Deep Reinforcement Learning for Voltage Stability: DDQN-Based Control with Real-Time Hardware Validation

Cem Haydaroğlu*‡, Heybet Kılıç**, Ahmet Top***

*Department of Electrical and Electronics Engineering, Engineering Faculty, Dicle University, Diyarbakir, 21280, Türkiye

**Department of Electric Power and Energy System, Dicle University, Diyarbakir, 21280, Türkiye

***Department of Electrical and Electronics Engineering, Technology Faculty, Fırat University, Elazığ, 23100, Türkiye

(cem.haydaroglu@dicle.edu.tr, heybet.kilic@dicle.edu.tr, atop@firat.edu.tr)

‡Corresponding Author: Cem Haydaroğlu, Dicle University, cem.haydaroglu@dicle.edu.tr

Received: 26.11.2025 Accepted: 15.12.2025

Abstract- This study proposes an autonomous voltage control framework based on the Double Deep Q-Network (DDQN) algorithm to enhance voltage stability in power distribution systems with high renewable penetration. The proposed controller learns adaptive voltage regulation policies by interacting with dynamic grid environments and observing voltage deviations, power flows, and generator-load dynamics. The method is evaluated on both IEEE 14-bus and IEEE 124-bus test systems and benchmarked against state-of-the-art DRL agents, including DQN, PPO, DDPG, and SAC. The results demonstrate that DDQN provides a favorable balance between control performance and computational efficiency, particularly in large-scale systems. Furthermore, the proposed approach is implemented and validated using an OPAL-RT real-time hardware-in-the-loop platform, confirming its practical applicability for real-time voltage control in next-generation smart grids.

Keywords: Deep reinforcement learning, double deep q-network, IEEE 14-bus and 124-bus systems, voltage stability, real-time simulation.

1. Introduction

The rapid increase in global energy demand, combined with climate change objectives, has made the transition from fossil fuels to renewable energy sources imperative [1]. In this transformation process, microgrids (MGs) and distributed generation systems offer significant opportunities in terms of sustainability, energy security, and grid flexibility [2]. However, the integration of variable energy sources such as solar and wind introduces new operational challenges, including voltage oscillations, frequency instabilities, and particularly fluctuations in DC-link voltage [3,4]. These issues can lead to the rapid propagation of faults and cascading outages, making the timely detection of abnormal conditions and appropriate corrective actions a critical necessity [5].

Power system stability (PSS) plays a central role in ensuring the reliable and efficient operation of electrical grids [6],[7]. However, the widespread integration of renewable energy sources, real-time uncertainties, and abrupt fluctuations in generation have made maintaining this stability increasingly complex [8]. Conventional voltage control methods, which commonly rely on mechanical equipment such as on-load tap changers (OLTCs), switchable capacitors, and voltage regulators, often respond too slowly to sudden load variations and rapid fluctuations in photovoltaic (PV) generation, rendering them insufficient [9], [10]. Furthermore, heuristic search techniques and game-theory-based optimization methods lack sufficient flexibility to cope with the complexity of grid dynamics and uncertainties. Additionally, heuristic search and theoretical game-based optimization techniques are not sufficiently flexible to cope with the complexity of grid dynamics and uncertainties; this limitation restricts their real-time applicability and leads to high computational costs [11].

In recent years, advances in data science and artificial intelligence have revealed the potential to enhance power

system stability by offering data-driven solutions to these challenges [12]. Deep learning (DL) and reinforcement learning (RL) algorithms can model complex operating conditions by learning system behaviors from large datasets [13]. Although these methods may carry the risk of not fully incorporating operational constraints, they offer significant advantages over traditional models in managing sudden changes. For instance, applications such as fault detection based on random vector functional link networks (RVFLNs) and load-frequency control using type-3 fuzzy logic have demonstrated promising results in scenarios characterized by high uncertainty [14],[15].

In this context, deep reinforcement learning (DRL) algorithms, in particular, have emerged as innovative approaches for model-free voltage and reactive power management. To effectively handle voltage deviations caused by sudden load changes and variable renewable generation, two-timescale control strategies have been proposed, enabling fast-responding inverters to operate in coordination with slower OLTCs and capacitors. For example, Sun and Qiu (2021) optimized voltage/reactive power control by integrating OLTCs and PV inverters through a multi-agent DRL architecture, demonstrating its effectiveness in real-world system scenarios [16]. Similarly, Wu et al. (2023) achieved faster regulation of reactive power flows by incorporating topological information into real-time optimization using a graph convolutional network-based DRL approach [17].

On the other hand, in active distribution networks where centralized control approaches fall short, multi-agent DRL (MADRL) models enable different agents to make decisions based on local data and operate in a coordinated manner. Cao et al. (2021) successfully implemented regional voltage regulation by developing a MADRL algorithm in a distribution system with high PV penetration [18]. Toubeau et al. (2020) enhanced the robustness of the DRL architecture against forecasting errors by accounting for model uncertainties [19]. Similarly, May et al. (2024) automated the participation of flexible loads in local energy markets using DRL, while Wang et al. (2020) demonstrated that the MADRL framework provided superior accuracy and efficiency compared to classical autonomous voltage control (AVC) in the Illinois 200-bus system [20],[21].

However, a major limitation of DRL-based control strategies is that the learned policies do not always guarantee system stability under all conditions. To address this issue, Shi et al. (2022) proposed the Stable-DDPG (Deep Deterministic Policy Gradient) algorithm, which ensures stability by incorporating Lyapunov-based constraints [22]. Petrusev et al. (2023) enhanced system stability by considering uncertainties in line impedances through a hybrid approach that combines offline training with online adaptation [23]. A selection of relevant studies on this topic is presented in Table 1.

High penetration of inverter-based renewable generation further complicates Volt-VAR optimization. Hossain et al. (2023) optimized Volt-VAR services by coordinating inverters with slower devices through a hybrid DDPG-DQN (Deep Q-Network) framework [34]. Xiang et al. (2023), on the other hand, developed a topology-aware voltage regulation method, successfully reducing voltage deviations through dynamic clustering of inverters and distributed energy storage devices [35].

In emergency scenarios where short-term voltage instabilities are prevalent, DRL-based control strategies offer faster and more adaptive solutions compared to traditional methods. Huang et al. (2020) proposed an adaptive DRL control scheme that incorporates generator braking and load shedding, and validated its performance on the IEEE 39-bus system [36]. Duan et al. (2020) demonstrated the integration of real-time DRL algorithms into grid operations by developing a closed-loop AVC architecture under the Grid Mind platform [37]. Li et al. (2022) further improved learning speed and accuracy in emergency voltage control by employing a DRL model augmented with expert supervision [38].

Innovative DRL applications in power systems are increasingly characterized by unique topologies and complex action spaces. Thayer and Overbye (2020) tested the Deep Q-Network (DQN) algorithm for automating human-based transmission voltage control and discussed its applicability in large-scale power networks [39]. Hagmar et al. (2023) incorporated a hybrid action space to enable real-time control of security margins, while Salehpour et al. (2025) achieved over 98% accuracy in fault diagnosis for PV systems using a two-stage DQN architecture [40], [41]. These pioneering examples demonstrate the broad applicability of DRL across diverse operational scenarios. Among them, the DF-SRL algorithm proposed by Hou et al. (2025) introduces an innovative approach that employs a DistFlow-based safety layer to ensure that agents learn policies without violating voltage limits. Compared to existing algorithms such as proximal policy optimization (PPO), twin delayed deep deterministic policy gradient (TD3), and soft actor-critic (SAC), this method yields lower voltage violations and shorter computation times [42].

In this study, an autonomous voltage control strategy based on deep reinforcement learning (DRL) is developed and implemented on the IEEE 14-bus test system. The model is deployed on the OPAL-RT real-time hardware platform, allowing the practical viability of the theoretical algorithm to be validated under real-world operational conditions. Using the Double Deep Q-Network (DDQN) algorithm, the system continuously monitors and regulates voltage deviations. Over time, the agent learns an optimal control policy by responding to voltage fluctuations, generator-load imbalances, and line power flows. The training and testing phases are extended to include diverse load scenarios and fault conditions, demonstrating that the proposed control structure is adaptive, stable, and capable of real-time response.

Overall, the study proposes a reliable, scalable, and hardware-deployable solution for autonomous voltage regulation in future active distribution grids. The original contributions of this work to the literature can be summarized as follows:

- The proposed DDQN-based control framework has been tested on both small-scale (IEEE 14-bus) and

large-scale (IEEE 124-bus) systems, demonstrating its scalability and generalizability.

- A comparative analysis with other popular DRL agents—such as PPO, SAC, Deep Deterministic Policy Gradient (DDPG), and classical DQN—has been conducted, showing that DDQN excels in terms of stable learning, reduced computation time, and higher reward performance.
- The entire control algorithm is implemented on a real-time hardware-in-the-loop (HIL) environment using OPAL-RT, validating the practical applicability of the academically proposed system.
- Subsystem performances such as voltage deviation suppression, generator-load coordination, and adaptation to load volatility are supported by comprehensive graphical and numerical results.
- In addition, a holistic and comparative analysis is presented for DRL-based voltage control through hyperparameter configurations and agent-level performance tables.

Table 1. Summary of related literature.

No	Author	Key Topic	Methodology	Application Area	Test System	Main Contribution
[24]	Yang et al. (2020)	Two-timescale voltage control	DRL + AC power flow	Distribution network	IEEE 47-bus, IEE 123-bus	Two-timescale voltage regulation using real data
[25]	Cao et al. (2022)	Two-timescale voltage control	MASAC DRL	Active distribution system	IEEE 33-, IEEE 123-, IEEE 342-bus	Two-timescale control without physical model
[26]	Karagiannopoulos et al. (2024)	Decentralized voltage control	DDPG MARL	Distribution network	Benchmark European LV grid	Emulating centralized solution via local DRL
[27]	Hu et al. (2022)	Active & reactive power coordination	EA-MAAC MARL	Distribution network	IEEE 33-bus, IEEE 123-bus	Two-timescale active/reactive optimization
[28]	Feng et al. (2024)	Stability-constrained RL	Lyapunov RL	Distributed voltage control	IEEE 13-bus, IEEE 123-bus	Ensures system stability
[29]	Hossain et al. (2024)	Model-based DRL learning	MB-DRL	Voltage control	IEEE 300-bus	97% improvement in training efficiency
[30]	Li et al. (2019)	PV inverter coordination	DRL agents	PV inverter grid	IEEE 37-bus	Maintains PV inverter voltage limits
[31]	Zhang et al. (2021)	Volt-VAR optimization	MARL	Smart grid	IEEE 13-, IEEE 123-bus	Effective Volt-VAR optimization
[32]	Zhang et al. (2023)	Load shedding	ConvLSTM-DQN	Large-scale power system	China Southern Grid	Short-term voltage recovery
[33]	Ruddick et al. (2024)	Home energy management	RL/MPC/Decision tree	Residential	Real residential home	Safe online RL implementation

2. Materials and Methods

2.1. Microgrid Model

In this study, the IEEE 14-bus test system, which is widely used for power system analyses, has been selected. The IEEE 14-bus network is extensively featured in the literature as a standard reference model for examining voltage stability, load flow analysis, critical node identification, and continuation power flow solutions [43]. The IEEE 14-bus system consists of 5 synchronous generators, 11 load centers, 14 transmission lines, and 4 transformers. This configuration allows for comprehensive analysis of both voltage profiles and reactive power balance [44]. The test system is typically modeled with a base voltage of 69 kV and a base power of 100 MVA. Within the system, one swing bus, four generator buses (PV buses), and the remaining buses as load buses (PQ buses) are defined [45].

In particular, for transient stability analyses, the behavior of the IEEE 14-bus network is evaluated by modeling the rotor angles of generators, critical clearing times, and various fault scenarios. Additionally, this system can be modeled using open-source MATLAB-based toolboxes such as PSAT or Matpower, and load flow solutions can be obtained using the Newton-Raphson method [44]. On the other hand, due to the nonlinear nature of power flow equations, the IEEE 14-bus system has also been widely discussed in the literature in the context of multiple or continuation power flow solutions [45]. In this study, the IEEE 14-bus system is modeled under both normal operating conditions and scenarios involving load increases and faults, providing a suitable testbed for critical bus identification, stability analysis, and power flow optimization. Thus, the applicability of the proposed methods to real-world systems is validated through testing on this standard benchmark network.

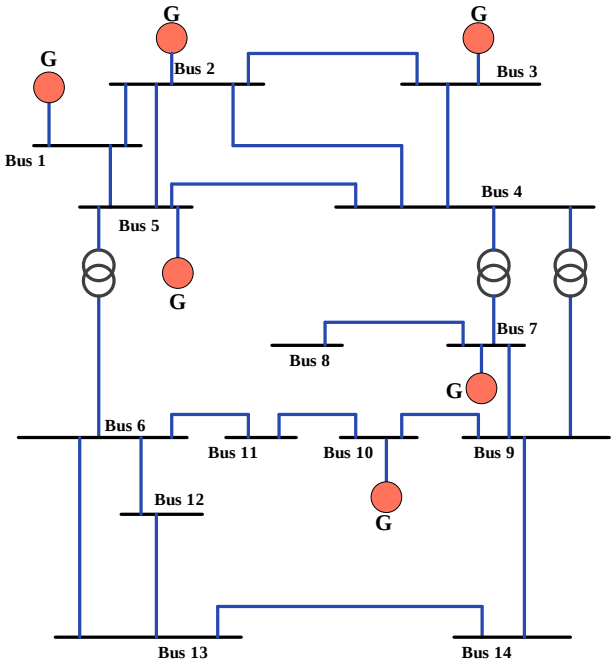

Fig. 1. IEEE 14-bus test system model.

2.2. Microgrid Model DDQN Algorithm

Reinforcement Learning (RL) is a machine learning approach in which an agent learns to maximize its cumulative reward by making sequential decisions within an environment. This method is inspired by the way humans and animals learn through trial-and-error interactions with their surroundings. In RL, the agent receives feedback from the environment in the form of rewards or penalties and learns which actions to take accordingly. Popular RL algorithms include Q-learning, SARSA, Deep Q-Network (DQN), and Proximal Policy Optimization (PPO). Q-learning is one of the most widely used algorithms in RL. It is a model-free learning algorithm that enables an agent to learn an optimal policy through a Markov Decision Process (MDP) [46]. Q-learning is particularly useful in scenarios where the agent has no prior knowledge of the environment's dynamics. The MDP framework is a mathematical formalism used to model decision-making problems in RL, consisting of a set of states, actions, and rewards. In Q-learning, the agent aims to learn an optimal action-value function, often referred to as the Q-value. The Q-value of a state-action pair (s, a) represents the expected cumulative reward obtained by the agent after taking action a in state s and subsequently following the optimal policy. This Q-value function is typically denoted as Q(s, a).

Mnih et al. introduced the Deep Q-Network (DQN), a learning method based on a newly defined value function, in 2015 [47]. DQN integrates neural networks with reinforcement learning (RL) to address complex problems that are otherwise difficult to solve [48]. This method utilizes function approximation to estimate values for state-action pairs, minimizing the need for extensive domain-specific knowledge in large-scale problems while optimizing computational resource usage. The main advantages of DQN are as follows:

i. Generalization Capability: DQN has the ability to generalize across vast state-action spaces. This allows the agent to make predictions even for previously unvisited states, eliminating the necessity of explicitly exploring every possible state and thereby significantly accelerating the learning process [49].

ii. Experience Replay: DQN employs an experience replay mechanism to break the correlation between sequential observations. The agent stores its experiences at each time step in a memory buffer (replay dataset). When updating the Q-learning estimates, random samples are drawn from this buffer. This approach reduces data correlation and leads to more stable and efficient learning [50].

iii. Target Network with Periodic Updates: Rather than updating the target network at every step, DQN updates it at fixed intervals (after a specified number of iterations). This strategy helps reduce instability and oscillations during training, enabling the algorithm to produce more stable and reliable outcomes [51].

In DQN, Q-values are estimated through a neural network; hence, DQN can be considered as a combination of Q-learning and a neural network. Artificial neurons transform one or more input values into output values. Each input is multiplied by a

weight that adjusts the strength of the input. These weighted input values are then passed through an activation function to produce the output value. An activation function is a mathematical function used by an artificial neuron to map its input values into an output. The input to the activation function is the linear combination of each input value multiplied by its associated weight. The activation function can be expressed mathematically as shown in Equation (1).

$$Out = \emptyset \sum_{i=1}^{n} w_i x_i \quad (1)$$

Here, φ represents the activation function, which can be a sigmoid, Rectified Linear Unit (ReLU), hyperbolic tangent (tanh), or similar function. The vector x denotes the input values, and w is the corresponding weight vector. The output layer contains the network's prediction. A deep learning model is essentially a neural network with multiple hidden layers. In deep Q-learning, the target for a given input state is calculated using a portion of the temporal difference formula, as shown in Equation (2):

$$T(s_t, a_t) = r_t + \gamma \max(Q(s_{t+1}, a)) \quad (2)$$

Here, T(st ,at) represents the target output value for the action taken in the previous state, rt is the reward received from performing that action, γ is a discount factor, and max(Q(st ,at)) denotes the maximum Q-value for any possible action in the current state [52].

An agent selects an action in the environment with the aim of achieving a specific goal. The environment then provides a reward based on the performed action and transitions to the next state. This process is typically modeled in reinforcement learning as a Markov Decision Process (MDP), represented as <S, A, P, r, γ> , where S denotes the state space, A the action space, P the state transition probability, r the reward, and γ the discount factor. One of the representative approaches to solving an MDP is the Q-learning algorithm, which estimates the expected value of the Q-function Q(s,a) for taking action aaa in state sss within an episode [52]. According to the Bellman optimality equation, this function is defined as shown in Equation (3):

$$Q(s,a) = E(r + \gamma P(s,s') \max\, Q(s',a')) \quad (3)$$

Here, E[·] denotes the expected value; P(s,s') represents the probability of transitioning from state sss to state s′; and a′ is the action taken in the subsequent state s′.

The traditional Q-learning algorithm stores the action-value estimates for each state-action pair in a table and updates them iteratively. However, when the state space is very large, storing and updating the values of all possible state-action pairs becomes computationally infeasible. To address this issue, Q-learning is combined with deep learning to form the Deep Q-Network (DQN) algorithm [49]. By using neural networks, the DQN algorithm learns to approximate the true reward function Q(s,a) with a parameterized value function Q(s,a|θ) , where θ represents the parameters of the neural network.

The DQN architecture, illustrated in Figure 2, is a feedforward neural network composed of an input layer, an output layer, and multiple hidden layers.

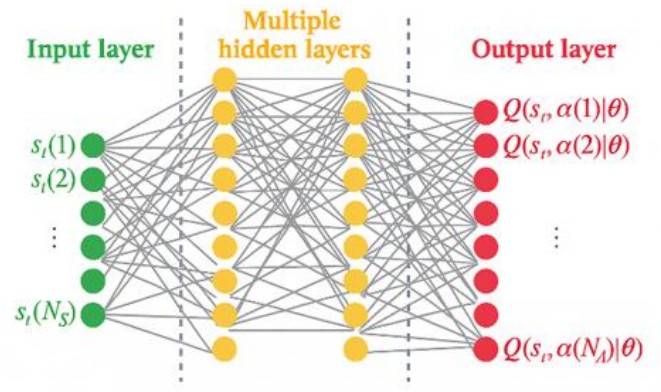

Fig. 2. Structure of DQN.

The input is the current state st' (st ∈ S) at time t, and the number of input neurons corresponds to the dimensionality of the state space NS. The output consists of the approximate action-value functions Q(st ,a|θ) for all possible actions a ∈A in that state st, where the output dimensionality is equal to the size of the action space NA. The agent selects the action associated with the highest Q-value from the output layer. The loss function used in DQN is defined in Equation (4) as follows:

$$L_{DQN}(\theta_i) = E\left(y_i^{DQN} - Q(s,a|\theta_i)\right)^2 \quad (4)$$

Here, y_i^{DQN} and θ_i represent the target function and the network parameters of Q(s,a|θ) at iteration i, respectively. The target function is defined by Equation (5) as follows:

$$y_i^{DQN} = r_i + \gamma Q(s', a^*_{DQN}|\theta_i^-) \quad (5)$$

Here, $Q(s', a^*_{DQN}|\theta_i^-)$, denotes the maximum Q-value in the next state s' ; ri is the reward at iteration i ; θ_i^- is the target network parameter; and a^*_{DQN} is the action corresponding to the maximum approximate Q-value in state s', as defined in Equation (6) [53].

$$a^*_{DQN} = argmax Q(s', a'|\theta_i^-) \quad (6)$$

Overestimation or underestimation of action values in DRL can significantly degrade learning performance. In the DQN algorithm, the same action-value estimate is used both for selecting and evaluating an action, which commonly leads to overestimation bias [49]. To address this issue, the DDQN algorithm was proposed in [49]. This algorithm mitigates overestimation by decoupling the action selection and evaluation processes when designing the reward target, resulting in better performance compared to standard DQN.

According to Equations (5) and (6), it can be observed that the action a^*_{DQN} in the DQN algorithm is both selected and evaluated using the same function $(s', a'|\theta_i^-)$. This increases the likelihood of choosing overestimated values, leading to the overestimation of action-value functions. To address this issue, the target function in the DDQN algorithm, denoted as y_i^{DDQN} is redefined as shown in Equation (7):

$$y_i^{DDQN} = r_i + \gamma Q(s', a^*_{DDQN}|\theta_i^-) \quad (7)$$

Here, a^*_{DDQN}, is the action with the maximum approximate value in state s', as defined in Equation (7).

$$a^*_{DDQN} = argmaxQ(s', a'|\theta_i^-) \quad (8)$$

From Equations (7) and (8), it can be observed that in the DDQN algorithm, the action a^*_{DDQN} is selected using $Q(s', a'|\theta_i^-)$ and evaluated using $Q(s', a'|\theta_i^-)$. The probability of overestimating the action value is significantly reduced through this dual-estimator design. However, in DDQN, the separation of action selection and evaluation may sometimes lead to underestimation issues, particularly in environments characterized by high stochasticity and uncertainty [54].

3. Finding

The distribution of generator and load buses for the IEEE 14-bus and IEEE 124-bus test systems used in this study is presented in Table 2. The IEEE 14-bus system was selected to test small-scale scenarios and examine the fundamental behavior of the control algorithm. In contrast, the IEEE 124-bus system features a larger topology, a higher number of load centers, and more complex power flow dynamics, and was therefore employed to evaluate the algorithm's performance in large-scale networks. Accurate identification of generator and load buses is critical for analyzing voltage profiles and assessing the effects of control actions. This ensures that the proposed DDQN-based control structure demonstrates consistent and scalable performance in both small- and large-scale grids.

In this study, simulations conducted on the IEEE 14-bus test system have demonstrated that the proposed DDQN-based voltage control framework delivers effective results under various operating scenarios. Voltage deviations across the system were successfully constrained both under nominal operating conditions and during sudden load changes and fault events.

Table 2. Comparison of average reward and computation time for different DRL agents on the IEEE 14-bus system.

Test System	Generator Buses	Load Buses
IEEE 14-bus	1, 2, 3, 5, 7, 10	1, 2, 3, 4, 5, 9, 11, 12, 13, 14
IEEE 124-bus	1, 12, 23, 45, 98, 117	3, 11, 34, 57, 83, 92, 98, 105, 113

Notably, as illustrated in Figure 3, minimum (green) and scaled maximum (red) voltage deviations for each bus are presented in a comparative manner. According to the results, voltage deviations were maintained at acceptable levels across all buses, with the highest deviation occurring at Bus 13. However, even this peak deviation did not exceed 0.6 p.u. The DDQN algorithm was found to minimize the overestimation problem compared to the classical DQN approach and exhibited a more stable learning curve. These findings confirm the capability of the developed control architecture to manage voltage profiles reliably and consistently in distributed power systems.

The impact of the DDQN-based control framework on active power flows within the system was observed for each transmission line over time steps.

The graph presented in Figure 4 illustrates the variation in power flows across 20 different lines throughout the simulation period. The overall regularity and bounded oscillation ranges observed in the figure indicate that the proposed algorithm provides both stable and adaptive energy dispatch. In particular, noticeable fluctuations were observed during the initial 1,000 steps due to the learning process; however, in the subsequent steps, power flows were balanced and high-frequency volatility was effectively suppressed. These results demonstrate that the DDQN approach offers an effective solution not only for voltage regulation but also for active power sharing across the network.

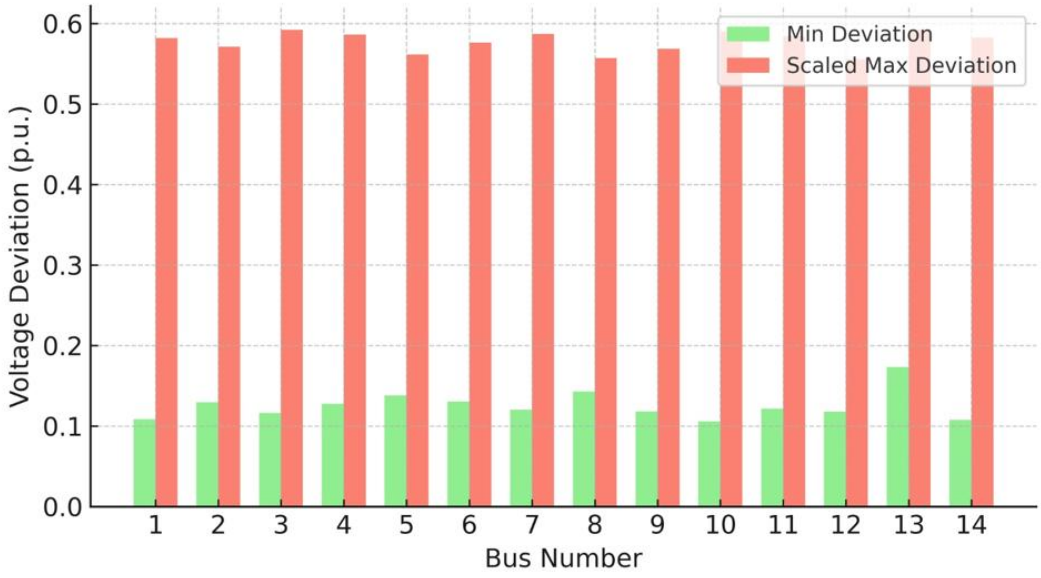

Fig. 3. Bus-wise minimum (green) and scaled maximum (red) voltage deviations (pu).

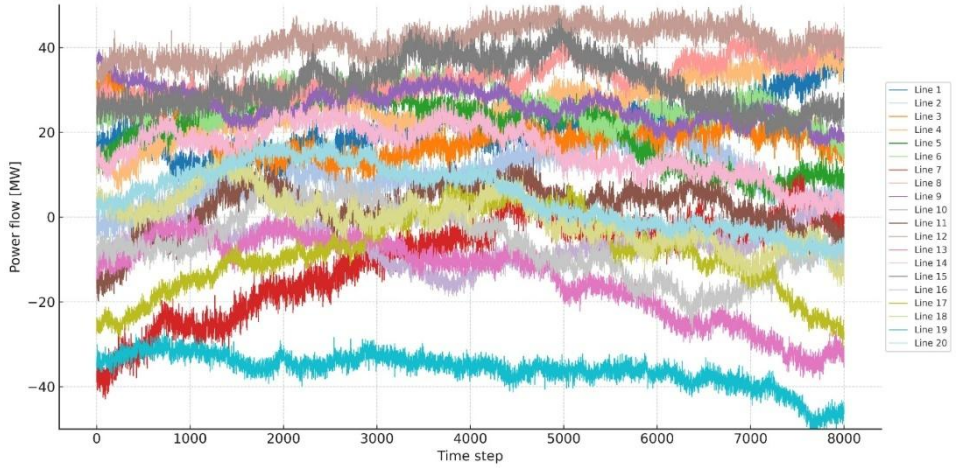

Fig. 4. Active power flows (MW, pu) over 20 transmission lines in the IEEE 14-bus system at different time steps.

Figure 5 illustrates the time-dependent variation in active power generation of the six generators within the system. The graph demonstrates that the DDQN algorithm effectively coordinates generator outputs. It is observed that high-capacity generators such as Gen 1 and Gen 2 respond rapidly to fluctuating demand, while lower-capacity generators such as Gen 5 and Gen 6 primarily play a supporting role. At certain time steps (e.g., around steps 100 and 300), the output of some generators approaches zero, indicating that load sharing within the system is dynamically optimized based on a learned strategy. This distribution highlights that the DDQN algorithm provides effective control not only for voltage regulation but also for generator coordination and production planning.

Figure 6 presents the time-dependent power demand trends of ten different load centers in the system. In this scenario, where the DDQN algorithm is implemented, the load profiles exhibit high volatility and randomness. The behavior of the loads is shaped particularly by scenario-based demand variations, which in turn necessitate dynamic adjustments in generator outputs. Fluctuating patterns observed in high-consumption centers such as Load 1 and Load 2 are among the key factors considered during the decision-making process of the DDQN controller. This figure demonstrates that, given the diversity and variability of loads over time, the proposed control algorithm successfully adapts to the system's changing conditions.

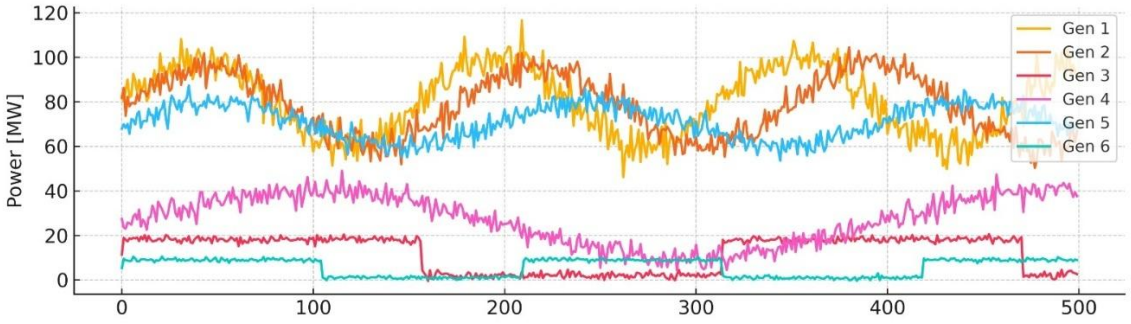

Fig. 5. Active power generation (MW, pu) of six different generators over time steps.

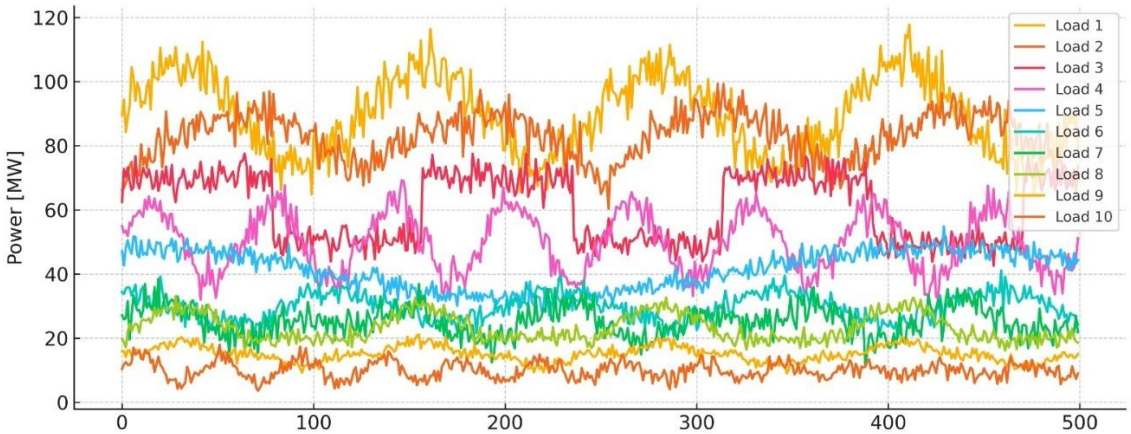

Fig. 6. Power demand variations (MW, pu) of 10 different loads in the system over time steps.

Figure 7 displays the voltage values of different buses over time under the control scenario where the DDQN algorithm is applied. The irregular fluctuations observed in the plot initially reflect the system's transient responses and the adaptation phase of the control algorithm. The concentration of data points within the 0.4–0.9 p.u. range indicates that bus voltages are generally kept under control across the system. Moreover, the presence of random variances in the voltage curves suggests that the simulation incorporates variability in both load and generation. This demonstrates that, despite sudden voltage deviations, the DDQN algorithm is able to effectively stabilize the overall voltage profile.

3.1. Comparative Performance Analysis of Different DRL Agents

In this study, the performance of the proposed DDQN algorithm is compared with other popular deep reinforcement learning (DRL) agents. The comparison is conducted on both the IEEE 14-bus and IEEE 124-bus test systems, and evaluation metrics include the average reward and computation time for each agent.

Table 3 presents the performance comparison of five different deep reinforcement learning (DRL) agents on the IEEE 14-bus test system. In terms of average reward, the PPO algorithm achieved the highest score with 92.410 points, followed by DDQN (78.920) and SAC (71.520). The DQN algorithm demonstrated the lowest performance, with an average reward of 63.540. Regarding computation time, DDPG produced results in the shortest time (40 seconds), whereas PPO incurred the highest computational burden with 82 seconds. DDQN achieved a significant improvement over DQN in both reward and computation time and exhibited a more balanced performance profile when compared to PPO.

Table 4 presents the performance results of the algorithms on the IEEE 124-bus system, which is a larger and more complex network. In this system as well, the PPO algorithm achieved the highest average reward, reaching 79.640 points. DDQN ranked second with a reward score of 63.120, while the DQN algorithm showed the lowest performance, with an average reward of 52.480. In terms of computation time, DDQN completed the task in 118 seconds—faster than DQN—and significantly more efficiently than PPO, which required 176 seconds. These findings indicate that the DDQN algorithm offers a meaningful improvement in reward while providing a computational advantage in large-scale systems.

Table 3. Comparison of average reward and computation time for different DRL agents on the IEEE 14-bus system.

Agent	Average Reward	Computation Time (s)
DQN	63.540	45
DDQN	78.920	53
PPO	92.410	82
DDPG	68.370	40
SAC	71.520	44

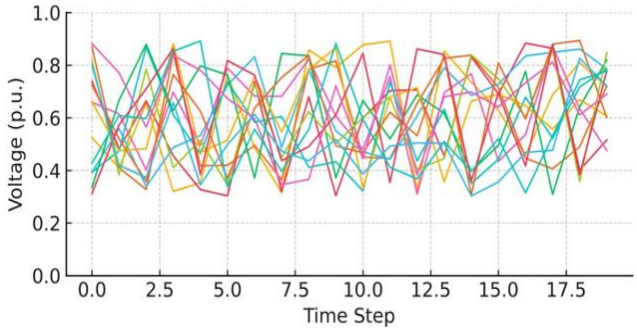

Fig. 7. Voltage variations (pu) of different buses in the system over time steps.

Table 4. Comparison of average reward and computation time for different DRL agents on the IEEE 124-bus system.

Agent	Average Reward	Computation Time (s)
DQN	52.480	150
DDQN	63.120	118
PPO	79.640	176
DDPG	59.350	100
SAC	62.280	105

3.2. OPAL-RT Hardware-in-Loop (HIL) Implementation and Control Execution Flow

Within the scope of this study, the real-time applicability of the proposed DDQN-based autonomous voltage control strategy is validated using the OPAL-RT OP4512 FPGA-based real-time simulator available at the Department of Electrical and Electronics Engineering, Dicle University. The OP4512 platform provides high computational capability and a low-latency communication infrastructure, enabling the real-time hardware-in-the-loop (HIL) execution of large-scale power system models. In this context, the IEEE 14-bus and IEEE 124-bus test systems are executed in real time on the OPAL-RT simulator, while the DDQN controller is configured as an external decision-making unit, forming a closed-loop control architecture. System states, including bus voltages, power flows, and generation–load imbalances, are synchronously measured through the OPAL-RT platform and transmitted to the controller. The optimal control actions computed by the controller are then fed back to the simulator without noticeable delay, thereby completing the real-time control loop.

The OPAL-RT system operates with a fixed sampling time of 50 μs, and all variables are scaled to the per-unit (p.u.) system to ensure numerical stability and compliance with real-time execution constraints. The developed control algorithm is designed by considering the FPGA-based processing limits of the OP4512 platform, and its reliable operation is verified by ensuring that all computations are completed within each simulation time step without any overrun. This real-time HIL infrastructure clearly demonstrates the practical applicability

and scalability of the proposed DDQN-based approach for power system voltage control. The real-time control loop consists of four stages:

- (i) acquisition of system states from the OPAL-RT simulator,
- (ii) action selection using an ε-greedy DDQN policy,
- (iii) application of the selected control actions to the simulated power system, and
- (iv) reward evaluation based on voltage deviation and system stability.

This structured workflow ensures consistent and reproducible real-time operation

3.3. Training Configuration and Hyperparameters

Within the scope of this study, to ensure a fair and systematic comparison among different deep reinforcement learning (DRL) algorithms, all agents were trained using carefully selected and widely adopted hyperparameters. The considered methods include Double Deep Q-Network (DDQN), Deep Q-Network (DQN), Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), and Soft Actor–Critic (SAC).

While a common training framework was maintained across all agents—such as the use of the Adam optimizer, identical batch sizes, and ReLU-based multilayer perceptron architectures—algorithm-specific parameters were tuned according to their underlying learning mechanisms. Value-based methods (DDQN and DQN) employ ε-greedy exploration strategies with gradual decay, whereas policy-based methods rely on entropy regularization or clipped surrogate objectives. For actor–critic architectures, soft target network updates were adopted to improve training stability.

The detailed hyperparameter settings and network configurations used in this study are summarized in Table 5.

Table 5. Hyperparameter settings for all models.

Hyperparameter	DDQN	DQN	PPO	DDPG	SAC
Learning Rate (α)	0.0005	0.0005	0.0003	0.0001 (Actor), 0.001 (Critic)	0.0003
Discount Factor (γ)	0.99	0.99	0.99	0.99	0.99
Replay Buffer Size	100,000	50,000	-	100,000	100,000
Batch Size	64	64	64	64	64
Target Network Update / Soft Update Rate	Every 500 steps / Hard update	Every 1,000 steps / Hard update	- / Clip range: 0.2	Soft update (τ=0.005)	Soft update (τ=0.005)
Exploration (ε-greedy) / Entropy Coefficient	ε: 1.0→0.01 over 20k steps	ε: 1.0→0.01 over 15k steps	-	-	α_entropy: 0.2
Optimizer	Adam	Adam	Adam	Adam	Adam
Network Architecture	2 layers, 128 neurons, ReLU	2 layers, 64 neurons, ReLU	2 layers, 128 neurons, ReLU	2 layers, 256 neurons, ReLU	2 layers, 256 neurons, ReLU
Update Frequency / Epochs	Every 4 steps	-	10 epochs/update	-	-

4. Discussion

This study demonstrates the effectiveness of a Double Deep Q-Network (DDQN)-based autonomous voltage control strategy for power distribution systems with high renewable penetration. Simulation results obtained from both the IEEE 14-bus and IEEE 124-bus test systems confirm that the proposed controller successfully maintains voltage stability under varying load and disturbance conditions while ensuring coordinated generator operation.

A comparative evaluation with other deep reinforcement learning (DRL) agents highlights the practical advantages of the proposed approach. Although PPO achieves higher reward values in some scenarios, its increased computational burden limits its suitability for real-time voltage control applications. Similarly, SAC provides robust performance but requires higher computational resources due to continuous action optimization. In contrast, DDQN offers a favorable trade-off between control performance and computational efficiency, enabling faster convergence and more reliable execution in real-time and large-scale power systems. These characteristics make DDQN particularly attractive for practical grid deployment.

The real-time hardware-in-the-loop implementation using the OPAL-RT platform further validates the feasibility of the proposed method under strict execution constraints. The closed-loop interaction between the DDQN controller and the real-time simulator demonstrates that the learned control policy can be effectively deployed in realistic operating environments.

5. Conclusion

This study presents an autonomous voltage control framework based on Deep Reinforcement Learning (DRL) to enhance voltage stability in modern power distribution systems. The increasing integration of renewable energy sources has rendered traditional rule-based and optimization-focused control strategies insufficient. In this context, the

proposed solution is structured around the Double Deep Q-Network (DDQN) algorithm and aims to regulate voltage deviations under dynamic operating conditions.

The developed control architecture was trained and tested on both the IEEE 14-bus and IEEE 124-bus test systems, enabling a comprehensive evaluation of its performance across different system scales. Notably, the algorithm was also implemented on the OPAL-RT real-time hardware platform, thereby validating the practical applicability and response capability of the theoretical model. By observing voltage deviations, generator-load interactions, and power flow patterns, the system successfully learned adaptive control policies.

The simulation results demonstrated that the DDQN-based controller effectively constrained voltage deviations across all buses under both nominal operating conditions and in scenarios involving faults and load variations. Furthermore, the proposed method was benchmarked against leading DRL agents such as DQN, PPO, DDPG, and SAC, showing superior performance on the large-scale IEEE 124-bus system in terms of lower computational cost, faster convergence, and higher reward achievement. Detailed graphical analyses of bus voltage levels, generator outputs, line power flows, and load demands confirmed the stability and adaptability of the developed control algorithm. The model not only ensured instantaneous voltage regulation but also enabled a coordinated and efficient operation across the entire power system.

In conclusion, this study demonstrates that the DDQN-based autonomous voltage regulation strategy offers a reliable, scalable, and hardware-compatible solution for real-time implementation in intelligent distribution networks. Additionally, by providing a comprehensive comparison of DRL agents and detailed hyperparameter references, the work contributes to the literature and establishes a solid foundation for future academic and industrial research.

Despite its effectiveness, the proposed approach has certain limitations. The current implementation focuses on a single-agent control structure, and scalability to fully decentralized multi-agent frameworks remains a topic for future research. In addition, communication delays and cyber-physical uncertainties were not explicitly modeled. Future studies will investigate multi-agent extensions, robustness under communication constraints, and field-scale implementations in real distribution networks.

Author Contributions

C.H., H.K. and H.K. A.T. conceptualized and designed the study; developed the methodology; performed the analysis; and wrote the original draft of the manuscript. All authors have read and approved the final version of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] C. Haydaroğlu, “Chaos-based optimization for load frequency control in Islanded airport microgrids with hydrogen energy and electric aircraft,” International Journal of Hydrogen Energy, no. October 2024, Jan. 2025.

[2] S. Shahzad, M. A. Abbasi, H. Ali, M. Iqbal, R. Munir, and H. Kilic, “Possibilities, Challenges, and Future Opportunities of Microgrids: A Review,” Sustainability (Switzerland), vol. 15, no. 8, 2023.

[3] H. Kilic, M. E. Asker, C. Haydaroglu, “Enhancing power system reliability: Hydrogen fuel cell-integrated D-STATCOM for voltage sag mitigation,” International Journal of Hydrogen Energy, no. March, Mar. 2024.

[4] S. Imtiaz, L. Yang, H. M. Munir, Z. A. Memon, H. Kilic, M. N. Naz, “DC-link voltage stability enhancement in intermittent microgrids using coordinated reserve energy management strategy,” IET Renewable Power Generation, vol. 19, no. 1, Jan. 2025.

[5] R. Diao, Z. Wang, D. Shi, Q. Chang, J. Duan, X. Zhang, “Autonomous Voltage Control for Grid Operation Using Deep Reinforcement Learning,” IEEE Power and Energy Society General Meeting, vol. 2019-Augus, 2019.

[6] N. Hatziargyriou, “Definition and Classification of Power System Stability – Revisited & Extended,” IEEE Transactions on Power Systems, vol. 36, no. 4, pp. 3271–3281, 2021.

[7] M. S. Massaoudi, H. Abu-Rub, A. Ghrayeb, “Navigating the Landscape of Deep Reinforcement Learning for Power System Stability Control: A Review,” IEEE Access, vol. 11, no. October, pp. 134298–134317, 2023.

[8] M. Z. Oskouei, B. Mohammadi-ivatloo, “Optimal Allocation of Renewable Sources and Energy Storage Systems in Partitioned Power Networks to Create Supply-Sufficient Areas,” IEEE Transactions on Sustainable Energy, vol. 12, no. 2, pp. 999–1008, 2021.

[9] B. A. Robbins, H. Zhu, A. D. Domínguez-garcía, “Optimal Tap Setting of Voltage Regulation Transformers in Unbalanced Distribution Systems,” IEEE Transactions on Power Systems, vol. 31, no. 1, pp. 256–267, 2016.

[10] W. Wang, N. Yu, Y. Gao, J. Shi, “Safe Off-Policy Deep Reinforcement Learning Algorithm for Volt-VAR Control in Power Distribution Systems,” IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 3008–3018, 2020.

[11] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, F. Blaabjerg, “Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review,” Journal of Modern Power Systems and Clean Energy, vol. 8, no. 6, pp. 1029–1042, 2020.

[12] D. Wu, X. Zheng, D. Kalathil, L. Xie, "Nested Reinforcement Learning Based Control for Protective Relays in Power Distribution Systems," 2019 IEEE 58th Conference on Decision and Control (CDC), no. Cdc, pp. 1925–1930, 2019.

[13] S. Mukherjee, T. L. Vu, "Reinforcement Learning of Structured Stabilizing Control for Linear Systems with Unknown State Matrix," IEEE Transactions on Automatic Control, vol. 68, no. 3, pp. 1746–1752, 2023.

[14] C. Haydaroğlu, H. Kılıç, B. Gümüş, M. T. Özdemir, "Advancing Fault Detection in Distribution Networks with a Real-Time Approach Using Robust RVFLN," Applied Sciences, vol. 15, no. 4, p. 1908, Feb. 2025.

[15] İ. Türk, H. Kılıç, C. Haydaroğlu, A. Top, "Robust Load Frequency Control in Hybrid Microgrids Using Type-3 Fuzzy Logic Under Stochastic Variations," Symmetry, vol. 17, no. 6, p. 853, May 2025.

[16] X. Sun, J. Qiu, "Two-Stage Volt/Var Control in Active Distribution Networks with Multi-Agent Deep Reinforcement Learning Method," IEEE Transactions on Smart Grid, vol. 12, no. 4, pp. 2903–2912, 2021.

[17] H. Wu, Z. Xu, M. Wang, J. Zhao, X. Xu, "Two-stage voltage regulation in power distribution system using graph convolutional network-based deep reinforcement learning in real time," International Journal of Electrical Power and Energy Systems, vol. 151, no. April, 2023.

[18] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, Z. Chen, F. Blaabjerg, "Data-Driven Multi-Agent Deep Reinforcement Learning for Distribution System Decentralized Voltage Control with High Penetration of PVs," IEEE Transactions on Smart Grid, vol. 12, no. 5, pp. 4137–4150, 2021.

[19] J. F. Toubeau, B. B. Zad, M. Hupez, Z. de Grève, F. Vallée, "Deep reinforcement learning-based voltage control to deal with model uncertainties in distribution networks," Energies, vol. 13, no. 15, 2020.

[20] D. C. May, M. Taylor, P. Musilek, "Decentralized Coordination of Distributed Energy Resources through Local Energy Markets and Deep Reinforcement Learning," Energy and AI, vol. 18, no. November, p. 100446, 2024.

[21] S. Wang, J. Duan, D. Shi, C. Xu, H. Li, R. Diao, Z. Wang, "A Data-Driven Multi-Agent Autonomous Voltage Control Framework Using Deep Reinforcement Learning," IEEE Transactions on Power Systems, vol. 35, no. 6, pp. 4644–4654, 2020.

[22] Y. Shi, G. Qu, S. Low, A. Anandkumar, A. Wierman, "Stability Constrained Reinforcement Learning for Real-Time Voltage Control," Proceedings of the American Control Conference, vol. 2022-June, pp. 2715–2721, 2022.

[23] A. Petrusev, M. A. Putratama, R. Rigo-Mariani, V. Debusschere, P. Reignier, N. Hadjsaid, "Reinforcement learning for robust voltage control in distribution grids under uncertainties," Sustainable Energy, Grids and Networks, vol. 33, p. 100959, 2023.

[24] Q. Yang, G. Wang, A. Sadeghi, G. B. Giannakis, J. Sun, "Two-Timescale Voltage Control in Distribution Grids Using Deep Reinforcement Learning," IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2313–2323, 2020.

[25] D. Cao, J. Zhao, W. Hu, N. Yu, F. Ding, Q. Huang, Z. Chen, "Deep Reinforcement Learning Enabled Physical-Model-Free Two-Timescale Voltage Control Method for Active Distribution Systems," IEEE Transactions on Smart Grid, vol. 13, no. 1, pp. 149–165, 2022.

[26] S. Karagiannopoulos, P. Aristidou, G. Hug, A. Botterud, "Decentralized control in active distribution grids via supervised and reinforcement learning," Energy and AI, vol. 16, no. October 2023, p. 100342, 2024.

[27] D. Hu, Z. Ye, Y. Gao, Z. Ye, Y. Peng, N. Yu, "Multi-Agent Deep Reinforcement Learning for Voltage Control with Coordinated Active and Reactive Power Optimization," IEEE Transactions on Smart Grid, vol. 13, no. 6, pp. 4873–4886, 2022.

[28] J. Feng, Y. Shi, G. Qu, S. H. Low, A. Anandkumar, A. Wierman, "Stability Constrained Reinforcement Learning for Decentralized Real-Time Voltage Control," IEEE Transactions on Control of Network Systems, vol. 11, no. 3, pp. 1370–1381, 2023.

[29] R. R. Hossain, T. Yin, Y. Du, R. Huang, J. Tan, W. Yu, Y. Liu, Q. Huang, "Efficient learning of power grid voltage control strategies via model-based deep reinforcement learning," Machine Learning, vol. 113, no. 5, pp. 2675–2700, 2024.

[30] C. Li, C. Jin, R. Sharma, "Coordination of PV smart inverters using deep reinforcement learning for grid voltage regulation," Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019, pp. 1930–1937, 2019.

[31] Y. Zhang, X. Wang, J. Wang, Y. Zhang, "Deep Reinforcement Learning Based Volt-VAR Optimization in Smart Distribution Systems," IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 361–371, 2021.

[32] J. Zhang, Y. Luo, B. Wang, C. Lu, J. Si, J. Song, "Deep Reinforcement Learning for Load Shedding Against Short-Term Voltage Instability in Large Power Systems," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 8, pp. 4249–4260, 2023.

[33] J. Ruddick, G. Ceusters, E. Van Kriekinge Gillesand Genov, T. Coosemans, M. Messagie, "Real-world validation of safe reinforcement learning, model predictive control and decision tree-based home energy management systems," arXiv [eess.SY], vol. 18, no. November, 2024.

[34] R. Hossain, M. Gautam, J. Thapa, H. Livani, M. Benidris, "Deep reinforcement learning assisted co-optimization of Volt-VAR grid service in distribution

networks," Sustainable Energy, Grids and Networks, vol. 35, p. 101086, 2023.

[35] Y. Xiang, Y. Lu, J. Liu, "Deep reinforcement learning based topology-aware voltage regulation of distribution networks with distributed energy storage," Applied Energy, vol. 332, no. September 2022, 2023.

[36] Q. Huang, R. Huang, W. Hao, J. Tan, R. Fan, Z. Huang, "Adaptive Power System Emergency Control Using Deep Reinforcement Learning," IEEE Transactions on Smart Grid, vol. 11, no. 2, pp. 1171–1182, 2020.

[37] J. Duan, D. Shi, R. Diao, H. Li, Z. Wang, B. Zhang, D. Bian, Z. Yi, "Deep-Reinforcement-Learning-Based Autonomous Voltage Control for Power Grid Operations," IEEE Transactions on Power Systems, vol. 35, no. 1, pp. 814–817, 2020.

[38] X. Li, X. Wang, X. Zheng, Y. Dai, Z. Yu, J. J. Zhang, G. Bu, F. Wang, "Supervised assisted deep reinforcement learning for emergency voltage control of power systems," Neurocomputing, vol. 475, pp. 69–79, 2022.

[39] B. L. Thayer, T. J. Overbye, "Deep reinforcement learning for electric transmission voltage control," 2020 IEEE Electric Power and Energy Conference, EPEC 2020, vol. 3, pp. 1–8, 2020.

[40] H. Hagmar, R. Eriksson, L. A. Tuan, "Real-time security margin control using deep reinforcement learning," Energy and AI, vol. 13, no. October 2022, p. 100244, 2023.

[41] S. Salehpour, A. Eskandari, A. Nedaei, M. Aghaei, "Two-stage deep Q-network reinforcement learning based ultra-efficient fault diagnosis and severity assessment scheme for photovoltaic protection," Energy and AI, vol. 20, no. April, 2025.

[42] S. Hou, A. Fu, E. M. S. Duque, P. Palensky, Q. Chen, P. P. Vergara, "DistFlow Safe Reinforcement Learning Algorithm for Voltage Magnitude Regulation in Distribution Networks," Journal of Modern Power Systems and Clean Energy, vol. 13, no. 1, pp. 300–311, 2024.

[43] B. Liu, F. Liu, B. Zhai, H. Lan, "Investigating continuous power flow solutions of IEEE 14-bus system," IEEJ Transactions on Electrical and Electronic Engineering, vol. 14, 219AD.

[44] P. K. Iyambo, R. Tzoneva, "Transient Stability Analysis of the IEEE 14-Bus Electric Power System," AFRICON 2007, pp. 1–9, 2025.

[45] P. R. Sharma, R. K. Ahuja, S. Vashisth, V. Hudda, "Computation of Sensitive Node for IEEE- 14 Bus system Subjected to Load Variation," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, vol. 2, no. 6, pp. 1603–1606, 2014.

[46] S. Jain, E. Fallon, "Leveraging Unstructured Data to Improve Customer Engagement and Revenue in Financial Institutions: A Deep Reinforcement Learning Approach to Personalized Transaction Recommendations," International Conference on Computer, Information and Telecommunication Systems (CITS), 2023, pp. 01–08.

[47] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[48] N. D. Nguyen, T. Nguyen, S. Nahavandi, "System Design Perspective for Human-Level Agents Using Deep Reinforcement Learning: A Survey," IEEE Access, vol. 5, pp. 27091–27102, 2017.

[49] H. Van Hasselt, A. Guez, D. Silver, "Deep Reinforcement Learning with Double Q-Learning," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30, no. 1, pp. 10695–10701, Mar. 2016.

[50] T. Schaul, J. Quan, I. Antonoglou, D. Silver, "Prioritized experience replay," arXiv preprint arXiv:1511.05952, 2015.

[51] K. Geevers, "Deep Reinforcement Learning in Inventory Management," Master Thesis, University of Twente, 2020.

[52] D. A. Zeleke, H. D. Kim, "A New Strategy of Satellite Autonomy with Machine Learning for Efficient Resource Utilization of a Standard Performance CubeSat," Aerospace, vol. 10, no. 1, 2023.

[53] Y. Wang, M. Mao, L. Chang, N. D. Hatziargyriou, "Intelligent Voltage Control Method in Active Distribution Networks Based on Averaged Weighted Double Deep Q-network Algorithm," Journal of Modern Power Systems and Clean Energy, vol. 11, no. 1, pp. 132–143, 2023.

[54] Z. Zhang, Z. Pan, M. J. Kochenderfer, "Weighted Double Q-learning," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, vol. 0, no. 2, pp. 3455–3461.