Safe Multi-Agent Reinforcement Learning Volt-Var Optimization in Active Distribution Networks

Mohammad Hashemnezhad and Petros Aristidou
Department of Electrical Engineering and Computer Engineering and Informatics
Cyprus University of Technology, Limassol, Cyprus
ma.hashemnezhad@edu.cut.ac.cy, petros.aristidou@cut.ac.cy

***Abstract*—Volt-Var Optimization (VVO) plays a critical role in Active Distribution Networks (ADNs) by ensuring voltage stability and minimizing power losses, particularly with the increasing integration of distributed photovoltaic (PV) systems. In this paper, we propose a decentralized control framework using Multi-Agent Reinforcement Learning (MARL), enabling PV inverters to independently control their reactive power to minimize power losses while maintaining voltage within safe operational limits. To mitigate unsafe actions, such as voltage violations and line overloading, a pre-trained Deep Neural Network (DNN) is integrated as a safety layer. The DNN projects unsafe MARL-generated actions into a feasible space, ensuring operational safety. Our approach is evaluated on a modified 33-bus medium-voltage test network across three scenarios: (1) a base case with no control, (2) MARL without a safety layer, and (3) MARL with a safety layer. The results demonstrate that MARL with the safety layer achieves the greatest reduction in power losses while ensuring voltage stability across all buses. This study underscores the potential of combining MARL with safety mechanisms to enhance the reliability and efficiency of ADNs.**

***Index Terms*—Volt-Var Optimization (VVO), Active Distribution Networks (ADNs), Multi-Agent Reinforcement Learning (MARL), PV Inverter Control, Voltage Stability Safety Layer, MADDPG**

I. Introduction

The transition to decentralized renewable energy has transformed traditional power distribution networks into Active Distribution Networks (ADNs) [1]. Distributed Energy Resources (DERs) such as PV systems, Wind Turbines (WT), and Energy Storage Systems (ESS) introduce variability and uncertainty, complicating voltage regulation, reactive power management, and overall stability [2]. Among these challenges, voltage regulation is critical for reliable and efficient ADN operation. Volt/VAR Control (VVC) and Volt/VAR Optimization (VVO) help maintain voltage limits, minimize power losses, and enhance network efficiency [3]. However, rising DER penetration increases complexity, necessitating adaptive, intelligent control solutions. Data-driven control schemes have become essential, providing scalable, real-time solutions to modern ADN challenges [4].

Data-driven methods leverage operational data with Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL) to develop dynamic control policies [5]. Unlike traditional model-based methods, they bypass detailed physical models, making them ideal for high-DER environments with complex dynamics. These approaches enable real-time decision-making, effectively managing renewable energy variability and fluctuating loads. RL and DL have been applied to various ADN challenges, including demand response [6], energy management [7], and congestion management [8].

VVO has gained attention for reducing power losses and maintaining voltage stability. A data-driven VVO method in [9] employs ML-based local controllers for real-time voltage adjustments. MARL-based VVO in [10] uses MASAC for autonomous voltage regulation with centralized training and distributed execution. Dual-timescale regulation in [11] integrates day-ahead optimization with real-time PV inverter control via Multi-Agent Soft Actor-Critic (MASAC), while [12] presents a decentralized approach combining offline supervised learning and Deep Deterministic Policy Gradient (DDPG) for adaptive DER management. However, these methods lack safety guarantees, as agents can take unsafe actions, compromising network stability.

To mitigate these risks, recent studies have integrated safety layers into control frameworks. [13] proposes constrained policy optimization within an off-policy RL framework to ensure actions comply with safety limits. [14] introduces a projection-based safety mechanism using a supervisor-projector-enhanced SAC algorithm to adjust unsafe actions in real time. A hybrid approach in [15] employs a Supervisor-Projector-Enhanced Soft Actor-Critic (S3AC) method with Gaussian Process Regression (GPR) to predict and ensure action safety. However, these studies often rely on centralized frameworks or extensive inter-agent communication, limiting scalability. Moreover, most evaluations are short-term (single-day), lacking long-term performance analysis and comprehensive consideration of line loading and sustained power loss minimization.

To address these gaps, this paper proposes a decentralized control framework for VVO in ADNs using MARL, enhanced with a robust pretrained safety layer. The framework enables PV inverters to operate independently, using local measurements to collaboratively minimize power losses, maintain voltage stability, and prevent line overloading. A pretrained Projection Deep Neural Network (DNN) serves as the safety layer, adjusting unsafe actions before execution to ensure operational constraints are met. Unlike previous studies that rely on online safety constraints or centralized safety mechanisms, our approach leverages an offline-trained DNN that efficiently projects unsafe actions into a feasible space with minimal computational overhead. This eliminates the need for real-time constraint enforcement, making the method more scalable and suitable for large-scale ADNs. Furthermore, our framework is evaluated over extended testing periods, providing deeper insights into its long-term reliability and effectiveness in maintaining network stability and minimizing losses. The contributions of this paper are as follows:

- A decentralized MARL-based VVO framework with a pretrained projection DNN ensuring voltage and line loading constraints.
- The pretrained DNN improves computational efficiency by reducing the need for real-time safety corrections during MARL execution.
- Evaluated over extended time horizons, demonstrating long-term reliability in reducing power losses and maintaining network stability.

The remainder of the paper is structured as follows. Section II describes the VVO model in ADNs. The proposed MARL algorithms are detailed in Section III. Section IV presents the numerical results, which are based on real validated data. The paper concludes with Section V.

II. Volt/Var Optimization (VVO)

VVO enhances modern ADNs by coordinating network-wide adjustments to minimize power losses, maintain voltage stability, and improve efficiency. Unlike VVC, which reacts locally to voltage deviations, VVO employs advanced optimization to balance real-time and long-term objectives across multiple devices, such as PV inverters. Data-driven methods like MARL enable decentralized, adaptive control in VVO but pose challenges in ensuring safe control actions under dynamic conditions. To mitigate this, a DNN-based safety layer adjusts unsafe MARL actions in real-time. The following subsections detail the MARL model and the role of the DNN safety layer.

A. Multi-Agent Reinforcement Learning (MARL)

MARL, an extension of RL, enables multiple agents to learn policies while interacting with their environment and considering other agents' actions. It is well-suited for ADNs, where DERs like PV inverters operate in decentralized settings. Through iterative interactions, MARL optimizes power loss and voltage regulation under uncertainty.

This study employs Multi-Agent Deep Deterministic Policy Gradient (MADDPG), a state-of-the-art MARL algorithm for continuous action spaces. MADDPG enhances DDPG for multi-agent settings, utilizing centralized training with decentralized execution. It was selected over alternatives for its reduced overestimation bias, improved stability, and better sample efficiency, making it ideal for reactive power control in ADNs [16]. The key components of the MARL model are as follows.

- *States*: Each agent (representing a PV inverter) observes local voltage magnitudes, active and reactive power demands, and local generation. The global state, used during training, aggregates information from all agents to provide a comprehensive system view.
- *Actions*: The agents' actions correspond to the reactive power outputs of PV inverters. These actions are bounded by physical constraints of the inverters and are adjusted to minimize losses while maintaining voltage stability.
- *Rewards*: The reward function is carefully designed to encourage behaviors that achieve the system objectives. That is, minimizing power losses while penalizing voltage violations and line overloading.

MADDPG employs an actor-critic framework with separate actor and critic networks for policy learning and value estimation. It enhances stability through policy delays and clipped double Q-learning, which mitigates overestimation bias using two critic networks. Experience replay further improves sample efficiency by storing and reusing past experiences. These features enable robust policy learning in multi-agent environments with high DER penetration and stochastic demand. When integrated with the DNN safety layer, this framework effectively balances performance optimization and operational safety in ADNs.

B. DNN-based Safety Layer

The pretrained DNN safety layer ensures adherence to operational constraints in ADNs by acting as a projection mechanism that maps unsafe MARL actions into a feasible action space, enforcing voltage and line loading limits. Trained offline on historical or simulated datasets of grid states and corresponding safe actions, it generalizes across diverse scenarios to provide reliable real-time corrections. The DNN receives a combined input vector of the current grid state, including bus voltages, active/reactive power demands, and RL-generated actions. Through its multilayer architecture, it produces adjusted actions that meet safety requirements, maintaining voltages within 0.95–1.05 pu and preventing line overload.

The projection DNN architecture consists of three fully connected layers:

- *Input Layer*: Processes a combined state-action vector (e.g., 150 dimensions in the studied scenario).
- *Hidden Layers*: Two layers, each with 128 neurons, activated by ReLU functions to capture complex relationships between states and safe actions.
- *Output Layer*: Produces the adjusted reactive power actions for the PV inverters, ensuring compliance with safety constraints.

The DNN is trained using a supervised learning approach, minimizing the mean squared error (MSE) between its predicted actions and optimal actions derived from offline Optimal Power Flow (OPF) solutions. This pretraining allows the DNN to serve as a reliable safety mechanism during run-time, adjusting unsafe RL actions to ensure that the network remains within safe operational boundaries. By integrating this safety layer, the system effectively balances performance optimization with operational reliability, addressing key challenges in real-time ADN control.

Algorithm 1: Pretraining the Projection DNN for Safe VVO

```
Input: Historical ADN data, optimal control actions (safe reactive power
adjustments), neural network hyperparameters.
Output: Projection DNN for correcting unsafe reactive power actions.
Load historical ADN dataset containing state-action pairs $(S, A)$.
Normalize the dataset using a standard scaler to improve training stability.
Initialize Projection DNN with input layer (state-action features), hidden
layers, and output layer (corrected actions).
for each training epoch do
    Shuffle dataset and divide it into mini-batches.
    for each mini-batch do
        Extract input features (state-action pairs) and corresponding labels
        (optimal safe actions).
        Forward propagate through DNN to compute predicted safe actions.
        Compute loss using MSE between predicted and true safe actions.
        Backpropagate gradients and update weights using Adam optimizer.
    end for
    Evaluate model performance on validation set.
    Apply early stopping if validation loss does not improve.
end for
Save trained DNN model and standard scaler for real-time deployment.
```

III. Proposed model

The proposed approach integrates a MARL framework with a pretrained Projection DNN to ensure safe and efficient VVO in ADNs. In the pretraining phase (Algorithm 1), data is collected from extensive power flow simulations under varying load conditions, PV generation levels, and control actions. This generates a dataset of state-action pairs, where states include bus voltages, power injections, and network topology, while actions correspond to reactive power adjustments by PV inverters. Optimal reactive power settings are computed via OPF for each state to provide labeled data.

The DNN safety layer is trained to map state-action pairs to safe reactive power values using an MSE loss function. Inputs are normalized and implemented within the PyTorch framework. The model architecture includes an input layer for state variables and RL actions, two hidden layers with ReLU activations, and an output layer predicting adjusted reactive power settings. After validation on unseen test cases, the trained DNN is deployed as a fixed safety mechanism, intervening only to replace unsafe actions during MARL execution.

During MARL operation (Algorithm 2), PV inverters function as independent agents, selecting reactive power actions based on local observations. These actions are evaluated through power flow analysis, and the Projection DNN adjusts them if violations occur. The MARL framework employs the MADDPG algorithm implemented using Pytorch to minimize power loss and maintain voltage stability, leveraging an experience replay buffer for sample efficiency. The effectiveness of this approach is demonstrated in the next section by comparing it with a baseline scenario (no control) and MARL without the safety layer, highlighting its superior safety and efficiency in ADNs.

IV. Numerical Analysis

A. Dataset and train setup

This study uses a modified IEEE 33-bus distribution network, incorporating six PV systems installed at various buses. The load profiles are derived from Portuguese electricity consumption data in real time collected over a 3-year period [17]. PV generation profiles are derived from real-time solar power data provided by Elia Group, a Belgian transmission system operator [18]. To facilitate high-resolution control, both load and PV data are interpolated at a resolution of 3 minutes, aligned with the real-time control period of the grid. The training process consists of two primary components: (1) MARL training for VVO and (2) pretrained DNN safety layer training for action correction. The range of action of reactive power for each agent is set to $[-0.8, 0.8]$ per unit, ensuring feasible control within operational constraints. The hyperparameters for both training processes are presented in Tables I and II. These values were selected based on previous studies, empirical tuning and stability considerations to balance learning efficiency, convergence speed, and policy robustness in the MARL and DNN training processes. All simulations are conducted on a Windows 10 PC equipped with a 2.80 GHz Intel i7-7700 CPU and 16 GB of RAM.

Algorithm 2: MARL with Pretrained DNN Safety Layer for VVO

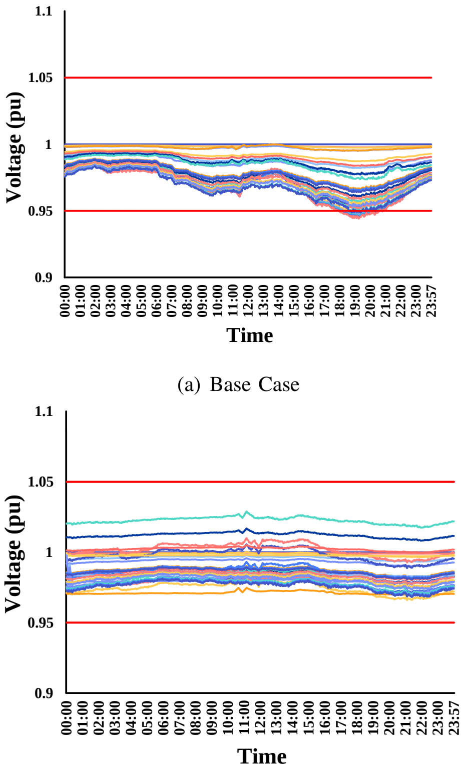
```
Input: ADN state, PV inverter specifications, pretrained Projection DNN,
and MARL agent parameters.
Output: Optimized reactive power ensuring voltage stability and power
loss minimization.
Initialize MARL agents with actor-critic networks and replay buffer.
Load pretrained Projection DNN and scaler for action adjustments.
for each episode do
    Reset the environment to obtain the initial state $S_t$.
    for each timestep in the episode do
        Agents observe local state $O_{t_i}$ and select actions $A_{t_i}$ by their policy.
        Take action $A_{t_i}$ and run power flow using the power flow solver.
        if actions result in unsafe conditions do
            Pass the state-action pair to the pretrained Projection DNN.
            Replace unsafe actions with safe actions provided by the DNN.
            Take the safe actions and simulate power flow.
        end if
        Observe the next state $S_{t+1}$ and calculate the reward $R_{t_i}$.
        Store the transition $(S_t, A_t, R_t, S_{t+1})$ in the replay buffer.
        sample from the replay buffer to update the actor-critic networks.
    end for
    Record cumulative rewards (minimizing power loss with voltage devi-
ation and line overloading penalties).
end for
Return trained policy networks for PV inverters.
```

TABLE I
Hyperparameters Used for MARL Training

Hyperparameter	Value
Learning Rate	0.001
Discount Factor (γ)	0.99
Batch Size	32
Replay Buffer Size	5×10^3
Target Network	$\tau = 0.1$
Gradient Clipping	$\epsilon = 1.0$
Training Episodes	400

TABLE II
Hyperparameters Used for DNN Training

Hyperparameter	Value
Input Dimension	150 (144 state + 6 action)
Output Dimension	6 (Adjusted actions)
Hidden Layers	2 layers (128 neurons each)
Learning Rate	0.001
Batch Size	32
Epochs	50



(a) Base Case

(b) MARL

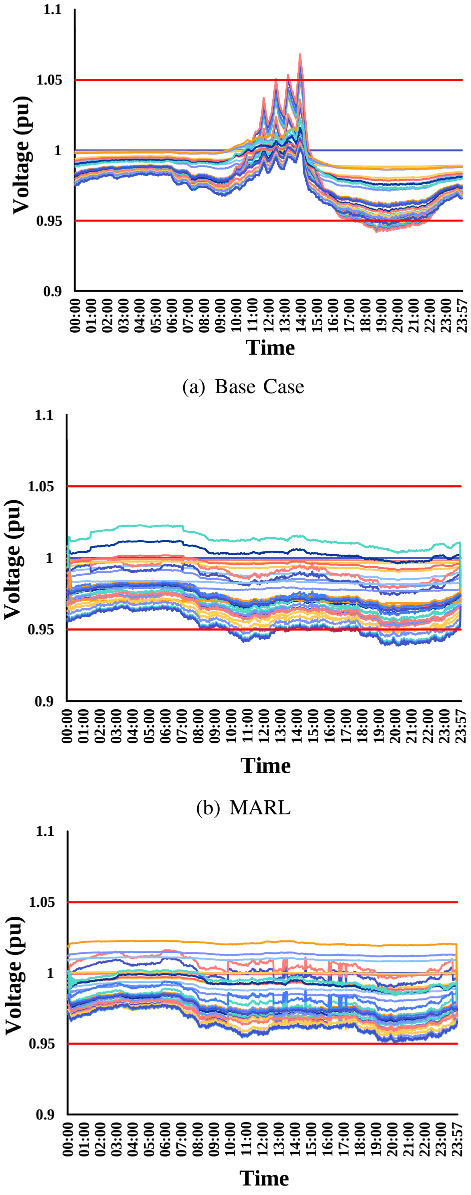
Fig. 1. Voltage profile of all buses in a random day (day 730)

B. Results

In this section, the proposed MARL strategy is implemented in the active distribution network to evaluate its performance in optimizing voltage profiles and minimizing power losses. Figure 1 illustrates the voltage profiles under two scenarios: (a) the base case without control actions and (b) the application of MARL-based VVO on a random day (day 730). In the base case (Figure 1a), significant voltage deviations occur between 18:00 and 21:00, with voltage levels falling outside the acceptable range of 0.95 pu to 1.05 pu. Furthermore, Table III shows a power loss of 6.68%, highlighting inefficiencies in network operation without reactive power control. In contrast, the MARL algorithm (Figure 1b) completely mitigates voltage deviations, maintaining stable voltages throughout the day. Moreover, power loss is reduced to 3.85%, demonstrating the effectiveness of the algorithm in both voltage regulation and loss minimization. Since the MARL strategy successfully stabilized the network without voltage violations, the DNN safety layer was not activated in this scenario.

TABLE III
PERFORMANCE OF THE PROPOSED APPROACH ON DAY 730

Day 730	**Base Case**	**MARL**
Total VD	293	0
Min Voltage (pu)	0.944	0.966
Max Voltage (pu)	1	1.028
Power Loss	6.68%	3.85%



(a) Base Case

(b) MARL

(c) MARL with DNN safety layer

Fig. 2. Voltage profile of all buses in a random day (day 736)

TABLE IV
PERFORMANCE OF THE PROPOSED APPROACH ON DAY 736

Day 736	**Base case**	**MARL**	**MARL-DNN**
Total VD	416	295	0
Min voltage (pu)	0.939	0.942	0.952
Max voltage (pu)	1.068	1.022	1.022
Power loss	7.13%	4.92%	3.68%

To further evaluate the effectiveness of the proposed approach, the implementation is carried out on another day (day 736), where the system exhibits greater instability compared to the previous case. Figure 2 presents the results for this scenario. Figure 2a shows the base case without any control actions. Compared to day 730, the system experiences more severe voltage deviations, with voltages frequently falling outside the acceptable range, indicating increased instability in the network. In Figure 2b, the MARL-based VVO is deployed. Although MARL was effective in stabilizing the network on

TABLE V
PERFORMANCE OF THE PROPOSED APPROACH ON A MONTH

Days 730 to 760	Base case	MARL	MARL-DNN
VD percentage	3.89%	1.45%	0%
Min voltage (pu)	0.936	0.938	0.952
Max voltage (pu)	1.068	1.026	1.026
Mean Power loss	5.98%	3.95%	3.38%

day 730, in this more challenging scenario, MARL alone fails to maintain voltages within the normal range, resulting in noticeable voltage deviations across several time steps and buses.

However, Figure 2c demonstrates the effectiveness of MARL combined with the DNN safety layer. The safety layer successfully corrects unsafe actions taken by the RL agents, ensuring that voltages across all buses and time steps remain within the safe operational range of 0.95 pu to 1.05 pu. In addition to voltage stability, Table IV highlights the impact of the safety layer on the reduction of power loss. The MARL with the safety layer achieves the greatest reduction in power losses, decreasing them from 7.13% in the base case to 3.68%, outperforming both the standalone MARL approach and the uncontrolled scenario. This demonstrates the dual benefit of the proposed safe MARL strategy in maintaining voltage stability and improving network efficiency.

To further evaluate the robustness of the proposed control strategies over an extended period, simulations were conducted for a month, covering days 730 to 760. Table V summarizes the results. The base case exhibited voltage violations in 3.89% of the total time steps across all buses. The application of MARL without the safety layer reduced the violations to 1.45%, while the integration of the DNN safety layer completely eliminated the voltage deviations, achieving no violations. In terms of power loss, the base case experienced an average loss of 5.98%, which was reduced to 3.95% with MARL and further decreased to 3.38% when the DNN safety layer was applied. These results highlight the effectiveness of the proposed MARL framework with the DNN safety layer in enhancing voltage stability and minimizing power losses over longer timescales.

V. CONCLUSION

This paper presented a decentralized Volt/VAR Optimization (VVO) strategy for Active Distribution Networks (ADNs) using Multi-Agent Reinforcement Learning (MARL) combined with a Deep Neural Network (DNN) safety layer. The proposed approach enables PV inverters to individually manage reactive power, aiming to minimize power losses while ensuring voltage stability and line loading within operational limits. The MARL framework was enhanced by a pretrained Projection DNN, which corrected unsafe actions in real-time, ensuring system reliability under highly volatile conditions.

The simulation results demonstrated that while MARL alone effectively reduced power losses and mitigated voltage deviations in moderately unstable conditions, it struggled to guarantee voltage safety in more challenging scenarios. The integration of the DNN safety layer addressed this gap, completely eliminating voltage violations and achieving the most significant reduction in power losses. The approach was validated over various timescales, including a month-long assessment, confirming its robustness and adaptability.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon Europe Framework Programme (HORIZON) under the GA n. 101120278 - DENSE.

REFERENCES

[1] Hidalgo, R., Abbey, C. and Joós, G., 2010, July. A review of active distribution networks enabling technologies. In IEEE PES General Meeting (pp. 1-9). IEEE.

[2] Abdelkader, S.M., Kinga, S., Ebinyu, E., Amissah, J., Mugerwa, G., Taha, I.B. and Mansour, D.E.A., 2024. Advancements in data-driven voltage control in active distribution networks: A Comprehensive review. Results in Engineering, p.102741.

[3] Alshehri, M. and Yang, J., 2024. Voltage Optimization in Active Distribution Networks—Utilizing Analytical and Computational Approaches in High Renewable Energy Penetration Environments. Energies, 17(5), p.1216.

[4] Allahmoradi, S., Afrasiabi, S., Liang, X., Zhao, J. and Shahidehpour, M., 2024. Data-Driven Volt/VAR Optimization for Modern Distribution Networks: A Review. IEEE Access, 12, pp.71184-71204.

[5] Radhoush, S., Whitaker, B.M. and Nehrir, H., 2023. An Overview of Supervised Machine Learning Approaches for Applications in Active Distribution Networks. Energies, 16(16), p.5972.

[6] Hashemnezhad, M., Delkhosh, H., Shahabi, A. and Moghaddam, M.P., 2024, May. Community Energy Management Using MARL: Synergy of Price-Based and Incentive-Based Demand Response. In 2024 32nd International Conference on Electrical Engineering (ICEE) (pp. 1-6). IEEE.

[7] Hashemnezhad, M., Delkhosh, H. and Moghaddam, M.P., 2025. Aggregator pricing strategy for community energy management based on multi-agent reinforcement learning considering customer loss or gain. Sustainable Energy, Grids and Networks, 41, p.101607.

[8] Ghazvini, M.A.F., Lipari, G., Pau, M., Ponci, F., Monti, A., Soares, J., Castro, R. and Vale, Z., 2019. Congestion management in active distribution networks through demand response implementation. Sustainable Energy, Grids and Networks, 17, p.100185.

[9] Huo, Y., Li, P., Ji, H., Yu, H., Yan, J., Wu, J. and Wang, C., 2022. Data-driven coordinated voltage control method of distribution networks with high DG penetration. IEEE Transactions on Power Systems, 38(2), pp.1543-1557.

[10] Wang, S., Duan, J., Shi, D., Xu, C., Li, H., Diao, R. and Wang, Z., 2020. A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning. IEEE Transactions on Power Systems, 35(6), pp.4644-4654.

[11] Yang, Q., Wang, G., Sadeghi, A., Giannakis, G.B. and Sun, J., 2019. Two-timescale voltage control in distribution grids using deep reinforcement learning. IEEE Transactions on Smart Grid, 11(3), pp.2313-2323.

[12] Karagiannopoulos, S., Aristidou, P., Hug, G. and Botterud, A., 2024. Decentralized control in active distribution grids via supervised and reinforcement learning. Energy and AI, 16, p.100342.

[13] Wang, W., Yu, N., Gao, Y. and Shi, J., 2019. Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems. IEEE Transactions on Smart Grid, 11(4), pp.3008-3018.

[14] Zhang, M., Guo, G., Zhao, T. and Xu, Q., 2023. Dnn assisted projection based deep reinforcement learning for safe control of distribution grids. IEEE Transactions on Power Systems.

[15] Yang, X., Liu, H., Wu, W., Wang, Q., Yu, P., Xing, J. and Wang, Y., 2024. Reinforcement Learning with Enhanced Safety for Optimal Dispatch of Distributed Energy Resources in Active Distribution Networks. Journal of Modern Power Systems and Clean Energy.

[16] Li, Q., Lin, T., Yu, Q., Du, H., Li, J. and Fu, X., 2023. Review of deep reinforcement learning and its application in modern renewable power system control. Energies, 16(10), p.4143.

[17] https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014.

[18] https://www.elia.be/en/grid-data/power-generation/solar-pv-power-generation-data.