©SHUTTERSTOCK.COM/FIT ZTUDIO

# Postdisaster Routing of Movable Energy Resources for Enhanced Distribution System Resilience

## A DEEP REINFORCEMENT LEARNING-BASED APPROACH

By Mukesh Gautam, Narayan Bhusal, and Mohammed Ben-Idris

*Digital Object Identifier 10.1109/MIAS.2023.3325046*
*Date of current version: 1 November 2023*

THE DEPLOYMENT OF MOVABLE ENERGY RESOURCES (MERs) CAN BE an effective strategy to restore critical loads to enhance power system resilience when no other energy sources are available after the occurrence of an extreme event. Since the optimal locations of MERs following an extreme event are dependent on system operating states (e.g., the loads at each node, on/off status of system branches, and so on), existing analytical and population-based approaches must repeat the entire analysis and calculation when the system operating states change. On the contrary, if deep reinforcement learning (DRL)-based algorithms are sufficiently trained with a wide range of scenarios, they can quickly find optimal or near-optimal locations irrespective of changes in system states. A deep Q-learning-based approach is proposed for optimal MER deployment to enhance power system resilience. MERs can be also utilized to complement other types of resources, if available. The proposed approach operates in two stages after the occurrence of extreme events.

1077-2618/23©2023IEEE

In the first stage, the distribution network is represented as a graph, and the network is then reconfigured using tie switches by using Kruskal's spanning forest search algorithm (KSFSA). To maximize critical load recovery, the optimal or near-optimal locations of MERs are chosen in the second stage. Case studies on a 33-node distribution system and a modified IEEE 123-node system demonstrate the effectiveness of the proposed approach for postdisaster routing of MERs.

## Introduction

Extreme events, both natural (such as hurricanes, wildfires, ice or hailstorms, and earthquakes) and man-made (such as cyber- and physical attacks), have become more frequent in the past 10 years [1]. Although the majority of loads can be kept in operation by modern power distribution systems in the face of typical weather-related disruptions, some high-impact low-probability extreme events, both man-made and weather-related, can still result in widespread power interruptions in distribution systems. For instance, Winter Storm Uri in Texas, USA, in February 2021, caused widespread power interruptions that left 4.5 million customers without power [2]. Also, during the year 2022 alone, there were 18 weather-related disasters in the United States, each of which cost more than $1 billion [3]. Such extreme events have caused damage to crucial power system components and prolonged power outages that affected the entire system. The goal of electric utilities to provide their consumers with a dependable and resilient electricity supply has been jeopardized by extreme weather events and subsequent outages. To minimize the impact of these events on end-user customers, efficient power distribution service restoration (PDSR) strategies must be implemented. By ensuring the optimal use of the available resources, PDSR's main objective is to minimize the duration of outages and load curtailments. The most efficient PDSR solutions in this context have been found to be based on smart grid technologies like microgrid (MG) formation, network reconfiguration, repair crew dispatch, distributed generation, energy storage, MERs, and combinations of these methods and techniques [4].

In the literature, several analytical and population-based intelligent search techniques for PDSR based on MERs have been developed to improve the reliability and resilience of the distribution system. A comprehensive survey of the literature on mobile energy storage for enhancing power system resilience has been provided in [5], where the modeling of power system constraints has also been discussed along with a cost–benefit analysis of MERs. A robust optimization framework based on two stages has been developed in [6] for routing and scheduling MERs to enhance the resilience of distribution systems. A two-stage PDSR strategy based on mixed-integer linear programming (MILP) has been proposed in [7] to enhance the seismic resilience of distribution systems with MERs. A MILP-based PDSR strategy has been proposed in [8] for an active distribution system, where routing and scheduling of mobile energy storage systems was performed for enhanced resilience. In [9], a two-stage optimization strategy has been proposed to enhance distribution system resilience with mobile energy storage units, where dynamic MG formation was also considered. In order to reduce the overall operating costs and enhance distribution system resilience, an innovative integrated restoration approach based on a stochastic MILP has been developed in [10] for the coordination of mobile energy storage fleets and MGs. A genetic algorithm-based approach has been developed in [11] to enhance distribution system reliability using MERs. The analytical and population-based intelligent search techniques utilized for PDSR based on MERs to enhance distribution system reliability and resilience have the following shortcomings. The accuracy and efficacy of analytical-based approaches are dependent on the accuracy of the models utilized, with accurate models imposing scalability challenges. Furthermore, mathematical models are typically derived using many approximations and require entire system information. Due to the enormous search space, population-based approaches, on the other hand, are computationally intensive, especially as system sizes increase.

Since learning-driven models can address uncertainty by extracting information from previous data, they have been utilized to overcome the shortcomings of analytical and population-based approaches [12], [13]. Furthermore, because of their capacity to employ information gathered from previous data to solve new scenarios, learning-driven models do not need to be solved whenever new scenarios are encountered [14]. RL-based systems are among the learning-driven approaches that can learn from experiences during online operations [15], [16]. Also, RL-based approaches are the best fit for online decision-making applications. Therefore, an RL-based approach for postdisaster MER dispatch is investigated in this article for distribution system resilience enhancement.

This article proposes a DRL-based framework for the postdisaster dispatch of MERs to enhance distribution system resilience. The proposed DRL approach is based on the training of a neural network (NN) that makes the best decision based on previous experiences [17], [18]. Given a specific decision, the sequential decision process gives a reward value as a function of the system outcome. The objective of the proposed DRL agent is to minimize critical load curtailment. To ensure realistic representation of distribution system operations, system constraints, including radiality and power balance constraints, are considered. In the training phase of the proposed framework, Q values are predicted using forward propagation of a deep NN (DNN). Actions are selected using the epsilon-greedy algorithm. When actions are passed through the training environment, the DRL agent gets rewarded (or penalized) based on its performance. Target Q values

are calculated based on the reward. The mean square error (MSE)—one of most commonly employed loss functions for regression—is computed using the predicted and target Q values. Errors are then backpropagated to update the weights of the DNN. The trained DRL agent is then used to find optimal or near-optimal locations for MER deployment. The proposed framework is validated through case studies on two different distribution test systems, and the results show that the proposed framework can effectively find an optimal network configuration and MER deployment locations, thereby minimizing critical load curtailment.

The remainder of the article is organized as follows. The mathematical formulation of the postdisaster reconfiguration and MER deployment problem is explained in the "Mathematical Modeling" section. The proposed framework and solution approach are described in the "RL for MER Routing" section. Case studies on two different distribution test systems are used to validate the proposed work in the "Case Study and Discussion" section. The "Conclusion" section provides some concluding remarks.

## Mathematical Modeling

This article combines network reconfiguration (the first stage) and MER routing (the second stage) to minimize load curtailments after extreme events. An introduction to graph theory, the graph theory-based modeling of the distribution network, and the mathematical formulation of the problem under study are presented in this section. In addition, states, actions, and the reward function are described in the context of the problem.

### *Graph Theory*

Graph theory refers to the study, modeling, and analysis of graphs. A graph is a framework built up of a collection of objects where some object pairs are conceptually "connected." The objects are represented by mathematical constructs known as *vertices* (sometimes known as *nodes* or *points*), and each pair of connected vertices is known as an *edge* (also referred to as a *link* or *line*) [19], [20]. A graph is typically shown diagrammatically as a collection of dots or circles representing the vertices and lines or curves representing the edges. The edges can be either directed or undirected. Mathematically, a graph is represented as a pair $G = (N, E)$, where $N$ is a set whose objects are called *nodes*, and $E$ is a set of connected nodes, whose objects are called *edges*.

The number of nodes in a graph determines the graph's size. In a graph, a path is a way that can be taken along edges and via nodes. A path's edges and nodes are all linked to one another. A cycle, also known as a *circuit*, is a path that starts and finishes at the same node. The length of a path or cycle is the sum of all of its edges. If there exists at least one edge connecting each pair of nodes and no node is directly connected to another node through an edge, the graph is referred to as a *connection graph* [21]. The term *tree graph* refers to a connection graph that is devoid of cycles. Equation (1) is satisfied in a tree graph with $|N|$ nodes and $|E|$ edges:

$$|N| = |E| - 1. \tag{1}$$

Equation (2) is used to calculate the number of cycles $N_{\text{cyc}}$ in a graph [21]:

$$N_{\text{cyc}} = (|E| + 1) - |N|. \tag{2}$$

### *Graph-Theoretic Modeling of Distribution Network*

Distribution systems are equipped with sectionalizing switches (normally closed) and tie switches (normally open). When all the switches of a distribution network are closed, a meshed network is formed, and the meshed network thus formed can be represented by an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is a set of nodes (or vertices) and $\mathcal{E}$ is a set of edges [22]. For the MER deployment problem proposed in this article, the status of tie switches is changed in such a way that radiality is always maintained, and MGs are formed after deployment of MERs.

#### Spanning Tree

A spanning tree (ST) is defined as a subset of the undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ that has a minimal number of edges linking all vertices (or nodes). In an ST, the number of edges is one less than the number of vertices. There are no cycles in an ST, and all the vertices are connected [23]. A linked graph can have many STs, each of which has the same number of edges and vertices. Each of the undirected graph $\mathcal{G}$'s edges has a specific value (or weights). The edge weights vary depending on the problem. The sum total of all edge weights of an ST is minimized when establishing the minimum cost ST. Figure 1 shows an ST of a hypothetical 12-node system. The ST shown in the figure consists of all system nodes (i.e., 12) and 11 closed branches (edges).

#### Spanning Forest

In graph theory, a forest is a disconnected union of trees. A spanning forest is a forest that covers all vertices of the undirected graph $\mathcal{G}$ and consists of a set of disconnected STs [23]. When all STs are connected, each vertex of the undirected graph $\mathcal{G}$ is included in one of the STs [24]. On the other hand, when a disconnected graph has many connected components, a spanning forest is formed, and it contains an ST of each component [25]. Figure 2 describes the spanning forest formed as a result of the disconnection of two addition branches (2–6 and 3–10) in the ST in Figure 1. The spanning forest in Figure 2 consists of three STs (ST-1, ST-2, and ST-3).

In this article, Kruskal's algorithm [26] is used to search for the optimal spanning forest. The KSFSA starts by constructing a forest $F$, with each graph vertex acting as a single tree based on the given undirected graph.

Since the KSFSA is a greedy algorithm, it goes on connecting the next least-weight edge that avoids a loop or cycle to the forest $F$ at each iteration. The resulting forest $F$ after the last iteration is the optimal spanning forest [22]. Figure 3 provides a flowchart of the KSFSA.

### *Problem Formulation*

This section presents the objective function and constraints of the problem under consideration.

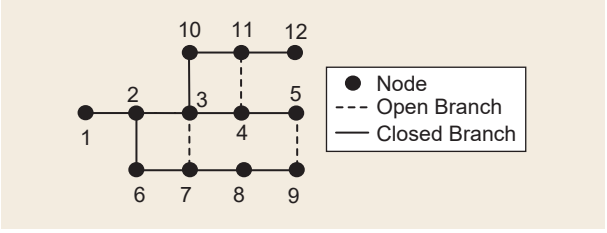


**FIGURE 1.** An ST of a hypothetical 12-node system.

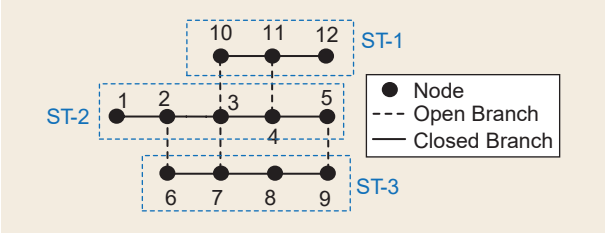


**FIGURE 2.** A spanning forest of the hypothetical 12-node system in Figure 1.

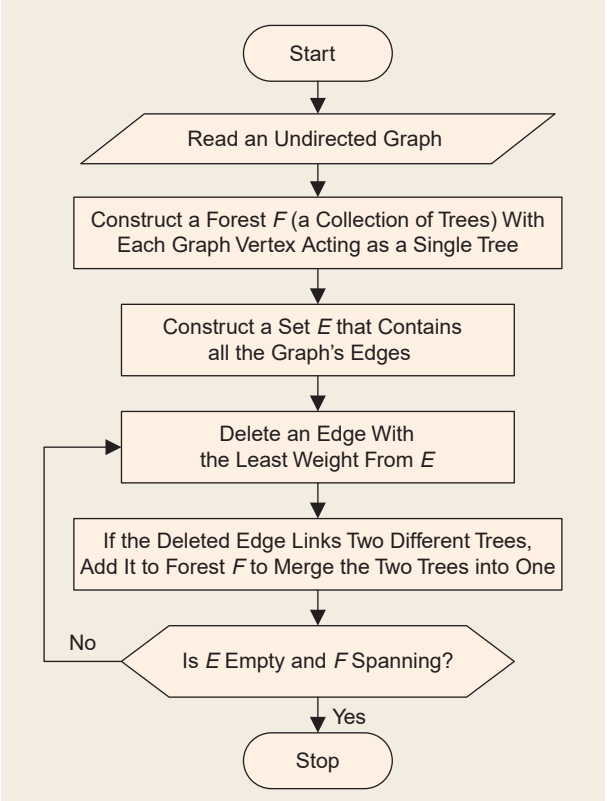


**FIGURE 3.** The KSFSA.

### Objective Function

As a result of an extreme event, some or all parts of the system may lose power. Under such a circumstance, tie switches should be used to reconfigure the network, and MERs should be deployed to enhance the distribution system's resilience. Therefore, the objective of the postdisaster MER routing problem under consideration is to minimize the critical load curtailment of the system since it can capture the severity of multiple line outages and is directly affected by the topology, or configuration, and MER deployment locations in a distribution system. Mathematically, the objective function of the critical load curtailment minimization is expressed as follows:

$$\text{Min} \sum_{i=1}^{N} \omega_i \Delta P_i \tag{3}$$

where $\Delta P_i$ is the load curtailment at node $I$, $\omega_i$ is the critical load factor at node $i$, and $N$ is the total number of nodes in the system.

### Constraints

The problem under consideration is subjected to various constraints, including nodal power balance constraints and a radiality constraint:

1) *Node power balance constraints*: The power balance constraint given by (4) ensures the balance of power at each node of the system (including MGs energized by MERs):

$$\sum_{j \in \Omega_g(j)} P_{g,j} + \sum_{l \in \Omega_L(j)} P_{l,j} = P_{D,j} \tag{4}$$

where $\Omega_g(j)$ is the set of sources (including MERs) connected to node $j$, $\Omega_L(j)$ is the set of lines connected to node $j$, $P_{g,j}$ is the power injected from source $j$, $P_{D,j}$ is the load at node $j$, and $P_{l,j}$ is the line power flow from node $l$ to node $j$.

2) *Radiality constraint*: A distribution system must always meet the radiality requirement. Therefore, each potential configuration should be radial (i.e., the radiality constraint should be met for each ST of the network). Each ST of the network is represented by a subgraph $\mathcal{G}_s = (\mathcal{N}_s, \mathcal{E}_s)$, where $\mathcal{N}_s$ is a set of nodes (or vertices) and $\mathcal{E}_s$ is a set of edges (or branches) in the subgraph. For the subgraph, a node branch incidence matrix should be constructed. If $n = |\mathcal{N}_s|$ denotes the number of nodes and $e = |\mathcal{E}_s|$ denotes the number of edges of a particular ST, then the node branch incidence matrix $A \in \mathbb{R}^{n \times e}$ is the matrix with element $a_{ij}$, calculated as follows [27]:

$$a_{ij} = \begin{cases} +1 & \text{if branch } j \text{ starts at node } i \\ -1 & \text{if branch } j \text{ ends at node } i \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

If the node branch incidence matrix $A$ is full ranked, then the radiality constraint is satisfied.

### States, Actions, and Reward Function

The choice of states, actions, and the reward function plays a critical role for the proper training of an RL agent. Therefore, states, actions, and the reward function must be chosen with careful consideration. For the MER routing problem under consideration, the state $S_t$ consists of the on/off status of each edge of the network and the amount of curtailed critical load at time step $t$:

$$S_t = \{s_{i,t} \mid i \in \Omega_E, LC_t\} \tag{6}$$

where $LC_t$ is the amount of curtailed critical load at time step $t$, $\Omega_E$ is the set of network edges, and $s_{i,t}$ denotes the status of network edge $i$ at time step $t$, determined as follows:

$$s_{i,t} = \begin{cases} 1 & \text{if network edge } i \text{ is closed} \\ 0 & \text{if network edge } i \text{ is open} \end{cases}, \forall i \in \Omega_E. \tag{7}$$

The action is a vector of MER deployment locations. The system state given by (6) consists of two different types of variables. The exogenous variable $s_{i,t}$ is independent of actions, whereas the endogenous variable $LC_t$ is affected by actions, i.e., MER deployment locations.

The RL agent is given a high positive reward when it reaches the optimal point and is not given any reward if the amount of curtailed critical load decreases. This motivates the agent to reach the optimal point faster instead of proceeding slowly. However, the RL agent is penalized by giving a negative reward when the amount of curtailed critical load remains constant or increases between two consecutive time steps. The total reward at time step $t$ is computed as follows:

$$R_t = \begin{cases} 100 & \text{if } LC_t^{\text{RL}} = LC^{\min} \\ 0 & \text{if } LC_t^{\text{RL}} < LC_{t-1}^{\text{RL}} \\ -5 & \text{if } LC_t^{\text{RL}} = LC_{t-1}^{\text{RL}} \\ -10 & \text{if } LC_t^{\text{RL}} > LC_{t-1}^{\text{RL}} \end{cases} \tag{8}$$

where $LC_t^{\text{RL}}$ and $LC_{t-1}^{\text{RL}}$ are critical load curtailments, respectively, at time steps $t$ and $t-1$ as a result of the action taken by the RL agent and $LC^{\min}$ is the minimum critical load curtailment for the particular line outage scenario.

## RL for MER Routing

This work leverages recently advanced RL techniques for post-disaster MER routing to minimize the amount of curtailed critical load. This section provides a brief overview of deep Q-learning and its training aspects.

### Deep Q-Learning

The four main integrands of an RL-based system are the policy, reward, value functions, and environment model. An agent decides what action to take based on the policy. The policy establishes a relationship between states and actions. When the agent performs a task, it is rewarded (or penalized). The value function determines the expected value of the cumulative reward when an agent follows a policy. There is a variety of algorithms for RL. A number of factors influence the choice of an algorithm, such as the nature of the states (continuous or discrete), the action space (continuous or discrete), and so on. The action space for the MER routing problem under consideration is discrete, making Q-learning an appropriate option for the task. Basic Q-learning, on the other hand, necessitates large lookup tables to store state–action values. As an action value function approximator, a DNN is employed to avoid the usage of large lookup tables. The addition of the DNN to the basic Q-learning framework transforms it into a deep Q-network (DQN). The goal of Q-learning is to simulate the Q-function, or, to put it another way, to predict the expected rewards for each action in a given state. The term *Q-function* here refers to a function that takes a state–action pair and returns the action value. The update rule for the action value function in Q-learning is defined as follows [15]:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \times [R_{t+1} + \gamma \times \max_a Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)] \tag{9}$$

where $A_t$ and $S_t$ are the action and state of an agent at the $t$th iteration, $Q(S_t, A_t)$ is the action value function at time step $t$, $Q(S_{t+1}, A_{t+1})$ is the action value function at time step $(t+1)$, $\alpha$ is the learning rate, and $\gamma$ is the reward discount factor.

Instead of updating the action value function iteratively, the DNN is trained, and the action value function's parameters are optimized to minimize the MSE

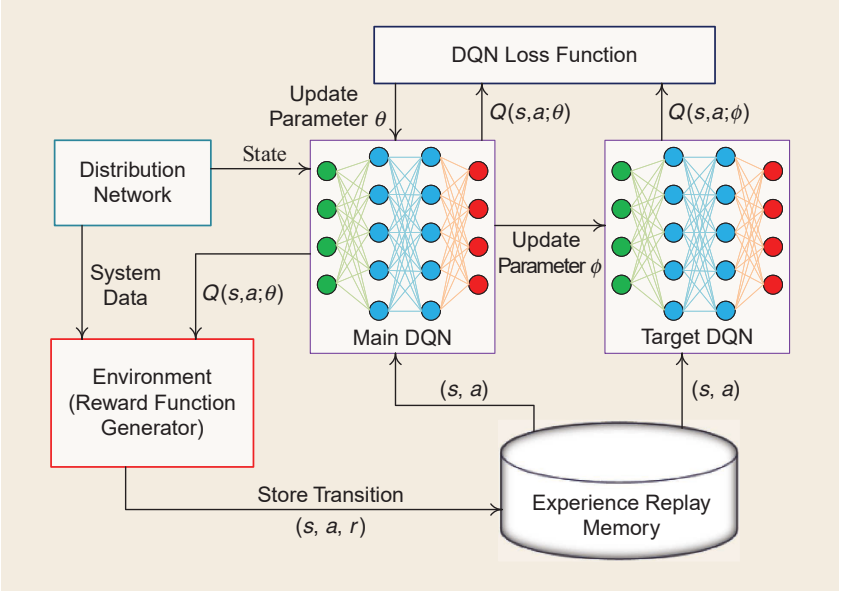


**FIGURE 4.** The training architecture of the proposed DRL model.

loss function (i.e., the regression loss function), which is expressed as follows [28]:

$$L(\theta) = \mathbb{E}[(Q(S_t, A_t \mid \theta) - R(S_t, A_t) + \gamma \times Q(S_t, A_t; \phi))^2] \quad (10)$$

**Algorithm 1. The training of the proposed DRL-based MER routing problem**

**Input**: System data, including line data, load data, the on/off status of branches, and so on.
Initialize experience replay memory $\mathcal{M}$.
Initialize parameters $\theta$ of the main DQN with random values.
Set target DQN parameters $\phi$ equal to the main DQN parameters; i.e., $\phi \leftarrow \theta$.
**for** *episode* $\leftarrow 1$ **to** $n_{ep}$ **do**
  Initialize the system with a random state (here, a vector of the line/branch status and load curtailment).
  **for** $t \leftarrow 1$ **to** $T$ **do**
    Generate action value function Q based on the current state.
    Calculate the reward function $R_{t+1}(S_t, A_t)$ after passing the state and action value function through the reward generator.
    Append the experience replay memory $\mathcal{M}$ with transition $(S_t, A_t, R_{t+1}(S_t, A_t))$.
    **if** *length*$(\mathcal{M}) >$ *batch_size* **then**
      Randomly select a minibatch.
      Calculate the DQN loss function based on the main Q-function and target Q-function.
      Perform backpropagation to update parameters $\theta$ of the main DQN.
      Periodically update parameters $\phi$ of the target DQN.
**Output**: MER deployment locations.

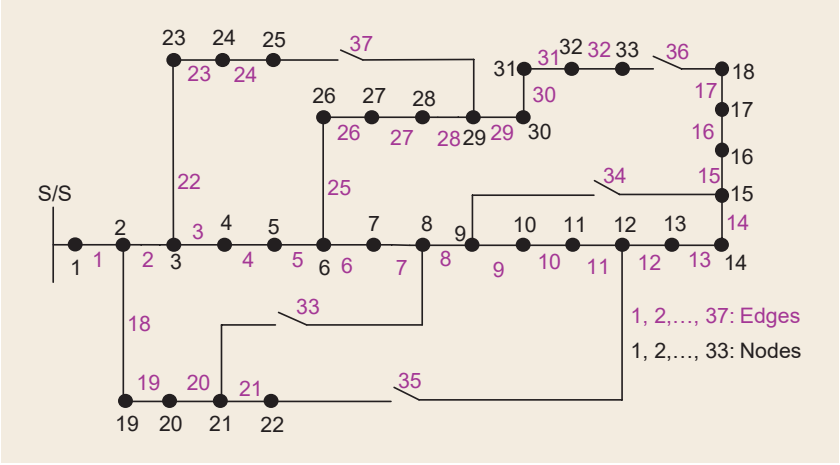


**FIGURE 5.** The 33-node distribution test system. S/S: substation.

where $\mathbb{E}$ denotes the expectation operator, $\theta$ denotes the parameter of action value function $Q(S_t, A_t)$, $R(S_t, A_t)$ denotes the reward function at time step $t$, $\phi$ denotes the parameter of the target DQN, and $Q(S_t, A_t; \phi)$ denotes the action value function of the target DQN.

### *Training Attributes*

When a nonlinear function approximator, like an NN, is employed to represent the Q-function, RL becomes unstable or divergent. The causal links in the series of observations, the possibility that tiny changes to Q may drastically alter the agent's behavior and the distribution of the data, and the relationships between predicted and target Q values all contribute to this instability. The technique is applicable to many different applications and areas of stochastic search [29]. Experience replay, a bioinspired process that chooses from a random sample of previous actions rather than the most recent action, is employed in Q-learning [30]. As a result, causality in the observational order is eliminated, and fluctuations in the data distribution are smoothed [31].

The experience replay memory-based training of the DQN is performed for a certain number of episodes ($n_{ep}$). The parameters $\theta$ of the main DQN are initialized with some random values, and the parameters $\phi$ of the target DQN are set equal to that of the main DQN. Each episode starts by initializing the system with a random state, which is a vector of the on/off status of the network branches after reconfiguration and the amount of curtailed critical load. In each time step, the predicted Q values corresponding to each action are computed based on forward propagation of the DNN. For the selection of actions, the epsilon-greedy (exploration–exploitation) algorithm [32] is used. The value of the exploration rate, $\varepsilon$, is initialized at $\varepsilon_{max} = 1$. The exploration rate is kept constant (i.e., 1) up to 10% of the total episodes, is decreased at a constant rate from 10% to 80% of the total episodes, and is again kept constant to a minimum value $\varepsilon_{min} = 0.001$ for the last 20% of episodes. This ensures that the DQN sufficiently explores in the initial phase of training and exploits (i.e., searches in the most

**Table 1. The locations of critical loads for the 33-node system**

| Node | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 18 | 19 | 20 | 21 | 22 | 23 | 26 | 27 | 28 | 29 | 30 | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Critical load (kW) | 60 | 30 | 60 | 200 | 200 | 60 | 30 | 25 | 45 | 45 | 45 | 45 | 45 | 45 | 60 | 60 | 60 | 60 | 60 | 30 |

prospective regions) during the last episodes of the training. The exploration rate $\varepsilon_j$ at the $j$th episode can be expressed as follows:

$$\varepsilon_j = \begin{cases} \varepsilon_{\max} & \text{if } j \le 0.1 n_{\text{ep}} \\ \varepsilon_{j-1} - \Delta\varepsilon & \text{if } 0.1 n_{\text{ep}} < j < 0.8 n_{\text{ep}} \\ \varepsilon_{\min} & \text{if } j \ge 0.8 n_{\text{ep}} \end{cases} \tag{11}$$

where

$$\Delta\varepsilon = \frac{\varepsilon_{\max} - \varepsilon_{\min}}{0.7 \times n_{\text{ep}}}. \tag{12}$$

The experience replay memory is appended with transition $(S_t, A_t, R_{t+1}(S_t, A_t))$. MSE losses for each time step $t$ are computed based on (10) using the predicted Q value of the main DQN and target Q values. The parameters of the main DQN are updated by back-propagating these MSE losses. After a certain number of iterations, the parameters of the target DQN are periodically updated. Figure 4 exhibits the training architecture of the proposed DRL model. Algorithm 1 provides the procedure of training the proposed DRL-based MER routing problem.

## Case Study and Discussion

To demonstrate the effectiveness of the proposed approach, a 33-node system and a modified IEEE 123-node system are used for numerical simulations.

### *The 33-Node System*

#### System Description

The 33-node distribution test system is a radial distribution system with 33 nodes, 32 branches, and five tie lines (37 branches) [33]. As displayed in Figure 5, all branches (including tie lines) are numbered from one to 37. The system's overall load is 3.71 MW. The deployment of four MERs, each of capacity 300 kW, is considered. The locations and amount of critical load considered for the 33-node system are listed in Table 1. The hyperparameter settings of the main and target DQNs of the proposed framework for the 33-node system are provided in Table 2.

#### Training

The training of the proposed model for the 33-node system is performed for 20,000 episodes. The parameters $\theta$ of the main DQN are initialized with random values, and the parameters $\phi$ of the target network are set equal to $\theta$. In each episode, the system is initialized with a random state, and the action value function is generated based on the current state. The reward function is calculated by passing the state and action value function through the reward generator. Initially, the rewards are very low, but then they increase as the number of episodes increases. Figure 6 presents the actual rewards and the running mean

**Table 2. The hyperparameter settings of the main and target DQNs**

| Hyperparameter | 33-Node System | 123-Node System |
|---|---|---|
| Number of hidden layers | 3 | 3 |
| Hidden layer neurons | 10, 10, 10 | 10, 10, 20 |
| Learning rate | $10^{-3}$ | $10^{-3}$ |
| Reward discount factor | 0.99 | 0.99 |
| Output layer activation | Linear | Linear |
| Hidden layer activation | ReLU | ReLU |
| Optimizer | Adam | Adam |
| Replay memory size | 20,000 | 20,000 |
| Batch size | 600 | 1,000 |
| Target update rate | 1,500 iterations | 2,500 iterations |

ReLU: rectified linear unit.

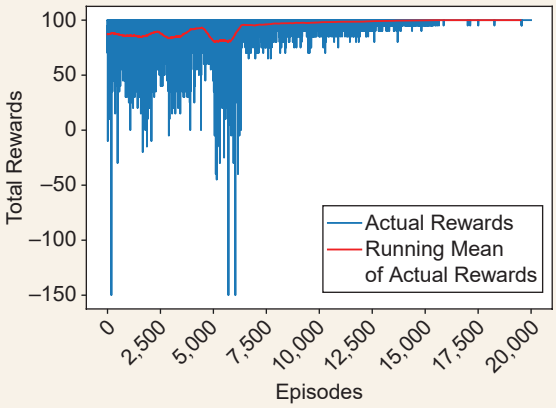


**FIGURE 6.** The learning curve of the proposed DRL model for the 33-node system.

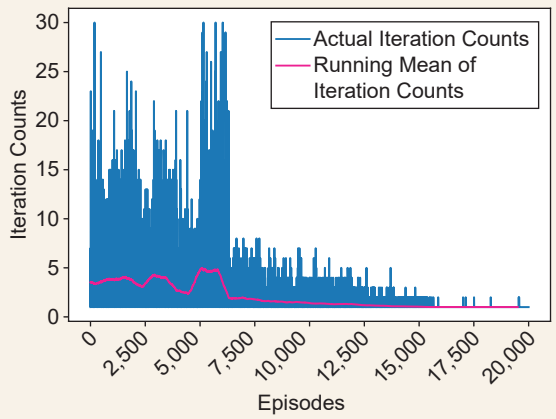


**FIGURE 7.** The total iteration counts during the training of the proposed DRL model for the 33-node system.

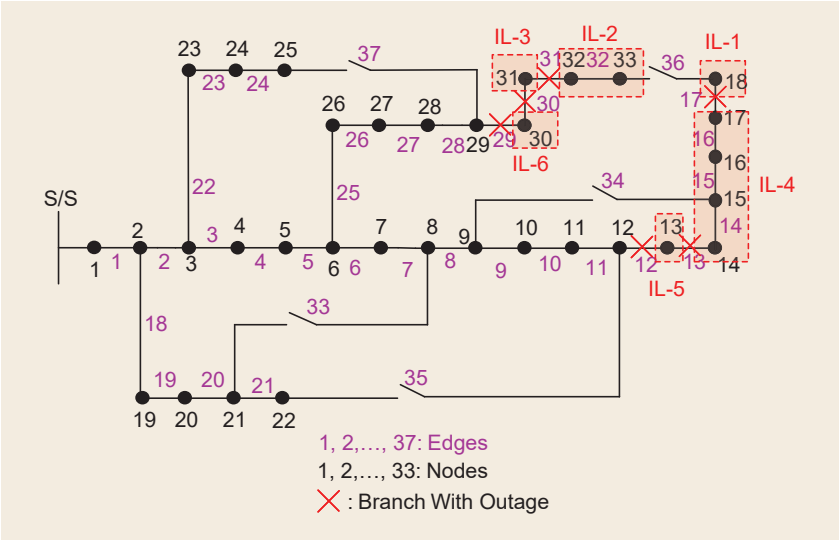


**FIGURE 8.** Test case 1 of the 33-node system before implementing the proposed approach.

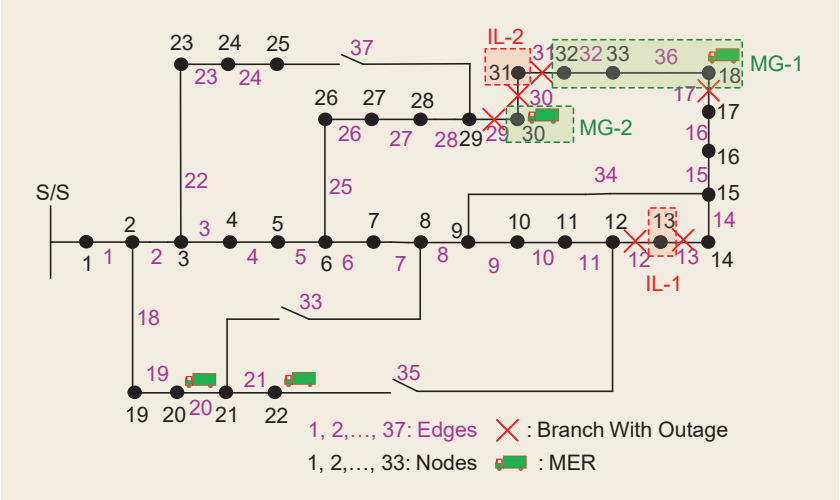


**FIGURE 9.** Test case 1 of the 33-node system after implementing the proposed approach.

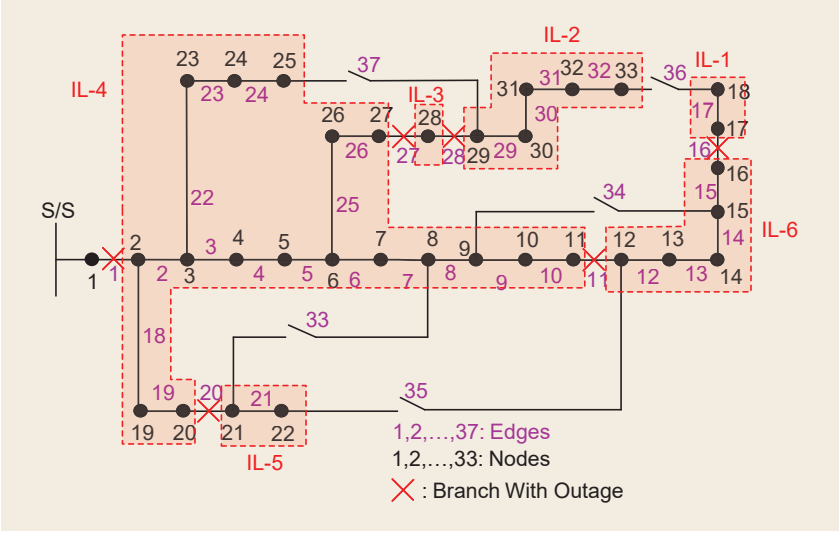


**FIGURE 10.** Test case 2 of the 33-node system before implementing the proposed approach.

(500-episode window) of the actual rewards as the episode progresses. It can be seen from the figure that as the number of episodes increases, the running mean of the reward increases and almost saturates after nearly 15,000 episodes. Similarly, Figure 7 shows the total iteration counts during the training of the proposed model. The total iteration count is the number of iterations taken by the DRL model to reach the optimal point, where the maximum iteration count is set to 30. The figure shows that the iteration count is very high up to 7,000 episodes. However, as the training continues, the total iteration count decreases and becomes almost constant after 17,000 episodes.

## Testing and Implementation

For the testing and implementation of the trained model, two test cases are devised with different line outage scenarios. The two test cases are explained in the following:

1) *Test case 1*: In this case, the outage of six lines, 12, 13, 17, 29, 30, and 31, is simulated. Due to the outage of these lines, six isolates (ILs) (IL-1, IL-2, IL-3, IL-4, IL-5, and IL-6) are formed, as in Figure 8. These ILs are devoid of power. This results in a total critical load curtailment of 135 kW.

   When the outage data are given as inputs to the proposed DRL model, two tie switches (34 and 36) are closed, and MERs are placed at nodes 18, 20, 22, and 30. This results in the formation of two MGs (MG-1 and MG-2) and two ILs (IL-1 and IL-2), as described in Figure 9. In both MG-1 and MG-2, the total generation exceeds the total critical loads, resulting in no critical load curtailments. The total critical loads in IL-1 and IL-1 are zeros as well. Therefore, the total amount of curtailed critical load after reconfiguration and MER deployment is 0 kW. The proposed approach is able to recover 135 kW of critical loads for the given outage scenario.

2) *Test case 2*: In this case, the proposed approach is tested with a more extreme outage scenario, where an outage of the line connected to the substation node (i.e., 1) is considered in addition to outage of lines 11, 16, 20, 27, and 28, as in Figure 10. Because of the outage of the line connected to the substation node, there is a power interruption at all system nodes, and the total critical load curtailment in this scenario is 1,265 kW.

After the implementation of the proposed DRL approach for this test case, four tie switches, 33, 34, 36, and 37, are closed, and MERs are placed at nodes 7, 22, 23, and 30. An MG, MG-1, and an IL, IL-1, are formed, as demonstrated in Figure 11. In MG-1, the amount of curtailed critical load is 5 kW. IL-1 consists of node 28, whose critical load is 60 kW. Therefore, the total amount of curtailed critical load is 65 kW. The total critical load recovered by the proposed approach is 1,200 kW for the given scenario.

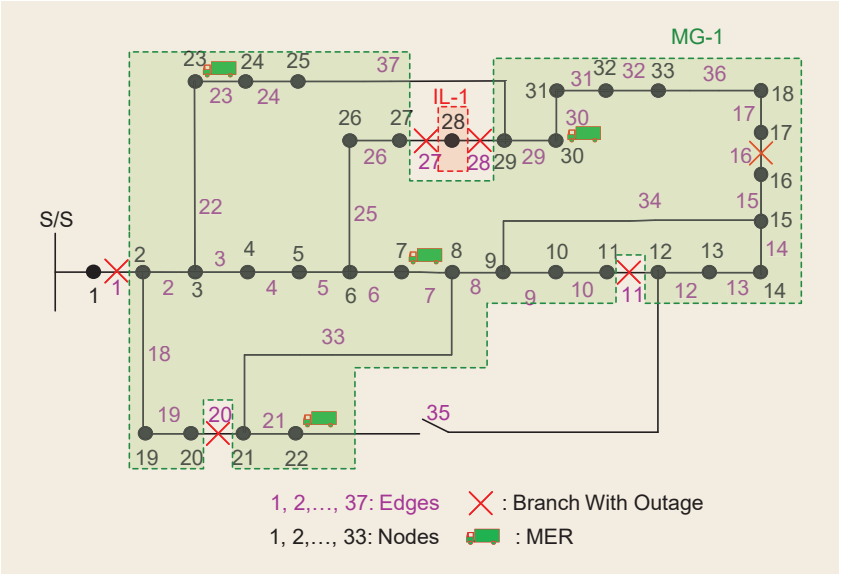


**FIGURE 11.** Test case 2 of the 33-node system after implementing the proposed approach.

## *IEEE 123-Node System*

### System Description

To further demonstrate the effectiveness and scalability of the proposed DRL-based approach, numerical simulation is also performed on the modified IEEE 123-node system

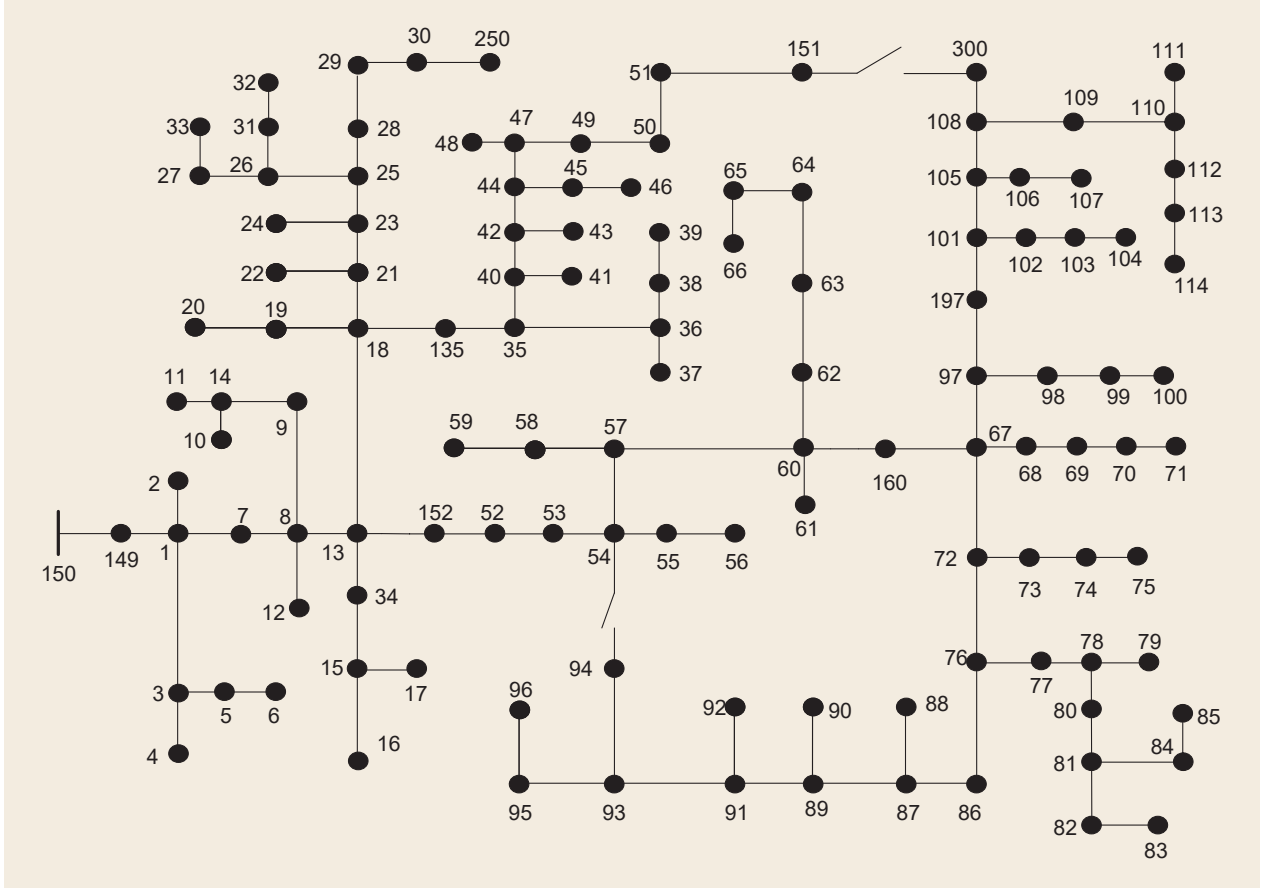

**FIGURE 12.** The modified IEEE 123-node system.

**Table 3. The locations of critical loads for the modified IEEE 123-node system**

| Node | 1 | 6 | 11 | 17 | 24 | 30 | 37 | 43 | 50 | 52 | 66 | 75 | 79 | 85 | 87 | 94 | 98 | 100 | 109 | 113 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Critical load (kW) | 40 | 40 | 40 | 20 | 40 | 40 | 40 | 40 | 40 | 40 | 75 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |

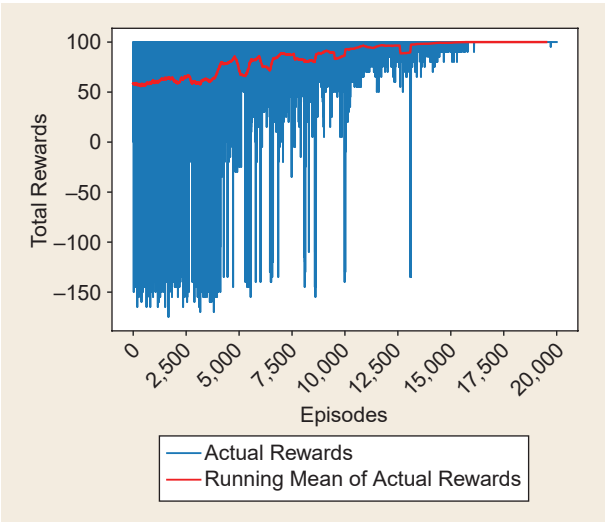


**FIGURE 13.** The learning curve of the proposed DRL model for the IEEE 123-node system.

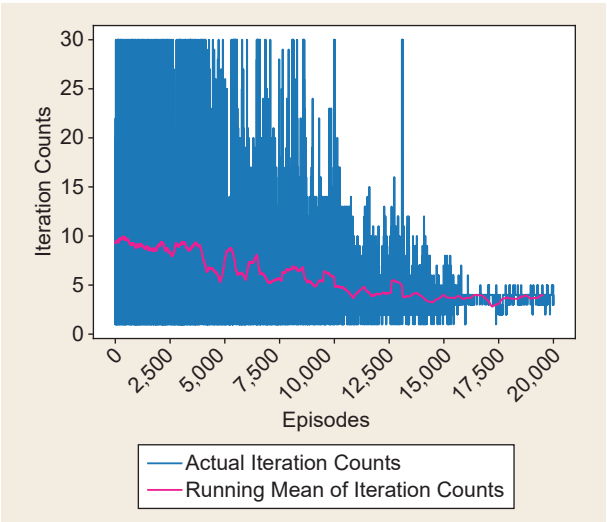


**FIGURE 14.** The total iteration counts during the training of the proposed DRL model for the IEEE 123-node system.

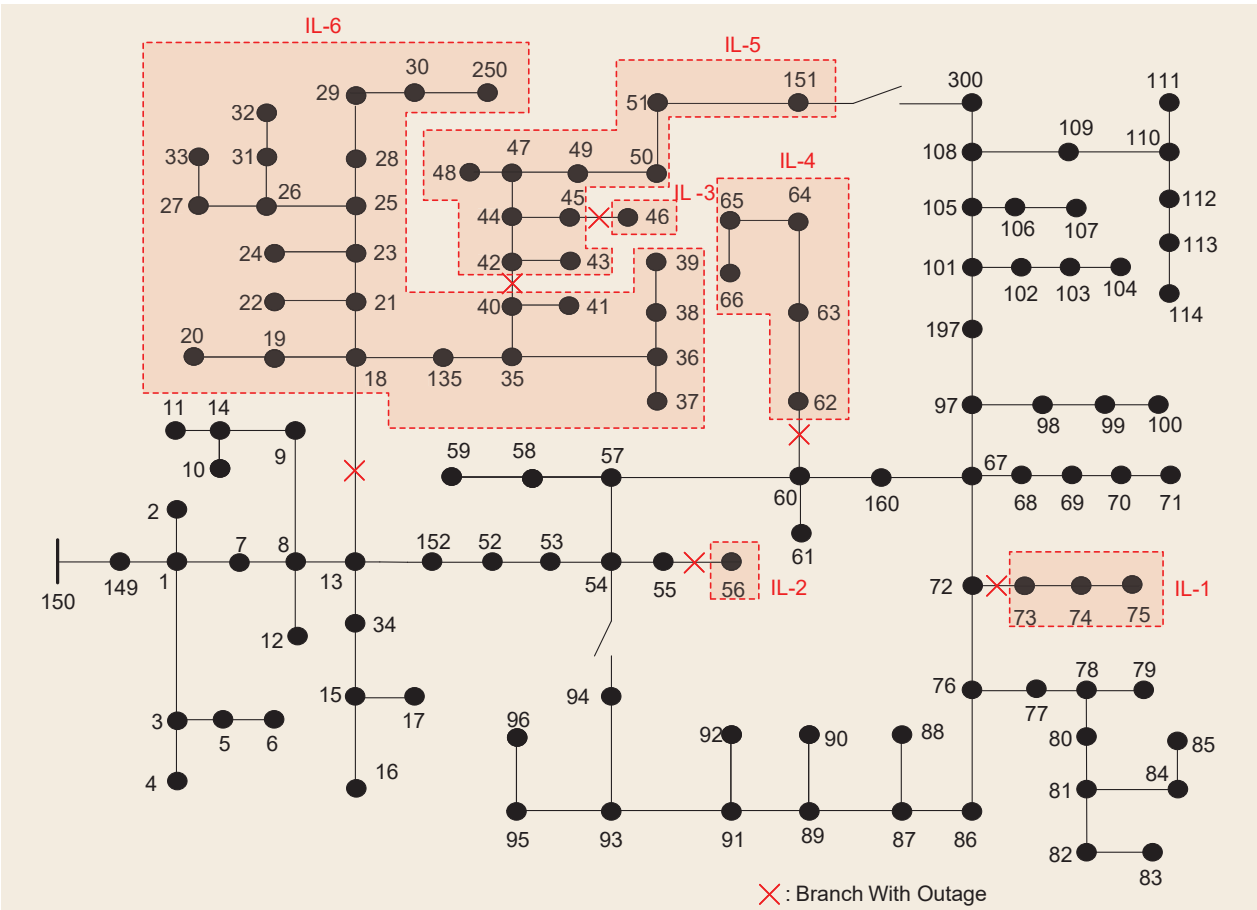


**FIGURE 15.** Test case 1 for the IEEE 123-node system before implementing the proposed approach.

in Figure 12. The modified IEEE 123-node system shown in the figure consists of 123 nodes and 126 branches. Out of the 126 branches, two of them (94–54 and 151–300) are equipped with tie switches. All branches and loads are assumed to be balanced. The deployment of five MERs, each of capacity 160 kW, is considered. The locations and amount of critical load considered for the system are given in Table 3.

### Training

Similar to the 33-node system, the proposed DRL-based model is trained for 20,000 episodes. Figure 13 displays the learning curve of the proposed DRL model. The figure shows that the total reward is very low in the early episodes of training, but it continuously increases as the episodes progress and becomes almost constant after 17,000 episodes. Similarly, Figure 14 shows the total iteration count during the training of the proposed DRL-based model. The figure shows that the total iteration count is very high up to 11,000 episodes. However, the total iteration count decreases as the training progresses and stays in the range of one to five after 16,000 episodes.

### Testing and Implementation

Similar to the 33-node system, two test cases with different line outage scenarios are devised for the testing and implementation of the trained DRL model in the case of the IEEE 123-node system. The two test cases are explained as follows:

1) *Test case 1*: In this case, the outage of six lines (13–18, 40–42, 45–46, 55–56, 60–62, and 72–73) is simulated. Due to the outage, six ILs (IL-1, IL-2, IL-3, IL-4, IL-5, and IL-6) are formed, as in Figure 15. The total amount of curtailed critical load under this scenario is 315 kW.

   After the implementation of the proposed approach, tie switch 151–300 is closed, and MERs are deployed at nodes 13, 27, 64, 70, 97, and 101, as illustrated in Figure 16. For this configuration, two MGs (MG-1 and MG-2) and three ILs (IL-1, IL-2, and IL-3) are formed as shown in the figure. Since the generation exceeds the critical load in both MG-1 and MG-2, they do not have any curtailed critical loads. IL-1 has a critical load of 40 kW at node 75, whereas IL-2 and IL-3 do not have any critical loads. Therefore, the total amount of curtailed critical load is 40 kW after

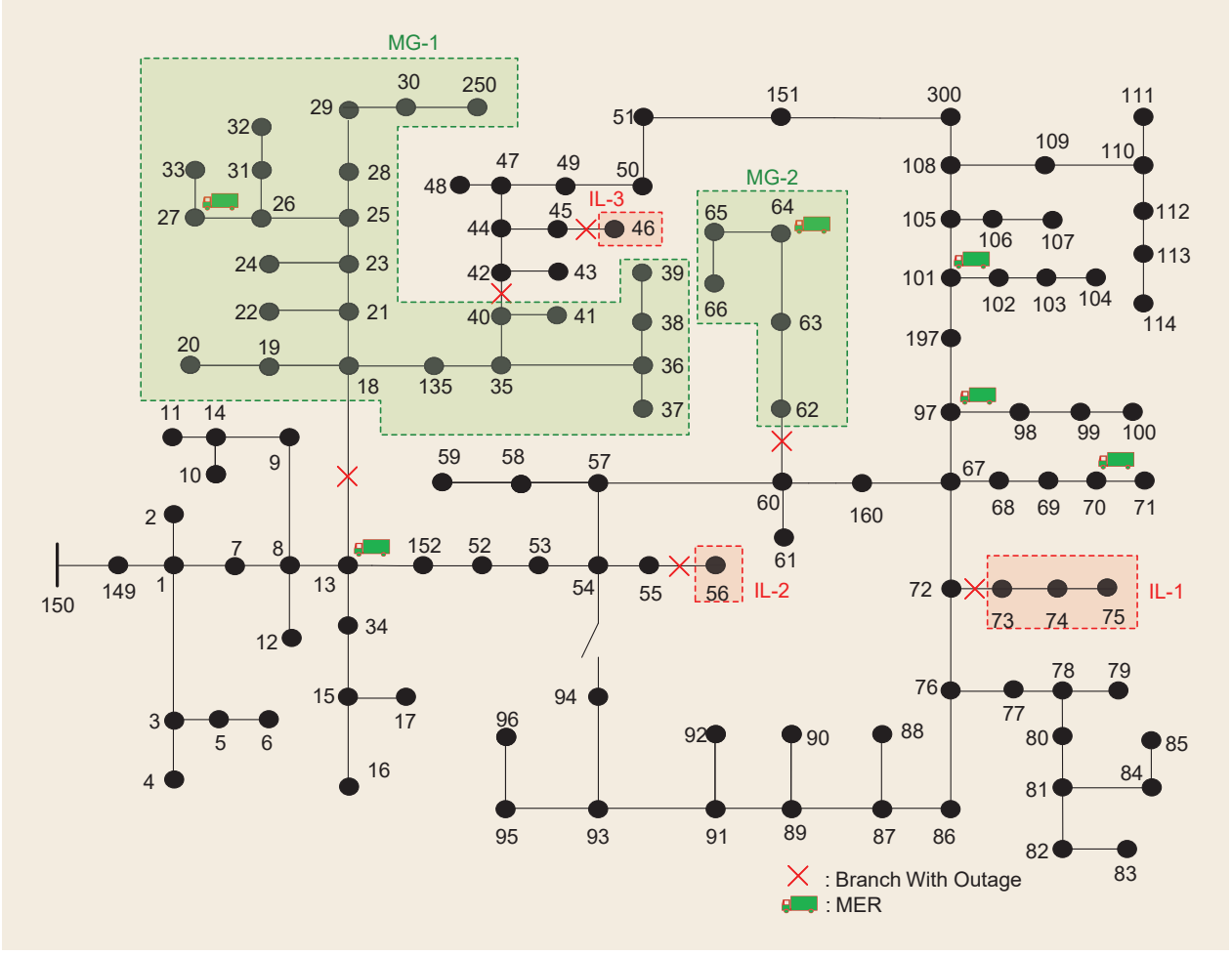


**FIGURE 16.** Test case 1 for the IEEE 123-node system after implementing the proposed approach.

implementing the proposed DRL model, recovering 275 kW of the critical loads.

2) *Test case 2*: In this test case, the outage of six lines (34–15, 52–53, 63–64, 67–68, 102–103, and 108–300) is simulated. As a result of the outage, six ILs (IL-1, IL-2, IL-3, IL-4, IL-5, and IL-6) are formed, as in Figure 17. The total amount of curtailed critical load under this scenario is 455 kW before the implementation of the proposed approach.

After the implementation of the proposed DRL-based model, tie switch 151–300 is closed, and MERs are deployed at nodes 1, 13, 64, 81, 93, and 101, as demonstrated in Figure 18. For this configuration, two MGs (MG-1 and MG-2) and three ILs (IL-1, IL-2, and IL-3) are formed. In each of the MGs, the total generation exceeds the total critical load. IL-1 and IL-2 do not have any critical loads, whereas IL-3 has a critical load of 20 kW at node 17. Therefore, the total amount of curtailed critical load is 20 kW, recovering 435 kW of the critical loads.

**The goal of electric utilities to provide their consumers with a dependable and resilient electricity supply has been jeopardized by extreme weather events.**

### *Comparison*

The proposed DRL-based approach is compared with an exhaustive search technique. In the exhaustive search technique, all possible candidates of the solution are enumerated, and the critical load curtailments are computed for each candidate solution to get the best solution. For each test case of both 33-node and 123-node systems, the exhaustive search technique is used to find the optimal MER deployment locations that result in minimum critical load curtailment. For each test case, the same values of critical load curtailments are obtained for both the exhaustive search technique and the proposed DRL-based approach.

Although the same values of critical load curtailments can be obtained using both approaches, the proposed approach outperforms the exhaustive search technique in terms of execution time. The analyses are

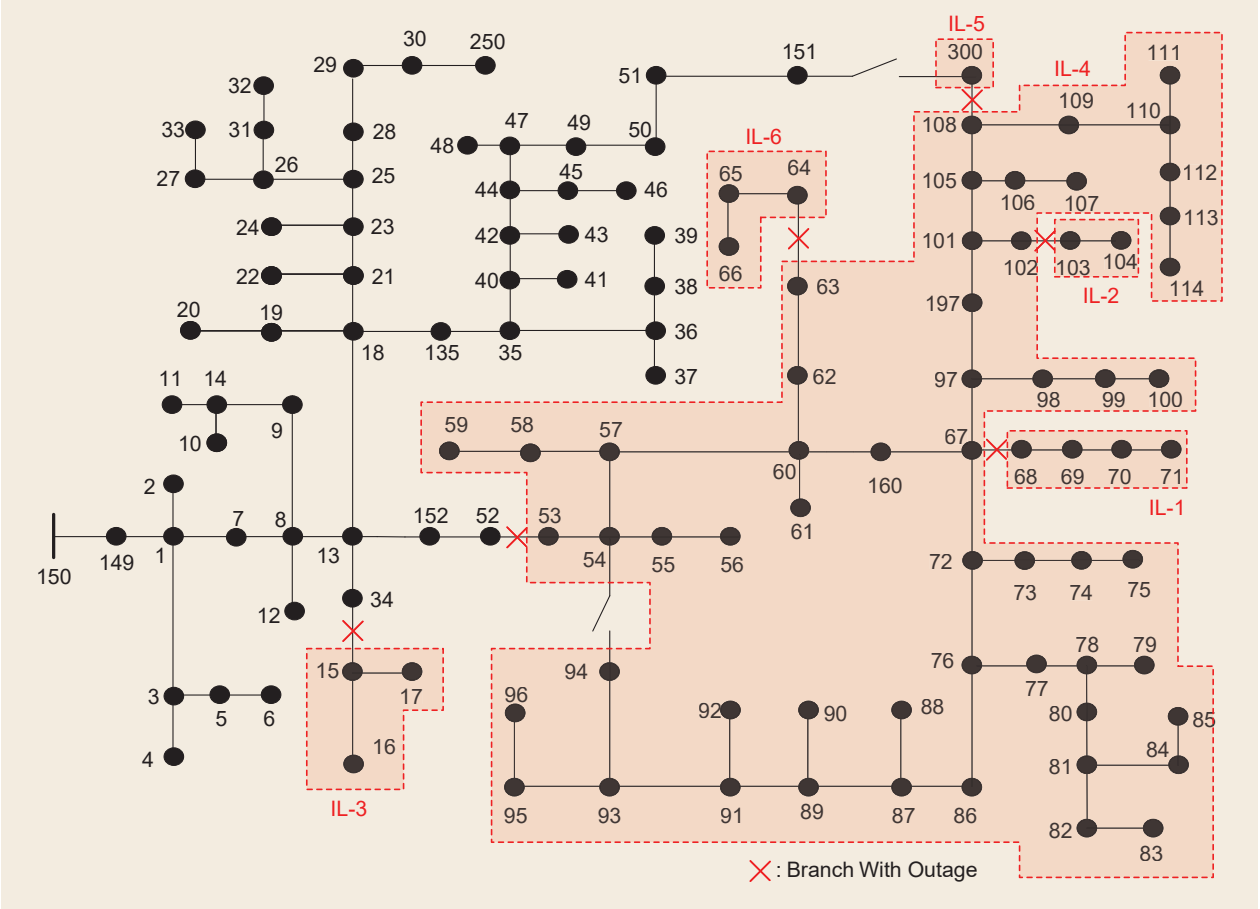


**FIGURE 17.** Test case 2 for the IEEE 123-node system before implementing the proposed approach.

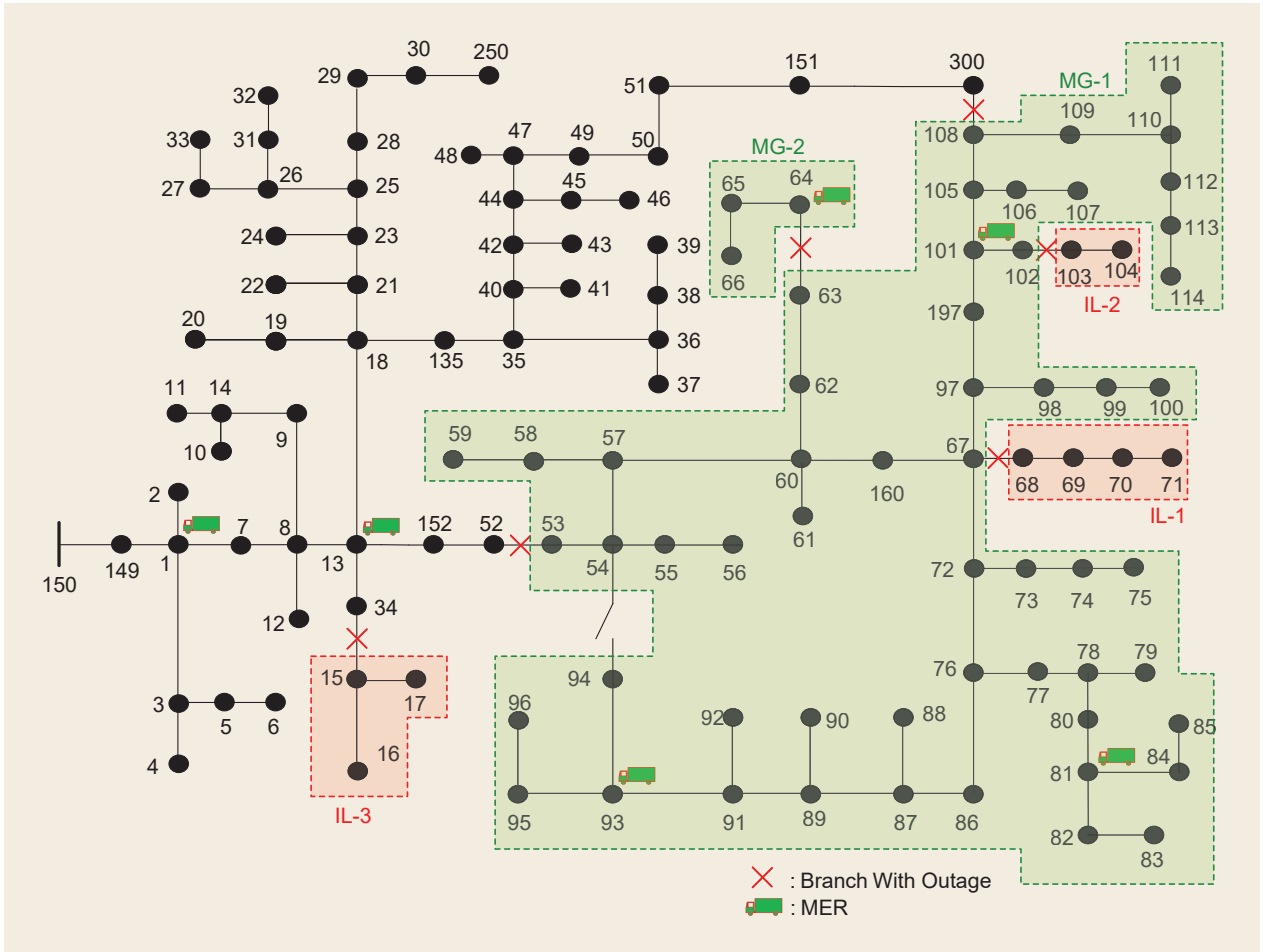


**FIGURE 18.** Test case 2 for the IEEE 123-node system after implementing the proposed approach.

**Table 4. A comparison of the execution times**

| System | Case | Execution Time (Exhaustive Search) | Execution Time (Proposed Approach) |
|---|---|---|---|
| 33-node system | Test case 1 | 0.248 s | 0.006 s |
| | Test case 2 | 0.239 s | 0.007 s |
| 123-node system | Test case 1 | 1.276 s | 0.017 s |
| | Test case 2 | 1.31 s | 0.022 s |

performed on a PC with a 64-b Intel i5 Core processor running at 3.15 GHz, with 8 GB of random-access memory and the Windows operating system. Table 4 displays the execution time for different test cases using both approaches. The execution time of the proposed approach ranges from 6 to 22 ms for both 33-node and 123-node systems, while the execution time of the exhaustive search technique is significantly higher than the proposed approach. The ratios of the execution time of the exhaustive search technique and the proposed approach are approximately 40 and 75, respectively, in the 33-node system and the 123-node system. This shows that the proposed approach is highly scalable compared to the exhaustive search technique.

## Conclusion

This article has proposed a DRL-based two-stage approach for network reconfiguration and MER routing to minimize the amount of curtailed critical load when multiple line outages occur following an extreme event. In the first stage, distribution network reconfiguration was performed using tie switches. In the second stage, MERs were utilized to form MGs. The distribution network was represented by an undirected graph, and the optimal spanning forest was formed. The proposed approach was tested and implemented on a 33-node system and a modified IEEE 123-node system. Two test cases in each of the systems exhibit the effectiveness of the proposed approach for recovering critical loads of the system by leveraging MERs and forming MGs, thereby enhancing system resilience. Additionally, the proposed DRL-based approach was found to be computationally efficient compared to the exhaustive search technique, indicating its

effectiveness for enhancing distribution systems' operational resilience.

## Acknowledgment

This work was supported by the U.S. National Science Foundation (NSF), under Grant NSF 1847578. This work was performed when Mukesh Gautam and Narayan Bhusal were with the Department of Electrical and Biomedical Engineering, University of Nevada, Reno, Reno, NV 89557 USA.

## Author Information

***Mukesh Gautam*** (mukesh.gautam@nevada.unr.edu) is with Idaho National Laboratory, Idaho Falls, ID 83415 USA. ***Narayan Bhusal*** is with Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA. ***Mohammed Ben-Idris*** is with Michigan State University, East Lansing, MI 48824 USA. Gautam is a Member of IEEE. Ben-Idris is a Senior Member of IEEE. This article first appeared as A Deep Reinforcement Learning-Based Approach to Post-Disaster Routing of Movable Energy Resources (doi: 10.1109/IAS54023.2022.9940073) at the 2022 IEEE Industry Applications Society (IAS) Annual Meeting. It was reviewed by the IAS Energy Systems Committee.

## References

[1] M. Gautam, N. Bhusal, and M. Benidris, "A deep reinforcement learning-based approach to post-disaster routing of movable energy resources," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, 2022, pp. 1–6, doi: 10.1109/IAS54023.2022.9940073.

[2] J. W. Busby et al., "Cascading risks: Understanding the 2021 winter blackout in Texas," *Energy Res. Social Sci.*, vol. 77, Jul. 2021, Art. no. 102106, doi: 10.1016/j.erss.2021.102106.

[3] "U.S. billion-dollar weather and climate disasters," NOAA National Centers for Environmental Information, Asheville, NC, USA, 2023. [Online]. Available: https://www.ncei.noaa.gov/access/billions/

[4] M. Gautam and M. Benidris, "Pre-positioning of movable energy resources for distribution system resilience enhancement," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, 2022, pp. 1–6, doi: 10.1109/SEST53650.2022.9898487.

[5] J. Dugan, S. Mohagheghi, and B. Kroposki, "Application of mobile energy storage for enhancing power grid resilience: A review," *Energies*, vol. 14, no. 20, 2021, Art. no. 6476, doi: 10.3390/en14206476.

[6] S. Lei, C. Chen, H. Zhou, and Y. Hou, "Routing and scheduling of mobile power sources for distribution system resilience enhancement," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5650–5662, Sep. 2019, doi: 10.1109/TSG.2018.2889347.

[7] Z. Yang, P. Dehghanian, and M. Nazemi, "Enhancing seismic resilience of electric power distribution systems with mobile power sources," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, 2019, pp. 1–7, doi: 10.1109/IAS.2019.8912010.

[8] M. Nazemi, P. Dehghanian, X. Lu, and C. Chen, "Uncertainty-aware deployment of mobile energy storage systems for distribution grid resilience," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3200–3214, Jul. 2021, doi: 10.1109/TSG.2021.3064312.

[9] J. Kim and Y. Dvorkin, "Enhancing distribution resilience with mobile energy storage: A progressive hedging approach," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, 2018, pp. 1–5, doi: 10.1109/PESGM.2018.8585791.

[10] S. Yao, P. Wang, X. Liu, H. Zhang, and T. Zhao, "Rolling optimization of mobile energy storage fleets for resilient service restoration," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1030–1043, Mar. 2020, doi: 10.1109/TSG.2019.2930012.

[11] F. M. Rodrigues, L. R. Araujo, and D. R. Penido, "A method to improve distribution system reliability using available mobile generators," *IEEE Syst. J.*, vol. 15, no. 3, pp. 4635–4643, Sep. 2021, doi: 10.1109/JSYST.2020.3015154.

[12] M. Gautam, M. Abdelmalak, M. MansourLakouraj, M. Benidris, and H. Livani, "Reconfiguration of distribution networks for resilience enhancement: A deep reinforcement learning-based approach," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, 2022, pp. 1–6, doi: 10.1109/IAS54023.2022.9939854.

[13] M. Gautam, R. Hossain, M. MansourLakouraj, N. Bhusal, M. Benidris, and H. Livani, "A deep reinforcement learning-based reserve optimization in active distribution systems for tertiary frequency regulation," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, 2022, pp. 1–5, doi: 10.1109/PESGM48719.2022.9916741.

[14] R. Hossain, M. Gautam, M. MansourLakouraj, H. Livani, M. Benidris, and Y. Baghzouz, "Soft actor critic based volt-var co-optimization in active distribution grids," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, 2022, pp. 1–5, doi: 10.1109/PESGM48719.2022.9916976.

[15] A. Zai and B. Brown, *Deep Reinforcement Learning in Action*. Shelter Island, NY, USA: Manning Publications, 2020.

[16] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.

[17] M. Gautam, N. Bhusal, and M. Benidris, "Deep Q-Learning-based distribution network reconfiguration for reliability improvement," in *Proc. IEEE/PES Transmiss. Distribution Conf. Expo. (T&D)*, 2022, pp. 1–5, doi: 10.1109/TD43745.2022.9817000.

[18] R. Hossain, M. Gautam, J. Thapa, H. Livani, and M. Benidris, "Deep reinforcement learning assisted co-optimization of VOLT-VAR grid service in distribution networks," *Sustain. Energy, Grids Netw.*, vol. 35, Sep. 2023, Art. no. 101086, doi: 10.1016/j.segan.2023.101086.

[19] R. J. Trudeau, *Introduction to Graph Theory*. North Chelmsford, MA, USA: Courier Corporation, 2013.

[20] M. Gautam, "Distribution system resilience enhancement using movable energy resources," Ph.D. dissertation, University of Nevada, Reno, USA, 2022.

[21] M. Assadian, M. M. Farsangi, and H. Nezamabadi-pour, "GCPSO in cooperation with graph theory to distribution network reconfiguration for energy saving," *Energy Convers. Manage.*, vol. 51, no. 3, pp. 418–427, 2010, doi: 10.1016/j.enconman.2009.10.003.

[22] M. Gautam and M. Benidris, "A graph theory and coalitional game theory-based pre-positioning of movable energy resources for enhanced distribution system resilience," *Sustain. Energy, Grids Netw.*, vol. 35, Sep. 2023, Art. no. 101095, doi: 10.1016/j.segan.2023.101095.

[23] A. Gibbons, *Algorithmic Graph Theory*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[24] A. Salles da Cunha, L. Simonetti, and A. Lucena, "Formulations and branch-and-cut algorithm for the K-rooted mini-max spanning forest problem," in *Proc. Int. Conf. Netw. Optim.*, Berlin, Germany: Springer-Verlag, 2011, pp. 43–50.

[25] I. K. Ahani, M. Salari, S. M. Hosseini, and M. Iori, "Solution of minimum spanning forest problems with reliability constraints," *Comput. Ind. Eng.*, vol. 142, Apr. 2020, Art. no. 106365, doi: 10.1016/j.cie.2020.106365.

[26] A. Katsigiannis, N. Anastopoulos, K. Nikas, and N. Koziris, "An approach to parallelize Kruskal's algorithm using helper threads," in *Proc. IEEE 26th Int. Parallel Distrib. Process. Symp. Workshops PhD Forum*, 2012, pp. 1601–1610, doi: 10.1109/IPDPSW.2012.201.

[27] M. Gautam and M. Benidris, "Distribution network reconfiguration using deep reinforcement learning," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, 2022, pp. 1–6, doi: 10.1109/PMAPS53380.2022.9810652.

[28] D. Cao et al., "Reinforcement learning and its applications in modern power and energy systems: A review," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1029–1042, Nov. 2020, doi: 10.35833/MPCE.2020.000552.

[29] B. Matzliach, I. Ben-Gal, and E. Kagan, "Detection of static and mobile targets by an autonomous agent with deep Q-learning abilities," *Entropy*, vol. 24, no. 8, 2022, Art. no. 1168, doi: 10.3390/e24081168.

[30] S. Adam, L. Busoniu, and R. Babuska, "Experience replay for real-time reinforcement learning control," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 2, pp. 201–212, Mar. 2012, doi: 10.1109/TSMCC.2011.2106494.

[31] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.

[32] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[33] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Power Eng. Rev.*, vol. 9, no. 4, pp. 101–102, Apr. 1989, doi: 10.1109/MPER.1989.4310642.