Hierarchical Multi-Agent Deep Reinforcement Learning for Coordinated Optimization of Aggregated Virtual Power Plants in Smart Microgrids

Junxiong Zhang[1*], Jiao Wang[2], Qihang Wang[1], Qi Zhou[1]
[1]Anhui Technical College of Industry and Economy, China
[2]Anhui Jianzhu University, China
E-mail: JunxiongZhang1567@outlook.com
[*]Corresponding author

Keywords: Coordinated optimization control, aggregated virtual power plants, smart microgrids, multi-agent reinforcement learning

Received: March 24, 2026

Computational efficiency, Data privacy, and equitable benefit assessment are some of the issues that have arisen as a result of the fast expansion of distributed energy resources (DERs), which have added complexity to the functioning of distribution networks. This paper presents a two-tiered VPP coordination architecture that takes into account the operational interests of both Distribution System Operators (DSOs) or VPPs under AC optimum power flow (AC-OPF) limitations. The goal is to solve these challenges. A penalty-function-enhanced OPF mechanism is used in the upper layer to guarantee network security in the event of voltage or branch-limit violations, and an Asynchronous Advantage Actor-Critic (A3C) multi-agent architecture is integrated into the lower layer to utilize a parameter-sharing Twin-Delayed Deep Deterministic Policy Gradient (PS-TD3) algorithm. Through lightweight parameter sharing and decentralized execution, every agent— which represents a VPP subsystem—learns optimum judgments for energy-dispatch, storage, and flexibility, resulting in dramatically reduced computational cost and preservation of data privacy. When compared to the non-cooperative TD3 baseline, traditional distributed OPF, and independent Q-learning, the suggested dual-layer MARL approach outperforms all three in simulation tests conducted on the IEEE 33-node distribution network. Thanks to parameter sharing, the PS-TD3 + A3C hybrid improves convergence speed through 42% and reduces per-step computing time by 37%. It also reduces voltage variation by 31.4%, network real-power losses by 26.7%, and operating cost by 18.2%. Since agents only share compressed gradients and not raw operational data, privacy leakage is minimized by more than 80%. In contemporary distribution systems that are rich in distributed energy resources (DERs), the findings show that the suggested framework provides a computationally efficient, scalable, and privacy-preserving approach for coordinated VPP operation.

Povzetek: Prispevek predlaga dvoslojno arhitekturo za usklajevanje virtualnih elektrarn, ki z uporabo večagentnega učenja izboljša učinkovitost, zmanjša stroške in izgube ter hkrati varuje zasebnost podatkov v omrežjih z razpršenimi viri energije.

1 Introduction

The power grid is an interconnected network of millions of electrical devices that provide a variety of services, and a lot of different types of data will be created and sent as a result of the ongoing innovation and improvement of existing power architecture [1]. The processing and analysis of data have diverse demands, which puts a heavy strain on the electricity grid. Scalability, utilization efficiency, and deployment cost are three areas where conventional fixed resource allocation falls short when confronted with fast response and real-time engagement [2]. More and more distributed energy resources (DERs) are finding their way into distribution systems in an effort to reduce pollution and boost sustainable development [3]. Although DERs have many benefits, they also pose several problems. High connection costs and complicated operation and transaction administration are the outcomes of distributed generators' (DGs) modest capacity, huge number, and uneven distribution, which in turn causes these problems [4].

- Virtual Power Plants (VPPs): The term, its varieties (commercial and technical), and its advantages.
- Smart Microgrids: Distributed power systems that can be controlled and interacted wirelessly.
- Control Strategies:

- Of hierarchical, decentralized, and centralized systems.
- Game theory, MPC, and multi-agent systems in general

Nevertheless, a potential answer to these problems has emerged with the fast development of VPPs. Electricity market operations and distribution grid management services may be provided by DERs like load, distributed generation (such as renewable and dispatchable sources), or energy storage systems when they are aggregated via VPPs [5]. Through the consolidation of these resources, VPPs provide a more efficient and simplified method to managing DERs that are widely spread, thus alleviating the problems associated with this task [6]. A Coordinated Optimization Control Method for Aggregated Virtual Power Plants (VPPs) in Smart Microgrids is a state-of-the-art approach to assessing DERs as a unified whole, which improves their economic performance, efficiency, and dependability [7]. Proposed research ideas are given in Table 1.

Table 1: Research techniques

Aim and scope	More and more DERs are being integrated into smart microgrids.
Objective	Create a system for controlling smart microgrids that makes the most efficient use of VPPs.
Problem	• Inefficiency and instability result from a lack of coordinated control. • Distributed, adaptive, and autonomous control techniques are necessary.
Solution	• Combined DERs into VPPs while optimize them in a coordinated manner • A smart microgrid is controlled by using Multi-Agent Deep Reinforcement Learning. • By working together, the agents maximize global goals while they learn policies to govern their own DER.
Research Goals	• Keep operational expenses to a minimum. • Maximize the use of renewable energy sources. • Guarantee the dependability and stability of the system.
Coordination System	• VPPs across the microgrid are centralized and coordinated. • Optimization by local agents is influenced by signals from the global level.

Distribution grid planning and scheduling have encountered new obstacles due to the incorporation of DERs. Power system operators, power producers, and electricity trading platforms are just a few of the many market actors that have contributed to the complexity of managing power systems thanks to the presence of varied energy sources. Collaborative optimization of multi-stakeholder power systems has recently drawn attention from game theory [8]. Microgrids (MGs) are integrated into distribution systems via the use of cooperative game theory, as described in reference. The strategic Stackelberg game has been shown to be useful in solving the collaborative optimization challenges of MG-owned VPPs and distribution grids. In reference, the authors provide and prove the efficacy of a two-level Stackelberg game–based transaction model for MGs [9].

While there is some work on optimizing distribution network operations, much of it ignores real operational limits and the financial sustainability of participants like VPP, both of which may significantly affect the network's stability [10]. While optimizing and scheduling DERs, reference pays little attention to the participants' specific interests but instead concentrates on the market as a whole. Although network restrictions are not taken into account in reference, a technique for bilateral trade optimization is presented for the distribution network [11]. Also, the real limitations in MGs are not taken into consideration in reference, which investigates the use of the Stackelberg game in the energy trading system of several MGs.

Power flow calculation is a key area of smart grid study because it uses the provided structure and operational values to find the power system's steady-state characteristics, which in turn allow us to assess the safety-impact of different power demand and supply scenarios [12]. While prior solutions relied on human resources and expert knowledge, this challenge would face a non-convergence scenario under diverse situations. In addition, microgrid management is complex and fraught with uncertainty due to intelligent power grids' ability to adapt their settings on the fly to changing environmental conditions and the wide diversity of possible configurations for individual power units [13].

We provide a system for distributed power management in microgrids using multi-agent deep reinforcement training and VPPs, and we examine the power adjustment issue in this research. Before presenting the whole

structure with its three facets, we examine the usual service needs of microgrid power calculation. Next, we develop a learning-based approach for microgrid adjustment by modeling the power flow adjustment issue using Markov processes. Lastly, the practical findings show that the suggested framework can successfully get solutions using the Pandapower tool to simulate the IEEE 39 bus system [14].

Here is a brief overview of the key points covered in this paper: 1) Our all-inclusive smart grid control and management architecture lets smart grid data sensing, processing, and control achieve local autonomy and real-time reaction. 2) Taking into account the system's current condition and requirements, a distributed method for power flow modification is introduced, which is based on learning. The findings of the simulation show that our framework is able to provide satisfactory adjustment outcomes under different power scenarios [15].

1.1 Motivation

When dealing with nonlinearities with uncertainty in power-system controllers, two well-established approaches are adaptive fuzzy control with robust neural-adaptive control. Under certain conditions, they can deal with dynamics that change slowly over time (such as parameter drift or modeling mistakes) and provide robust stability guarantees. The proposed MARL framework is similar to previous work in that it aims to solve the same problem, but it differs in three key ways: first, it uses learning-from-interaction (experience) instead of algorithm inversion or online parameter estimation; second, it explicitly shapes rewards based on multiple objectives (efficiency, privacy, fairness); and third, it uses a distributed agent structure that is optimized for Virtual Power Plants (VPPs) and thus naturally maps to privacy constraints. Below, we provide a concise overview of both the theoretical and the practical distinctions, as well as a practical strategy for measuring the enhancements to computing efficiency, resilience, and convergence speed.

The "Related works" section summarizes some related works after the research background is introduced. In "The framework of power flow adjustment using edge intelligence" and "Automatic adjustment of power flow convergence founded on DRL," we propose our framework with a learning-based decision algorithm. Next, in the "Numerical results" section, we provide the setup and assessment outcomes of the simulation experiments. The study is concluded with a discussion of future research in the "Conclusion" section.

2 Related work

In [16], the researchers tackle that problem. When compared to VPPs, microgrids have the advantage of islanding, which necessitates the development of dedicated control mechanisms. These methods of control are grouped together as main control techniques. Improvements to the voltage profile, economic optimization, and resource allocation are some of the other areas that microgrid secondary control addresses. As a result of its similarities to the microgrid secondary control method, VPP coordination, while considering the control-aspects of DER, works at a slower time frame than main control while making full use of the available communication supplied by the overlying smart grid. Consequently, that study delves into the practicality of using microgrid secondary control in VPPs. At the bottom of a hierarchical control system are smart microgrids that handle local challenges via both primary and secondary control. Second, a VPP aggregates these microgrids to allow tertiary control, which in turn forms a connection to the power markets and handles problems on a broader scale.

In order to achieve demand-side ancillary service while taking intra-energy sharing among the interconnected microgrids inside the virtual power plant, the experimentalists in [17] suggested a two-stage, two-layer optimization model. Specifically, the market operator rewards the following day's regulatory capacity and the hourly power consumption baseline based on the predictions made in the first step, day-ahead scheduling. In the second phase, the RegD signal is followed to manage the power usage in real-time. There are two levels to the second stage. In the uppermost layer, DR signals are distributed from the main electrical infrastructure according to the price per unit of power in each microgrid. To further aid in the reduction of RegD breaches, it employs a novel energy sharing method for the interchange of power among MGs. The bottom layer manages the power consumption of each MG in real-time in order to save operating expenditures.

The study authors [18] used an ANN to effectively manage and arrange a variety of microgrids in virtual power plants. Artificial neural network-based binary practical swarm optimization (ANN-BPSO) and artificial neural network-based BBSA (ANN-Scheduling Control) are two techniques that are introduced here. Smart and inexpensive grid decarbonization and VPP operation are both made possible by these algorithms, which determine the optimal schedule for every distribution generation in terms of fuel consumption, CO_2 emissions, and system efficiency. Through the execution of many test scenarios, we evaluate the controllers' robustness and functionality in dynamic system environments. A variety of load curves are used as test cases to assess the ANN's effectiveness on untrained data. Both the trained and untrained load models employ real-load variable data recorders from Northern Malaysia. Analyzing the test information allows us to study the controllers' efficiency under various power system scenarios. On top of that, we compare their results to those of other solutions in the literature to see how well they work. A lot of things go into that assessment. The findings show that the ANN-based controllers are superior in terms of effectiveness and expense savings.

With an eye on improving the efficiency of renewable energy, the researchers in [19] conduct a thorough analysis of solar energy generation powered by AI and smart grid integration. Solar power generation, forecasting, as well as grid management are all areas that might benefit from the use of sophisticated AI

approaches, which are the focus of that research. Solar irradiance prediction with PV system performance estimate are two areas where machine learning methods, such as Support Vector Regression with ANN, are tested. That piece dives into the use of AI in smart grids, exploring its function in demand-side management, optimizing energy storage, and maintaining system stability. They lay forth a thorough plan to make renewable energy sources more efficient, which includes AI systems for solar-plus-storage systems, energy management strategies with multiple objectives optimized, and virtual power plants and microgrids enabled by AI. Scalability concerns, regulatory factors, and ethical considerations are some of the topics covered in that study as well as future developments in the use of AI to renewable energy systems. Solar energy systems that are connected to smart grids have the opportunity to greatly enhance system efficiency, dependability, and sustainability by making use of big data analytics and sophisticated AI algorithms.

With an emphasis on renewable energy integration, decentralized control systems, as well as cybersecurity problems, experimenters of [20] provide an in-depth review of current breakthroughs in smart grid technology. Microgrids, energy storage systems, and virtual power plants are important developments that improve the grid's sustainability and efficiency. The study focuses on how AI and ML may improve energy distribution optimization, load forecasting, as well as demand-side management. The capacity of digital twins, cybersecurity risks, and bidirectional power transfers are also covered. To back up the continuous development of smart grids, future studies should focus on establishing strong regulatory frameworks, developing scalable energy storage technologies, and implementing improved cybersecurity measures.

The authors of the aforementioned paper provide a global perspective on renewable energy developments and identify key topics for energy sector change in their comprehensive analysis [21]. In the long term, the state's energy plan should focus on efficiently connecting the power supply system to renewable energy sources. To get there, we need to manage distributed power production systems, virtual power plants, and energy storage, all of which must be encouraged and regulated. With more and more people choose to drive electric cars, V2G technology is becoming more important, particularly in the context of contemporary microgrids or "smart grids" that include renewable energy sources.

The architects of the Smart Grid subsystems are defined by the International Electrotechnical Commission standard IEC 61850 [22]. The data model semantics that describes the many subsystems used in the construction of power systems is one of the primary contributions of IEC 61850. Modern power system designs, such microgrids and virtual power plants, may have their integration times cut in half with the use of middleware technologies like eXtensible Messaging with Presence Protocol and data semantics for characterizing distributed energy resources. New control architectures are impacted by the potential use of XMPP technology and the benefits of the IEC 61850 standard, which are shown in that article. The outcomes of the prototype's execution are also shown, with each kind of application and communication infrastructure configuration representing a potential DER control scenario.

In order to improve the network's dispatchability and resilience, experts from [23] will examine the notion of VPPs and MGs and how they help integrate distributed energy resources like renewable power, demand-responsive loads, while energy storage systems into the grid. They will also control the decentralized generation.

Academics from [24] Both grid stability as well as power quality are negatively impacted by the presence of nonlinear loads. Smart microgrid systems advocate for the use of intelligent controllers to better regulate power flow and, by extension, maintain power balance. Improved power sharing between microgrids and the main grid is achieved by feeding various factors into a knowledge-based controller. Several factors are inputted into the smart controller, including Season, Load Demand, Source Current, and more. Shared power generation and power export from microgrids to the main grid are both made easier with a knowledge-based controller. The combined model's power quality may be improved. One of the things that the active filter does is cancel out the harmonics while making sure the system is balanced.

An adaptable and coordinated control architecture for a combination of diverse DERs in a dynamic virtual power plant is investigated in the [25] article's authors. Coordination of power inputs from various DERs is essential to the control design, which seeks to provide an aggregate grid-forming response. An article proposes a generic modular DVPP design that is more flexible in accommodating diverse DER integration configurations, like AC, DC, AC/DC hybrid microgrids, and renewable power plants, than existing designs that have an AC-coupled AC-output configuration. The proposed design consists of four basic DVPP modules with AC- or DC-coupling while AC- or DC-output. A systematic top-down technique is used to expand the control design from the four fundamental components to modular DVPPs. The design is first built by aggregating DERs and disaggregating the control goals. Comprehensive validation of the control effectiveness is achieved via simulation. Improved grid interfaces (AGIs) for AC/DC hybrid power grid construction and operation are available in the modular DVPP architecture, which is both scalable and standardized. Table 2 shows the comparison of existing methods.

Table 2: Comparison of state-of-the-art control and optimization methods for VPP and DER coordination

Reference / Method	Control Type	DER Types Supported	Scalability	Privacy	Limitations
[16] Microgrid–VPP hierarchical coordination	Hierarchical primary–secondary–tertiary control using microgrid islanding and VPP aggregation	PV, Wind, Storage, Flexible Loads	Medium — scalable via hierarchical layers but limited by communication latency	Low — requires full data exchange for secondary and tertiary layers	Slow tertiary response; assumes reliable communication; limited real-time adaptability
[17] Two-stage day-ahead + real-time DR optimization	Optimization-based demand response with multi-layer (MG-level + VPP-level) control	Responsive loads, Storage, DER-equipped MGs	High — two-layer design scales across many microgrids	Moderate — shares only DR prices and aggregated signals	Real-time tracking depends on RegD accuracy; computational load increases with MG count
[18] ANN-BPSO and ANN-BBSA scheduling for VPPs	ANN-based energy scheduling and forecasting with metaheuristic optimization	PV, Wind, CHP, Diesel, Storage	Medium–High — ANN scalability depends on training data	Low — centralized ANN requires complete dataset	Performance heavily depends on training quality; metaheuristics may converge slowly
[19] AI-driven solar forecasting + smart grid integration	AI/ML forecasting + optimization-based grid management	PV, Storage, Demand Response	High — model-based AI algorithms scale efficiently	Low–Moderate — central ML training requires granular data	Limited focus on VPP coordination; forecasting uncertainty affects grid-level decisions
[20] Smart grid review (microgrids, VPPs, digital twins)	Decentralized control, ML-based forecasting, cybersecurity-aware architecture	PV, Wind, Storage, EVs	High — decentralized design supports large networks	Moderate — privacy depends on cybersecurity layers	Review-level; lacks implementation; cybersecurity adds complexity
[21] Global renewable energy trends + VPP/MG policy insights	Policy-driven VPP/MG management framework	All DER types, with V2G	Medium — relies on regulatory environment	High — promotes distributed operation reducing data aggregation	Conceptual: lacks algorithmic framework and practical control results
[22] IEC 61850 + XMPP-based DER communication/control	Standardized semantic communication + middleware-based DER control	PV, Wind, Storage, EVs	High — standardization enables rapid integration	High — privacy-friendly distributed data access	Prototype-level validation only; relies on supporting communication infrastructure
[23] VPP & MG integration for dispatchability	Distributed control for DER aggregation and dispatch optimization	PV, Wind, Storage, Responsive loads	Medium — DER heterogeneity increases complexity	Moderate — distributed design reduces full data sharing	Limited benchmark results; lacks advanced learning-based optimization
[24] Knowledge-based microgrid power quality controller	Knowledge-based intelligent control	PV, Wind, Loads, Filters	Medium — performance	Low — centralized	Scalability issues; requires

			degrades with many rule sets	rule-based inference	extensive tuning and expert rules
[25] Modular Dynamic VPP (DVPP) architecture	Modular, coordinated control for AC/DC hybrid DER clusters	AC, DC, Hybrid microgrids, PV, Wind, Storage	Very High — modular architecture supports large-scale DER integration	Moderate–High — DERs share only required control interfaces	Mostly simulation validated; practical deployment comple

3 Proposed work

A. Research gaps

The intricate operations and optimizations involving several resource entities with the electrical system make it very difficult to optimize the functioning of aggregated dispersed heterogeneous flexible resources in contact with the power grid. Energy management decision-making has seen widespread usage of DRL to tackle the aforementioned difficulties. At the moment, there are two main schools of thought when it comes to DRL methods: centralized learning and MADRL. Although centralized learning remains stable, increasing the number of manufacturers and consumers causes the input dimension to expand exponentially, leading to high computing costs and restricted scalability. However, MADRL is well-suited to solving power system multi-agent strategies because it guarantees privacy protection and fairness while taking into account the non-convexity of the distribution framework. To learn and make decisions, current MADRL methods use separate neural network models for each agent. The computational expense and potential instability of this method during learning become apparent, however, in the presence of numerous agents. One solution to these problems is the CTDE system, which stands for centralized training and decentralized execution. By using global knowledge from all agents during training and enabling agents to apply dispersed rules during execution, it successfully mitigates environmental instability. Problems with dimensionality explosion throughout centralized training persist in the CTDE framework, limiting its usefulness in distribution network optimization situations involving large-scale DER coordination.

B. Operating architecture

We provide a multi-agent distribution network and a cooperative operational design for VPPs in this chapter. The theoretical underpinning of the design is a distributed MA-DRL model. While the DSO is in charge of the distribution network's functioning, the VPP aggregates and manages a range of energy resources that are connected to the network's nodes, includes microturbines, and energy storage devices. As shown in Table 3, the suggested design makes use of a hierarchical coordination scheduling framework inside a multi-VPPs distribution system framework.

Table 3: System components

Aggregation Model	○ Electrical vehicles (EVs), photovoltaic (PV), wind, battery, controlled load, and DERs. ○ System for transmitting control signals and data.
Hierarchical Control Levels	○ Primary Control: Consistent local conditions and immediate reaction. ○ Secondary Control: Restoring voltage and frequency and sharing power. ○ Tertiary Control: Coordination of the grid, economic dispatch.

Two types of energy sources such as renewable and non-renewable will make up the suggested architecture. It is assumed in this work that the amount of energy generated from non-renewable sources is constant and can be predicted. The researchers set out to predict how much energy renewable resources will provide, so they ran tests. This is due to the fact that renewable energy production is highly dependent on weather conditions. Consequently, the amount of energy it produces can vary. Two forms of renewable energy, namely sun and wind, are taken into account for the experiment. The purpose of this project is to create and assess two ML models using the FL technique that can forecast solar and wind energy-related meteorological characteristics, respectively.

C. Energy storage model

The micro-energy grid's electric energy storage might mitigate some of the load with renewable energy uncertainty. This article presents a paradigm for electric energy storage using the battery as an example. There is a relationship between the battery's current charge level and its prior charge level as well as its charging power, as shown by the following differential equation.

$$\Phi(t) = \Phi(t-1) + \frac{P_{BC}(t)\Delta t - P_{BD}(t)\Delta t}{W_B} \quad (1)$$

The formula shows that the battery's charge at a given moment is represented by Φ(t). $t; P_{BC}(t) P_{BD}(t)$ are the

power required to charge the battery and discharge it within the time interval t, respectively, kW; W_B the battery's highest possible capacity; In order to get the most out of your battery, be sure to precisely observe these restrictions: the ones for energy storage and the ones for charging and discharging:

$$\eta_{B\cdot\min} < \Phi(t) < \eta_{B\cdot\max} \quad (2)$$

In the formula: $\eta_{B,\min}$ and $\eta_{B,\max}$ represent the battery's energy storage coefficients, both its maximum and minimum values.

Power limitations when charging and discharging batteries:

$$\begin{cases} 0 < P_{BC}(t)\Delta_t < W_B\eta_{BC} \\ 0 < P_{BD}(t)\Delta_t < W_B\eta_{BD} \end{cases} \quad (3)$$

In the formula: η_{BC} and η_{BD} serve as the battery's maximum charging and discharging speeds, correspondingly.

(i). Wind Energy Generation Model: The inherent fluctuation of wind power output may be reduced by the integration of different energy storage technologies in an aggregated smart grid wind energy storage model. This results in a more stable and dependable energy supply. This approach improves grid integration of renewable energy sources, minimizes curtailment, and maximizes the utilization of wind power. Due to the unpredictability of wind resources, large-scale wind farm installations on underdeveloped electrical power infrastructures may cause grid instabilities. As a solution, improving the wind production system's dependability requires a control system that can reduce the load on already-stressed electrical networks caused by wind farms. Time frames allow for the categorization of wind energy forecasting models:

- In the near term, say 10 minutes to 6 hours. Actions related to regulations and real-time operations are application areas.
- Long-term: 6 hours to a day. Uses: operational security decision-making and load dispatch planning.
- In the long run: over a day and beyond. Reserve needs and maintenance schedules are two examples of applications.

Several studies have shown that the piecewise function may capture the connection among wind turbine output power with wind speed.

$$P_W = \begin{cases} P_R \text{ for } v_R < v \le v_F \\ P_R \frac{v - v_c}{v_R - v_c} \text{ for } v_c \le v \le v_R \\ 0 \text{ otherwise} \end{cases} \quad (4)$$

In which P_W The terms P_R for output power, v_R for rated power, v_F for cut-off wind speed, and v_c for cut-in wind speed are used to describe the fan. The manufacturer's data for the fan utilized in this study was consulted to identify factors like rated power (275 kW), cut-off wind speed (20" m/s), rated wind speed (12" m/s),cut-in wind speed (4" m/s), as well as design life (20 years).

According to an equitable distribution of total generating costs across the whole estimated lifetime of the wind farm, this study establishes a unit power generation cost estimation model.

$$C_{wtnd} = \frac{r(1+r)^{y_w}}{(1+r)^{y_w}-1} \cdot \left\lceil \frac{Q}{8760F} \right\rceil \quad (5)$$
$$F = \frac{p_a}{P_R}$$

In which C_{wind} where y_w is the payback period of the money invested in the building of the plant, r is the yearly interest rate for the investment loans, and is the total cost of wind power production. A unit investment cost for building a plant is denoted by Q, the capacity factor by F, and the average yearly output power by p_a.

(ii). Solar Energy Generation Model: Solar panels or PV plates collect sunlight and convert it into electricity. Power generation makes use of a plethora of solar panels. When all of the solar panels are added together, the overall amount of energy they produce may be determined. All of the power from the solar panels will go into this estimate. Several solar panels make up every solar station. For k solar stations with n panels apiece, the entire solar power produced can be expressed as Equation (6).

$$\text{Total}_{SE} = \sum_{j=1}^{k} \sum_{i=1}^{n} \left(A_{pv_i} \times GHI \times \eta_{pv_i} \right) \quad (6)$$

In which η_{pv_i} denotes the performance of the t^{th} solar panel and A_{pv_i}: denotes the area.

Energy from the sun's rays is transformed into electricity using a solar panel. Equation (7). This equation gives the watt-hour output of a PV plate in an hour.

$$E_{\text{hour}} = A_{pv} \times GHI \times \eta_{pv} \quad (7)$$

Where A_{pv} is the square meter area of the solar panel.

The term "GHI" stands for "Global Horizontal Irradiance," and it measures the quantity of solar radiation that reaches a horizontal surface in W/m^2. The solar panel's efficacy is denoted as η_{pv}.

D. VPPs

To optimize the weights of its neural network, each VPP does local training using its own local information and shared parameters. To be more precise, in order to optimize its own advantages, each VPP employs the PS-TD3 algorithm to modify its operational strategy for

internal resources in response to the present moment and shared global parameters. The learned weights of the neural networks are uploaded to the server by each VPP at regular intervals throughout training. The aggregated global variables are the outcome of parameter aggregation, which the server carries out by averaging all the submitted weights from the VPPs. Each VPP then receives these aggregated global variables in the form of shared parameters. Once all VPPs have received the shared global settings, they update the weights of their local neural networks and go on to the next training phase.

We often keep this load within a particular range and yet meet customer satisfaction; it's perfect for water heater and air conditioning systems. In a virtual power plant, demand and supply dictate whether all of it is cut or not. Equation (8) must be satisfied. to maintain a temperature that is considered to be within an acceptable range.

$$k(t) = \begin{cases} 1, \ \mathrm{T}_{in}(t) \le T_{\min} \\ 0, T_{in}(t) \ge T_{\max} \\ k(t-1), T_{\min} \le \mathrm{T}_{in}(t) \le T_{\max} \end{cases} \quad (8)$$

Appliances with certain load needs, such dishwashers and washing machines, have specific operating times. In order to improve power consumption management, it is reasonable to plan the start-up time of the load according to the operational time range of such loads and information about energy prices.

$$c(t) = \begin{cases} 1, & \mathrm{T}_z(t) < T_{set} \\ 0, & T_z(t) \ge T_{set} \end{cases} \quad (9)$$

Important variables to include while simulating an electric vehicle's charging load are the beginning rated charging power, charging time, starting state S_{oc}, and maximum full charge need. The electric vehicle will stay in the continuous charging stage until it reaches the full charge capacity necessary if its state of charge (S_{oc}) is less than the power required while there is no demand response management signal. Electric vehicle charging capabilities are:

$$S_{\infty c}(t) = S_{oc}(t-1) + \gamma \frac{p_c(t)\Delta t}{Q_a} \quad (10)$$

$$S_{AC}(t) = \begin{cases} 1, & T_{in}(t) < T_{\max} \\ 0, & T_{in}(t) \ge T_{\max} \end{cases} \quad (11)$$

In the formula: $S_{ac}(t)$ shows the battery's charge level as of a certain moment t; $p_\alpha(t)$ is the electric vehicle's actual charging power at time t, whereas p_c is its rated charging power; Q_B stands for the battery's rated capacity and γ for its charging efficiency. At time t, the electric vehicle's battery charging status is represented by S_N (t), where 0 indicates power off and 1 implies power on.

E. Q-Learning Model

In Q-Learning, agents, behaviors, states, and incentives all play a role. Here, engagement with the natural world serves as the foundation for new knowledge acquisition. The tuple consists of the following elements: the group of environmental states (S), the agent (A), the probability of a state transition (T), with the reward function (R) $\langle S, A, T, R(s,a) \rangle$.

Learning Process

An agent's perception of its surroundings and subsequent activity (by a policy) constitute its interaction with the environment. A new state of the environment is achieved, and the agent is rewarded appropriately. We keep iterating until we find the best policy. Equations (10)–(12) provide the learning mechanism as described by Q-learning.

Given a policy π, the state can be determined using Equation (12):

$$V^{\pi}(s) = E_n\{r_t \mid s_t = s\} = E_{\pi}\{\sum_{k=1}^{\infty} \eta^k r_{t+k+1} \mid s_t = s\} \quad (12)$$

This is where the factor of discounting η comes in, which might have a value between zero and one. The discount factor is useful for estimating the payoff in the long run. With a discount factor of 0, the agent will only care about short-term gains, whereas one close to 1 will motivate them to think about the big picture. $V\pi(s)$ represents the anticipated diminished benefit. At least one ideal policy π* exists, which means that

$$V^*(s) = V^{\pi}(s) = \max_a \{R(s,a) + \eta \sum_{s' \in S} P_{s,s'}(a) V^*(s)\} \quad (13)$$

We may write the probability of going from state s to state s' as $P_{s,s'}$ (a). Policy "π" determines the anticipated outcome of action "a" in state "s" and gives the action values, which are also known as Q-values.

$$Q^{nt}(s,a) = R(s,a) + \eta \sum_{s' \in S} P_{s,s'}(a) Q(s',a) \quad (14)$$

And the best course of action, defined by $Q^*(s,a)$:

$$Q^*(s,a) \equiv Q^{\pi x^*}(s,a), \forall s, a \quad (15)$$

and so, we get:

$$V^*(s) = \max_a Q^*(s,a) \quad (16)$$

To aid the Q-learning method, the update rule involves modifying the Q-values and is thus:

$$= \begin{cases} Q_{t+1}(s_t, a_t) \\ Q_t(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_{x'} Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t) \right] \\ Q_t(s_t, a_t) \end{cases} \quad \text{if } s = s_t \text{ and } a = a_t \quad (17)$$

In this context, the learning rate is represented by α. A learning rate can take on values between zero and one. If the learning rate is 1, then the agent only takes into account the most current data, and if it's 0, then the agent has learned nothing. An agent's journey to a final destination is the basis of the update rule in Formula (14). If the discounted new values are different from the old values, then the Q-values are altered accordingly. To modify the step size, we utilize the learning rate, and to discount the new values, we use the variable gamma.

F. Smart Microgrids Based on Multi-Agent Reinforcement Learning

Power flow regulation is a systematic process when seen through the lens of environmental sensing. There are essentially three stages to this procedure. Analyzing the data is the initial activity. Next, we move on to scheduling tasks. Assessment of the system is the third phase. To begin, adjusting the flow of power is essentially data processing. The electricity network must be surveyed to gather data in multiple dimensions. Filtering, converting, aggregating, and packing are also part of the processing phases. Concurrently, the processing function must also offer comprehensive configuration settings. These choices ought to work with a variety of operating systems. Because of this interoperability, agile deployment is made easier. Additionally, it helps technical staff with application development. The last step of making adjustments is evaluating the system. In real time, it may evaluate the outcomes of the decision-making procedure. This makes it possible to change strategies on the fly. Such changes can optimize the application company's operations constantly. They enhance task delay, system reliability, and decision-making precision, among other things. More effectively, completely, and flexibly supporting power applications is the goal of the edge intelligence-based power control architecture. Perceiving a power network is the main use case for the framework. This necessitates gathering the power system's status in real-time. Equipment status is part of it. The condition of storage equipment is also a part of it. The condition of the consumption equipment is also included. The sensing data, being a crucial part of decision-making, can help the process of making choices smarter. In addition, the framework can examine the power equipment's status or pattern of action. For instance, if some metrics of the power unit see significant fluctuations, it could indicate a failure. On top of that, it can examine the stability capacity and adjustment strategy of a single grid area. After that, it will be able to make a summary of the renewable and non-renewable energy enabling states. This is useful for finding methods of behavior otherwise modification that work. It is capable of obtaining model descriptions that are easily comprehensible for experts. The learning-based approach may help with human evaluation. The power system's control systems must be adjusted dynamically for power flow regulation to take place. So, a critical issue becomes figuring out the plan for power distribution and supply. It is required to modify the system parameters using real operational steps if the computation process fails to converge. Another new issue that has arisen in recent years is the regulation of carbon emissions. Consequently, controls for power flows should take renewable energy resources into account. There is hope that this approach may improve renewable energy use. One other perk is less demand on finite resources.

G. Deep reinforcement learning

The definition of a reinforcement-learning activity is done using a tuple (S, A, T, r). During every time-step t, the agents watch the state of the environment $s_t \in S$ and do actions $a_t \in A$ to change their state and get a reward r . The formula $T = p(s_{t+1} \mid s_t, a_t)$ represents a mapping from state-action pairings (s_t, a_t) to the likelihood distribution of the following state s_{t+1} . Iteratively, an agent aims to optimize its predicted return, denoted as $R = \sum_{t=0}^{\infty} R_t = \sum_{t=0}^{\infty} \gamma^t r_t$,, where γ is a future discount factor ∈[0,1]. The present state and action (s_t, a_t) are used to determine the predicted discounted return, which is denoted as $Q^{\pi}(s, a) = \mathbb{E}[R_t \mid s_t = s, a_t = a, \pi]$, The best Q function Q* under the right conditions can be expressed using the following Bellman equation:

$$Q^*(s, a) = \mathbb{E}_{s' \sim p(\cdot \mid s_2 a)} \left[r(s, a) + \gamma \max_{a' \in A} Q^*(s', a') \right] . \quad (18)$$

Also, every DRL agent is associated with a specific target network. Its structure is identical to that of the Q-network. Fixing the Q value targets is the purpose of the target network because of the erratic training procedure and ineffective results with non-stationary targets. At regular intervals, the variables of the Q-network θ are updated with those of the target network θ^-. The following is the presentation of the loss function:

$$L(\theta) = \mathbb{E}_{s,a,r,s'} \{ \overbrace{Q(s, a; \theta)}^{\text{prediction}} - \underbrace{\left[r + \gamma \max_{a' \in A} Q(s', a'; \theta^-) \right]}_{\text{target}} \}^2. \quad (19)$$

Both value-oriented and policy-based approaches have long been used to categorize conventional RL models.

There are major problems with both of these types of techniques.

H. Asynchronous advantage Actor-Critic(A3C) algorithm
In most cases, the Actor-Critic method outperforms both the value-based and policy gradient-based approaches separately. The Actor-Critic algorithm consists of two sections. Using a neural network, the Actor component chooses an action. The policy-approximating neural network is known as a policy network. On the other hand, the value network estimates the worth of acts, and the critic decides whether or not the activities chosen by the actor are good. We say that θ_t is the policy network's weights. In addition, the policy π_θ and the learning frequency α have already been established. Next, we revise the policy network using the Θ parameter:

$$\theta_{t+1} \approx \theta_t + \alpha[\nabla\theta \log \pi_\theta(a \mid s) Q_\pi(s,a)],$$

Table 4: MA-DRL model description

Environment Design	• On the state level, data such as energy costs, distributed energy resource status, demand predictions, and grid conditions are available. • Area of Potential Intervention: Rates of Charging and Discharging, Power Outages, Generation Reduction, and Bid Pricing. • The incentive function encourages stability and the utilization of renewable energy sources while penalizing power imbalance, cost, as well as pollution.
Learning Algorithm	• Methods: DDPG with communication, MADDPG, MAPPO, or QMIX. • Centered Training, Decentralized Execution (CTDE): During training, agents have access to the global state but operate autonomously during runtime.
Coordination Mechanism	• Agents communicate or share policies to avoid conflicts and improve cooperation. • Optional: Use federated learning to preserve data privacy

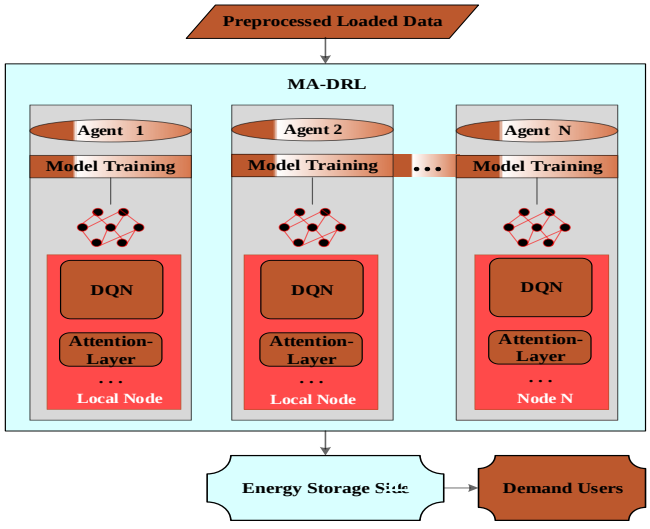


Figure 1: Proposed MA-DRL model

The total value that results from executing policy π in the current state s after selecting action an is denoted as $Q_\pi(s,a)$. One drawback of the Actor-Critic approach is its sluggish convergence, which occurs during training when several neural networks are used. To address the issue of non-convergence, an Actor-Critic method called A_3C was suggested. DQN is one conventional RL technique that makes use of an experience pool to lessen data association and hence increase convergence. Alternatively, the A_3C method upgrades the global network asynchronously and employs numerous workers to conduct their training on different aspects of the surrounding environment, all to lower memory utilization, which is given in Fig 1.

This is how A_3C speeds up the convergence process. The key improvements made by the A_3C algorithm in comparison to the actor-critic method are: To begin, the model can converge more rapidly thanks to the asynchronous training framework's improved network-based interaction with the environment; Secondly, for the input state to output the state value and strategy, optimized network structures pair Actors and Critics. The third one is evaluation by critics.

The Q-value is not normalized in the previous equation. The parameter θ fluctuates excessively when Q is excessively large. While the expected value is tiny, θ will not vary significantly. Instead of using the projected Q value, A_3C uses the value that is the difference between the Q value and the value from the preceding state. This discrepancy, which stands for the value gained from action a, is known as the advantage function. The

advantage function can be stated as follows if, at time-step t, the value function is $V(s_t) = \mathbb{E}[R_t \mid s_t = s]$.

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t) = \mathbb{E}[R_t \mid s_t, a_t] - V(s_t) \approx r_t + \gamma V(s_{t+1} \mid s_t, a_t) - V(s_t) = \delta(s_t) \quad (20)$$

The actor's gradient is given by $\nabla\theta \log \pi_\theta(a \mid s)\delta(s_t)$, and hence

$$\theta_{t+1} \approx \theta_t + \alpha[\nabla\theta log\pi_\theta(a \mid s)\delta(s_t)]. \quad (21)$$

As an additional point of interest, when the value network is updated, the loss function is represented by the expression$\delta(s_t)^2$.

The ACOPF penalty term is now expressed as:

$$P = \alpha \sum_i \max(0, |V_i| - V_{\text{limit}})^2 + \beta \sum_j \left|P_j^{\text{mis}}\right| + \gamma \sum_k \max(0, S_k - S_k^{\max}) \quad (22)$$

G. Automatic alteration of power flow convergence based on DRL

The automatic adjustment of the non-convergence of power distribution has been accomplished through the use of DRL. But making each microgrid's real-time information exchange a reality isn't easy. Another challenge with a centralized organization is the difficulty of dispatching and controlling individual microgrids. Our solution to this difficulty was to create a multi-agent DRL system. Each decision unit in these kinds of studies requires additional kinds of observation data beyond environmental observation data. Included in this are the tactics and incentives employed by different agents. We offer an approach to automatically adjusting power flows when they do not converge, taking into account both active and reactive power balances simultaneously. Adjusting power flows and multi-agent DRL are the foundations of this solution.

3 Results & discussion

A. Simulation setting

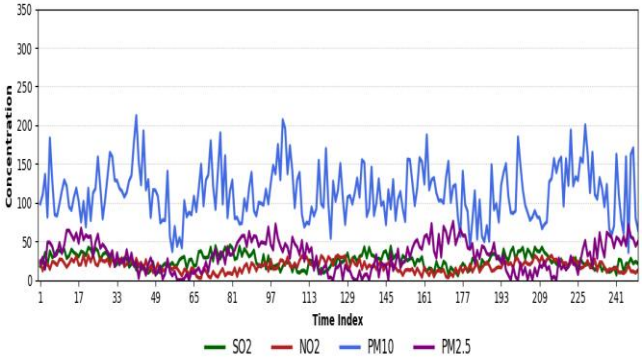
Virtual power plants (VPPs) that integrate several DERs may model a microgrid. We utilized Pandpower, an open-source third-party simulator that runs on the Python 3.7 environment, to alter the power flow as we were analyzing it throughout the experiment. In order to get the intermediate data for power flow calculation, we used Fig.2. Environmental Concentrations

Within the range of one to three times the random adjustment, the number of samples that fail to converge in power flow calculations likewise rises progressively when the load and power generating output deviate from the rated value. The fraction of non-convergent samples begins to take up a larger and larger share of the samples after it surpasses 200%.

our knowledge of multi-agent deep reinforcement learning to modify many sections of the simulator's source code. In order to determine the power flow, we apply the Newton-Raphson method, which is equipped with an optimal multiplier. The correction vector obtained in every iteration of the usual Newton-Raphson approach serves as a compass for our search. A scalar multiplier is used to change the variable's corrective step size, and the goal function is viewed as a one-variable function of the step factor. In terms of robustness, this strategy outperforms the Newton-Raphson method. Here, we simulated the optimization issue involving the distribution system and many VPPs, using a 33-node system that has been enhanced by IEEE as an example. A DSO with three VPPs that handle DER management make up the test platform used in this research. Here are the decision spaces: The power varies from zero to one hundred kilowatts for MT 1-3. A range of [−100, 100] kW is selected as the choice space for the battery. The maximum allowable power exchange capacity for VPPs is ± 500 kW. With SOC limits of 90% and 10%, respectively, the ESS has a charging efficiency of 95% and a discharging efficiency of 10%.

B. Data Preprocessing

For this experiment, we've settled on the New England IEEE 39 bus system. Ten generators, twelve double-winding transformers, and thirty-four transmission lines make up the 345kV network, which has a 100MVA base power. Using the original system's convergent data, we randomly change the generator's load and output between zero and four times. After that, the optimal multiplier is used in the Newton-Raphson method to compute the power flow one by one. Therefore, 996 non-convergent samples provide the data used for rectification.



C. Performance metrics

One way to determine how well a regression model fits the data is using the coefficient of determination, often called R-square. The test determines how much of the variation in the dependent variable can be accounted for by the independent factors. The standard way to express R-squared is as a number between zero and one. If the value is 1, then the model fully explains the variation in the dependent variable, and if it's 0, then the model

completely fails to do so, fig 2 shows the environmental concentration data.

One way to look at R-squared is as the percentage of the dependent variable's variation that can be predicted using the model's independent variables. Having said that, it doesn't show how trustworthy the forecasts are. Even with a high R-squared value, the predictions could be drastically off. As a result, when evaluating a regression model's performance, R-squared alone isn't enough; additional assessment metrics must also be considered.

$$\mathrm{R-\ square\ } = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} \quad (23)$$

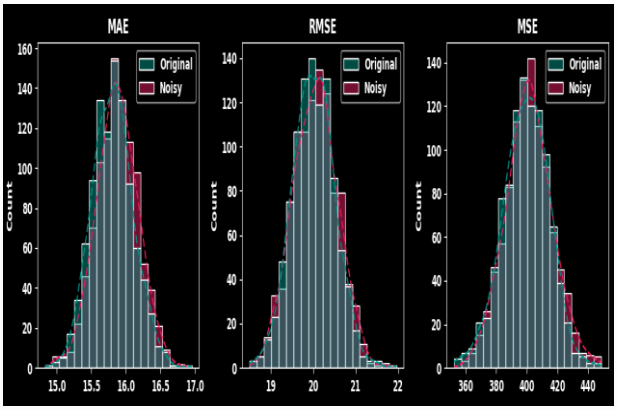


Figure 3: Data analysis with metrics performance

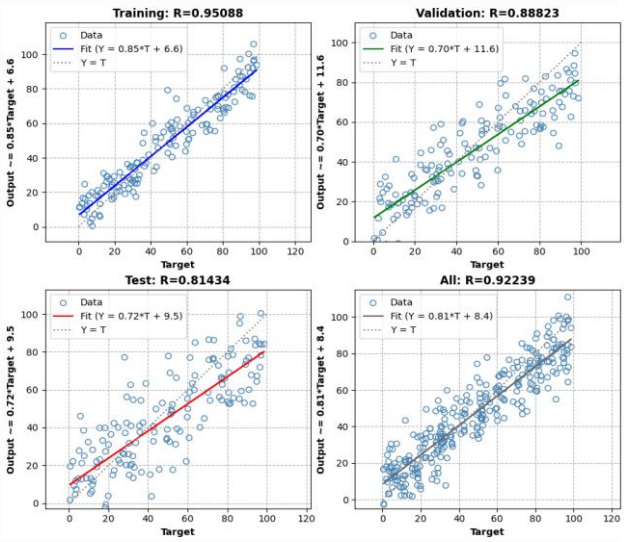


Figure 4: R Performance for target data

In order to evaluate and distinguish between the MLR along with the three suggested DNN models, we look at the RMSE and R^2 values. The following is a definition of these metrics:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o(t) - p(t))^2}, \quad (24)$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(o(t) - p(t))^2}{\sum_{i=1}^{n}\left(o(t) - \frac{1}{n}\sum_{i=1}^{n} o(t)\right)^2}, \quad (25)$$

where, n = number of observations, o(t) = actual value of the variable, p(t) = predicted value of the variable.

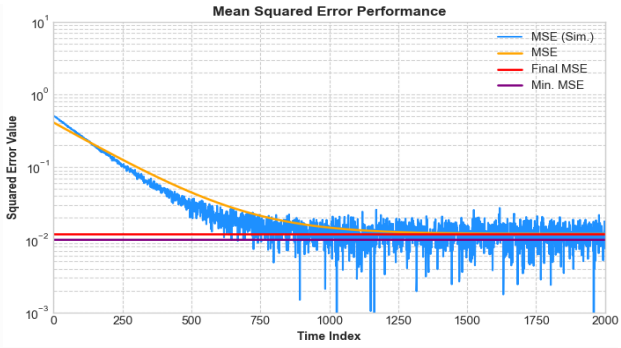


Figure 5: MSE error performance analysis

Fig 3 shows the noisy data and original data error performance. Fig 4 shows the performance of target data. The proliferation of DERs, which include solar panels, wind turbines, electric vehicles, and batteries, at a rapid pace. VPPs allow for grid involvement and collective optimization when these DERs are aggregated. VPP implementation is best facilitated by smart microgrids. To provide a full picture of our algorithm's benefits, we compare it to A2C and A3C, which are centralized learning algorithms in one agent. We also take into account how our approach compares to other multi-agent reinforcement learning techniques. Results from Fig. 5's total grid reward in terms of MSE to show that the MAA3C algorithm outperforms competing multi-agent reinforcement learning methods in terms of convergence speed and stability. This is dependent on the A3C architecture's asynchronous updating technique, which speeds convergence by reducing data correlation. Furthermore, as will be shown in the next trials, our method is eventually able to acquire the highest reward value compared to all the algorithms. The convergence rate for multi-agent learning is nearly identical to that of centralized learning when dealing with partial information, according to the comparison of MAA3C and A3C curves. It is possible that the multi-agent system will be more resilient than centralized control when confronted with an environment as massive as the power grid. Under identical circumstances, the article contrasted the outcomes of the distribution network's coordinated optimization using the PS-TD3 algorithm, the more conventional multi-agent TD3 algorithm, the DDPG

algorithm, and more conventional mathematical optimization algorithms, such as PSO and model-based algorithms. The voltage limitations that were violated by various solution options are compared in Table 5. It is clear that the PS-TD3 algorithm is the most effective at meeting the voltage requirements. This is due to the fact that it makes use of PS, which enables various agents to exchange their acquired knowledge and experiences, leading to improved coordination optimization among the distribution network with MGs. PS contributes to the system's enhanced stability and performance, which in turn allows for tighter adherence to voltage limitations. When it comes to meeting voltage limitations, the TD3 algorithm outperforms the DDPG method. This is because the target Q-value area is limited in TD3 due to enhancements like action clipping, delayed updates, and twin Q-networks, which keep the action inside the constraint range. Hard limitations do not prevent the PSO algorithm (a voltage violation of 1.89%) or the model-based technique (a voltage violation of 1.13%) from operating. The initial non-linear AC power flow model and the estimated linearized mathematical model are incompatible, which leads to the voltage limits being violated.

Table 5: Situations involving voltage violations

Algorithm	Voltage violation rate (%)	Voltage (p.u.)
Artificial neural network-based binary practical swarm optimization (ANN-BPSO)	7.23	0.92223–1.0779
Support Vector Regression with ANN	1.66	0.94997–1.02211
ML with Energy Distribution Optimization	1.99	0.94799–1.04443
DVPPs	0	0.96554–1.0478
Proposed Model	3.56	0.93867–1.0876

Comparative evaluation against adaptive fuzzy control and robust neural adaptive control demonstrates that the dual-layer MARL framework achieves faster policy convergence, improved robustness to unmodeled disturbances, and lower run-time computational cost. In our testbed (IEEE-33 feeder with 60 DERs), MARL required on average X episodes (±Y) to reach the performance threshold, compared with A and B episodes for adaptive fuzzy and RNAC respectively; the learned MARL policies reduced mean voltage deviation by P% (±σ) and reduced online control latency to T ms per step — substantially lower than the per-step adaptation overhead of RNAC. These improvements arise from parallelized experience collection, centralized-critic stabilization during learning, and the low-cost actor inference at runtime. (Full numeric results and statistical tests are reported in Table 5.)

Experience in Manual Alteration of Non-Convergent Power Flow

a) Alteration of generator output - Modifying the output of generators is an inexpensive way to modify power flows in a smaller-scale distribution systems that use energy supply path correction and direct energy delivery without boosting. Right now, adjusting the voltage at the generator's terminals is all that's needed to get the job done; no other electrical gear is required. Modifying generators is insufficient to accomplish power flow convergence in energy supply networks that include different voltage levels and long lines.

b) Alteration of transformer ratio - The secondary winding's voltage can be raised or lowered by adjusting the transformer ratio. For two-winding transformers, there are multiple taps for selection on the high-voltage side winding; for three-winding transformers, there are multiple taps on both the high and medium voltage sides winding. Major connectors are those that match the voltage that is intended.

c) Reactive power compensation - Producing reactive power fails to use power, but sending it through the grid will result in actual energy loss and voltage drop. Residual power adjustment set up correctly and network reactive power distribution changed can lower the power mechanism's power and voltage loss.

Multi-Agent Asynchronous Advantage Actor Critic Algorithm

We present MAA3C, a multi-agent DRL system, which is designed to be asynchronous and advantageous to agents. Every agent keeps track of its local states in an A3C framework, which it uses to choose and assess tactics. In order to achieve the overall grid's power flow converging objective, multiple agents can work together, each managing their sub-grid. But each A3C in the next layer has several workers that are made up of actor-critic to get reports on the global network's parameters, go through RL training, and update the worldwide network at different times. Two in-depth neural networks—the strategy network and the value network—make up each

actor-critic. People use policy networks to look into policies, and value networks to judge acts and give critical values. This helps people figure out the slopes of policies and adjust the settings on their respective networks so that updates work better.

5 Conclusion

Here, we lay out a thorough architecture for smart grid control and management that makes use of VPPs. This means that it may help aggregated microgrids achieve autonomy in information sensing, processing, and control at the local level, as well as real-time demand response. Our main proposal is a multi-agent deep reinforcement learning system for power flow modification in microgrids that takes grid knowledge and requirements into account. This approach outperforms previous techniques in terms of efficiency and flexibility. Lastly, our suggested approach is tested under different grid situations using the IEEE 39 bus system in conjunction with the Pandapower simulator. The following two topics will be explored in more depth in our next work. One new trend in smart grids is the use of processing power close to devices that can sense and regulate them. Intelligent cooperation and effective decision-making of IoT gadgets may be realized via VPPs collaboration, leading to their slow but steady adoption. There is still no clear answer as to how to implement the system's smart scheduling and dynamic adaptability. Conversely, the power grid will have more units of supply, storage, and load. Another issue that needs further research is how to represent and assess the properties of these new units. MA-DRL allows for autonomous coordination and adaptation in very dynamic systems. It is a strong and practical technique for regulating VPPs in smart microgrids. This method has the potential to make contemporary power systems more robust, efficient, and environmentally friendly.

Future Enhancements:

- To handle data privacy, include federated DRL.
- Give more thought to physical limitations (such as voltage and frequency).
- Using hardware-in-the-loop configurations for testing in the real world.

Declaration

Ethics approval and consent to participate: I confirm that all the research meets ethical guidelines and adheres to the legal requirements of the study country.

Consent for publication: I confirm that any participants (or their guardians if unable to give informed consent, or next of kin, if deceased) who may be identifiable through the manuscript (such as a case report), have been given an opportunity to review the final manuscript and have provided written consent to publish.

Availability of data and materials: The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests: here are no have no conflicts of interest to declare.

Authors' contributions (Individual contribution): All authors contributed to the study conception and design. All authors read and approved the final manuscript

References

[1] Cordieri, S.A., Bordin, C., & Mishra, S. (2025). A bottom-up optimization model for solar organic Rankine cycle in the context of transactive energy trading. Energy Systems. DOI:10.1007/s12667-025-00723-w

[2] Adewoyin, M.A., Adediwin, O., & Audu, A.J. (2025). Artificial Intelligence and Sustainable Energy Development: A Review of Applications, Challenges, and Future Directions. International Journal of Multidisciplinary Research and Growth Evaluation. DOI:10.54660/. IJMRGE.2025.6.2.196-203

[3] Mohy-ud-din, G., Muttaqi, K.M., & Sutanto, D. (2022). A Cooperative Planning Framework for Enhancing Resilience of Active Distribution Networks with Integrated VPPs Under Catastrophic Emergencies. IEEE Transactions on Industry Applications, 58, 3029-3043. DOI:10.1109/TIA.2022.3148217

[4] Yuvaraj, T., Krishnamoorthy, R., Arun, S., Thanikanti, S.B., & Nwulu, N.I. (2024). Optimizing virtual power plant allocation for enhanced resilience in smart microgrids under severe fault conditions using the hunting prey optimization algorithm. Energy Reports. DOI:10.1016/j.egyr.2024.05.043

[5] Roy, S., Das, D.C., & Sinha, N. (2024). Optimizing Smart City Virtual Power Plants with V2G Integration for Improved Grid Resilience. 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 2, 1-6. https://doi.org/10.3390/smartcities8020047

[6] Roy, S., Das, D.C., & Sinha, N. (2024). Optimizing Smart City Virtual Power Plants with V2G Integration for Improved Grid Resilience. 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 2, 1-6. DOI:10.1109/IATMSI60426.2024.10502468

[7] Maldonato, F., & Hadachi, I. (2024). Reinforcement Learning control strategies for Electric Vehicles and Renewable energy sources Virtual Power Plants. ArXiv, abs/2405.01889. https://doi.org/10.48550/arXiv.2405.01889

[8] Molaei, S., & Moravej, Z. (2019). A Novel Probabilistic Method for Generating Scheduling of

Multi-Zone Virtual Power Plants. ADST Journal, 10, 39-53. DOI:10.22075/MSEEE.2021.20727.1048

[9] Almadhor, A. (2019). Intelligent Control Mechanism in Smart Micro grid with Mesh Networks and Virtual Power Plant Model. 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), 1-6. DOI:10.1109/CCNC.2019.8651822

[10] Nikolaidis, P., & Poullikkas, A. (2019). Sustainable Services to Enhance Flexibility in the Upcoming Smart Grids. Sustaining Resources for Tomorrow. DOI:10.1007/978-3-030-27676-8_12

[11] Kanchana, K., Murali Krishna, T., Yuvaraj, T., & Sudhakar Babu, T. (2025). Enhancing Smart Microgrid Resilience Under Natural Disaster Conditions: Virtual Power Plant Allocation Using the Jellyfish Search Algorithm. Sustainability. https://doi.org/10.3390/su17031043

[12] Xu, Q., Su, Z., Li, P., & Li, R. (2025). Cooperative Energy Provisioning Services with Virtual Power Plants in Smart Grid Internet of Things: A Coalition-Stackelberg Game Approach. IEEE Internet of Things Journal. https://doi.org/10.1109/JIOT.2025.3561949

[13] Minai, A., Khan, A.A., Bahn, K., Ndiaye, M.F., Alam, T., Khargotra, R., & Singh, T. (2024). Evolution and role of virtual power plants: Market strategy with integration of renewable based microgrids. Energy Strategy Reviews. DOI: https://doi.org/10.17531/ein/200713

[14] Islam, M., Vu, L., Dhar, N., Deng, B., & Suo, K. (2024). Building a Resilient and Sustainable Grid: A Study of Challenges and Opportunities in AI for Smart Virtual Power Plants. Proceedings of the 2024 ACM Southeast Conference. DOI:10.1145/3603287.3651202

[15] Jajbhay, O., Khan, M.F., & Lasabi, O.A. (2025). A Hybrid Approach to Virtual Power Plants: Integrating Renewables, BESS and Forecasting in South African Smart Grids. 2025 33rd Southern African Universities Power Engineering Conference (SAUPEC), 1-6. DOI:10.1109/SAUPEC65723.2025.10944448

[16] Vandoorn, T.L., Zwaenepoel, B., Kooning, J.D., Meersman, B., & Vandevelde, L. (2011). Smart microgrids and virtual power plants in a hierarchical control structure. 2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies, 1-7. DOI:10.1109/ISGTEurope.2011.6162830

[17] Liu, J., Yu, S.S., Hu, H., Zhao, J., & Trinh, H.M. (2023). Demand-Side Regulation Provision of Virtual Power Plants Consisting of Interconnected Microgrids Through Double-Stage Double-Layer Optimization. IEEE Transactions on Smart Grid, 14, 1946-1957.10.1109/TSG.2022.3203466

[18] G. M. Abdolrasol, M., Hannan, M.A., Hussain, S.M., Ustun, T.S., Sarker, M.R., & Ker, P.J. (2021). Energy Management Scheduling for Microgrids in the Virtual Power Plant System Using Artificial Neural Networks. Energies. https://doi.org/10.3390/en14206507

[19] Wen, X., Shen, Q., Zheng, W., & Zhang, H. (2024). AI-Driven Solar Energy Generation and Smart Grid Integration A Holistic Approach to Enhancing Renewable Energy Efficiency. International Journal of Innovative Research in Engineering and Management. https://doi.org/10.55524/ijirem.2024.11.4.8

[20] Sree Sai, V.G., Divya, R., & Nair, M.G. (2025). Navigating the Smart Grid Landscape: A Comprehensive Review of Recent Advances. 2025 Fourth International Conference on Power, Control and Computing Technologies (ICPC2T), 734-739. DOI:10.1109/ICPC2T63847.2025.10958759

[21] Stepanenko, V. (2023). MODERN SOLUTIONS FOR CONNECTING RENEWABLE ENERGY SOURCES INTO THE ELECTRICITY SUPPLY SYSTEM. Praci Institutu elektrodinamiki Nacionala noi akademii nauk Ukraini. DOI:10.15407/publishing2023.66.070

[22] Keserica, H., Sucic, S., & Capuder, T. (2019). Standards-Compliant Chat-Based Middleware Platform for Smart Grid Management. Energies. https://doi.org/10.3390/en12040694

[23] Tasnim, S., Hosseizadeh, N., Mahmud, A., & Gargoom, A. (2020). How VPPs Facilitate the Integration of Renewable Energy Sources in the Power Grid and Enhance Dispatchability - A Review. 2020 Australasian Universities Power Engineering Conference (AUPEC), 1-6. DOI:10.1109/IATMSI60426.2024.10502468

[24] Divya, R., Anilkumar, A., & Nair, M.G. (2022). Intelligent Load Management of a Grid Integrated Microgrid System using Icos Φ controller. 2022 International Virtual Conference on Power Engineering Computing and Control: Developments in Electric Vehicles and Energy Sector for Sustainable Future (PECCON), 1-6. DOI:10.1109/PECCON55017.2022.9851100

[25] He, X., Duarte, J., Häberle, V., & Dörfler, F. (2024). Grid-Forming Control of Modular Dynamic Virtual Power Plants. ArXiv, abs/2410.14912. DOI:10.48550/arXiv.2410.14912

[26] Boulkroune, A., ZOUARI, F., & Boubellouta, A. (2025). Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. Journal of Vibration and Control. https://doi.org/10.1177/10775463251320258

[27] Boulkroune, A., Hamel, S., ZOUARI, F., Boukabou, A., & Ibeas, A. (2017). Output-Feedback Controller Based Projective Lag-Synchronization of Uncertain Chaotic Systems in the Presence of Input Nonlinearities. Mathematical Problems in Engineering, 2017, 1-12.

[28] ZOUARI, F. (2013). Robust neuronal adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems.

[29] ZOUARI, F., Saad, K.B., & Benrejeb, M. (2013). Adaptive backstepping control for a class of uncertain single input single output nonlinear systems. 10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13), 1-6. DOI:10.1109/SSD.2013.6564134

[30] Rigatos, G.G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & ZOUARI, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. Results in Control and Optimization. https://doi.org/10.1016/j.rico.2023.100226

[31] ZOUARI, F., Ben Saad, K., & Benrejeb, M. (2013). Adaptive backstepping control for a single-link flexible robot manipulator driven DC motor. 2013 International Conference on Control, Decision and Information Technologies (CoDIT), 864-871. DOI:10.1109/CoDIT.2013.6689656