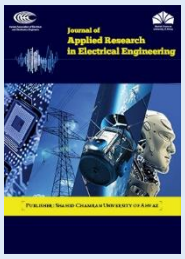
Shahid Chamran University of Ahvaz

Iranian Association of Electrical and Electronics Engineers

Journal of Applied Research in Electrical Engineering

E-ISSN: 2783-2864
P-ISSN: 2717-414X
Homepage: https://jaree.scu.ac.ir/



Research Article

An Analytical Approach to Distribution Network Reconfiguration Based on Deep Reinforcement Learning

Amirhossein Ghaemipour[1], Habib Rajabi Mashhadi[1,2,*], and Seyed Hossein Mostafavi[1]

[1] *Department of Electrical Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*

[2] *Center of Excellence on Soft Computing and Intelligent Information Processing (SCIIP), Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*

** Corresponding Author: h_mashhadi@um.ac.ir*

Abstract: This paper proposes a deep reinforcement learning (DRL) framework based on the dueling deep Q network (Dueling DQN) algorithm for distribution network reconfiguration (DNR), aiming to minimize power losses and voltage deviations. Since the training process in DRL relies heavily on trial-and-error, accurate environment modeling is essential. Therefore, this study comprehensively investigates various DRL modeling strategies, including different state representations and reward function formulations. Individual DRL agents are trained for each configuration to determine the most effective modeling approach for the DNR problem. The proposed methodology is evaluated on the modified IEEE 16-, 33-, and 69-bus distribution networks. Furthermore, the impact of uncertainties in distributed generation and load demand is analyzed to demonstrate the flexibility and robustness of the proposed method.

Keywords: Deep reinforcement learning, Distribution network reconfiguration, Various modeling, Analytical Approach.

Article history
Received 06 February 2026; Revised 18 May 2026; Accepted 10 June 2026; Published online 1 August 2026.
© 2026 Published by Shahid Chamran University of Ahvaz & Iranian Association of Electrical and Electronics Engineers (IAEEE)
How to cite this article
A. Ghaemipour, H. Rajabi Mashhadi, and S. H. Mostafavi, " An Analytical Approach to Distribution Network Reconfiguration Based on Deep Reinforcement Learning," *J. Appl. Res. Electr. Eng.*, vol. 4, no. 2, pp. 216-229, 2026.
DOI: https://doi.org/10.22055/jaree.2026.51747.1292

Check for updates

1. INTRODUCTION

Electricity is one of the most essential human needs [1][2]. The electricity generation process consists of three basic stages: generation, transmission, and distribution. Numerous studies have been conducted to increase the performance of these sections [3][4]. The distribution network (DN) is the closest part of the electricity transmission process to customers. As a result, they have a high impact on customer satisfaction. Therefore, various solutions are employed to improve the performance of this section, including distributed generation (DG) and batteries [5], tap changers [6], demand response planning, and distribution network reconfiguration (DNR) [7]. DNR is a practical approach to controlling and optimizing DNs, in which the distribution network operator (DSO) uses switches within the DN to optimize its operation under varying conditions [8]. Numerous studies have been conducted in this field, using various algorithms to find the optimal arrangement of DNs. Some of these studies are mentioned below.

DNR in [9] is implemented to minimize power losses and voltage deviations. This model is tested on a DN using a hybrid of the honey bee mating optimization (HBMO) and the shuffled frog leaping algorithm (SFLA). DNR aims to reduce power losses and voltage deviations, which are modeled in [10]. This approach is simulated on a DN using the bacterial foraging optimization algorithm (BFOA). DNR in [11] is modeled using adaptive modified particle swarm optimization (AMPSO). This method aims to minimize power losses, voltage deviations, and operating costs on two DNs. Uncertainty in load and wind turbines is also considered in this article. Dynamic reconfiguration for two DNs has been performed in [12]. This approach uses the modified honey bee mating optimization (MHBMO) algorithm to minimize energy losses and operating costs and to increase the voltage stability index (VSI). The DGs, including photovoltaic (PV) panels and batteries, are incorporated into this model. In [13], the DNR was performed utilizing a genetic algorithm (GA) and an enhanced genetic algorithm (EGA). This study aims to reduce power losses and improve the reliability of DN by

modeling and testing two DN. In [14], dynamic reconfiguration of the DN is performed to minimize operating costs, energy losses, and energy not supplied (ENS). This method is implemented using the hybrid improved particle swarm optimization-artificial bee colony optimization algorithm, accounting for DGs. The fireworks algorithm (FWA) was used in [15] to simultaneously and separately perform DNR and determine DG locations. This modeling was conducted to reduce power losses and voltage deviations on the DN. Various loadings were employed in this article. In [16], dynamic distribution network reconfiguration (DDNR) is formulated as a multi-objective problem. This model is designed using the improved particle swarm optimization (IPSO) algorithm to minimize energy losses, ENS, and operating cost. In this method, uncertainty in resources, the time-of-use (TOU) mechanism, and the demand response program (DRP) are also considered.

Mixed-integer quadratically constrained programming (MIQCP) and mixed-integer linear programming (MILP) algorithms are employed in [17] for DNR. These models have been tested across seven DNs to reduce power losses and voltage deviations. In [18], DNR is employed to reduce power losses and voltage deviations in a DN. This article uses the modified plant growth simulation algorithm (MPGSA) and employs DGs, such as wind turbines and PV panels. The DNR and DG location in [19] are used to reduce power losses and voltage deviations. It has been exploited using the uniform voltage distribution-based constructive algorithm (UVDA). This approach has been tested on two DNs. DNR is modeled using the runner-root algorithm (RRA) in [20]. This technique is analyzed under different objective functions, including minimizing power losses and voltage deviation, load balancing between lines and feeders, and minimizing the number of switching operations across two DNs. In [21], a novel graph-based bi-level optimization model is used. In the first step, DNR seeks to decrease power losses and voltage deviation. In the second step, the locations of the disconnector switches and the telecommunication load breaker switches are tested across two DNs to minimize power losses, voltage deviations, and ENS. The DNR and finding capacitor locations for two DNs have been determined using the adaptive whale optimization algorithm (AWOA) in [22]. This approach seeks to reduce power losses, voltage deviations, and operating costs. Reference [23] introduces the reconfiguration of several DNs using feasibility-preserving evolutionary optimization (FPEO). This method was evaluated with the goal of decreasing voltage deviation and power loss. In [24], the optimal configuration for two DNs was determined to minimizing voltage deviation and power loss. This modeling was simulated using binary group search optimization (BGSO). Several meta-heuristic algorithms, such as ant colony (AC), have been proposed in [25] to reduce power loss and voltage deviation. This model was applied to finding the optimal configuration for several DNs.

In [26], the DNR has been modeled to minimize power losses and voltage deviations. This procedure is applied to two DNs and programmed using the enhanced binary cuckoo search algorithm (EBCSA). In [27], two DNs were reconfigured utilizing a discrete particle swarm optimization (DPSO) algorithm. This experiment was conducted to reduce voltage deviation and power losses. The reconfiguration of six DNs in [28] was performed to minimize power losses and voltage deviations. This article employs an analytical method called the H-matrix. DNR has been performed in [29] along with scheduling for electric vehicle charging and discharging. This model has been analyzed using modified grey wolf optimization (MGWO) to minimize power losses and voltage deviations in two DNs. DNR is described in [30] employing switch opening and exchange (SOE) and mathematical programming (MP) methods. This paper aims to minimize power losses and voltage deviations on several DNs. DNR has been simulated in [31] using a modified sequential switch-opening and exchange (MSSOE). This approach is implemented to reduce power losses and voltage deviations and has been applied to several DNs employing DGs. DNR has been performed in [32] using a chaotic stochastic fractal search algorithm (CSFSA). This technique has been applied to several DNs to minimize voltage deviation and power losses. Reference [33] is a comprehensive study on the DNR applied to various DNs. This model is formulated as a mixed-integer conic programming (MICP) problem to reduce power losses and voltage deviations. DNR and on-site generation's location are done in [34]. This seeks to decrease power loss and voltage deviation and is evaluated on several DNs. This model was evaluated by mixed-integer quadratic programming (MIQP). The DNR and DG location has been developed in [35]. This approach is done by the harmony search algorithm (HSA) and tested on two DNs.

With the development of optimization algorithms and their use across various fields of electricity, artificial intelligence (AI) is a powerful method for optimizing multiple aspects of the electricity industry. Several studies have applied AI to the electricity industry, such as forecasting DGs [36][37], fault location and detection [38], and DN management optimization. The following studies that have used AI algorithms for DNR are discussed.

In [39], the reconfiguration of two DNs is modeled to minimize power losses and voltage deviations. This research employs several deep reinforcement learning (DRL) algorithms, including deep Q network (DQN), double deep Q network (DDQN), and NoisyNet deep Q network (NoisyNet-DQN). DNR employs several DRL algorithms as described in [40]. The list of modeled algorithms in this article includes proximal policy optimization (PPO), dueling deep Q network (Dueling DQN), DQN, soft actor-critic (SAC), and DQN with prioritized experience replay. This approach was tested on two DNs to minimize power loss and voltage deviation, using action sampling to reduce the action space. DNR was modeled utilizing DQN and Dueling DQN algorithms in [41]. This method employs action-space reduction, called power-flow-based action sampling, and applies it to three DNs to reduce voltage deviation and power loss. DNR is modeled on a DN in [42]. This model is trained using the DQN algorithm to decrease power loss and voltage deviation. In [43], DNR is modeled using a hybrid of an explainer neural network and a reinforcement learning (RL) agent. This method was evaluated on two DNs with four loadings to decrease power losses and voltage deviations. Online reconfiguration of DN is modeled in [44] using the DQN algorithm. This modeling aims to maintain DN stability by controlling its constraints. This method has been tested on two DNs equipped with PV panels and wind turbines. Dynamic reconfiguration of two DNs in [45] has been performed to reduce switching and power loss costs. This modeling has been investigated and simulated using different loadings using the DQN algorithm. Restorative dynamic distribution network reconfiguration

(RDDNR) is modeled in [46] for two DNs under variable conditions. This modeling is simulated using DQN and SAC algorithms to minimize switching operations, voltage fluctuations, and line overload, and to minimize the areas affected by outages. A convolutional neural network (CNN) model for stochastic distribution network reconfiguration (SDNR) is suggested in [47]. This essay seeks to decline power losses and enhance the VSI on two DNs.

In [48], a new method called a loss-based experience replay buffer was introduced to improve the learning speed of DRL algorithms for DNR. This method was applied to three DNs while considering uncertainty and probabilistic analysis, with the objective of minimizing voltage deviation and power loss.

In [49], the long short-term memory (LSTM) algorithm is utilized to determine the optimal microgrid structure. This modeling aims to improve reliability and VSI while minimizing voltage deviation, power losses, and operating costs across two DNs. In this modeling, uncertainty in load, DGs, and cyber-attacks is considered. The DDNR for two DNs in [50] has been performed using the guaranteed convergence particle swarm optimization (GCPSO) and honeybee-graph (HG) algorithms, as well as DQN. This modeling has been done to improve reliability and reduce power losses and voltage deviations. Also, uncertainty in the load and DGs has been considered. In [51], DRL was used for simultaneous DNR and radial topology identification in DNs. This method was evaluated on two DNs with the aim of minimizing power loss and voltage deviation.

In DRL, formulating an accurate environment is crucial; suboptimal modeling can mislead the agent, whereas a well-designed framework significantly enhances learning performance. While prior studies on DNR typically describe their finalized models, they often fail to justify the selection of specific components. To address this gap, this paper systematically investigates which state representations and reward functions are most effective for the DNR problem. Across our comparative analysis, the underlying Dueling DQN agent, the physical DN, and the action space are kept identical, with variations introduced strictly in the state space definitions and reward formulations. We examine state variables adopted from existing literature alongside newly proposed features to determine their specific impact on the agent's learning efficiency. Furthermore, several reward function designs, targeting power loss minimization and voltage deviation reduction, are rigorously evaluated. The proposed methodology is tested on the modified IEEE 16-, 33-, and 69-bus DNs to identify the optimal modeling strategy.

The rest of the paper is organized as follows. The second section describes the mathematical modeling. The third section describes the DRL algorithm and its components. The fourth section introduces the simulation and results, and the fifth section analyzes the simulation results. The final section presents the conclusion.

2. MATHEMATICAL FORMULATION OF DNR

In this section, the mathematical formulations and physical constraints governing the DNR are detailed.

2.1. Power Flow Analysis

Two popular methods for power flow in power systems include:

- Gauss-Seidel method
- Newton-Raphson method

The Newton-Raphson method is the most commonly used because it converges rapidly with fewer iterations. Initially, the method is grounded in KCL, and estimates are then made of the magnitudes and angles of the voltages at the load and generator buses. Then, for each equation in the system representing power balance, a Taylor series is developed, neglecting higher-order terms. This results in a linear system of equations, as shown below [52][53].

$$[I_{bus}] = [Y_{bus}].[V_{bus}] \quad (1)$$

$$Y_{ik} = G_{ik} + jB_{ik} \quad (2)$$

$$\begin{bmatrix} \Delta\phi \\ \Delta|V| \end{bmatrix} = -J^{-1}\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} \quad (3)$$

$$\Delta P_i = -P_i + \sum_{k=1}^{N_{bus}} |V|_i\,|V|_k(G_{ik}\cos\phi_{ik} + B_{ik}\sin\phi_{ik}) \quad (4)$$

$$\Delta Q_i = -Q_i + \sum_{k=1}^{N_{bus}} |V|_i\,|V|_k(G_{ik}\sin\phi_{ik} - B_{ik}\cos\phi_{ik}) \quad (5)$$

$$J = \begin{bmatrix} \frac{\partial\Delta P}{\partial\phi} & \frac{\partial\Delta P}{\partial|V|} \\ \frac{\partial\Delta Q}{\partial\phi} & \frac{\partial\Delta Q}{\partial|V|} \end{bmatrix} \quad (6)$$

Where I_{bus} represents the current matrix, Y_{bus} represents the admittance matrix, and V_{bus} represents the voltage matrix of the buses. And G_{ik} and B_{ik} represent the conductance and susceptance between buses i and k, respectively. ΔP and ΔQ represent the mismatch equation, and J represents the Jacobian matrix. The next estimation *(m+1)* of voltage magnitude and angles is obtained by solving the linearized system of equations shown below [54].

$$\phi^{(m+1)} = \phi^m + \Delta\phi \quad (7)$$

$$|V|^{(m+1)} = |V|^m + \Delta|V| \quad (8)$$

This process continues until a termination condition is met. A common stopping point is when the norm of the mismatch equations falls below a specified threshold. In this article, the PandaPower package is employed for power flow analysis [55].

2.2. Constraints

2.2.1. Bus Voltage Limit

The bus voltage must be between the the minimum and maximum allowable values as follows.

$$V_{Min} \le V_i \le V_{Max};\ i = 1,2,\dots,N_{bus}. \quad (9)$$

2.2.2. Line Current Limit

The line current must be between its upper and lower limits as follows.

$$0 \le I_i \le I_{Max};\ i = 1,2,\dots,N_{branch}. \quad (10)$$

2.2.3. Radial structure Limit

One of the most basic constraints when operating DNs is the radial structure of its components. The DSO must choose the DN arrangement to ensure this constraint is satisfied. In the absence of network radiality, errors may occur in the DN's protection devices. The number of tie switches in DNs must be constant and equal to the number of DN loops. As a result, the number of tie switches in DNs is calculated from the following relationship [32].

$$N_{ts} = N_{branch} - (N_{bus} - N_s) \quad (11)$$

Where N_{ts} represents the number of tie switches, N_{branch} and N_{bus} denote the number of lines and buses, respectively, and N_s shows the number of resources in the DNs.

2.2.4. Wind turbine

The power output of a wind turbine depends on the turbine model and wind speed. It is calculated using the following equation [49]:

$$P_t^{WT} = \begin{cases} 0; & 0 \le v_{w,t} < v_{ci} \\ P_{ra}\left(A + Bv_{w,t} + Cv_{w,t}^2\right); & v_{ci} \le v_{w,t} < v_{ra} \\ P_{ra}; & v_{ra} \le v_{w,t} < v_{co} \\ 0; & v_{w,t} \ge v_{co} \end{cases} \quad (12)$$

which P_t^{WT} represents the turbine's power output. v_{ci} , v_{co}, and v_{ra} denote the cut-in, cut-out, and rated wind speeds, respectively. P_{ra} is the rated power output, while A, B, and C are the specific coefficients for the turbine model.

3. RL ALGORITHM

RL is a machine learning paradigm that aims to maximize a predefined reward signal, which represents the objective specified by the problem designer. In addition to the reward, RL comprises four main components: the agent, the environment, the action, and the state. The agent is the decision-making entity that interacts with the environment and selects actions according to a policy in order to maximize the cumulative reward. The environment represents the external system with which the agent interacts and determines the system's response to the agent's actions. An action refers to any decision or operation performed by the agent that affects the environment.

The state represents the current configuration of the environment as perceived by the agent. In a Markov state, the current state contains all the information required to predict the next state and reward, such that the future is conditionally independent of the past given the present state. This property is known as the Markov property. If the state does not fully capture the system dynamics, the problem is considered non-Markovian, and additional historical information may be required for optimal decision-making. The overall structure of the RL framework is depicted in Fig. 1.

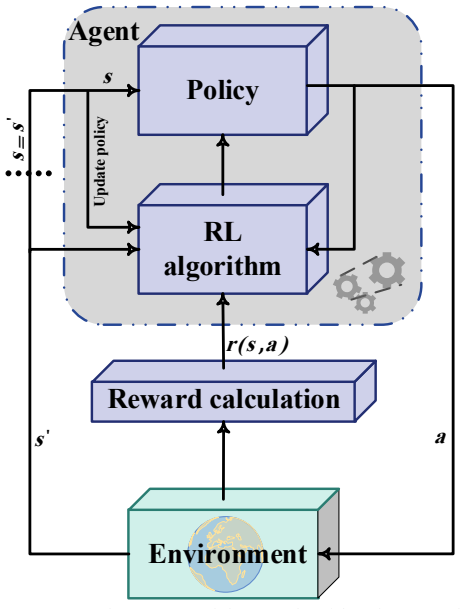


Fig. 1: General structure of the RL algorithm framework.

RL algorithms are generally categorized as on-policy or off-policy. In on-policy methods, the policy used to interact with the environment is the same as the policy being learned. In contrast, in off-policy methods, the behavior policy differs from the target policy being optimized. A well-known on-policy algorithm is SARSA, in which the state–action value function is updated as follows:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma Q(s',a') - Q(s,a)) \quad (13)$$

An example of an off-policy algorithm is Q-learning, where the update rule is given by:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r + \max_{a'} Q(s',a') - Q(s,a)) \quad (14)$$

In these equations, α represents the learning rate.

3.1. DRL Algorithm

In this study, the state of the problem is defined in various forms. A critical point is that the state dimensionality in the DNR problem is high, which makes the use of DRL essential. One of the DRL algorithms is DQN. This algorithm utilizes an experience replay buffer, where experiences are stored in the form of (s_t, a_t, r_t, s_t'). At each time step, a random mini-batch is sampled from the buffer, and the loss function is computed using the following equation [56][57]:

$$L(\theta) = \frac{1}{K}\sum_{i=1}^{K}(y_i - \hat{y}_i)^2 \quad (15)$$

Where K is the batch size, and y_i and $\hat{y}_i$ are defined as follows:

$$\hat{y}_i = Q_\theta(s_i, a_i) \quad (16)$$

$$y_i = \begin{cases} r_i & if\ s'\ is\ terminal \\ r_i + \gamma \max_{a'} Q_{\theta'}(s_i', a') & if\ s'\ is\ not\ terminal \end{cases} \quad (17)$$

Where θ and θ' show the main and target parameters of the neural network, respectively.

One extension of DQN is the Dueling DQN algorithm [58]. In this architecture, the neural network branches into two separate streams to estimate the state value function (V-function) and the advantage function (A-function), respectively. The final action-value function (Q-function) is then computed as follows:

$$Q_{[\theta_1,\theta_2,\theta_3]}(s,a) = V_{[\theta_1,\theta_3]}(s) + \left(A_{[\theta_1,\theta_2]}(s,a) - \frac{1}{\mathcal{A}}\sum_{a'} A_{[\theta_1,\theta_2]}(s,a')\right) \quad (18)$$

In these equations, $\theta_1$represents the weights of the shared layers, θ_2 denotes the weights of the A stream, and θ_3 denotes the weights of the V stream. Additionally, $\mathcal{A}$ indicates the number of possible actions in a given state. In this research, the Dueling DQN algorithm is employed as the baseline. This algorithm was selected primarily due to its high sample efficiency compared to other methods, its straightforward implementation, and its ability to deliver excellent performance in problems involving discrete action spaces, given that DNR is inherently a discrete problem. The components of the proposed method will be elaborated upon in the following section.

3.2. Problem Modeling using DRL

In the DNR problem, the agent is the DSO, responsible for opening and closing switches as actions. The environment is

the DN itself, which is linked to the PandaPower package in this article. Most prior studies have employed a single reward function for the DNR problem. However, in this research, to precisely investigate the impact of the reward function on the final solution, three distinct reward functions are introduced and compared. The total power loss is incorporated into the reward function via the following equation:

$$r_{loss,t} = -\frac{P_{total,t}}{P_{total,init}} \quad (19)$$

where $P_{total,t}$ represents the total line power losses in the current configuration, and $P_{total,init}$ denotes the total line power losses in the initial DN state. Essentially, this term aims to minimize the current power losses relative to the DN's initial power losses. To incorporate voltage deviation into the reward function, the following two options are available:

$$r_{mae,t} = -\frac{1}{N_{bus}} \sum_{i=1}^{N_{bus}} |v_i - v_{ref}| \quad (20)$$

$$r_{mse,t} = -\frac{1}{N_{bus}} \sum_{i=1}^{N_{bus}} (v_i - v_{ref})^2 \quad (21)$$

where v_i represents the voltage at the i-th bus, and v_{ref} denotes the reference voltage, which is typically one p.u. Based on the three terms introduced above, three reward functions are defined as follows:

$$Reward\ 1: \quad r_t = r_{loss,t} \quad (22)$$

$$Reward\ 2: \quad r_t = r_{loss,t} + \lambda_1 r_{mae,t} \quad (23)$$

$$Reward\ 3: \quad r_t = r_{loss,t} + \lambda_2 r_{mse,t} \quad (24)$$

Another important factor that significantly affects the agent's learning process is the state. The learning process depends on the agent's state at any given moment. In previous research, only three states were available to the DSO, whereas in DNs, data from those states may be unavailable under different conditions. Therefore, examining the variable effective states for DRL-based modeling of DNR is sensitive. In this paper, we modeled several states that the DN can experience under different conditions, as shown as follows:

$$S1 = [p_t, q_t] \quad (25)$$

$$S2 = [p_t, q_t, \alpha_t] \quad (26)$$

$$S3 = [v_t] \quad (27)$$

$$S4 = [v_t, \alpha_t] \quad (28)$$

$$S5 = [v_t, \phi_t] \quad (29)$$

$$S6 = [v_t, \phi_t, \alpha_t] \quad (30)$$

As a result, the Dueling DQN algorithm is trained under six modeled states. In these models, α_t denotes the connected/disconnected status of lines; v_t and ϕ_t denote the buses' voltage and voltage angle, respectively; p_t and q_t denote the active and reactive power of all buses, respectively. The investigated states essentially determine the number of input neurons in the utilized neural network. As the length of the state vector increases, the number of these input neurons correspondingly increases. In practice, the differences among the various state representations lead to structural changes in the neural network's input layer during implementation. Furthermore, when storing experiences in the replay buffer, more complex state representations require additional memory space.

As explained above, a detailed comparison of the proposed method's modeling of the structures of different DRL components with previous articles is presented in Table 1. An examination of the modeled sections reveals numerous similarities across the DRL models used in prior DNR research. The DN operator is an agent that seeks to find its optimal configuration in different conditions by interacting with the DN.

To do this, it performs the action selection process by changing the structure of the DN by disconnecting or connecting its lines, and receives a response from the environment accordingly. In this research, because the DSO seeks to minimize power losses and voltage deviations, these parameters are incorporated into the reward function. As a result, through trial and error, the DSO moves toward the optimal policy and ultimately identifies the optimal DN structure. The difference between previous research is the state in DRL. Therefore, in this research, various states that can affect the agent's learning process are modeled, including those identified in prior research. In addition, three reward function formats were designed to minimize power losses and voltage deviation, as discussed above. In summary, Fig. 2 illustrates the learning process of the proposed model within the DN. Furthermore, the pseudocode for the proposed algorithm is provided in Algorithm 1.

Table 1: A comparison of the various parts modeled in DRL with previous research.

Ref	Agent	ENV	Action	State	Reward
[39]	DSO	DN	lines	$[p_t, q_t]$	$P_{loss} + VD$
[40]	DSO	DN	lines	$[p_t, q_t, \alpha_t]$	$P_{loss} + VD$
[41]	DSO	DN	lines	$[p_t, q_t, \alpha_t]$	$P_{loss} + VD$
[42]	DSO	DN	lines	$[v_t, \alpha_t]$	$P_{loss} + VD$
[43]	DSO	DN	lines	$[p_t, q_t, \alpha_t]$	$P_{loss} + VD$
[51]	DSO	DN	lines	$[p_t, q_t, \alpha_t]$	$P_{loss} + VD$
[48]	DSO	DN	lines	$[v_t, \alpha_t]$	$P_{loss} + VD$
This research	DSO	DN	lines	Eq 25-30	P_{loss}, $P_{loss} + VD$

ENV= Environment, lines= Disconnecting or connecting lines.

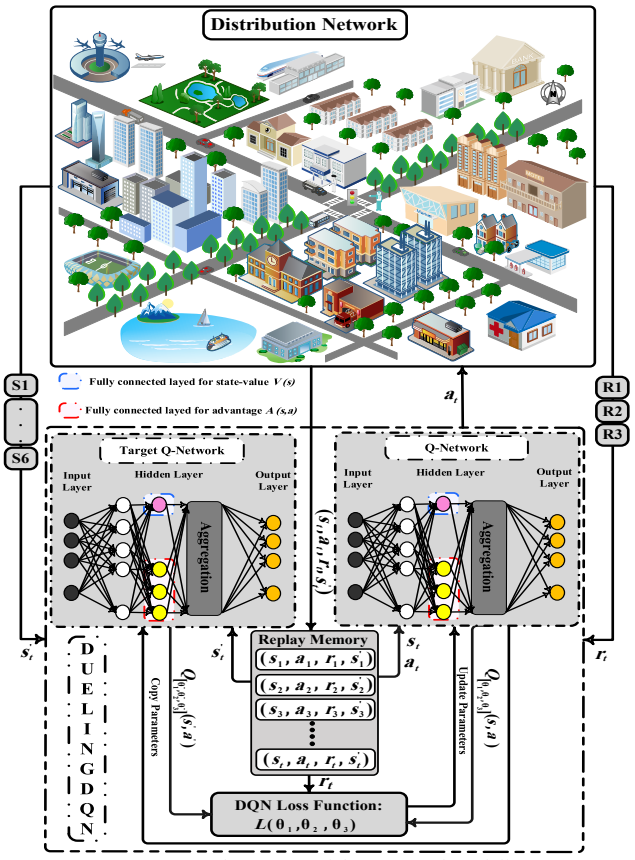


Fig. 2: General structure of the proposed modeling.

Algorithm 1 Dueling DQN for DNR

Input: modeled state and reward.
Output: optimal policy.
Initialize:
  Online Q-network with random parameters $\theta = [\theta_1, \theta_2, \theta_3]$
  Target Q-network with parameters $\theta^- = [\theta_1^-, \theta_2^-, \theta_3^-]$
  Replay buffer D
  Exploration rate ε_0
Main loop (for episode = 1 to M):
  Reset the environment and receive s_0
  for each time step:
    Select and apply action a_t according to $\varepsilon - greedy$ policy and receive s_t' and r_t
    Store (s_t, a_t, r_t, s_t') in replay buffer
    Select a mini batch from the replay buffer and train the agent according to Equation 15
    Set $s_t = s_t'$ and update ε value
  Every C episodes set $\theta^- = \theta$

4. Simulation and Results

The IEEE modified 16- and 33-bus DNs were used as the first and the second test DNs, and the IEEE 69-bus DN as the third. The specifications of these DNs are shown in Table 2, and their general structure is shown in Figs. 3 to 5.

Table 2: Characteristics of the tested DNs.

DN specifications	16-bus	33-bus	69-bus
Number of buses	15	33	69
Number of lines	17	37	73
Rated voltage (kV)	12.66	12.66	12.66
Tie switches	S15-S17	S33-S37	S69-S73
Power loss (kW)	511.390	202.677	225
AVD (p.u.)	0.0141	0.0515	0.0266
Minimum voltage (p.u.)	0.9693	0.913	0.909

AVD: Average voltage deviation.

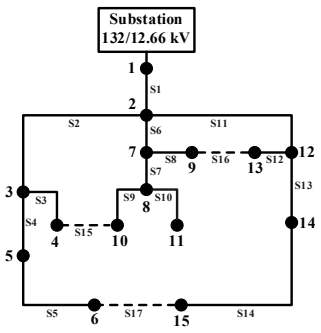


Fig. 3: Structure of the first test DN.

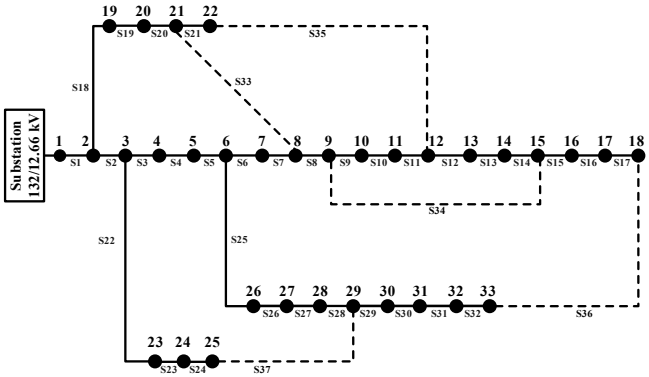


Fig. 4: Structure of the second test DN.

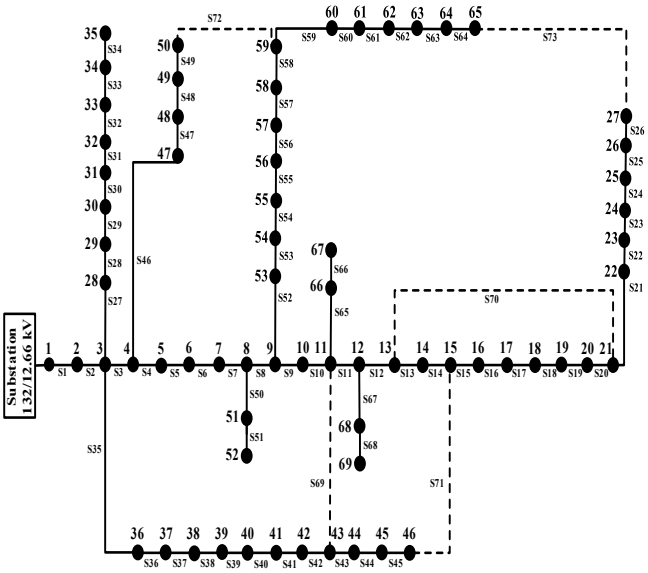


Fig. 5: Structure of the third test DN.

An important point regarding the simulations in this section is that, for each reported result, the proposed algorithm was executed ten times. In addition, to generate the reward curves, the plots were averaged across the ten runs and then smoothed with a 100-run moving average. Furthermore, the standard error is illustrated as lightly shaded regions in all plots, and the most frequent outcome across these ten runs was selected as the final configuration. The hyperparameters used in this study are listed in Table 3. Additionally, λ_1 and λ_2 are supposed to be 10 and 50, respectively. These values are selected by empirical tests.

4.1. First Case Study's Results

The convergence curves of the agents trained on the 16-bus DN using different models are shown in Fig. 6. As can be observed from this figure, almost all of the proposed reward functions converge smoothly to their final solution. The lower values of the Reward 2 and Reward 3 curves at most time steps are due to the presence of an additional negative term in their reward functions. Overall, in this DN, all models converge stably to their final solutions. Moreover, all assumed states in this DN preserve the Markov property, ensuring appropriate behavior of all reward functions.

Table 4 presents the final switching combinations for this DN under different states and reward functions. As can be observed, all models have converged to the same solution. In addition, Fig. 7 illustrates the average voltage deviation and power losses of the obtained combinations for each model, which again indicates that all considered states and reward functions are appropriate for this DN. Moreover, a significant improvement relative to the DN's initial condition is observed in all models.

Table 3: Hyperparameters of the proposed algorithm.

Hyperparameter	Value		
	16-bus	33-bus	69-bus
Hidden units	{32,32}	{32,32}	{32,32}
Activation function	ReLU	ReLU	ReLU
Replay buffer capacity	500	2000	2000
Batch size	16	128	128
Learning rate	0.005	0.0005	0.0005
Optimizer	RMSprop	RMSprop	RMSprop

4.2. Second Case Study's Results

The proposed algorithm was also applied to the second test DN across different state-modeling schemes. The resulting convergence curves, representing the learning process of the agent under these modeled states, are shown in Fig. 8. According to this figure, the performance of Reward 1 is generally superior, as it reaches a near-stable state in a shorter period of time. In addition, the converged policies obtained for different models are presented in Table 5. As shown in this table, due to the complexity of this DN, the converged policies across models are not identical.

Fig. 9 illustrates the power losses and the average voltage deviation for different models. In terms of power losses, the best performance is achieved with Reward 1 in combination with states S2 and S6, whereas the minimum average voltage deviation is obtained with Reward 2 and Reward 3. Overall, all reward functions perform best under state S6 and worst under state S1. These results clearly highlight the critical role of state representation in DRL-based control of DNs. As DN complexity increases, richer, more informative state definitions are required to preserve the Markov property, thereby yielding more stable learning and superior control performance.

4.3. Third Case Study's Results

The learning process of the DRL agent under different state models for the third test DN is shown in Fig. 10. These curves indicate that all three proposed reward functions exhibit stable performance in this DN. In addition, Table 6 presents the converged policies for the different models, indicating that the models perform differently in this DN. The numerical results from each of these models are shown in Fig. 11. According to this figure, the lowest power losses are again achieved by combining Reward 1 with states S2 and S6. In addition, the reward functions Reward 2 and Reward 3 can achieve optimal power loss performance in certain states. In contrast, the minimum average voltage deviation is obtained from the combination of Reward 2 and S6, indicating that adding a voltage-related term to the reward function can effectively reduce average voltage deviation. Overall, in this DN, all models significantly reduce both voltage deviation and power losses relative to the DN's initial configuration, and the percentage variations in the results are smaller than those observed in the 33-bus DN.

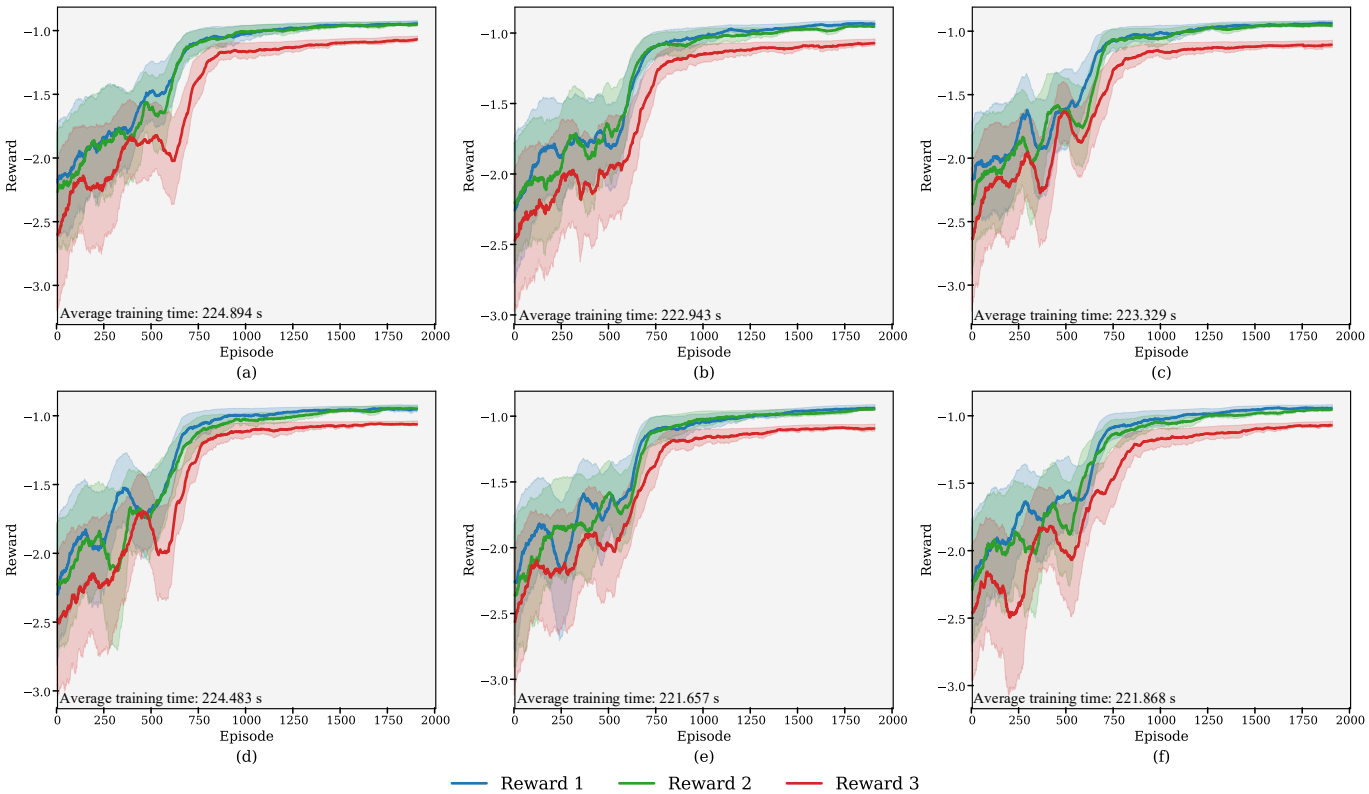


Fig. 6: Converge diagram in the first DN under different modeling: (a) state S1, (b) state S2, (c) state S3, (d) state S4, (e) state S5, (f) state S6.

Table 4: Policies learned by the agent in the first DN.

Reward 1		Reward 2		Reward 3	
State	**Open switches**	**State**	**Open switches**	**State**	**Open switches**
S1	10-8, 7-9, 6-15	S1	10-8, 7-9, 6-15	S1	10-8, 7-9, 6-15
S2	10-8, 7-9, 6-15	S2	10-8, 7-9, 6-15	S2	10-8, 7-9, 6-15
S3	10-8, 7-9, 6-15	S3	10-8, 7-9, 6-15	S3	10-8, 7-9, 6-15
S4	10-8, 7-9, 6-15	S4	10-8, 7-9, 6-15	S4	10-8, 7-9, 6-15
S5	10-8, 7-9, 6-15	S5	10-8, 7-9, 6-15	S5	10-8, 7-9, 6-15
S6	10-8, 7-9, 6-15	S6	10-8, 7-9, 6-15	S6	10-8, 7-9, 6-15

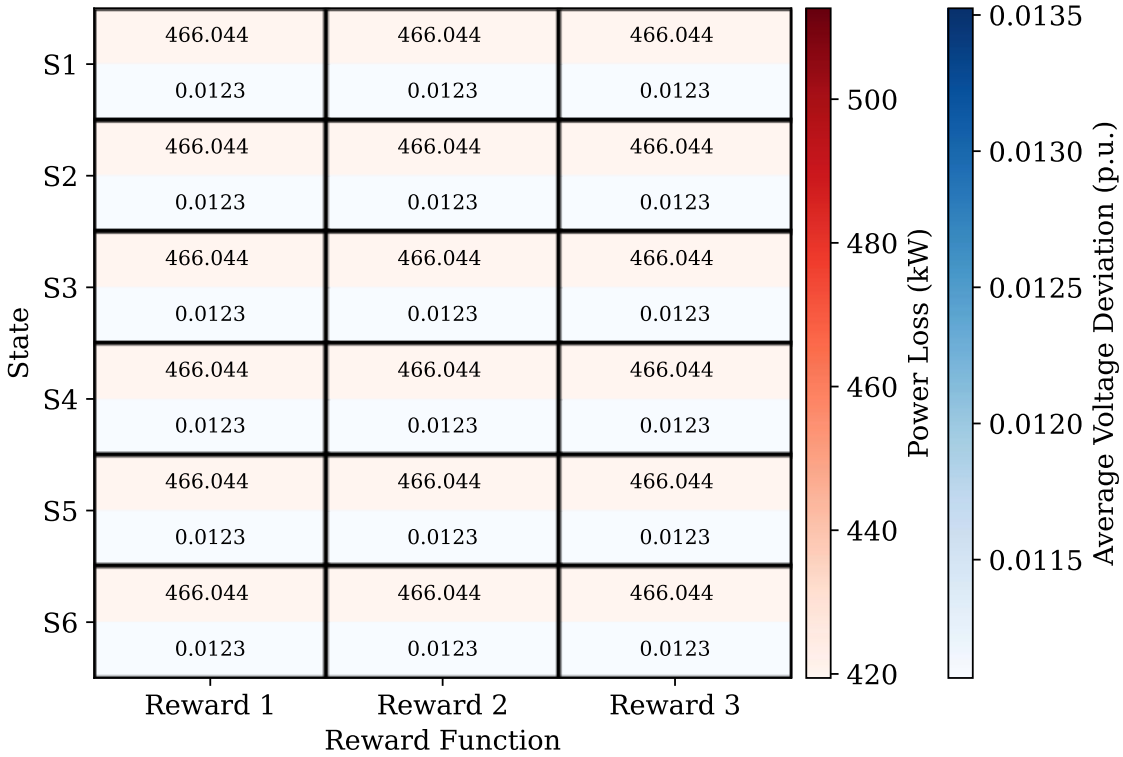


Fig. 7: Different modeling results in the first DN by the proposed method.

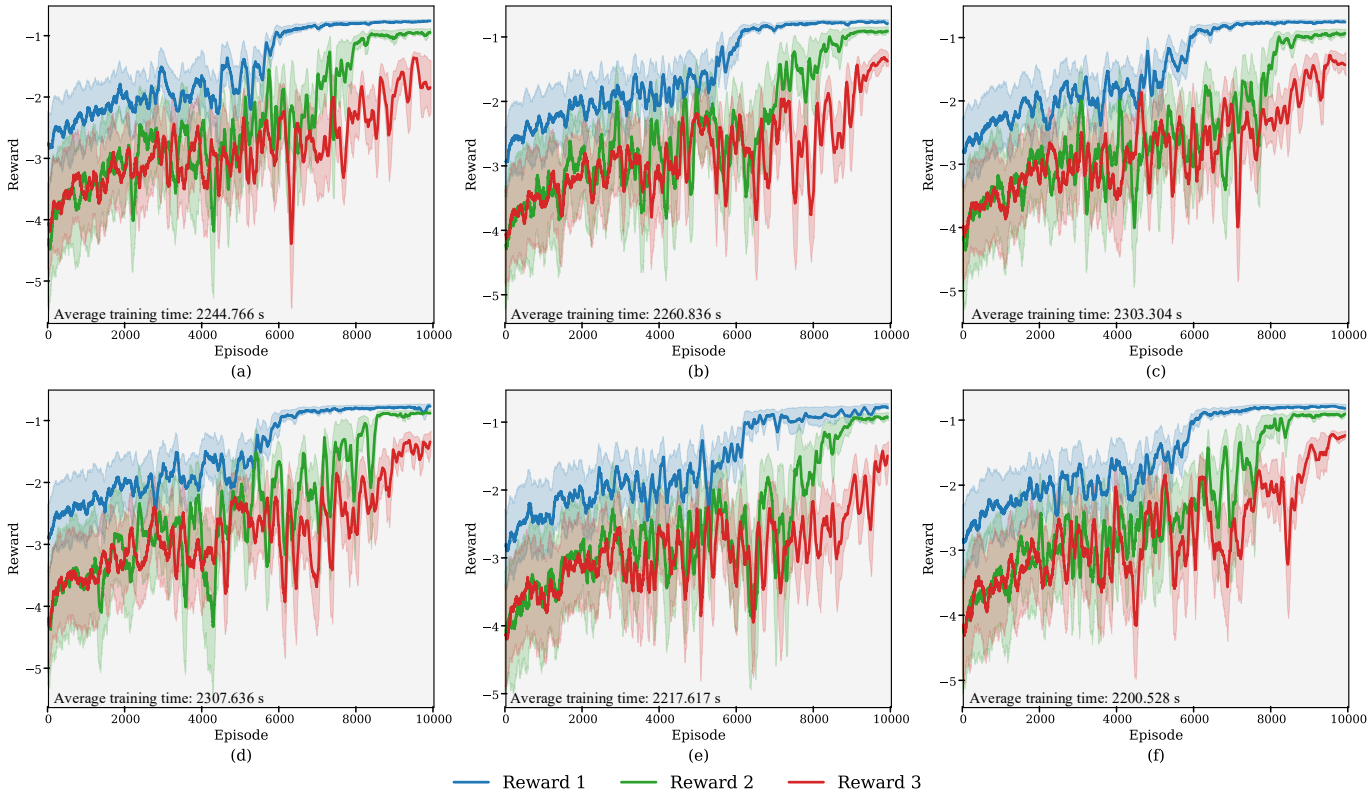


Fig. 8: Converge diagram in the second DN under different modeling: (a) state S1, (b) state S2, (c) state S3, (d) state S4, (e) state S5, (f) state S6.

Table 5: Policies learned by the agent in the second DN.

Reward 1		Reward 2		Reward 3	
State	**Open switches**	**State**	**Open switches**	**State**	**Open switches**
S1	9-10, 14-15, 32-33, 7-8, 28-29	S1	11-12, 14-15, 18-33, 6-7, 27-28	S1	10-11, 14-15, 17-18, 7-8, 27-28
S2	9-10, 14-15, 32-33, 7-8, 25-29	S2	9-10, 13-14, 18-33, 7-8, 27-28	S2	9-10, 14-15, 18-33, 7-8, 27-28
S3	9-10, 14-15, 32-33, 7-8, 28-29	S3	9-10, 14-15, 18-33, 7-8, 27-28	S3	9-10, 14-15, 18-33, 7-8, 28-29
S4	9-10, 14-15, 32-33, 7-8, 28-29	S4	9-10, 14-15, 18-33, 7-8, 28-29	S4	9-10, 14-15, 18-33, 7-8, 28-29
S5	9-10, 14-15, 32-33, 7-8, 28-29	S5	9-10, 14-15, 18-33, 7-8, 28-29	S5	10-11, 14-15, 32-33, 7-8, 27-28
S6	9-10, 14-15, 32-33, 7-8, 25-29	S6	9-10, 14-15, 18-33, 7-8, 28-29	S6	9-10, 14-15, 32-33, 7-8, 28-29

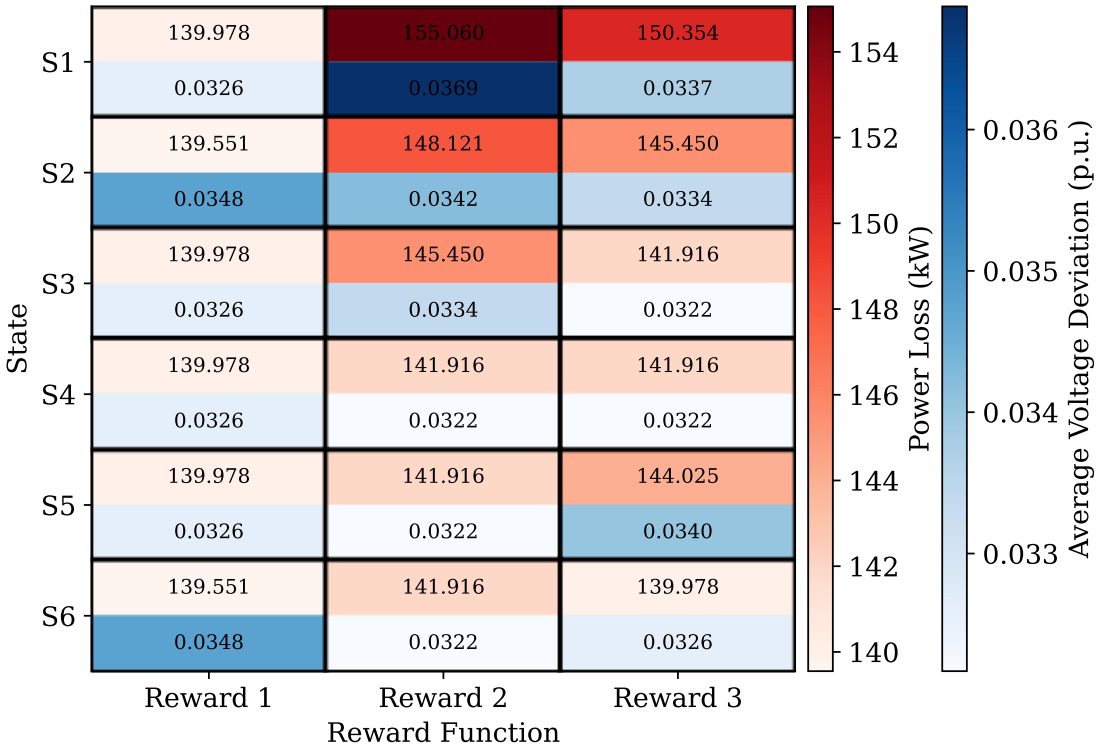


Fig. 9: Different modeling results in the second DN by the proposed method.

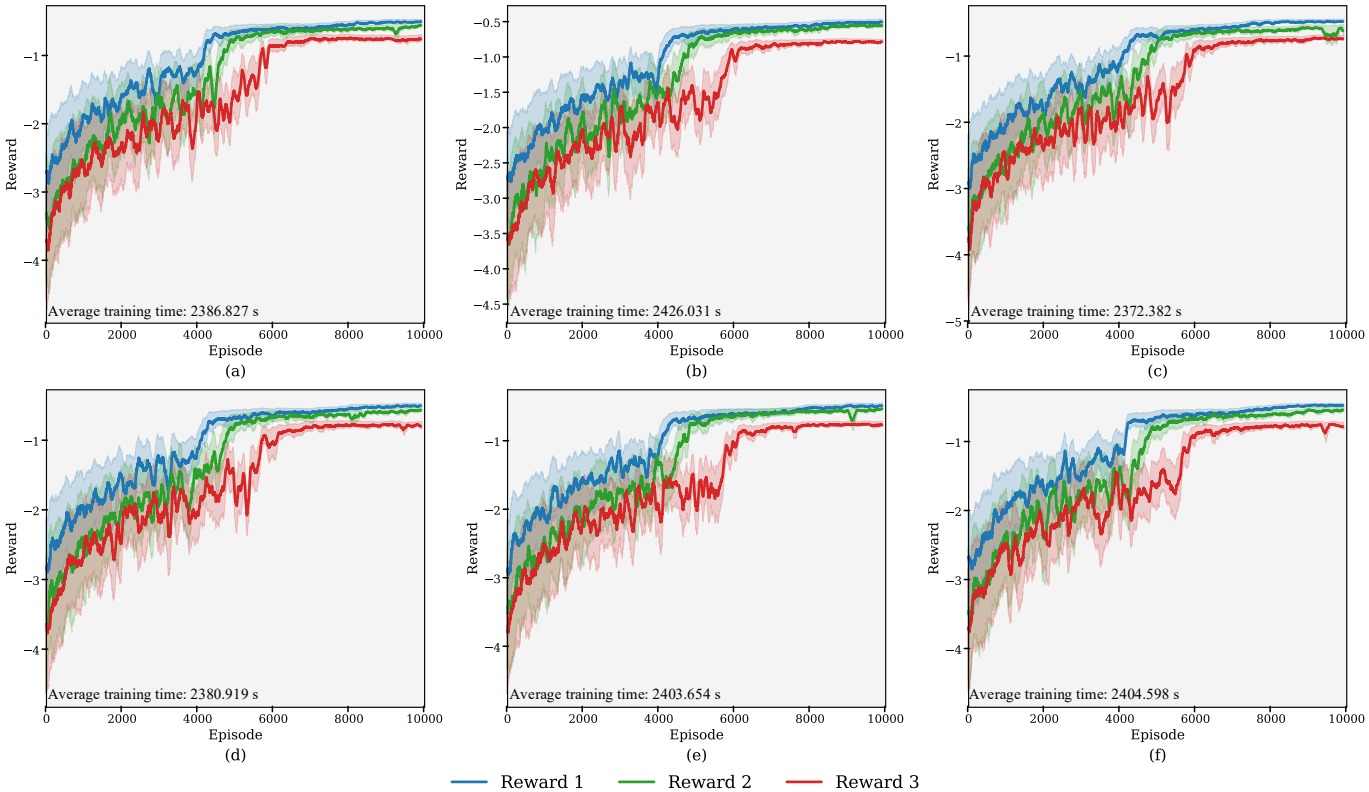


Fig. 10: Converge diagram in the third DN under different modeling: (a) state S1, (b) state S2, (c) state S3, (d) state S4, (e) state S5, (f) state S6.

Table 6: Policies learned by the agent in the third DN.

Reward 1		Reward 2		Reward 3	
State	**Open switches**	**State**	**Open switches**	**State**	**Open switches**
S1	13-14, 13-21, 62-63, 11-43, 57-58	S1	13-14, 13-21, 63-64, 11-43, 55-56	S1	13-14, 13-21, 63-64, 11-43, 56-57
S2	13-14, 13-21, 61-62, 11-43, 56-57	S2	12-13, 14-15, 62-63, 11-43, 54-55	S2	13-14, 13-21, 62-63, 11-43, 55-56
S3	13-14, 13-21, 62-63, 11-43, 55-56	S3	13-14, 13-21, 61-62, 11-43, 57-58	S3	13-14, 13-21, 62-63, 11-43, 56-57
S4	14-15, 13-21, 63-64, 11-43, 55-56	S4	13-14, 13-21, 61-62, 11-43, 53-54	S4	13-14, 13-21, 62-63, 11-43, 55-56
S5	13-14, 13-21, 62-63, 11-43, 57-58	S5	13-14, 13-21, 62-63, 11-43, 55-56	S5	13-14, 13-21, 61-62, 11-43, 57-58
S6	13-14, 13-21, 62-63, 11-43, 57-58	S6	13-14, 13-21, 62-63, 11-43, 58-59	S6	14-15, 13-21, 63-64, 11-43, 55-56

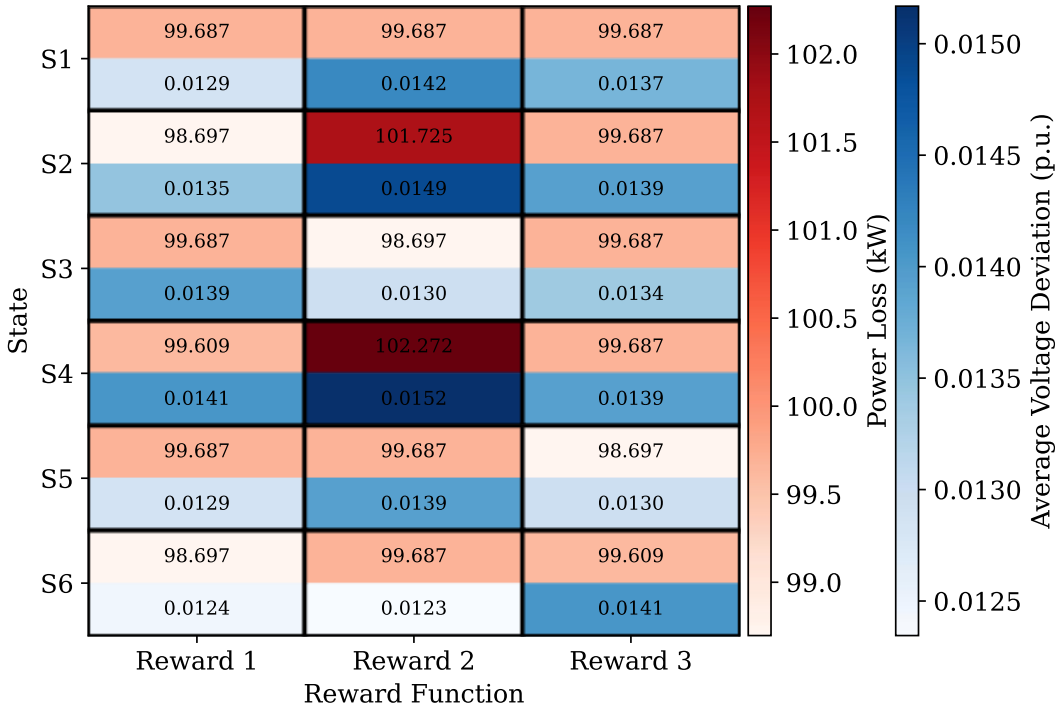


Fig. 11: Different modeling results in the third DN by the proposed method.

4.4. Comparative Results

A comparison with non-DRL methods, including mathematical and metaheuristic approaches, for the second and third case studies is presented in Tables 7 and 8, respectively. In these evaluations, state representation S6 and reward function 1 were utilized. As shown in Table 7, it achieved better results compared to FWA, Graph, FPEO, EBCSA, and H-matrix. The SOE and MP methods obtained results similar to the proposed method. In addition, in Table 8, it performed better than MSSOE, HSA, FWA, BGSO, and AC. Notably, regarding power loss, the proposed algorithm achieves the minimum value in both tables. These results indicate that the proposed modeling framework is not only effective compared to other DRL configurations, it also demonstrates highly competitive and superior performance against conventional optimization algorithms.

Table 7: Algorithms comparative in second case study.

Algorithm	PL (kW)	AVD (p.u)	MV (p.u)
Initial	202.677	0.0515	0.913
Proposed method	139.551	0.0348	0.9378
SOE [30]	139.551	0.0348	0.9378
MP [30]	139.551	0.0348	0.9378
FWA [15]	139.978	0.0326	0.9413
Graph [21]	139.978	0.0326	0.9413
FPEO [23]	139.978	0.0326	0.9413
EBCSA [26]	139.978	0.0326	0.9413
H-matrix [28]	139.978	0.0326	0.9413

AVD: Average voltage deviation, PL: Power loss, MV: Min voltage.

Table 8: Algorithms comparative in third case study.

Algorithm	PL (kW)	AVD (p.u)	MV (p.u)
Initial	225	0.0266	0.909
Proposed method	98.697	0.0124	0.9495
MSSOE [31]	98.697	0.0129	0.9494
HSA [35]	105.221	0.0140	0.9494
FWA [15]	116.486	0.0175	0.9495
BGSO [24]	98.800	0.0142	0.9495
AC [25]	98.800	0.0142	0.9495

AVD: Average voltage deviation, PL: Power loss, MV: Min voltage.

4.5. The Effect of Uncertainty on Proposed Method

In this section, it is assumed that wind-based DGs are integrated into two more complex DNs: the second and third case studies. For the second case study, these DGs are installed at buses 8, 18, 24, and 30, while in the third case study, they are located at buses 10, 22, 48, and 60. Fig. 12 illustrates the considered uncertainties regarding the DN loading levels and wind speed variations over a daily period. From this daily profile, hours 9, 15, 19, and 24 have been specifically selected for detailed evaluation. These distinct operating hours are subsequently analyzed and referred to as Scenarios 1, 2, 3, and 4, respectively.

The simulation results for the second case study are presented in Fig. 13. As can be observed, the proposed method successfully reduces both power loss and average voltage deviation across all examined scenarios compared to the initial configuration. This demonstrates the robust performance and adaptability of the model when subjected to uncertainties. Furthermore, the results indicate that the improvements achieved by the Dueling DQN algorithm are particularly significant during the more severe operating conditions (Scenarios 1 and 3), highlighting the algorithm's effectiveness in managing stressed DN states.

Fig. 14 illustrates the simulation results across the four defined scenarios for the third case study. Similar to the previous test case, the proposed Dueling DQN achieves substantial reductions in both power loss and average voltage deviation. It is noteworthy that the most dramatic improvements occur during the highly stressed conditions of Scenarios 1 and 3. This consistent behavior across a larger system underscores the scalability of the proposed method and its capability to maintain optimal grid operation even under severe DN stress.

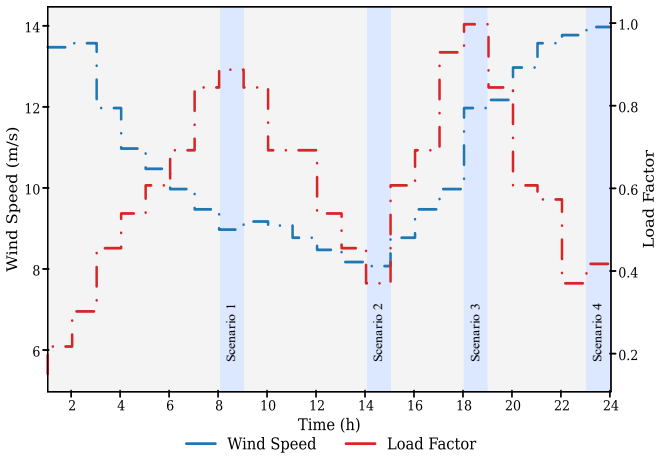


Fig. 12: Daily profiles of wind speed and load factor.

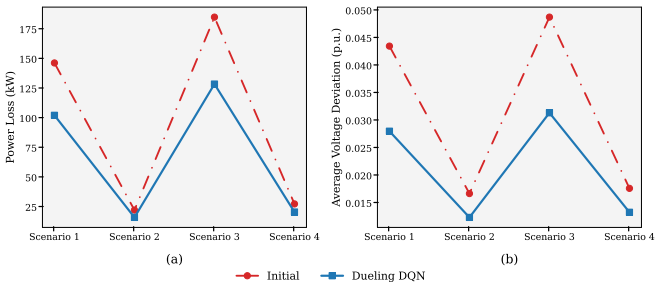


Fig. 13. Simulation results for the second case study under various operating scenarios: (a) Power loss and (b) Average voltage deviation.

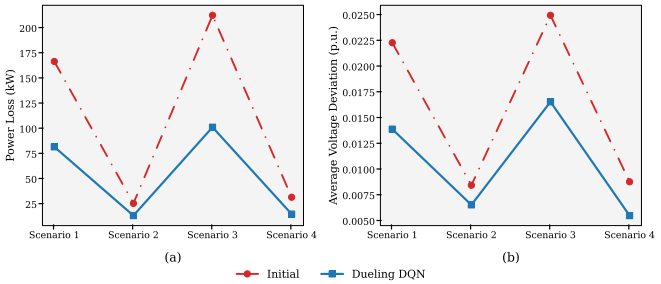


Fig. 14. Simulation results for the third case study under various operating scenarios: (a) Power loss and (b) Average voltage deviation.

5. COMPARATIVE ANALYSIS OF SIMULATION RESULTS

In the simulation section, the Dueling DQN method is evaluated on three different IEEE DNs. To investigate the impact of modeling choices, six different state representations and three different reward functions are considered. The key findings obtained from the simulations are summarized as follows:

- For the 16-bus DN, all models converge to the same solution. However, as the complexity of the DN increases, the proper selection of state representations and reward functions becomes increasingly important.
- In general, increasing the amount of information included in the state representation leads to a process with stronger Markovian properties. The simulation results support this claim, as in most cases on the 33-bus and 69-bus DNs, the S6 state representation yields the best performance across the corresponding reward functions.
- A strong correlation is observed between the average voltage deviation and power losses in DNs. As a result, even when the reward function is defined solely in terms of power losses, a significant reduction in voltage deviations can still be expected.
- Adding a voltage-related term to the reward function can lead to solutions with improved voltage deviation performance. However, this improvement may come at the cost of increased power losses.
- Incorporating a voltage term into the reward function increases the learning process's sensitivity to the selected state representation. Therefore, when such a reward function is used, it is recommended that the chosen state representation include as much DN information as possible.
- In general, the designed states for the DNR problem in this article are helpful because they significantly reduce power losses and average voltage deviation relative to the initial DN configuration.
- Although increasing the complexity of the state representation theoretically adds to the computational burden, the convergence plots demonstrate that the average training time across different states remain remarkably similar. This is primarily because, in the proposed algorithm, the majority of the processing time is consumed by experience gathering (interaction with the environment). Furthermore, since the neural network training is accelerated using a GPU, expanding the input dimensions has a negligible impact on the overall training time.

6. CONCLUSION

This study explores the application of the Dueling DQN algorithm for distribution network reconfiguration (DNR) to effectively minimize power losses and voltage deviations in three distribution networks. Furthermore, the impact of uncertainty in distributed generation and load profile has been simulated to demonstrate the flexibility of the proposed method. A major focus is placed on the underlying learning formulation, revealing that the success of the deep reinforcement learning (DRL) agent is highly dependent on how state representations and reward functions are carefully designed. The evaluations demonstrate that including a broader range of network parameters in the state representation generally leads to better optimization outcomes. Additionally, adding voltage-specific terms to the reward function improves the system's voltage profile, though it makes the agent's performance more sensitive to the chosen state variables. Overall, the findings indicate that the optimal state formulation is not universal and must be tailored to the specific complexity of the distribution network; however, in general, a state representation containing more elements usually better preserves the Markov property.

Although the proposed method performs well across various distribution networks, it may require significant training time. Future research will focus on implementing transfer learning techniques to leverage the agent's learned experiences from smaller distribution networks for application in larger ones, as well as reducing training time with action space reduction methods.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Amirhossein Ghaemipour: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Roles/Writing - original draft. **Habib Rajabi Mashhadi**: Supervision, Writing - review & editing. **Seyed Hossein Mostafavi**: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Roles/Writing - original draft,

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The ethical issues; including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy has been completely observed by the authors.

REFERENCES

[1] A. Rouhipour, E. Mashhour, and M. Saniei, "Optimal Switch Placement in Radial Distribution Networks to Reduce Energy Loss and Improve Network Security in Khuzestan Province Conditions," *J. Appl. Res. Electr. Eng.*, vol. 1, no. 2, pp. 149–158, 2022.

[2] A. Bagheri, S. Mobayen, S. Behzadi, and S. A. Miri, "Probabilistic optimal co-planning of distributed series reactor and dynamic line thermal rating for congestion management of highly-renewable-penetrated electric power systems," *Electr. Power Syst. Res.*, vol. 248, p. 111924, 2025.

[3] P. Li, E. Zahedi, Y. Shi, Y. Deng, and L. Liu, "Dynamically tuning and reconfiguring microwave bandpass filters using optical control of switching elements," *Microw. Opt. Technol. Lett.*, vol. 66, no. 2, p. e34082, Feb. 2024.

[4] E. Zahedi, M. Palei, and A. J. Hoffman, "Anomalous diffraction patterns from nonlocal two-dimensional metacrystals," in *2025 Conference on Lasers and Electro-Optics (CLEO)*, IEEE, 2025, pp. 1–2.

[5] M. Najjarpour, B. Tousi, and A. H. Karamali, "Optimizing Reactive Power for DG Units to Minimize Power System Losses Using Stochastic Modeling," *J. Green Energy Res. Innov.*, vol. 1, no. 4, pp. 35–46, 2024.

[6] S. A. Miri, M. Rahimi, and A. Bagheri, "An Optimal Conservative Voltage Reduction Strategy in Active Distribution Networks Using Coordinated Operation of RES, OLTC, ESS, and SCB," in *2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG)*, 2025, pp. 1–6.

[7] H. Lotfi, M. H. Nikkhah, and M. E. Hajiabadi, "Dynamic Reconfiguration for Energy Management in EV and RES-Based Grids Using IWOA," 2025.

[8] S. A. Miri, Z. Azimi, and A. Bagheri, "Dynamically Reconfigurable Active Distribution Networks for Maximum Renewables Integration at the Presence of Voltage-Dependent Loads," in *2024 28th International Electrical Power Distribution Conference (EPDC)*, IEEE, 2024, pp. 1–6.

[9] B. Bahmanifirouzi, E. Farjah, T. Niknam, and E. A. Farsani, "A new hybrid HBMO-SFLA algorithm for multi-objective distribution feeder reconfiguration problem considering distributed generator units," *Iran. J. Sci. Technol. Trans. Electr. Eng.*, vol. 36, no. E1, p. 51, 2012.

[10] K. S. Kumar and T. Jayabarathi, "Power system reconfiguration and loss minimization for an distribution systems using bacterial foraging optimization algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 36, no. 1, pp. 13–17, 2012.

[11] T. Niknam, A. Kavousifard, and J. Aghaei, "Scenario-based multiobjective distribution feeder reconfiguration considering wind power using adaptive modified particle swarm optimisation," *IET Renew. Power Gener.*, vol. 6, no. 4, pp. 236–247, 2012.

[12] H. Lotfi and R. Ghazi, "Optimal participation of demand response aggregators in reconfigurable distribution system considering photovoltaic and storage units," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 2, pp. 2233–2255, 2021.

[13] D.-L. Duan, X.-D. Ling, X.-Y. Wu, and B. Zhong, "Reconfiguration of distribution network for loss reduction and reliability improvement based on an enhanced genetic algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 88–95, 2015.

[14] M. Noruzi Azghandi, A. A. Shojaei, S. Toosi, and H. Lotfi, "Optimal reconfiguration of distribution network feeders considering electrical vehicles and distributed generators," *Evol. Intell.*, vol. 16, no. 1, pp. 49–66, 2023.

[15] A. M. Imran, M. Kowsalya, and D. P. Kothari, "A novel integration technique for optimal network reconfiguration and distributed generation placement in power distribution networks," *Int. J. Electr. Power Energy Syst.*, vol. 63, pp. 461–472, 2014.

[16] H. Lotfi and A. A. Shojaei, "A dynamic model for multi-objective feeder reconfiguration in distribution network considering demand response program," *Energy Syst.*, vol. 14, no. 4, pp. 1051–1080, 2023.

[17] H. Ahmadi and J. R. Martı, "Linear current flow equations with application to distribution systems reconfiguration," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2073–2080, 2014.

[18] R. Rajaram, K. S. Kumar, and N. Rajasekar, "Power system reconfiguration in a radial distribution network for reducing losses and to improve voltage profile using modified plant growth simulation algorithm with Distributed Generation (DG)," *Energy Reports*, vol. 1, pp. 116–122, 2015.

[19] A. Bayat, A. Bagheri, and R. Noroozian, "Optimal siting and sizing of distributed generation accompanied by reconfiguration of distribution networks for maximum loss reduction by using a new UVDA-based heuristic method," *Int. J. Electr. Power Energy Syst.*, vol. 77, pp. 360–371, 2016.

[20] T. T. Nguyen, T. T. Nguyen, A. V. Truong, Q. T. Nguyen, and T. A. Phung, "Multi-objective electric distribution network reconfiguration solution using runner-root algorithm," *Appl. Soft Comput.*, vol. 52, no. C, pp. 93–108, Mar. 2017.

[21] H. Parsadust, M. E. Hajiabadi, and H. Lotfi, "Bi-Level Graph-Based Optimisation for Distribution Network Reconfiguration and Optimal Placement of TCLBS and DC Switches," *IET Gener. Transm. Distrib.*, vol. 19, no. 1, p. e70144, 2025.

[22] M. R. Babu, C. V. Kumar, and S. Anitha, "Simultaneous reconfiguration and optimal capacitor placement using adaptive whale optimization algorithm for radial distribution system," *J. Electr. Eng. Technol.*, vol. 16, no. 1, pp. 181–190, 2021.

[23] A. Landeros, S. Koziel, and M. F. Abdel-Fattah, "Distribution network reconfiguration using feasibility-preserving evolutionary optimization," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 3, pp. 589–598, 2019.

[24] S. Teimourzadeh and K. Zare, "Application of binary group search optimization to distribution network reconfiguration," *Int. J. Electr. Power Energy Syst.*, vol. 62, pp. 461–468, 2014.

[25] S. H. Mirhoseini, S. M. Hosseini, M. Ghanbari, and M. Ahmadi, "A new improved adaptive imperialist competitive algorithm to solve the reconfiguration problem of distribution systems for loss reduction and voltage profile improvement," *Int. J. Electr. Power Energy Syst.*, vol. 55, pp. 128–143, 2014.

[26] T. T. Nguyen, T. T. Nguyen, and B. Le, "Optimization of electric distribution network configuration for power loss reduction based on enhanced binary cuckoo search algorithm," *Comput. Electr. Eng.*, vol. 90, p. 106893, 2021.

[27] M. Rahimi Pour Behbahani, A. Jalilian, and M. Amini, "Reconfiguration of distribution network using discrete particle swarm optimization to reduce voltage fluctuations," *Int. Trans. Electr. Energy Syst.*, vol. 30, no. 9, p. e12501, 2020.

[28] H. Karimianfard and H. Haghighat, "An initial-point strategy for optimizing distribution system reconfiguration," *Electr. Power Syst. Res.*, vol. 176, p. 105943, 2019.

[29] J. Singh and R. Tiwari, "Real power loss minimisation of smart grid with electric vehicles using distribution feeder reconfiguration," *IET Gener. Transm. Distrib.*, vol. 13, no. 18, pp. 4249–4261, 2019.

[30] J. Zhan, W. Liu, C. Y. Chung, and J. Yang, "Switch Opening and Exchange Method for Stochastic Distribution Network Reconfiguration," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 2995–3007, 2020.

[31] V. Vai, S. Suk, R. Lorm, C. Chhlonh, S. Eng, and L. Bun, "Optimal reconfiguration in distribution systems with distributed generations based on modified sequential switch opening and exchange," *Appl. Sci.*, vol. 11, no. 5, p. 2146, 2021.

[32] T. Tran The, D. Vo Ngoc, and N. Tran Anh, "Distribution network reconfiguration for power loss reduction and voltage profile improvement using chaotic stochastic fractal search algorithm," *Complexity*, vol. 2020, no. 1, p. 2353901, 2020.

[33] M. Mahdavi, H. H. Alhelou, N. D. Hatziargyriou, and A. Al-Hinai, "An Efficient Mathematical Model for Distribution System Reconfiguration Using AMPL," *IEEE Access*, vol. 9, pp. 79961–79993, 2021.

[34] N.-V. Pham and H.-N. Nguyen, "A Modified Distribution Flow-Based Quadratic Programming with Integer Variables for Co-Optimization of On-Site Generation Placement and Network Topology in Power Distribution Grids," *Results Eng.*, p. 107347, 2025.

[35] R. S. Rao, K. Ravindra, K. Satish, and S. V. L. Narasimham, "Power loss minimization in distribution system using network reconfiguration in the presence of distributed generation," *IEEE Trans. power Syst.*, vol. 28, no. 1, pp. 317–325, 2012.

[36] A. Karamali, A. Daeichian, and S. Rezaei, "Using Long Short-Term Memory Networks as Virtual Wind Direction Sensors for Improved Wind Farm Turbines Orientation," in *2024 9th International Conference on Technology and Energy Management (ICTEM)*, 2024, pp. 1–5.

[37] A. H. Karamali, A. R. Danesh, M. Kalantar, M. R. Moghadasian, and S. Rezaei, "Enhancing Wind Power Forecasting Through a Hybrid Deep Learning Approach: Long Short-Term Memory Integrated with Attention Mechanism," in *2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG)*, 2025, pp. 1–6.

[38] S. Ansari, H. Samet, and T. Ghanbari, "Fault Location in Solar Farms," *IEEE Syst. J.*, vol. 15, no. 3, pp. 4003–4012, 2021.

[39] B. Wang, H. Zhu, H. Xu, Y. Bao, and H. Di, "Distribution Network Reconfiguration Based on NoisyNet Deep Q-Learning Network," *IEEE Access*, vol. 9, pp. 90358–90365, 2021.

[40] N. Gholizadeh, N. Kazemi, and P. Musilek, "A Comparative Study of Reinforcement Learning Algorithms for Distribution Network Reconfiguration With Deep Q-Learning-Based Action Sampling," *IEEE Access*, vol. 11, pp. 13714–13723, 2023.

[41] N. Gholizadeh and P. Musilek, "A Generalized Deep Reinforcement Learning Model for Distribution Network Reconfiguration with Power Flow-Based Action-Space Sampling," 2024.

[42] A. Ghaemipour, H. R. Mashhadi, and S. H. Mostafavi, "Distribution Network Reconfiguration in the Presence of Distribution Generation Using Deep Reinforcement Learning," in *2025 6th International Conference on Optimizing Electrical Energy Consumption (OEEC)*, 2025, pp. 1–5.

[43] N. Gholizadeh and P. Musilek, "Explainable reinforcement learning for distribution network reconfiguration," *Energy Reports*, vol. 11, pp. 5703–5715, 2024.

[44] S. H. Oh, Y. T. Yoon, and S. W. Kim, "Online reconfiguration scheme of self-sufficient distribution network based on a reinforcement learning approach," *Appl. Energy*, vol. 280, p. 115900, 2020.

[45] O. B. Kundačina, P. M. Vidović, and M. R. Petković, "Solving dynamic distribution network reconfiguration using deep reinforcement learning," *Electr. Eng.*, vol. 104, no. 3, pp. 1487–1501, 2022.

[46] S. Jo, J.-Y. Oh, Y. T. Yoon, and Y. G. Jin, "Self-healing radial distribution network reconfiguration based on deep reinforcement learning," *Results Eng.*, vol. 22, p. 102026, 2024.

[47] W. Huang and C. Zhao, "Deep-Learning-Aided Voltage-Stability-Enhancing Stochastic Distribution Network Reconfiguration," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 2827–2836, 2024.

[48] A. Ghaemipour, H. R. Mashhadi, and S. H. Mostafavi, "Accelerating the learning process of deep

reinforcement learning algorithms in distribution network reconfiguration using an innovative replay method," *Sci. Rep.*, 2026.
[49] E. Yaghoubi, E. Yaghoubi, Z. Yusupov, and M. R. Maghami, "A Real-Time and Online Dynamic Reconfiguration against Cyber-Attacks to Enhance Security and Cost-Efficiency in Smart Power Microgrids Using Deep Learning," 2024.
[50] S. Malekshah, A. Rasouli, Y. Malekshah, A. Ramezani, and A. Malekshah, "Reliability-driven distribution power network dynamic reconfiguration in presence of distributed generation by the deep reinforcement learning method," *Alexandria Eng. J.*, vol. 61, no. 8, pp. 6541–6556, 2022.
[51] A. Ghaemipour, H. Rajabi Mashhadi, and S. H. Mostafavi, "Distribution Network Reconfiguration and Radiality Detection using Deep Reinforcement Learning," *J. Appl. Res. Electr. Eng.*, 2026.
[52] A. A. Chirani, "Identification of the optimal location and capacity of STATCOM using Grey Wolf Optimizer," *Majlesi J. Electr. Eng.*, vol. 18, no. 3, pp. 1–13, 2024.
[53] C. R. Fuerte-Esquivel and E. Acha, "A Newton-type algorithm for the control of power flow in electrical power networks," *IEEE Trans. power Syst.*, vol. 12, no. 4, pp. 1474–1480, 1997.
[54] S. Ghosh, S. P. Ghoshal, and S. Ghosh, "Optimal sizing and placement of distributed generation in a network system," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 8, pp. 849–856, 2010.
[55] L. Thurner *et al.*, "Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6510–6521, 2018.
[56] V. Mnih *et al.*, "Playing atari with deep reinforcement learning," *arXiv Prepr. arXiv1312.5602*, 2013.
[57] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[58] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, "Dueling network architectures for deep reinforcement learning," in *International conference on machine learning*, PMLR, 2016, pp. 1995–2003.

BIOGRAPHY



Amirhossein Ghaemipour was born in Mashhad, Iran, in 2000. He received his B.Sc. degree in Electrical Power Engineering from Shahid Montazeri University, Mashhad, Iran, in 2021. He received the M.Sc. degree in power systems at Ferdowsi University of Mashhad, Mashhad, Iran, in 2025. He received the prestigious Best Paper Award at the 6th International Conference on Optimizing Electrical Energy Consumption (OEEC) on 25-26 February 2025. His research interests include the application of AI in electrical engineering, deep reinforcement learning, deep learning, smart grid, power systems reliability and resilience, power system operation and planning, and power markets.



Habib Rajabi Mashhadi was born in Mashhad, Iran, in 1967. He received the B.Sc. and M.Sc. degrees with honors from the Ferdowsi University of Mashhad, Mashhad, Iran, both in electrical engineering, and the Ph.D. degree from the Department of Electrical and Computer Engineering of Tehran University, Tehran, Iran, under joint cooperation of Aachen University of Technology, Germany, in 2002. He is a professor of electrical engineering at Ferdowsi University of Mashhad and is with the Center of Excellence on Soft Computing and Intelligent Information Processing, Ferdowsi University of Mashhad. His research interests are power system operation and planning, power system economics, and biological computation.



Seyed Hossein Mostafavi received the B.Sc. degree in electrical engineering from Shahid Rajaee University, Tehran, Iran, in 2022, and the M.Sc. degree in electrical engineering from Ferdowsi University of Mashhad, Mashhad, Iran, in 2025. His research interests include power systems, deep learning, deep reinforcement learning, and control theory.

Copyrights
© 2026 by the author(s). Licensee Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution –NonCommercial 4.0 International (CC BY-NC 4.0) License (http://creativecommons.org/licenses/by-nc/4.0/).