

A Q-Learning Based Approach For Voltage Control In Unbalanced Distribution System

Muhammed Turhan Çakır¹, Abdullah Atak¹, Hasan Özeren², and Oğuzhan Ceylan¹

¹Department of Electrical Engineering, Marmara University, Turkey
Email: muhammed.turhan@marun.edu.tr, abduallahatak@marun.edu.tr, oguzhan.ceylan@marmara.edu.tr

²Department of Computer Engineering, Marmara University, Turkey
Email: hasanozeren@marun.edu.tr

Abstract—In recent years, with the increasing use of renewable energy sources, voltage regulation in power distribution systems has become a greater challenge due to these sources' intermittent and variable nature. To address this problem, this paper proposes a Q-learning-based method for voltage regulation in power distribution systems that incorporates tap changer voltage regulators when renewable energy sources are present. The voltage deviation problem is formulated as an optimization problem, and Q-learning is used to find near-optimal solutions that can effectively control voltage deviations and improve the overall performance of the power distribution system. The proposed method is applied to power distribution system using OpenDSS incorporated into Python. The simulations are performed on base case IEEE-123 Bus distribution network and with modifications by adding Photovoltaics. From the simulation results we observed that the proposed algorithm is able to solve the overvoltage and undervoltage problems.

Index Terms—tap changer voltage regulator, unbalanced distribution system, Q-learning

I. INTRODUCTION

There has been an increase in the number of facilities for decentralized power generation from renewable energy sources (DG). They bring environmental and economic benefits. However, this increase in variable resources such as photovoltaics (PV) can cause problems in power systems especially in deviations of voltage magnitudes.

Traditional methods such as fixed/switched capacitors and voltage regulators are used to control voltage levels. The authors of [1] determine the optimal tap position of voltage regulators by applying an algorithm based on the alternating direction method of multipliers (ADMM). In [2] a game theory based algorithm was used in voltage regulators for improving voltage regulation. In [3] a coordinated voltage control approach based on whale optimization algorithm utilizing voltage regulators and PV units was proposed. In [4], the use of capacitors, and voltage regulators was simulated by aiming to minimize the voltage deviations and maximizing power output from DGs using harmony search algorithm. Due to their limitations, these control methods cannot respond quickly and optimally to the highly fluctuating nature of PV. According to IEEE 1547 standard [5], smart inverter technology (SI) can be used to control voltage with smart voltage regulation functions. Two of these smart functions include injecting and absorbing reactive power (Volt-Var) and limiting the

active power generated by PV systems (Volt-Watt). SIs can respond more quickly to changes in grid conditions than conventional voltage regulation devices. However, the widely used Volt-Var function in [5] is to improve the local bus voltage by injecting/absorbing reactive power. Without coordination between SIs and a high penetration of PVs in the system, performance problems may occur. In [6] coordinated use of all the mentioned devices together with the reactive power support from DGs was simulated.

In recent years, artificial intelligence has made revolutionary progress. The power of AI is being applied in areas such as robotics [7], self-driven cars, etc. An OpenAI AI team defeated professional players in the very complex video game Dota 2 [8]. [9] shows that machine learning is also being used in medicine.

Over the years, this popularity of AI has attracted the attention of power and energy society. Paper [10] summarizes reinforcement learning (RL) methods used in power grid operations. In [11] the deep deterministic policy gradient (DDPG) algorithm is used to successfully solve voltage problems of a modified IEEE 37-node test network. In [12], a linearization method is used to balance voltage levels with SI and voltage regulators on the highly unbalanced IEEE 37-bus test feeder. In [13] a variant of Q-Learning called Double Dueling Q-Learning was used to coordinate the OLTC-fitted transformer at primary substations and the power factor of PV inverters. The authors of [14] developed deep reinforcement learning for solving voltage regulation problem in unbalanced distribution systems by considering capacitors, voltage regulators, and smart inverters. In [15] the authors have developed an procedure for voltage regulation by using nonlinear programming approach and a local learning algorithm by using voltage regulators and reactive power support from DGs.

In this paper we aim to apply Q-learning method to solve voltage deviation problem in unbalanced power distribution system. The paper will utilize the Q-learning method by incorporating python and OpenDSS which is an open source power systems simulation software. The main contributions of the paper are summarized as given below.

- We propose a Q-Learning based optimization approach that incorporates tap changer voltage regula-

tors.

- We apply the proposed methodology on a base case IEEE 123 Bus Test System, and provide the simulation results.
- We also apply the proposed method to a modified IEEE 123Bus Test System with PVs and provide the simulation results.

The rest of the paper is organized as follows. In Section II we briefly introduce Q-Learning, and provide a general algorithm. Section III is devoted to the implementation of Q-Learning algorithm in an unbalanced test system. Proposed algorithm for solving over-voltage and under-voltage problems is provided. In Section IV we provide the simulation test results by using the modified and base case IEEE 123 Bus Test System. Finally, in Section V we conclude the paper.

II. Q-LEARNING

A. A Brief Introduction

Q-Learning [16] introduced by Watkins and Dayan in 1992 is a model-free reinforcement learning algorithm. Reinforcement learning is an area of machine learning focused on how an artificial agent would act on certain states of an environment. In Q-Learning, an agent acts in the environment by experiencing consequences and returning them as punishment and/or reward. In a finite discrete space, these values can be stored for that state, the action value as a reward. The goal is to first explore the environment by going from state to state, performing random actions, and storing the reward values in a Q-table. After sufficient exploration, the agent will begin to exploit its previously returned reward values and choose the most valuable action for the state it's in.

B. Q-Table

Q-Table is a technique used to form a table with rows equal to the number of states and columns equal to the number of actions. Values of states are stored in Q-Table. Note that for a discrete state and action space, Q-Table is finite, thus Q-Table can be constructed beforehand to store the Q-values.

C. The Bellman Equation

Bellman Equation given in (1) is used to determine each state's value and how valuable it is to stay or change the state [16].

$$V(s) = \max_a (R(s, a) + \gamma V(s)) \quad (1)$$

where, $V(s)$ is the expected return at the current state s and $R(s, a)$ is the expected reward for taking action a in state s . Note that, γ represents the discount factor which is multiplied with the value of the next state, $V(s)$.

D. General Algorithm For Q-Learning

At step n , the agent is in state s_n and can choose an action a_n . For this action at that state, the agent receives a reward r_n . R_n is mean value of this r_n and only dependent on action a_n and state s_n . The state changes

by probabilistic to a state y_n according to equation given below:

$$\text{Prob}[y_n = y | s_n, a_n] = P_{s_n, y}[a_n] \quad (2)$$

For a policy π , Q values are defined as following [16]:

$$Q^n(s, a) = R_s(a) + \gamma \sum_y P_{sy}[\pi(x)] V^n(y) \quad (3)$$

In a Q-Learning algorithm [16], the agent in the n^{th} episode performs following actions:

- Observe the current state x_n ,
- Take an action a_n ,
- Observe the following state y_n
- Get the corresponding reward
- Update the Q_{n-1} with learning a_n , using equation: $Q_n(x, a) = (1 - \alpha_n)Q_{n-1}(s, a) + \alpha_n[r_n + \gamma V_{n-1}(y_n)]$ where, $V_{n-1}(y) = \max_b \{Q_{n-1}(y, b)\}$, is the best value the agent can get from state y . Initial Q values for every state, action (s, a) pair is assumed to be know when starting the process.

In the power distribution voltage regulation problem, the policy will be towards minimizing voltage deviation from 1 pu. The reward function is given below in Eq. (4).

$$R = - \sum_{i=1}^N (V_i - 1)^2 \quad (4)$$

where $\mathbb{N} = \{1, 2, \dots, n\}$ and n is the number of the nodes. Note that, the voltage level interconnected between the electrical systems of the supplier and consumer is called service voltage. According to American National Standards Institute (ANSI) standard [17], the service voltage should be within a $\pm 5\%$ range of the nominal voltage values. In other words, this voltage level corresponds to 0.95 to 1.05 per unit (p.u.) voltage. Lower or higher voltage magnitudes are considered as violations.

Aim of the used reward function is to reward best tap settings of the voltage regulators that makes the system's voltage p.u. magnitudes to stay close to 1. Taking the square helps us to negate sign difference of summations for different simulations. Reason behind taking the minus of the sum is to use The Bellman Equation (1) properly.

III. IMPLEMENTING Q-LEARNING IN IEEE 123 NODE TEST FEEDER

In the simulations, we will first implement a Q-Learning function to IEEE 123 Node Test Feeder [18] to determine the near optimal tap settings of the voltage regulators for minimizing the voltage deviations. Then in the next phase, we will integrate the PVs by using the functionalities of OpenDSS, an open-source software [19].

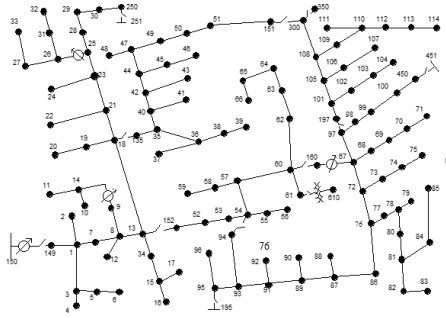


Fig. 1. Single phase diagram of the IEEE 123 Node Test Feeder [18].

The IEEE 123 Node Test Feeder single line diagram of which is given in Fig. 1 [18] has seven voltage regulators. The first regulator is located between nodes 150 and 149. The others are located as follows; between nodes 9 and 14, one regulator; between nodes 25 and 26, two regulators; between nodes 160 and 67, three regulators. The tap positions of the regulators may vary between -16 to 16. By changing these values, voltage levels in system can be adjusted. During the implementation, a power flow run will be performed on the system at every tap setting change using OpenDSS and Python library py-dss-interface. Using (1) and (2), Q-Values will be placed in proper locations on the Q-Table.

A. Q-Learning Algorithm Implementation

Implementation of the Q-Learning Algorithm (See Algorithm 2 below) will be as following:

- 1) Perform power flow using python interfaced OpenDSS. Get reward values of three phases for selected regulator's current tap position.
- 2) Increase the tap setting of the voltage regulator by 1. Perform power flow and get reward values.
- 3) Perform step 2 again. This time instead of increasing, decrease the tap setting by 1. Perform power flow and get reward values.
- 4) Check which reward values are greater. Move to the tap setting with the greatest reward values.
- 5) Perform the same steps with other voltage regulators available in the distribution feeder.

After performing the steps, additional punishment for the cases with over-undervoltages is required. So we implemented an algorithm to assist our reward function (2). Idea of the algorithm is to draw very strict lines at points according to ANSI standard [17] with some tweaks. These strict lines will be at $y = 0.95$, $y = 0.90$, $y = 1.05$ and $y = 1.10$ p.u values.

As an example, if voltage magnitude at a node is 1.07 p.u., impact of this point to the reward function is -1. If that point exceeds 1.1, that impact will become -2. With the help of this approach (given in Algorithm 1) implemented to the reward function, actions taken will be highly dependent on these strict lines.

Algorithm 1 Assisting Reward Function

```

1: for Node = 1, 2, ... N do
2:   if NodeV oltage ≥ 1.05 p.u. then
3:     if NodeV oltage ≥ 1.10p.u. then
4:       Reward := Reward - 2
5:     else
6:       Reward := Reward - 1
7:     end if
8:   else if NodeV oltage ≤ 0.95 p.u. then
9:     if NodeV oltage ≤ 0.90 p.u. then
10:      Reward := Reward - 2
11:    else
12:      Reward := Reward - 1
13:    end if
14:   end if
15: end for
16: return Reward

```

Algorithm 2 Q Learning Algorithm

```

for iteration = 1, 2, ... do
2:   for every regulator do
3:     Tap setting := random integer from -16 to 16
4:     Perform power flow for (using tap setting)
5:     Get the reward function return
6:     Middle := Reward
7:     Perform power flow for (Tap setting + 1)
8:     Get the reward function return
9:     Upper := Reward
10:    Perform power flow for (Tap setting - 1)
11:    Get the reward function return
12:    Lower := Reward
13:    if Middle < Upper then
14:      Tap number := Tap setting + 1
15:    else if Middle < Lower then
16:      Tap number := Tap setting - 1
17:    end if
18:   end for
end for

```

IV. RESULTS AND ANALYSIS

This section provides the simulation results by using the base case of IEEE 123 Bus Test System and the modified IEEE 123 Bus Test System with adding five PVs.

A. IEEE 123 Node Test Feeder (Base Case) Simulations

The base case data of the IEEE 123 Node Test Feeder can be found in [18]. We used the data of the the base case for loads and set the tap positions of the voltage regulators randomly between -16 and 16 and performed simulations. In figures 2, 3, and 4 the orange line represent the voltage magnitudes for initial tap settings (-6, 5, -10, -8, 4, -15, 3) as given in the first line of the Table I, for Phase A, B and C, respectively. The blue lines represent the voltage magnitudes after the implemented Q-Learning based control algorithm for three phases. One can see the improvement in the voltage magnitudes after the implementation of the control algorithm, as the voltage

magnitudes out of the allowed voltage ranges are brought between 0.95 and 1.05 pu.

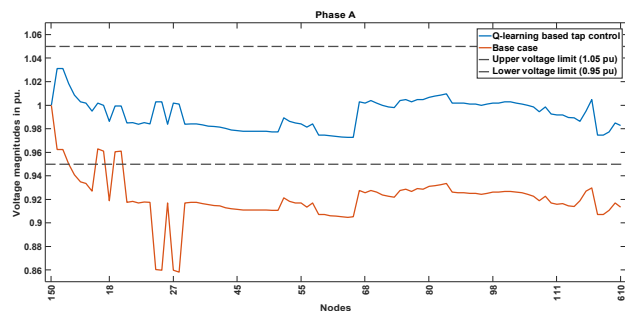


Fig. 2. Voltage magnitudes of Phase A nodes for base case and after control using Q-Learning.

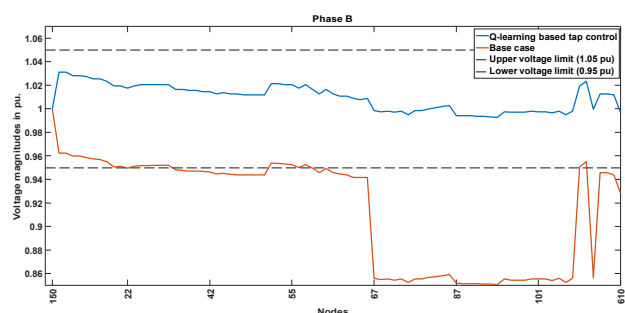


Fig. 3. Voltage magnitudes of Phase B nodes for base case and after control using Q-Learning.

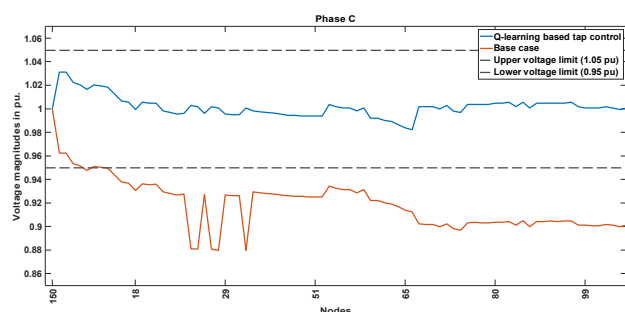


Fig. 4. Voltage magnitudes of Phase C nodes for base case and after control using Q-Learning.

Table I is representing a random case with the specified initial voltage regulator tap positions (the first row of the table). The progress over the iterations are shown, as one can see, the reward function starts with an higher absolute value, as the the iterations emerge, this value gets closer to zero, which represents the near optimal tap setting value.

B. IEEE 123 Node Test Feeder with PV Integration

In the second set of simulations we modified the IEEE 123 Node Test Feeder and added 5 PVs on nodes 28, 47, 67, 95 and 105, respectively. Rated real power output for each PV is selected to be 400kW. We took the initial tap positions of voltage regulators from [18] as 7, -1, 0, -1, 8, 1, 5, respectively for reg1a, reg2a, reg3a, reg3c, reg4a,

TABLE I
TAP VALUES OF REGULATORS AND CORRESPONDING REWARD VALUE

reg1a	reg2a	reg3a	reg3c	reg4a	reg4b	reg4c	R
-6	5	-10	-8	4	-15	-3	-287.1
-5	6	-9	-7	5	-16	-2	-253.8
-4	7	-8	-6	6	-15	-1	-232.5
-3	8	-7	-5	7	-14	0	-195.2
-2	7	-6	-4	8	-13	1	-162.9
-1	6	-5	-3	9	-12	2	-101.7
0	5	-4	-2	10	-11	3	-59.5
1	4	-3	-1	9	-10	4	-40.3
2	3	-2	0	8	-9	5	-31.2
3	2	-1	1	7	-8	4	-30.2
4	1	0	2	6	-7	3	-0.1
5	0	1	1	5	-6	2	-0.07
6	-1	2	0	4	-5	1	-0.06
7	-2	1	-1	3	-4	0	-0.07
6	-1	2	0	4	-3	1	-0.05
5	0	3	1	5	-2	2	-0.04

reg4b, reg4c. In the simulations we assumed that all the PVs are giving their maximum active power to the system.

The simulation results are illustrated on Figures 5, 6, and 7. Similar to the previous case, the orange lines represent the voltage magnitudes without control and with Q-learning based control, respectively. In Fig. 5, initially there over-voltage issue which is solved by implementing the Q-Learning based approach. Phase B doesn't have the over-voltage problem, however the Q-learning based algorithm is able to provide a better voltage profile (closer to 1 pu). Initially, in Phase C, the over-voltages are less compared to Phase A, again with using Q-learning based control the algorithm is able to provide a better voltage profile.

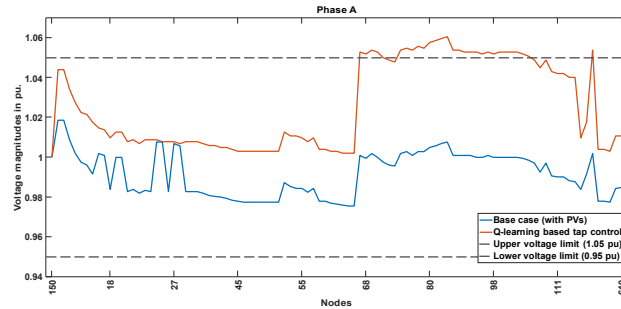


Fig. 5. Case with PVs: voltage magnitudes of Phase A nodes and after control using Q-Learning.

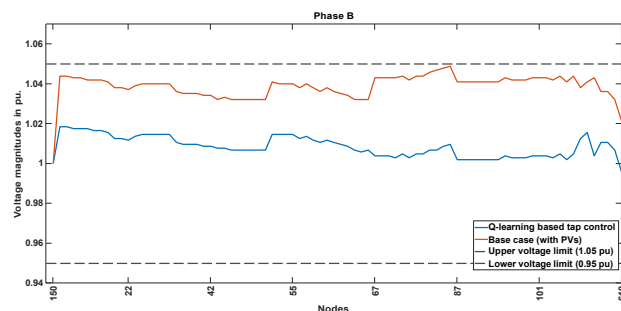


Fig. 6. Case with PVs: voltage magnitudes of Phase B nodes and after control using Q-Learning.

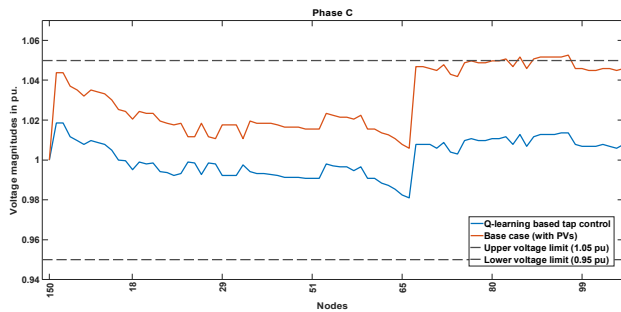


Fig. 7. Case with PVs: voltage magnitudes of Phase C nodes and after control using Q-Learning.

The changes of the tap positions of the voltage regulators and their corresponding reward function values are presented in Table II. Note that, the algorithm requires less number of iterations compared to previous simulation.

TABLE II
TAP POSITIONS OF THE VOLTAGE REGULATORS AND
CORRESPONDING REWARD VALUES

reg1a	reg2a	reg3a	reg3c	reg4a	reg4b	reg4c	\bar{R}
6	-2	1	-2	7	0	4	-0.10
4	0	3	0	5	-2	2	-0.04
3	1	4	1	4	-1	-3	-0.03

V. CONCLUSION

In this paper we propose a Q-Learning based approach for determining the near optimal tap positions of voltage regulators in unbalanced distribution networks. Two set of simulations were performed using the base case and modified IEEE 123 Node Test Feeder with PVs.

The results showed that the near optimal tap settings of the IEEE 123 Node Test System have eliminated the over-voltages and under-voltages for both base case and case with PVS. This algorithm is adaptable to coordinate the SIs with voltage regulators for better operation of the unbalanced distribution networks. In future, we are planning to incorporate the use of reactive power support from PVs by using daily simulations.

ACKNOWLEDGMENT

This research is funded as a part of “120N996 Implementing digitalization to improve energy efficiency and renewable energy deployment in Turkish distribution networks” project under the framework of 2551 Project organized by TUBITAK and British Council.

REFERENCES

- [1] B. A. Robbins, H. Zhu, and 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.
- [2] S. Tasnim, C. R. Sarimuthu, B. L. Lan, and C. P. Tan, “A game theory approach for oltc voltage control operation in an active distribution network,” *Electric Power Systems Research*, vol. 214, p. 108861, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378779622009142>
- [3] A. Nakadomari, N. Krishnan, A. M. Hemeida, H. Takahashi, E. Omine, and T. Senjyu, “Coordinated voltage control of three-phase step voltage regulators and smart inverters to improve voltage profile and energy efficiency in unbalanced distribution networks,” *Energy Reports*, vol. 9, pp. 234–241, 2023, 2022 The 3rd International Conference on Power Engineering. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352484723006406>
- [4] O. Ceylan, G. Liu, Y. Xu, and K. Tomsovic, “Distribution system voltage regulation by distributed energy resources,” in *2014 North American Power Symposium (NAPS)*, 2014, pp. 1–5.
- [5] “Ieee standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces,” *IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003)*, pp. 1–138, 2018.
- [6] O. Ceylan, G. Liu, and K. Tomsovic, “Coordinated distribution network control of tap changer transformers, capacitors and pv inverters,” *Electrical Engineering*, vol. 100, no. 2, pp. 1133–1146, 2018.
- [7] P. W. Huang and K.-J. Chung, “The prediction of positioning shift for a robot arm using machine learning techniques,” in *2019 14th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT)*, 2019, pp. 58–61.
- [8] OpenAI, :, C. Berner, G. Brockman, B. Chan, V. Cheung, P. Debiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, R. Józefowicz, S. Gray, C. Olsson, J. Pachocki, M. Petrov, H. P. d. O. Pinto, J. Raiman, T. Salimans, J. Schlatter, J. Schneider, S. Sidor, I. Sutskever, J. Tang, F. Wolski, and S. Zhang, “Dota 2 with large scale deep reinforcement learning,” 2019. [Online]. Available: <https://arxiv.org/abs/1912.06680>
- [9] T. Lin, L. Cervino, X. Tang, N. Vasconcelos, and S. B. Jiang, “Tumor targeting for lung cancer radiotherapy using machine learning techniques,” in *2008 Seventh International Conference on Machine Learning and Applications*, 2008, pp. 533–538.
- [10] D. Zhang, X. Han, and C. Deng, “Review on the research and practice of deep learning and reinforcement learning in smart grids,” *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362–370, 2018.
- [11] C. Li, C. Jin, and R. Sharma, “Coordination of pv smart inverters using deep reinforcement learning for grid voltage regulation,” in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 1930–1937.
- [12] C. Li, V. R. Disfani, H. V. Haghi, and J. Kleissl, “Optimal voltage regulation of unbalanced distribution networks with coordination of oltc and pv generation,” in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, 2019, pp. 1–5.
- [13] G. Custodio, L. F. Ochoa, T. Alpcan, F. C. L. Trindade, and R. A. de Godoy Rosolen, “Using dueling double q-learning for voltage regulation in pv-rich distribution networks,” in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–6.
- [14] Y. Zhang, X. Wang, J. Wang, and 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.
- [15] D. Villacci, G. Bontempi, and A. Vaccaro, “An adaptive local learning-based methodology for voltage regulation in distribution networks with dispersed generation,” *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1131–1140, 2006.
- [16] C. J. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3-4, p. 279–292, 1992.
- [17] National Electrical Manufacturers Association, “American national standard for electric power systems and equipment—voltage ratings (60 hertz),” Washington, DC, USA, 2016.
- [18] K. P. Schneider, B. A. Mather, B. C. Pal, C.-W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. Pereira, L. F. Ochoa, L. R. de Araujo, and et al., “Analytic considerations and design basis for the ieee distribution test feeders,” *IEEE Transactions on Power Systems*, vol. 33, no. 3, p. 3181–3188, 2018.
- [19] D. Montenegro and R. C. Dugan, “Opendss and opendss-pm open source libraries for ni labview,” in *2017 IEEE Workshop on Power Electronics and Power Quality Applications (PEPQA)*, 2017, pp. 1–5.