders and avoids the unrealistic non-radial solutions that can appear if learning agents close tie switches without topology correction.

2.3 Power Quality and Operational Constraints

Power quality in distribution networks covers voltage adequacy, waveform quality, loading limits, and continuity of supply. Zheng et al. (2024) presented a comprehensive assessment method for power quality in active distribution networks, while Samuel et al. (2025) examined evaluation and improvement of distribution-network power quality in a practical case. Wang et al. (2025) discussed power quality control methods for active distribution networks with large-scale renewable energy integration. These studies reinforce the idea that restoration decisions should be assessed not only by load recovery but also by the electrical quality of the restored state.

Voltage control studies are also closely related to the present work. Suchithra et al. (2024) applied deep reinforcement learning-based coordinated voltage control to enhance PV hosting capacity in distribution networks. Su et al. (2024), Zhao et al. (2023), and Zhang et al. (2024) applied multi-agent reinforcement learning to Volt-VAR or distributed voltage control. Although these works focus mainly on voltage regulation rather than fault restoration, they demonstrate that learning agents can use power-system feedback to maintain operational limits.

Technical losses and line loading are additional indicators of restoration quality. Parvizi et al. (2025) reviewed mechanisms and mitigation strategies for technical losses in power networks, and Xia et al. (2025) linked voltage-violation and transformer-overload risk assessment to short-term load forecasting. These studies motivate the inclusion of line loading and power loss in the proposed reward function. In this paper, power quality is represented through minimum voltage,

voltage violations, maximum line loading, overloaded-line count, active power loss, and power-flow convergence.

2.4 Reinforcement Learning and MARL in Power Systems

Deep reinforcement learning has been applied to reconfiguration because the switching problem can be expressed as a decision process in which an agent selects actions and receives a reward based on the resulting power-flow state. Gholizadeh and Musilek (2024) proposed a generalized deep RL model for DNR with power-flow-based action-space sampling. Guo et al. (2025) combined graph attention networks with reinforcement learning for dynamic reconfiguration of active distribution networks. Ghaemipour et al. (2026) proposed an innovative replay method to accelerate the learning process of deep RL algorithms in DNR. These studies show that RL can learn reconfiguration policies from repeated simulation feedback.

Graph and data-driven approaches are especially relevant because distribution networks have explicit topological structure. Jacob et al. (2024) used graph reinforcement learning for outage management, while Liu et al. (2025) proposed graph-enhanced deep RL for distribution network fault recovery. These works are close to the present paper because they connect fault recovery with topology-aware learning. The present study differs by emphasizing a compact, reproducible Python-pandapower-PyTorch workflow with explicit radiality repair and a power-quality-aware reward.

MARL extends RL to systems where multiple decision makers or controllable devices share a common objective. In power systems, MARL has been used for secondary voltage control in microgrids (Wang et al., 2023), model-free voltage control in AC/DC distribution networks (Zhao et al., 2023), distributed voltage control with soft open points (Zhang et al., 2024), and data-driven voltage control under weak-grid support (Wu et al., 2025). These applications demonstrate the value of coordinated learning when local actions interact through the network state.

A persistent difficulty in MARL is coordination. If each switching agent learns independently, the joint action can be inconsistent or suboptimal. This paper addresses the coordination issue using a centralized-training and decoded multi-agent execution structure. The model learns over the joint action space and then decodes the selected action into five binary tie-switch decisions. This retains a multi-switch execution interpretation while improving coordination across agents.

2.5 Research Gap

The reviewed studies provide strong foundations in self-healing restoration, distribution network reconfiguration, power quality control, RL, and MARL. Yet three gaps remain. First, many restorations and DNR studies either rely on online optimization or do not explicitly integrate learning-based coordination with radiality correction. Second, several RL-based studies focus on reconfiguration or voltage control, but fewer combine post-fault restoration, radiality-preserving switching, and power-quality-aware reward design in one workflow. Third, independent multi-agent structures can struggle with coordination when the reward depends on the full joint switching pattern.

The proposed framework responds to these gaps by integrating enhanced MARL, radiality-preserving repair, pandapower-based power-flow evaluation, and explicit power quality constraints in one simulation-based study. The intended contribution is a reproducible, technically constrained learning framework rather than a purely algorithmic demonstration.

2.6 Design Rationale for the Proposed Framework

The design of the proposed framework follows directly from the limitations observed in the reviewed literature. Self-healing studies show the need for rapid post-fault switching decisions, DNR studies show the importance of radiality and loss-aware reconfiguration, power quality studies show that restored service must remain electrically acceptable, and MARL studies show that coordinated learning is useful when several devices interact through a shared network state. The proposed method therefore combines these four requirements in a single workflow: post-fault restoration, radial topology enforcement, power-quality evaluation, and coordinated learning.

A central modeling decision is to use power-flow feedback after each candidate restoration action. This avoids treating the learning environment as a purely abstract control problem. Every reward value is tied to a physical network state evaluated by pandapower. The approach is consistent with DNR studies that rely on load-flow evaluation to judge feasibility and performance, while still allowing the learning policy to be trained offline. In this sense, the method uses simulation not only to generate results but also to define the learning environment.

Another design decision is to keep the action space interpretable. Each final action corresponds to the status of five tie switches, and any additional branch opening is produced by the radiality repair mechanism. This makes the result easier to audit than a black-box switching policy with no topological explanation. The reported repair openings show which sectionalizing branches are opened to remove loops. Such interpretability is important for power-system operation, where a restoration plan must be reviewed by engineers and cannot be justified only by a numerical reward.

The proposed enhanced MARL structure is therefore positioned between two extremes. It is not a fully decentralized independent-agent design, because such designs may struggle with coordination. It is also not merely an exhaustive optimizer, because the trained network learns a policy that can be evaluated quickly once trained. This middle position is appropriate for the present problem: the model benefits from centralized training over the joint action space, but the resulting action remains a coordinated set of individual switch commands.

3. System Model and Problem Formulation

The system model is built around the IEEE 33-bus radial distribution feeder, which is widely used as a benchmark test system in distribution network reconfiguration and self-healing studies. The feeder data were implemented in pandapower following the standard IEEE 33-bus configuration commonly adopted in recent distribution network reconfiguration and restoration research (Jo et al., 2024; Gholizadeh & Musilek, 2024; Meteab et al., 2025; Lotfi et al., 2024). The test feeder includes 33 buses, 32 normally closed sectionalizing branches, five normally open tie branches, and one slack/source bus. The total connected active load is 3.715 MW. Each restoration action is evaluated through a backward-forward sweep power-flow calculation. The test system and simulation parameters are summarized in Table 1.

Table 1. IEEE 33-Bus Test System Parameters

Parameter	Value
Test system	IEEE 33-bus radial distribution network
Number of buses	33
Normally closed sectionalizing branches	32
Normally open tie branches	5
Slack/source bus	Bus 1
Base voltage	12.66 kV
Base apparent power	10 MVA
Total connected active load	3.715 MW
Power system simulation tool	pandapower
Learning framework	PyTorch
Programming environment	Python

Source: Authors' pandapower implementation based on the standard IEEE 33-bus benchmark configuration commonly used in recent distribution network reconfiguration and self-healing studies (Jo et al., 2024; Gholizadeh & Musilek, 2024; Meteab et al., 2025; Lotfi et al., 2024), with numerical values verified from the final simulation output files.

Five line-fault scenarios are defined to represent different levels of restoration difficulty. The F1 case is a near-source fault, while the remaining cases occur in mid-feeder, lateral, far-end, and high-load sections. The faulted line is isolated in every scenario and remains out of service. Table 2 lists the simulated fault scenarios. The interpretation of F1 is important because a fault close to the source can structurally limit the amount of load that can be restored by the available tie switches.

Table 2. Fault Scenario Description

Scenario	Faulted Line	Fault Location Description	Restoration Interpretation
F1	L2-3	Near-source fault	Severe upstream contingency with limited restoration capability
F2	L7-8	Mid-feeder fault	Recoverable fault through tie-switch restoration
F3	L3-23	Lateral-branch fault	Recoverable lateral feeder fault
F4	L30-31	Far-end fault	Recoverable downstream fault
F5	L24-25	Heavy-load-area fault	Recoverable fault affecting a high-load section

Source: Authors' fault-scenario design in the simulation notebook.

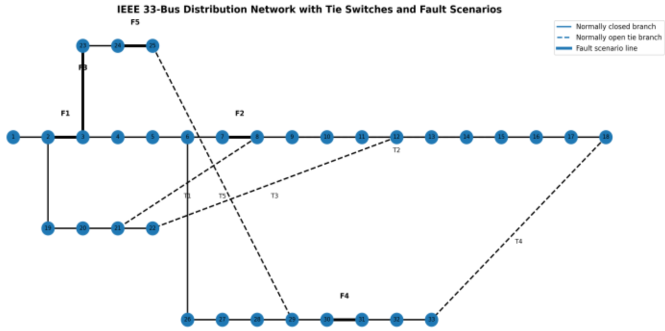

Figure 1. IEEE 33-bus distribution network with fault locations and tie switches. Source: Authors' diagram generated from the pandapower implementation of the standard IEEE 33-bus benchmark feeder, following recent distribution network reconfiguration and self-healing studies (Jo et al., 2024; Gholizadeh & Musilek, 2024; Meteab et al., 2025; Lotfi et al., 2024), with fault scenarios defined by the authors.

The restoration objective is to maximize the amount of healthy load restored after fault isolation while satisfying power quality and topology constraints. The restored load percentage is computed as the ratio between the supplied active load and the total connected active load:

$$L_restored(\%) = \left(sum_{iinOmega_S} P_i / sum_{iinOmega_L} P_i\right) x100 \quad (1)$$

where $Omega_S$ is the set of supplied load buses, $Omega_L$ is the set of all load buses, and P_i is the active load at bus i. This metric is used in both the reward function and the final evaluation tables.

The voltage constraint requires all supplied buses to remain within the adopted operating range:

$$V_min \leq V_i \leq V_max, foralliinOmega_B \quad (2)$$

In this study, V_min is set to 0.90 $p.u.$ and V_max to 1.05 $p.u.$ The lower limit is selected because the uncompensated IEEE 33-bus feeder may operate below 0.95 $p.u.$ under standard loading, as also reflected in the baseline simulation. A stricter 0.95 $p.u.$ case would require explicit voltage-support devices, such as capacitors, distributed generation, or reactive power support.

Line loading is constrained by the allowable current limit of each branch:

$$Loading_l = (|I_l|/I_l^max)x100 \leq 100\%, foralllinOmega_E \quad (3)$$

Active power loss is evaluated as an operational efficiency indicator:

$$P_loss = sum_{linOmega_E} R_l |I_l|^2 \quad (4)$$

A radiality constraint is imposed because the restored distribution network must not contain energized cycles:

$$C(G) = 0 \quad (5)$$

where $C(G)$is the number of cycles in the energized network graph G. For a connected supplied radial component, the number of active branches must satisfy:

$$|Omega_E^active| = |Omega_B^supplied| - 1 \quad (6)$$

The fault-isolation constraint keeps the faulted line disconnected during restoration:

$$x_f = 0 \quad (7)$$

Table 3. Power Quality and Operational Constraints

Constraint	Mathematical Form	Adopted Limit / Condition
Voltage lower limit	$V_i \geq V_m in$	$V_m in = 0.90 p.u.$
Voltage upper limit	$V_i \leq V_m ax$	$V_m ax = 1.05 p.u.$
Line loading limit	$Loading_l \leq Loading_m ax$	$Loading_m ax = 100\%$
Radial topology	$C(G) = 0$	No cycles allowed
Fault isolation	$x_f = 0$	Faulted line remains disconnected
Power-flow feasibility	$PF_c onv = 1$	Power flow must converge

Source: Authors' problem formulation and simulation settings.

4. Proposed Enhanced MARL Methodology

The proposed framework treats restoration as a coordinated switching-control problem. Once a fault is isolated, the learning model selects a restoration action over the five tie switches. Each tie-switch action is binary, where 0 denotes an open switch and 1 denotes a closed switch. The joint action vector is expressed as:

$$A = [a_1, a_2, a_3, a_4, a_5], a_k in 0,1 \quad (8)$$

The total joint action space is therefore: $|A| = 2^5 = 32$ (9)

Figure 2. Proposed Enhanced MARL self-healing framework. Source: Authors' conceptual design based on the implemented simulation workflow.

4.1 Radiality-Preserving Restoration Repair

A direct tie-switch action can create a loop. Instead of accepting such a non-radial solution, the proposed method applies a radiality-preserving repair mechanism. The mechanism detects cycles after tie-switch closure, identifies candidate sectionalizing branches in the cycle, evaluates their effect through power flow, and opens the branch that provides the highest reward while restoring radial topology. This step is essential because a learning model that only closes tie switches may produce technically invalid restoration states.

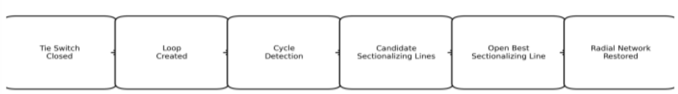

Figure 3. Radiality-preserving restoration repair mechanism. Source: Authors' algorithmic design implemented in the simulation notebook.

4.2 Reward Function

The reward function is designed to prioritize service restoration while penalizing violations of power quality and operating constraints. The enhanced reward function is written as:

$$R = w_L L_r estored - w_V N_V - w_O N_O - w_P(100 P_l oss) - w_S N_S - w_R C(G) - w_F I_f ail \quad (10)$$

SJST
sjst.scst.edu.ly
Surman Journal for Science and Technology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol 8, No.2. Jun – Dec. 2026
Pages: 018 ~ 040

where $L_restored$ is the restored load percentage, N_V is the number of voltage violations, N_O is the number of overloaded lines, P_loss is active power loss in MW, N_S is the number of switching actions, $C(G)$ is the number of cycles, and I_fail is a binary indicator of power-flow failure. The selected weights are shown in Table 4. The high non-radial and power-flow-failure penalties act as safety terms, while the restored-load weight encourages the learning model to prefer complete service recovery whenever feasible.

Table 4. Enhanced Reward Function Weights

Reward Component	Role	Weight
Restored load	Positive reward	5.00
Voltage violation	Penalty	15.00
Line overload	Penalty	15.00
Power loss	Penalty	0.50
Switching action	Penalty	0.25
Non-radial topology	Hard penalty	1000.00
Power-flow failure	Hard penalty	1000.00

Source: Authors' enhanced reward design used in the final training run.

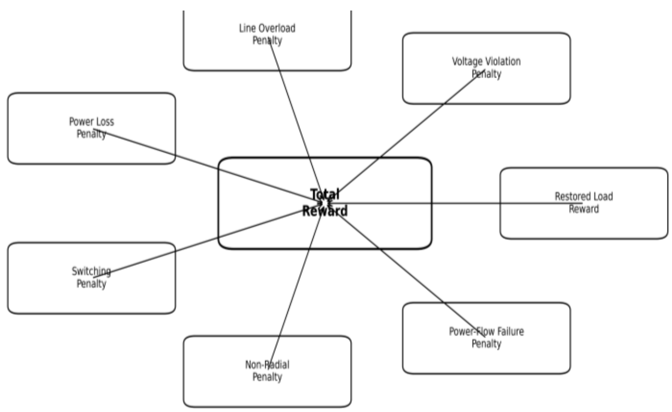

Figure 4. Power-quality-aware reward function components. Source: Authors' reward-function design derived from the proposed formulation.

4.3 Enhanced MARL Structure

The final learning model uses a centralized-training and decoded multi-agent execution structure. During training, the Q-network learns over the joint action index. The selected action index is then decoded into the five binary tie-switch actions. This structure improves coordination because the value function evaluates the effect of the complete switching pattern rather than five isolated local decisions. The approach remains compatible with multi-agent execution because the final joint action is implemented as individual tie-switch statuses.

The state vector includes the active fault scenario encoded as a one-hot vector, restored load ratio, minimum bus voltage, maximum line loading, active power loss, and radial status:

$$s_t = [F_one - hot, L_restored, V_min, Loading_max, P_loss, R_status] \quad (11)$$

The DQN target and loss follow the standard deep Q-learning form:

$$y_t = r_t + gammamax_{a'} Q_{theta^-}(s_{t+1}, a') \quad (12)$$

$$L(theta) = \left(y_t - Q_theta(s_t, a_t)\right)^2 \quad (13)$$

The joint action decoding is expressed as:

$$a^joint \rightarrow [a_1, a_2, a_3, a_4, a_5] \quad (14)$$

Table 5. Enhanced MARL Training Parameters

Parameter	**Value**
Learning approach	Enhanced MARL with CTDE-style joint-action learning
Action representation	Joint action decoded into five tie-switch decisions
Number of tie switches	5
Joint action space size	2^5 = 32
Training episodes	10,000
Discount factor gamma	0.95
Learning rate	0.001
Batch size	64
Replay buffer size	10,000
Target network update frequency	50 episodes
Epsilon start	1.00
Epsilon end	0.02
Epsilon decay portion	85% of training episodes
Neural network hidden layers	128 neurons, 128 neurons
Activation function	ReLU
Optimizer	Adam

Source: Authors' PyTorch training configuration in the final experiment.

5. Simulation Setup

The experimental workflow was implemented entirely in Python. pandapower was used to model the network, run power-flow calculations, and evaluate voltage, line loading, and active power loss after each restoration candidate. PyTorch was used to implement the Q-network, replay buffer, target network update, and epsilon-greedy training policy. NetworkX was used for connectivity and cycle detection in the radiality-preserving repair mechanism.

For computational efficiency, all feasible combinations of fault scenario and tie-switch action were precomputed. Since there are five fault scenarios and 32 tie-switch combinations, 160 candidate cases were evaluated with pandapower. Each candidate case was repaired for radiality where necessary, power-flow results were recorded, and the reward was computed. During training, the model used this restoration cache rather than repeatedly running a power flow for identical cases. The cache does not alter the numerical basis of the results because every cached reward and metric was originally obtained from pandapower power-flow evaluation.

The proposed method was compared with three benchmarks. The first benchmark, fault isolation only, represents the post-fault network after the faulted line is isolated and no restoration action is taken. The second benchmark, exhaustive radial search, evaluates all 32 tie-switch combinations for each fault and selects the one with the highest reward after radiality repair. The third benchmark, single-agent DQN, uses one centralized DQN agent to select among the 32 joint switching actions. The enhanced MARL model is then evaluated using the same scenarios and metrics.

The performance indicators are restored load percentage, minimum voltage, number of voltage violations, maximum line loading, number of overloaded lines, active power loss, radiality status, cycle count, power-flow convergence, and reward. These indicators provide a combined view of service restoration, power quality, network topology, and operating efficiency.

5.1 Reproducibility and Experimental Controls

Several controls were used to keep the experiment reproducible. The same network data, faulted lines, reward weights, voltage limits, and line-loading limits were used for all methods. The fault isolation only case, single-agent DQN, exhaustive radial search, and enhanced MARL were evaluated using the same pandapower-based metric calculations. This prevents performance differences from being caused by different load-flow models or inconsistent constraint definitions. The comparison also uses the same radiality repair logic for restoration candidates. This point is important because a method that is allowed to keep non-radial loops could appear to restore more load while producing an invalid distribution topology. By applying the same radiality-preserving mechanism before reward evaluation, the comparison focuses on the quality of the selected switching plan rather than on differences in feasibility screening.

The training process was performed using an epsilon-greedy exploration strategy, replay memory, and a target network update procedure. These elements were included to stabilize Q-learning and reduce the likelihood that the policy would converge to a narrow set of early exploratory actions. The training curve is therefore interpreted together with the final benchmark comparison, because a high training reward alone is not sufficient unless the final selected actions also satisfy the engineering constraints.

5.2 Evaluation Protocol

The evaluation protocol was designed to separate feasibility, restoration quality, and learning performance. Feasibility was checked through power-flow convergence, radiality status, voltage violations, and overloaded-line counts. Restoration quality was assessed through restored load percentage and power loss. Learning performance was assessed through the training reward curve and the final deterministic policy selected with zero exploration. This separation is important because a learning model may show a high training reward but still fail engineering feasibility if the reward is not properly constrained.

For the final comparison, each method was evaluated once for each of the five predefined fault scenarios using the trained or selected deterministic policy. The resulting tables therefore report operational outcomes rather than exploratory training samples. This makes the comparison closer to a practical post-training deployment setting, where the controller receives a fault scenario and must return a final restoration action.

6. Results and Discussion

6.1 Baseline Operating Condition

The baseline power-flow analysis was performed before any fault was introduced. The full connected active load of 3.715 MW was supplied, corresponding to 100% load restoration. The minimum bus voltage was 0.9038 p.u., the maximum bus voltage was 1.0000 p.u., and the maximum line loading was 52.72%. Under the adopted voltage range of 0.90-1.05 p.u., no voltage violations and no line overloads occurred. The baseline active power loss was 0.2110 MW, and the network remained radial. These results establish the normal operating reference for the post-fault restoration tests.

6.2 Training Performance

The enhanced MARL model was trained for 10,000 episodes. The training curve in Figure 5 shows that the reward improved and stabilized as exploration decreased. The moving-average reward indicates that the learning process converged toward high-reward restoration policies rather than remaining dominated by low-quality or infeasible switching actions. This behavior is consistent with the purpose of the enhanced reward function, which prioritizes restored load while penalizing poor power-quality and topology outcomes.

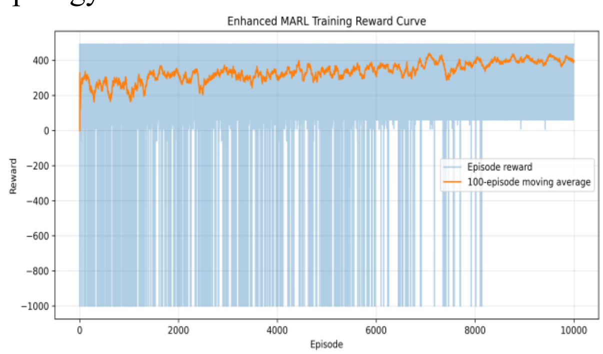

Figure 5. Enhanced MARL training reward curve. Source: Authors' PyTorch training output from the final simulation run.

6.3 Overall Comparative Performance

Table 6 summarizes average performance across the five fault scenarios. The enhanced proposed MARL achieved an average restored load of 82.48%, equal to the exhaustive radial search and single-agent DQN averages. This value is substantially higher than the 68.24% obtained by fault isolation only. The result indicates that the learning-based restoration policy recovered the same amount of load as the exhaustive benchmark, while preserving all imposed constraints.

Table 6. Overall Comparative Performance

Method	Average Restored Load (%)	Average Power Loss (MW)	Average Reward	All Cases Radial
Fault Isolation Only	68.24	0.1234	335.02	Yes
Single-Agent DQN	82.48	0.1529	404.24	Yes
Exhaustive Radial Search	82.48	0.1430	404.53	Yes
Enhanced Proposed MARL	82.48	0.1455	404.31	Yes

Source: Authors' final comparative results computed from the simulation output files.

SJST
sjst.scst.edu.ly
Surman Journal for Science and Technology
ISSN: Online (2790-5721) - Print (2790-5713)
مجلة صرمان للعلوم والتقنية
Vol 8, No.2. Jun – Dec. 2026
Pages: 018 ~ 040

The average reward of enhanced MARL was 404.31, very close to the exhaustive radial search value of 404.53. This small difference shows that the trained learning model reached a near-optimal switching policy under the reward structure. The enhanced MARL average power loss was 0.1455 MW, slightly above exhaustive search but lower than the single-agent DQN average of 0.1529 MW. Thus, the enhanced multi-agent design matched the restoration level of the single-agent baseline while producing lower average losses.

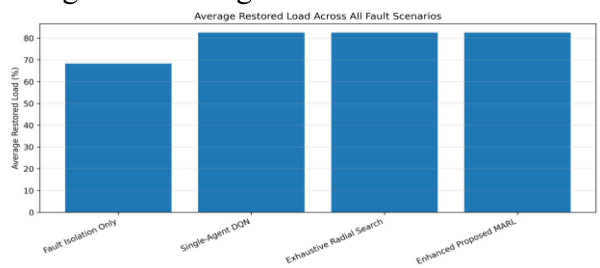

Figure 6. Average restored load summary across all fault scenarios. Source: Authors' final comparative results.

6.4 Scenario-Based Restoration Results

Table 7 presents the restored load percentage for each fault scenario. In scenarios F2, F3, F4, and F5, the enhanced proposed MARL restored 100% of the load. These cases represent recoverable mid-feeder, lateral, far-end, and heavy-load-area faults. The results show that the model learned restoration actions that fully re-energized all healthy loads in these scenarios while remaining radial and satisfying power quality limits.

Table 7. Scenario-Based Restored Load Performance

Scenario	Faulted Line	Fault Isolation Only (%)	Single-Agent DQN (%)	Exhaustive Radial Search (%)	Enhanced Proposed MARL (%)
F1 near-source	L2-3	12.38	12.38	12.38	12.38
F2 mid-feeder	L7-8	76.45	100.00	100.00	100.00
F3 lateral branch	L3-23	74.97	100.00	100.00	100.00
F4 far-end	L30-31	88.69	100.00	100.00	100.00
F5 heavy-load area	L24-25	88.69	100.00	100.00	100.00

Source: Authors' scenario-level simulation results.

The F1 case behaves differently. All methods restored only 12.38% of the total load when line L2-3 was isolated. Since even exhaustive radial search could not improve this value, the limitation is structural rather than algorithmic. The fault is close to the source, and the available tie-switch configuration does not provide sufficient alternative paths to restore most of the feeder. In practical systems, this type of severe upstream fault would require additional resources such as alternative feeders, distributed generation, or network reinforcement.

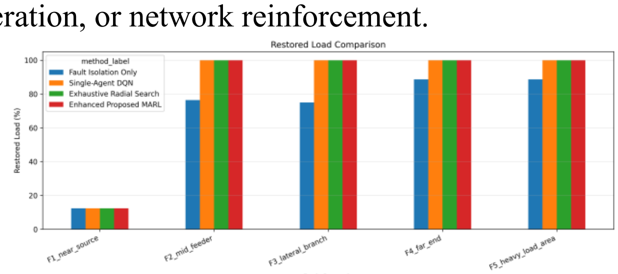

Figure 7. Restored load comparison under different fault scenarios. Source: Authors' final comparative results.

6.5 Power Quality and Topology Validation

The final restored configurations satisfied all considered power quality and topology constraints. As shown in Table 8, power flow converged in all cases, radial topology was maintained, voltage violations were zero, overloaded lines were zero, and cycle count was zero. These outcomes confirm that the proposed method did not achieve service restoration by violating operating limits.

Table 8. Validation Summary of Final Restored Configurations

Validation Indicator	Fault Isolation Only	Single-Agent DQN	Exhaustive Radial Search	Enhanced Proposed MARL
Power-flow convergence	Yes	Yes	Yes	Yes
Radial topology maintained	Yes	Yes	Yes	Yes
Voltage violations	0	0	0	0
Overloaded lines	0	0	0	0
Cycle count	0	0	0	0

Source: Authors' validation summary from the final experiment.

The minimum voltage for enhanced MARL ranged from 0.9061 p.u. to 0.9942 p.u. The lowest values appeared in the F3 and F4 scenarios, but they remained above the 0.90 p.u. lower limit. The maximum line loading also remained below 100% in every scenario, with the highest enhanced MARL value equal to 53.23%. The voltage and line-loading figures provide visual confirmation of these results.

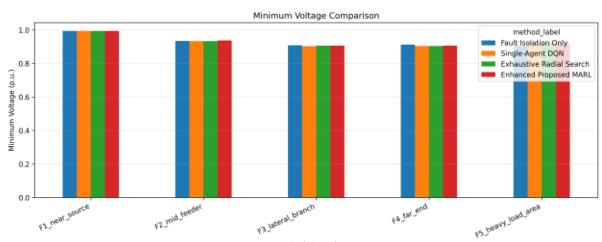

Figure 8. Minimum voltage comparison under different fault scenarios. Source: Authors' final simulation results.

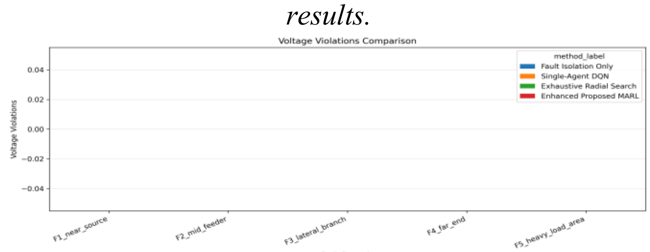

Figure 9. Voltage violations comparison. Source: Authors' final simulation results.

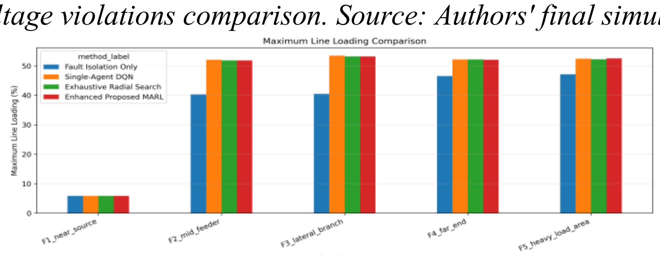

Figure 10. Maximum line loading comparison. Source: Authors' final simulation results.

6.6 Power Loss Analysis

Although the primary objective is load restoration, power loss remains an important operating indicator. The enhanced MARL model produced an average power loss of 0.1455 MW. This is slightly higher than the exhaustive radial search average of 0.1430 MW, as expected because exhaustive search directly selects the highest reward among all feasible combinations. However, enhanced MARL achieved lower average loss than the single-agent DQN baseline while restoring the same amount of load. This suggests that coordinated joint-action learning improved the balance between service restoration and loss performance.

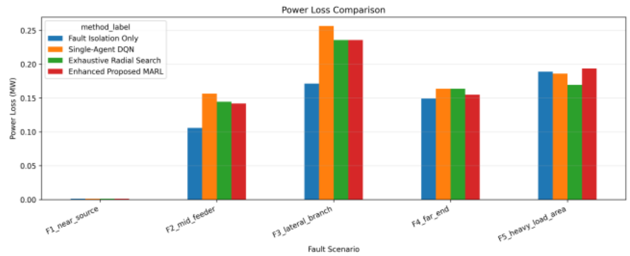

Figure 11. Power loss comparison under different restoration methods. Source: Authors' final simulation results.

6.7 Voltage Profile of the F2 Scenario

The F2 mid-feeder fault is useful for visual inspection because the fault isolation only case supplies less than the full load, while the restoration methods recover the disconnected healthy loads. Figure 12 shows the voltage profile for the selected F2 case. The restored voltage profile remains within the adopted operating range and confirms that full service restoration was achieved without violating the voltage constraint.

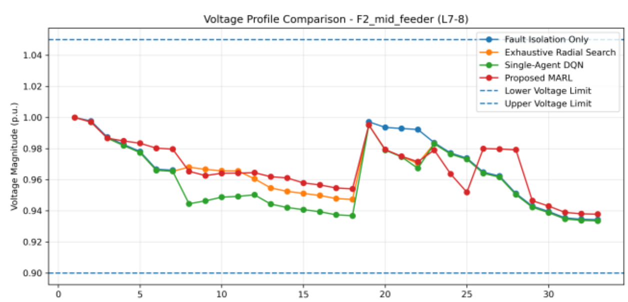

Figure 12. Voltage profile for the F2 mid-feeder fault scenario. Source: Authors' final simulation output for the selected scenario.

6.8 Discussion

The results support three main observations. First, restoration decisions must be evaluated under radiality and power quality constraints. The earlier problem of non-radial candidate configurations was eliminated through the radiality-preserving repair mechanism. This confirms the importance of combining learning decisions with topology validation in distribution restoration problems.

Second, the enhanced MARL policy achieved near-optimal performance relative to exhaustive radial search. The exhaustive benchmark evaluates all 32 joint switching combinations for each fault and therefore represents a strong comparison for this small test system. Matching its restored load performance and approaching its reward value indicates that the trained model learned a high-quality restoration policy.

Third, the learning-based method is more scalable in concept than exhaustive enumeration. For five tie switches, exhaustive search is manageable. However, in larger feeders with many controllable switches, the action space grows exponentially. A trained learning policy can provide fast decision support after training, whereas exhaustive enumeration may become computationally expensive. This advantage aligns with recent RL and graph-RL studies that seek fast outage management and reconfiguration policies for active distribution networks (Gholizadeh & Musilek, 2024; Jacob et al., 2024; Liu et al., 2025).

The severe F1 case also provides an important practical lesson. A learning model cannot restore loads that are structurally unreachable through the available topology. The identical F1 results across all methods demonstrate that restoration potential depends on network connectivity and available tie paths. Therefore, future deployment of learning-based self-healing should be accompanied by planning studies that evaluate whether the feeder has sufficient switching flexibility and alternative supply paths.

6.9 Interpretation of Restoration Actions

The scenario-level results also show that the selected restoration actions are not identical across methods even when the restored-load percentage is the same. This is expected because several radial switching plans may restore all load while differing in loss, voltage profile, or number of switching actions. In the F2 case, for example, exhaustive radial search selected tie actions (1, 0, 1, 0, 0) with one repair opening, whereas enhanced MARL selected (1, 1, 1, 0, 1) with three repair openings. Both restored all loads without voltage or loading violations, but the final losses and reward differed slightly. This illustrates that restoration is a multi-objective engineering decision rather than a binary success-or-failure outcome.

The F3 case is particularly useful because enhanced MARL and exhaustive radial search selected the same tie-action pattern and the same radial repair openings. Both methods used all five tie switches and opened L14-15, L9-10, L7-8, and L30-31 during radiality repair. The equality of the reward and loss values in this scenario indicates that the learning policy can reproduce the best exhaustive plan when the reward landscape strongly favors a particular configuration. This finding supports the suitability of the joint-action learning structure for coordinated restoration.

The F4 and F5 scenarios show a different but still acceptable pattern. Enhanced MARL restored 100% of the load and satisfied all constraints, but it selected restoration plans with slightly different losses compared with exhaustive radial search. Such small differences are acceptable in practical decision-support contexts, especially when the trained policy can provide rapid decisions after offline training. The trade-off is between the exact optimality of exhaustive enumeration and the speed and scalability of a learned restoration policy. This trade-off is consistent with the motivation for RL-based reconfiguration in recent studies (Gholizadeh & Musilek, 2024; Ghaemipour et al., 2026; Guo et al., 2025).

6.10 Relationship to Prior Studies

The obtained results are consistent with the broader self-healing literature, which stresses that restoration should integrate fault isolation, switching control, and power-flow feasibility. Jo et al.

(2024) showed that deep reinforcement learning can support radial distribution network reconfiguration for self-healing operation, while Jacob et al. (2024) emphasized the role of topology-aware reinforcement learning for outage management. The present study follows the same learning-based direction but adds an explicit radiality repair stage and evaluates power quality constraints after every candidate restoration action.

Compared with conventional DNR optimization studies, the proposed method does not replace the need for accurate power-flow evaluation. Instead, it uses pandapower results as the engineering feedback that shapes the learning policy. This is important because optimization and learning methods should not be separated from the physical network model. Studies such as Jia et al. (2024), Puma et al. (2024), Choobdari et al. (2024), and Muñoz et al. (2025) demonstrate the value of optimization for high-quality reconfiguration, while the present work uses exhaustive radial search as a benchmark to test whether the learned policy can approach a strong enumerative solution.

The MARL element of the proposed method also connects to voltage-control studies where multiple devices coordinate through shared objectives. Su et al. (2024), Zhao et al. (2023), Zhang et al. (2024), and Wu et al. (2025) demonstrate that MARL can be effective for coordinated voltage or converter control. The present work adapts this coordination concept to post-fault switching restoration. The difference is that switching restoration has discrete topology-changing actions, and therefore radiality preservation becomes a central issue that is less prominent in continuous voltage-control problems.

Power quality studies support the decision to include voltage violations, line loading, losses, and convergence in the evaluation. Zheng et al. (2024) and Samuel et al. (2025) emphasize that power quality assessment requires more than one indicator, while Parvizi et al. (2025) highlight the relevance of technical losses. The proposed reward function follows this multi-indicator logic. The results show that the enhanced MARL policy restored load without producing voltage violations or overloads, which is essential for treating the solution as operationally meaningful rather than merely numerically successful.

6.11 Computational and Practical Considerations

The restoration cache used in this study has two purposes. First, it reduces training time by avoiding repeated evaluation of the same fault-action pairs. Second, it preserves physical interpretability because each cached entry is based on a pandapower power-flow calculation after radiality repair. For the IEEE 33-bus case, the cache contains 160 cases. For larger systems, the number of possible switch combinations can become much larger, and direct enumeration may no longer be practical. In such cases, the same framework can be extended with action screening, graph-based state representation, or hierarchical restoration policies.

The exhaustive radial search benchmark is feasible in the present test system because the action space is small. However, its role in this paper is not to claim that exhaustive enumeration is a scalable operational solution. It is used as a rigorous reference to measure how close the enhanced MARL policy comes to the best available action within the defined action space. The small reward gap between enhanced MARL and exhaustive radial search is therefore an important validation result.

The single-agent DQN baseline provides a useful comparison because it also learns over the joint action space. The enhanced MARL structure differs in interpretation: the joint action is decoded into individual tie-switch decisions, preserving the connection to multi-agent switching execution. The lower average loss obtained by enhanced MARL compared with single-agent DQN suggests that the enhanced training formulation and reward tuning improved the quality of selected solutions, although the difference remains modest.

From an implementation perspective, the proposed workflow is reproducible because it is built on open Python-based tools and deterministic scenario definitions. The results can be regenerated by using the same IEEE 33-bus data, fault scenarios, reward weights, and training parameters. This reproducibility is important for research in learning-based power-system control, where results may otherwise be difficult to compare due to differences in environments, reward scaling, or action definitions.

6.12 Limitations of the Present Study

Several limitations should be acknowledged. First, the experiment uses a deterministic IEEE 33-bus test system and five predefined fault locations. Although this is appropriate for a controlled proof of concept, practical feeders may have uncertain load patterns, distributed generation variability, communication delays, and protection constraints. These factors should be included in future extensions before field-level deployment is considered.

Second, the voltage lower limit is set to 0.90 p.u. because the uncompensated IEEE 33-bus feeder has a baseline minimum voltage close to this value. This choice is transparent and consistent with the simulated feeder condition, but it also means that the study does not claim compliance with a stricter 0.95 p.u. lower limit. A stricter study would require additional voltage-control resources such as capacitors, voltage regulators, distributed generation reactive support, or flexible interconnection devices. Recent voltage-control studies using RL and MARL provide useful directions for such integration (Suchithra et al., 2024; Su et al., 2024; Zhang et al., 2024).

Third, the restoration decision is represented as a one-step switching decision followed by radiality repair. Real restoration may involve sequential switching with operational safety checks after each switching operation. Extending the model to multi-step restoration would allow switching order, transient constraints, and operator approval rules to be represented more realistically. Nevertheless, the one-step formulation is suitable for evaluating the quality of final restoration configurations under the defined action space.

Fourth, the proposed cache-based training approach relies on precomputing feasible cases. This is efficient for the present network but may require adaptation for larger feeders. Future work can combine the cache strategy with graph neural networks, action-space pruning, or model-based screening to avoid exhaustive precomputation while preserving physical feasibility. These extensions would align with recent graph-enhanced and topology-aware RL directions in outage management and fault recovery (Jacob et al., 2024; Liu et al., 2025).

7. Conclusion and Future Work

This paper proposed an enhanced multi-agent reinforcement learning framework for self-healing distribution networks under power quality constraints. The IEEE 33-bus radial distribution system was modeled in pandapower, and the learning model was implemented in PyTorch. The proposed method combined coordinated joint-action learning, radiality-preserving restoration repair, and a power-quality-aware reward function that prioritizes restored load while penalizing voltage violations, line overloads, power losses, excessive switching, non-radial topology, and power-flow failure.

The final simulation results show that the enhanced proposed MARL restored all recoverable loads in the tested scenarios, maintained radial topology in every final configuration, produced zero voltage violations, produced zero-line overloads, and achieved an average restored load of 82.48%. This matches the exhaustive radial search benchmark and improves substantially over the 68.24% average achieved by fault isolation only. The enhanced MARL average reward of 404.31 was very close to the exhaustive radial search reward of 404.53, indicating near-optimal restoration performance under the adopted reward design.

The comparison with single-agent DQN showed that both approaches restored the same average load, but enhanced MARL produced lower average power losses. This supports the value of coordinated multi-agent learning for restoration decisions where switching actions interact through the network topology.

Future work should extend the proposed framework to larger distribution systems with more switches and distributed energy resources. Additional work should also consider stochastic load and renewable generation, multi-step switching sequences, communication delays, cyber-physical constraints, and stricter voltage support scenarios. The framework can also be extended by using graph neural networks or actor-critic MARL methods to improve generalization across unseen feeder topologies.

Data Availability Statement

The data used in this study were generated through the Python-based simulation workflow described in the methodology. The final numerical outputs include baseline metrics, fault-scenario results, benchmark comparisons, validation summaries, and training histories. These outputs can be reproduced using the described IEEE 33-bus model, pandapower power-flow evaluation, and PyTorch learning configuration.

Conflict of Interest

The author declares no conflict of interest.

References

Choobdari, M., Samiei Moghaddam, M., Davarzani, R., Azarfar, A., & Hoseinpour, H. (2024). Robust distribution networks reconfiguration considering the improvement of network resilience considering renewable energy resources. Scientific Reports, 14, 23041. https://doi.org/10.1038/s41598-024-73928-1

Chu, Y., Zhou, R., Cui, Q., Wang, Y., Li, B., & Zhou, Y. (2025). A reinforcement learning-based dynamic network reconfiguration strategy considering the coordinated optimization of SOPs and traditional switches. Processes, 13, 1670. https://doi.org/10.3390/pr13061670

Ghaemipour, A., Rajabi Mashhadi, H., & Mostafavi, S. H. (2026). Accelerating the learning process of deep reinforcement learning algorithms in distribution network reconfiguration using an innovative replay method. Scientific Reports. https://doi.org/10.1038/s41598-026-40508-4

Gholizadeh, N., & Musilek, P. (2024). A generalized deep reinforcement learning model for distribution network reconfiguration with power flow-based action-space sampling. Energies, 17, 5187. https://doi.org/10.3390/en17205187

Guo, C., Jiang, C., & Liu, C. (2025). Dynamic reconfiguration method of active distribution networks based on graph attention network reinforcement learning. Energies, 18, 2080. https://doi.org/10.3390/en18082080

Huang, W., Chen, G., Jiang, X., Xiao, X., Chen, Y., & Liu, C. (2025). Research on self-healing distribution network operation optimization method considering carbon emission reduction. Processes, 13, 1850. https://doi.org/10.3390/pr13061850

Jacob, R. A., Paul, S., Chowdhury, S., Gel, Y. R., & Zhang, J. (2024). Real-time outage management in active distribution networks using reinforcement learning over graphs. Nature Communications, 15, 4766. https://doi.org/10.1038/s41467-024-49207-y

Jia, H., Zhu, X., & Cao, W. (2024). Distribution network reconfiguration based on an improved arithmetic optimization algorithm. Energies, 17, 1969. https://doi.org/10.3390/en17081969

Jiang, L., Wang, C., Li, C., & Xiang, C. (2025). Research on fault self-healing strategy of a distribution network considering wind and solar accommodation capacity. Frontiers in Energy Research, 13, 1563243. https://doi.org/10.3389/fenrg.2025.1563243

Jo, S., Oh, J.-Y., Yoon, Y. T., & Jin, Y. G. (2024). Self-healing radial distribution network reconfiguration based on deep reinforcement learning. Results in Engineering, 22, 102026. https://doi.org/10.1016/j.rineng.2024.102026

Li, W., Xu, Z., Chen, M., & Wu, Q. (2025). Smart grid self-healing enhancement E-SOP-based recovery strategy for flexible interconnected distribution networks. Energies, 18, 3358. https://doi.org/10.3390/en18133358

Liu, Y., Liao, P., & Wang, Y. (2025). Using graph-enhanced deep reinforcement learning for distribution network fault recovery. Machines, 13, 543. https://doi.org/10.3390/machines13070543

Lotfi, H., Hajiabadi, M. E., & Parsadust, H. (2024). Power distribution network reconfiguration techniques: A thorough review. Sustainability, 16, 10307. https://doi.org/10.3390/su162310307

Mallala, B., Manyam, R. R., Saini, J., Tajuddin, M. F. N., & Thanikanti, S. B. (2025). IoT enabled smart power quality analysis in three phase electrical systems with practical implementation. Scientific Reports, 15, 23317. https://doi.org/10.1038/s41598-025-23317-z

Meteab, F., Tousi, B., & Omran, M. (2025). Efficient self-healing framework for smart distribution networks. Scientific Reports, 15, 31124. https://doi.org/10.1038/s41598-025-16929-y

Muñoz, J., Tipán, L., Cuji, C., & Jaramillo, M. (2025). Resilient distribution system reconfiguration based on genetic algorithms considering load margin and contingencies. Energies, 18, 2889. https://doi.org/10.3390/en18112889

Ni, X., Lu, Y., Ding, C., Shang, Y., & Wang, Z. (2025). Self-healing strategy for distribution networks with AC flexible interconnection devices. Frontiers in Energy Research, 13, 1572606. https://doi.org/10.3389/fenrg.2025.1572606

Pang, J., Li, J., Wang, X., Jiang, Y., & Xue, F. (2025). Distribution network fault self-healing scheme based on network partition and flexible resource aggregation. Frontiers in Energy Research, 13, 1538033. https://doi.org/10.3389/fenrg.2025.1538033

Parvizi, P., Jalilian, M., Amidi, A. M., Zangeneh, M. R., & Riba, J.-R. (2025). Technical losses in power networks: Mechanisms, mitigation strategies, and future directions. Electronics, 14, 3442. https://doi.org/10.3390/electronics14173442

Puma, D. W., Molina, Y. P., Atoccsa, B. A., Luyo, J. E., & Ñaupari, Z. (2024). Distribution network reconfiguration optimization using a new algorithm hyperbolic tangent particle swarm optimization (HT-PSO). Energies, 17, 3798. https://doi.org/10.3390/en17153798

Samuel, I. A., Toyin, D. A., Somefun, T., Awelewa, A. A., & Abba-Aliyu, S. (2025). Evaluation and improvement of power quality of distribution networks: A case study of Covenant University, Ota. Frontiers in Energy Efficiency, 2, 1389622. https://doi.org/10.3389/fenef.2024.1389622

Su, S., Zhan, H., Zhang, L., Xie, Q., Si, R., Dai, Y., Gao, T., Wu, L., Zhang, J., & Shang, L. (2024). Volt-VAR control in active distribution networks using multi-agent reinforcement learning. Electronics, 13, 1971. https://doi.org/10.3390/electronics13101971

Suchithra, J., Rajabi, A., & Robinson, D. A. (2024). Enhancing PV hosting capacity of electricity distribution networks using deep reinforcement learning-based coordinated voltage control. Energies, 17, 5037. https://doi.org/10.3390/en17205037

Wang, T., Ma, S., Tang, Z., Xiang, T., Mu, C., & Jin, Y. (2023). A multi-agent reinforcement learning method for cooperative secondary voltage control of microgrids. Energies, 16, 5653. https://doi.org/10.3390/en16155653

Wang, Y., Guo, Y., Ning, H., Li, P., Cen, B., Zhao, H., & Zou, H. (2025). Research on power quality control methods for active distribution networks with large-scale renewable energy integration. Processes, 13, 1469. https://doi.org/10.3390/pr13051469

Wu, J., Wang, Z., Han, J., Li, Q., Sun, R., Li, C., Cheng, Y., Zhou, B., Guo, J., & Long, B. (2025). A novel data-driven multi-agent reinforcement learning approach for voltage control under weak grid support. Sensors, 25, 7399. https://doi.org/10.3390/s25237399

Wu, Y., Wang, L., Wan, Z., Liu, J., Fu, D., An, Y., & Zhang, X. (2024). Dynamic reconfiguration of multiobjective distribution networks considering the variation of load and DG using a novel LDEDBO algorithm. Scientific Reports, 14, 31524. https://doi.org/10.1038/s41598-024-83307-5

Xia, T., Lan, H., Fu, T., Hao, L., Wang, Q., & Wang, S. (2025). Ultra-short-term load forecasting and risk assessment method for distribution networks based on the VMD-DeepAR model. Frontiers in Energy Research, 13, 1692222. https://doi.org/10.3389/fenrg.2025.1692222

Zhang, L., Yang, F., Yan, D., Qian, G., Li, J., Shi, X., Xu, J., Wei, M., Ji, H., & Yu, H. (2024). Multi-agent deep reinforcement learning-based distributed voltage control of flexible distribution networks with soft open points. Energies, 17, 5244. https://doi.org/10.3390/en17215244

Zhao, Q., Han, Z., Wang, S., Dong, Y., & Qian, G. (2023). Coordinated control of multiple converters in model-free AC/DC distribution networks based on reinforcement learning. Frontiers in Energy Research, 11, 1202701. https://doi.org/10.3389/fenrg.2023.1202701

Zheng, J., Chen, R., Zhang, Z., Wei, X., & Xue, J. (2024). Analysis and comprehensive assessment method of power quality in advanced distribution networks based on complex network theory. Frontiers in Energy Research, 12, 1501006. https://doi.org/10.3389/fenrg.2024.1501006