Received 23 December 2024, accepted 1 January 2025, date of publication 3 January 2025, date of current version 10 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3525843

APPLIED RESEARCH

Energy Management in Microgrids Using Model-Free Deep Reinforcement Learning Approach

ODIA A. TALAB, (Member, IEEE), AND ISA AVCI
Department of Computer Engineering, Faculty of Engineering, 78050 Karabük, Türkiye

Corresponding author: Odia A. Talab (e.uday4@gamile.com)

ABSTRACT Electric power systems are undergoing rapid modernization driven by advancements in smart-grid technologies, and microgrids (MGs) play a crucial role in integrating renewable energy sources (RESs), such as wind and solar energy, into existing grids. MGs offer a flexible and efficient framework for accommodating dispersed energy resources. However, the intermittent nature of renewable sources, coupled with the rising demand for Electric Vehicles (EVs) and fast charging stations (FCSs), poses significant challenges to the stability and efficiency of microgrid (MG) operations. These challenges stem from the uncertainties in both energy generation and fluctuating demand patterns, making efficient energy management in MG a complex task. This study introduces a novel model-free strategy for real-time energy management in MG aimed at addressing uncertainties without the need for traditional uncertainty modeling techniques. Unlike conventional methods, the proposed approach enhances MG performance by minimizing power losses and operational costs. The problem is formulated as a Markov Decision Process (MDP) with well-defined objectives. To optimize decision-making, an actor-critic-based Deep Deterministic Policy Gradient (DDPG) algorithm is developed, leveraging reinforcement learning (RL) to adapt dynamically to changing system conditions. Comprehensive numerical simulations demonstrated the effectiveness of the proposed strategy. The results show a total cost of 51.8770 €ct/kWh, representing a reduction of 3.19% compared to the Dueling Deep Q Network (Dueling DQN) and 4% compared to the Deep Q Network (DQN). This highlights the robustness and scalability of the proposed model-free approach for modern MG energy management.

INDEX TERMS DDPG, RESs, energy management, FCSs, microgrid, EVs.

NOMENCLATURE

RESs	Renewable Energy Sources.
PV	Photovoltaic.
WT	Wind Turbine.
MG	Microgrid.
MPC	Model Predictive Control.
ESS	Energy Storage System.
RL	Reinforcement Learning.
DRL	Deep Reinforcement Learning.
DDPG	Deep Deterministic Policy Gradient.
MDPs	Markov Decision Process.
DQN	Deep Q Network (DQN).
P_t^{DG}	Active power output of the i^{th} Distributed Generator (DG) at time t.
$P_{i,min}^{DG}$	Minimum output power of thei^{th} DG.
$P_{i,max}^{DG}$	Maximum output power of the i^{th}. DG
C^{pv}	Cost function of photovoltaic panels.
P^{pv}	Power generated by photovoltaic panels.
C^{wind}	Cost function of wind turbines.
P^{wind}	Power generated by wind turbines.
C^{MT}	Cost function of microturbines.
P^{MT}	Power generated by microturbines.

The associate editor coordinating the review of this manuscript and approving it for publication was Ning Kang.

C^{FC}	Cost function of fuel cells.
P^{FC}	Power generated by fuel cells.
P_{max}^{grid}	Maximum power that the MG may either purchase from or sell to the main grid.
$B^{grid}\left(t\right)$	Buying price associated with the active power consumption from the main grid at time t .
$S^{grid}\left(t\right)$	**Selling** price during period t.
$P^{grid}\left(t\right)$	Active power of the main grid in period t.
$\left(COST\right)^{Grid}$	Price of electricity obtained from the grid.
EMS	Energy Management Systems.
$\left(COST\right)^{Grid}$	Price of electricity obtained from the grid.
I_j	Current flowing through branch j.
R_j	Resistance of branch j.
P_{Loss}	Power loss in the distribution system.
$V_{i,t}$	Voltage of the bus i at time t.
V_{min}	Lower limit of the voltage level.
V_{max}	Upper limit of the voltage level.
P_t^{wind}	Power generated by wind turbine at time t.
P_t^{pv}	Power generated by PV system at time t.
P_t^{demand}	Total demand.
s_t	System state at time t.
a_t	Actions taken at time t.
$r_t\left(s_t, a_t\right)$	Reward at time step t.
$\mu(s, \theta^{\mu})$	Actor function.
$Q(s, a\vert\theta^{Q})$	Critic function.
J	Expected outcome of the initial distribution.
ρ	Discounted state visitation distribution for the policy.
β	Alternative stochastic behavior policy.
SoC_{min}	Minimum State of Charge (SoC) of each EV battery.
SoC_{max}	Maximum State of Charge (SoC) of each EV battery.
$SoC_{i,des}$	Minimum desired level of the final SoC of each EV battery before departure.
N_{bus}	Number of buses in the distribution system.
N_{ev}	Number of electric vehicles.
N	Noise process used in DDPG for exploration.
Δt	Time interval.

I. INTRODUCTION

The era of traditional power systems is rapidly transitioning, giving way to more productive and environmentally friendly electricity from various RESs, including photovoltaic (PV), Wind Turbines (WT), and other innovative sources. Microgrids (MGs) harness these RESs, functioning independently or supporting the main grid. Typically, MGs utilize solar panels, WT, or hydropower, which are inherently irregular and variable. Despite these fluctuations, effective energy management ensures a stable and reliable energy supply [1].

The variability observed in energy generation and consumption patterns within MGs is significant. Timely adjustments are essential to address shifts in user demand, weather conditions affecting renewable energy production, and changes in stored energy levels. Effective responses to these variations require advanced energy management systems (EMS) to guarantee dependable, secure, adaptable, and cost-effective operations for grid-connected and standalone MGs [2]. A proper control strategy prioritizes critical loads and manages Distributed Generation (DG) resources to maintain a continuous, stable, and high-quality power supply [3]. MGs, characterized by their ability to operate autonomously or in conjunction with the main grid, offer flexibility, reliability, and efficiency in energy utilization [4]. Figure 1 illustrates the structure and functionality of an EMS.

However, the dynamic and stochastic nature of energy supply and demand in MGs poses significant challenges for effective energy management [5]. These issues are exacerbated by the necessity to guarantee system stability, economic efficiency, and environmental sustainability. The energy management problem currently faces the following difficulties.

- Managing MG energy is challenging due to the erratic nature of solar and wind generation, which depends on weather conditions, making precise supply-demand balancing difficult [6].
- Fluctuations from EV charging and distributed loads add to operational challenges.
- Conventional optimization relies on accurate system models, which struggle with non-linearities, uncertainties, and the dynamic nature of microgrids, limiting scalability [7]. Addressing these dynamic conditions while optimizing costs underscores the need for advanced energy management solutions.

Recent studies have explored innovative approaches to tackle challenges in renewable energy integration, energy management, and electricity markets. For example, researchers in [8] proposed a hybrid strategy integrating PV and solar PV systems into residential grids, achieving up to 40% cost savings and reducing grid stress through advanced energy optimization techniques. Similarly, energy consumption analyses of independent systems, such as those described in [9], highlight the importance of flexible and efficient monitoring mechanisms for optimizing energy use in agriculture and other sectors. Moreover, studies on electricity markets, such as [10], underscore the challenges posed by limited intraday fluidity and real-time price volatility, recommending structural reforms to improve market efficiency and support renewable energy integration.

In the domain of energy management in MGs, traditional approaches have predominantly relied on model-based methodologies. These methods typically involve developing a well-defined model that represents MG dynamics, a predictor to account for uncertainties, and a scheduling algorithm to optimize energy distribution and planning. Among such methods, Model Predictive Control (MPC), also referred to as rolling horizon optimization, has gained significant attention due to its capability to adjust control sequences dynamically and handle model uncertainties effectively. This adaptability

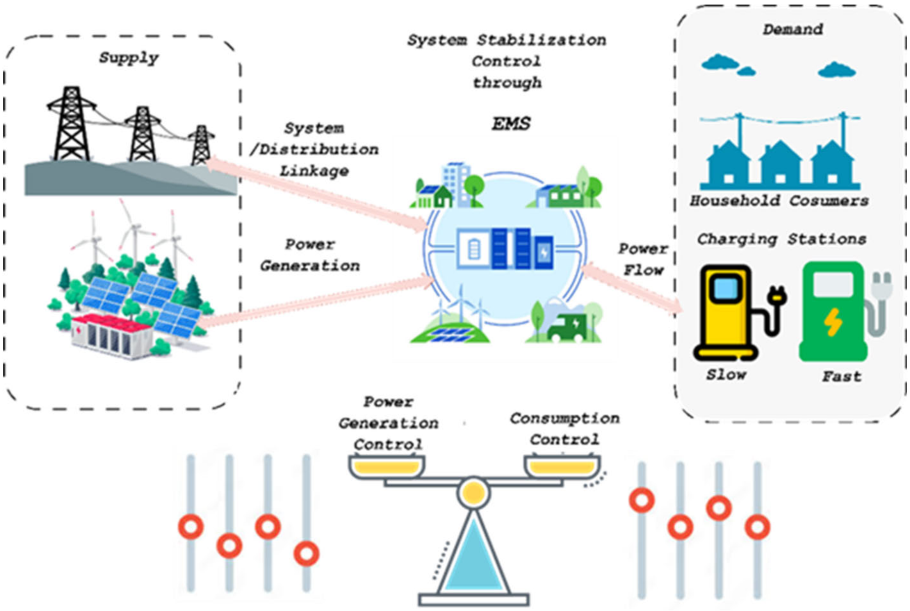

FIGURE 1. Energy management system for renewable integration and EVs charging.

is achieved by continuously refining the predictive model over time. For instance, researchers in [11] successfully implemented MPC for a hydrogen-powered MG to enhance the efficiency of generation schedules.

In another study [12], a robust framework employing rolling horizon optimization and weather forecasts was developed to improve MG operations, and two-step stochastic decision-making was employed for load demand and power cost [13]. Other techniques, such as convex forecast control [14], have optimized power flow in alternating current (AC) MGs with Energy Storage Systems ESSs. While model-based approaches, including those highlighted above, have demonstrated effectiveness, they exhibit several critical shortcomings:

These approaches rely heavily on detailed system modelling and precise parameter estimations. Inaccuracies in these models can reduce operational efficiency.

- The development, maintenance, and periodic updates of these models are resource-intensive. Adjusting model-driven systems to accommodate changing MG configurations or uncertainties can increase costs and complexity.
- Rapidly changing scenarios, such as fluctuations in renewable generation or load demands, are challenging to address in static models.
- Updating the framework, forecasting methods, and optimization strategies for evolving MG configurations is cumbersome and can be unsustainable in large-scale systems.

To overcome these limitations, learning-based approaches have gained prominence in recent years. Unlike model-based methods, these techniques do not require predefined system models. Instead, they utilize a data-driven approach, treating MG as a black box and learning nearly optimal strategies through iterative interactions. For example, batch RL was introduced in [15] to optimize battery scheduling in MGs. Other studies [16], [17] demonstrated RL's capability to manage resources effectively and minimize total costs. Multi-agent RL systems, such as the one proposed in [18], used multiple agents for advanced resource allocation in islanded MGs. However, these approaches often face challenges, such as the curse of dimensionality struggling to handle high-dimensional state spaces and uncertainties effectively.

In [19], a hybrid strategy employing a Neural Network (NN) with a Variable Fractional Order Proportional Integral-Derivative (VFOPID) controller improved frequency stability, resilience, and adaptability in isolated MGs with high-RES penetration.

While model-based techniques are effective, they have significant limitations. These methods rely on precise system models, increasing development and maintenance costs. Moreover, they often fail to adapt to dynamic and high-dimensional environments, such as those observed in modern MGs with RESs variability, load uncertainties, and battery degradation [20], [21].

Despite advancements, RL approaches face challenges related to dimensionality, struggling with high-dimensional state spaces and uncertainties. Nonetheless, RL has proven effective in addressing power system challenges like renewable energy output prediction, load demand forecasting [22], frequency regulation [23], and energy resource management [24]. Earning-based methods leverage data-driven strategies to extract models or policies from available data, reducing reliance on system models [25], [26]. This is accomplished through the adoption of a data-driven approach to implementation. In the context of MGs, addressing energy management issues using a learning-based approach typically involves formulating the problem using transition probability or MDP [27], [28], [29].

The MG energy management problem was addressed by [30] by introducing an RL approach, which utilizes a state representation with limited dimensions and a discrete set of actions. The study in [31] tackled the issue of unpredictability in an MG that incorporated PV systems and batteries by devising a Q-learning strategy. References [32] and [33] introduced a Q-learning method incorporating fuzzy control to enhance the economic efficiency of an MG equipped with an Energy Storage System ESSs.

In [34], RL with artificial neural networks (ANN) for frequency control in isolated AC MGs to ensure stability with the system's reliable performance even under challenging conditions, such as communication delays. The study in [35] presented an efficient RL technique for scheduling an MG. This study introduces a combination of the Monte Carlo tree search and RL method. Although power models are not required for this research on RESs, comprehending uncertainty distribution is still crucial.

MGs are based on continuously collecting and interpreting diverse high-dimensional data from sources and optimally managing the processes from power generation to power delivery to the end user. Therefore, integrating advanced metering infrastructure, control technology, and information and communication technology suitable for the current requirements into the grid is a very important issue. In contrast to traditional optimization methods, DRL acquires optimal policies through direct interaction with the environment, enhancing its adaptability to uncertainties like power cuts and changes in demand. For instance, [36] introduced a DRL-based real-time economic energy management technology for MGs, effectively dealing with the uncertainties related to RES and demand variability. Also, [37] has proposed a DRL-based optimal energy management structure for multi-energy MGs, showcasing its leverage in managing systems with substantial integrated RES. Consequently, the approaches of DRL in MGs have gained increasing popularity.

Despite these advancements, a significant research gap exists in the development of scalable and robust solutions for energy management in MGs. All the above-stated methods are considered model-based, requiring the formulation of a precise environmental model. Most model-based RL methods use deterministic dynamics models, which fail to account for stochastic variations in MG environments. MGs involve uncertainties from RESs (e.g., solar and wind generation), dynamic electricity prices, load demand, and battery degradation. Model-based RL often struggles to accurately model these uncertainties due to the complexity and non-linearity of the underlying processes. The application of model-free RL is employed to address the constraints mentioned above and resolve the MG optimal EM problem while taking into account uncertainties [38].

Additionally, MGs consist of multiple interconnected components, such as DERs, storage systems, and charging stations, resulting in high-dimensional state-action spaces. Model-based approaches often face scalability challenges in such complex environments [39]. Consequently, these approaches are impractical for real-world applications, leading to inefficiencies and limited scalability. This limitation underscores the need to develop more advanced and scalable solutions to manage complex, high-dimensional environments like MGs effectively.

The suggested model considers the inherent uncertainties in RESs, the varying demand at charging stations, fluctuations in charging load, and dynamic tariff rates. It does not rely on meteorological data or the precise modeling of RES output distributions. The energy management method is enhanced by utilizing action factors to find the most efficient allocation of power acquisition from various sources. The reward function is designed to minimize operational costs while complying with essential operational constraints, such as voltage and power limitations. An effective DDPG technique is proposed in this study to train the policy network. This network utilizes immediate feedback to generate continuous scheduling outcomes. The input and output layers of the network dynamically match the dimensions of the system's state and action spaces. This capability enables the network to effectively handle complex systems with high-dimensional state-activity representations. Such adaptability eliminates the need for manual adjustments, thereby facilitating the application of the algorithm to intricate systems. Based on the above discussion.

In summary, this study addresses the identified research gap by proposing a scalable and model-free RL-based solution for energy MG. Unlike traditional model-based approaches, the proposed method eliminates the need for precise system models, making it adaptable to uncertainties in RESs, load demands, and dynamic system conditions. The main contributions of this work are as follows:

- The study comprehensively analyzes different system components, including slow and FCSs and multiple DGs. The approach enhances operational efficiency by evaluating their interactions and impacts on system dynamics.
- The proposed approach utilizes an actor-critic network architecture, where the actor-network determines the optimal power acquisition from various sources, and the critic network evaluates the effectiveness of these decisions, ensuring continuous and robust performance in complex, high-dimensional environments.

In the subsequent section of the study, Section II, a comprehensive system overview is provided, including the state and action variables and the objectives and constraints. This is followed by a detailed explanation of the suggested strategy and a description of the proposed algorithm. Optimal Results are ultimately assessed in Section IV, which includes comparison results. Finally, Section V consists of the concluding remarks of the work.

II. SYSTEM DESCRIPTION

A. SYSTEM DESIGN

The grid-tied MG system is utilized to assess our proposed model-free real-time energy management, which incorporates several alternative energy sources. The MG, located in an urban setting, integrates RESs to harness sustainable energy and decrease reliance on the main power grid. Additionally, it has both slow and FCSs to minimize carbon emissions. The load profile of charging stations encompasses diverse energy consumption patterns, encompassing the demand for EVS charging. The test data combines variations in renewable energy generation, EVS charging patterns, and grid costs to simulate real-world conditions and uncertainties. This enables a comprehensive evaluation of the suggested energy management algorithm's capacity to optimize MG operations in dynamic and uncertain circumstances while assessing its resilience and efficacy.

The environment for deep reinforcement learning (DRL) is designed using an IEEE 33 bus distribution system integrated with DG such as PV, WT, etc. Household load demand and charging station demands are also considered for consumption. The schematic diagram of the modified IEEE 33 test system is presented in Figure 2. The test system consists of 33 buses with a total active power load of 3.72 MW and a reactive power load of 2.30 MVar. Power is supplied from a single source located at bus 1, also known as the infinite bus. Voltage limits for each bus are set between 0.9 (p.u.) and 1.1 (p.u.), with a base voltage of 12.66 kV, ensuring voltage levels remain within prescribed limits.

RESs are integrated as follows: a WT connected to a bus. WT_i= 12, four microturbines (MTs) connected at buses MT_i= 13, 17, 24, and 27, one PV system at bus PV_i= 21, and one fuel cell (FC) at bus FC_i= 29. Household loads and charging station loads are taken into consideration. Both fast and slow charging stations are connected to different buses. A total of ten charging stations, including slow and fast, are connected to buses. $E_i \in$ [2, 6, 10, 14, 19, 22, 23, 25, 29, 31]. The EMS determines DG and utility grid electricity purchases each hour.

The placement of DGs within the test system was determined based on renewable resource availability, load profiles, and system balance requirements, ensuring realistic modeling of interactions between DGs, loads, and charging stations. The specific placements are summarized in the Table below.

TABLE 1. Placement of DGs in the IEEE 33-bus system.

Bus Number	DGs Type	Rational for Placements
12	WT	High simulated wind resource availability, representing areas with suitable wind speeds.
21	PV	High solar irradiance profile, ensuring realistic PV generation potential.
13, 17, 24, 27	MTs	Strategic distribution to balance loads and ensure redundancy in generation.
29	FC	Placed near buses with critical demand to provide consistent energy supply.

These placements allow for a comprehensive evaluation of the energy management strategy, accounting for diverse scenarios and dynamic interactions, including the integration of slow and fast charging stations (FCSs) across buses. This study also uses PV and WT as DGs for electricity generation. Due to the random and uncertain nature of their power generation, accurate power models are challenging to obtain; instead of traditional prediction algorithms, historical data from [40] generate power patterns for PV, WT, and loads. The DGs, such as MTs and FC, are also put into the distribution system to ensure sufficient energy supply in cases where the power generation from DERs is not enough to meet the demand. The power output from MT and FC is denoted as P^{MT} and P^{FC} Respectively. Furthermore, the upper limit and lower limit of RES power output are denoted as $P^{MT}_{max,i}$, $P^{MT}_{min,i}$, $P^{FC}_{max,i}$ and $P^{FC}_{min,i}$, $P^{WT}_{max,i}$, $P^{,WT}_{min,i}$, $P^{PV}_{max,i}$, and $P^{PV}_{min,i}$ respectively. The value of power fetch from these sources will update every time based on load demand. We used the MATPOWER 7.0 function to calculate power flow each time, which is available in MATLAB.

B. PROBLEM FORMULATION

One of the primary goals of EMS is to reduce the amount of money spent on operations. Another objective we have considered is power loss reduction to improve the performance of the given system. MGs are becoming increasingly popular because they offer a more ecologically sustainable and resilient alternative to electricity. However, the intermittent nature of alternative sources and energy consumption makes the EM task more challenging. This includes minimizing the costs of the conventional sources associated with power exchange, as well as the operational costs of non-conventional sources.

1) OBJECTIVE FUNCTIONS FOR EM

The first objective can be mathematically represented as follows:

$$minimize\ cost = [COST^{DG\in\{PV,WT,FC,MT\}} + COST^{grid}] \tag{1}$$

The total cost incurred to fulfill the load demand at the current time step. This includes the extra cost incurred for purchasing extra power from the utility grid as well as from DG sources.

The $COST^{DG}$ is defined as the cost of DG, C^{pv} Is the cost function of PV panels, P^{pv} represents the power output of solar PV systems, C^{wind} represents the cost of WT, whereas P^{wind} Defines the output generation of WT. Similarly, C^{MT} and P^{MT}Represents the cost coefficient and Power of MT, respectively. C^{FC} represents the cost function of fuel cells, P^{FC} It is power generated from the fuel cell.

The upcoming section will provide the value of the cost coefficients, which represent the bidding cost of DGs. Multiplying the power of these sources by cost coefficients will

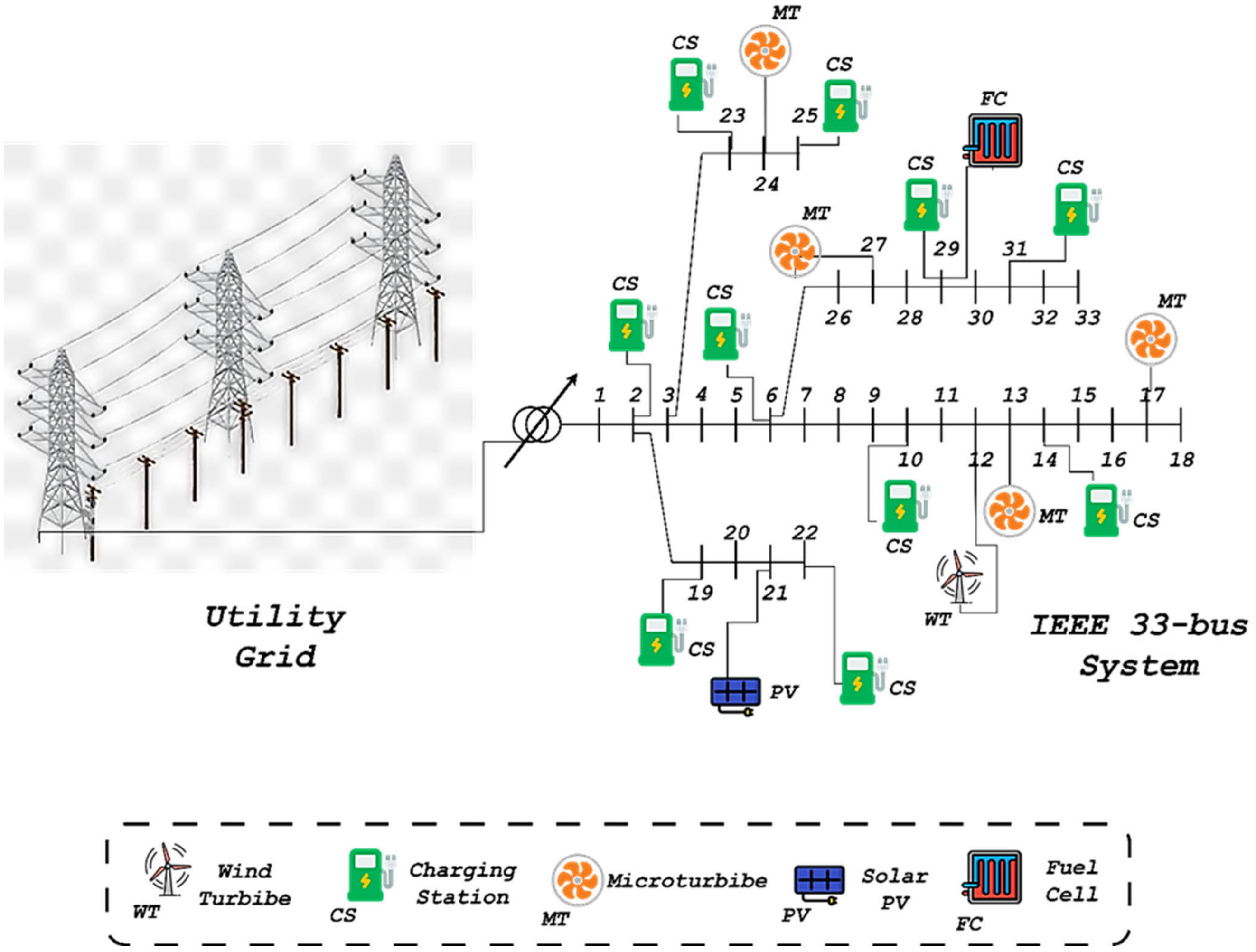

FIGURE 2. System environment and component details.

yield the cost of each DG.

$$\mathrm{COST}^{DG} = \left(C^{pv} \times P^{pv}\right) + \left(C^{\text{wind}} \times P^{\text{wind}}\right) + \left(C^{MT} \times P^{MT}\right) + \left(C^{FC} \times P^{FC}\right) \tag{2}$$

The utility grid power is denoted by P^{grid}. The MG cannot procure and sell power at the same time. The power output of the grid P^{grid}, must be within the range of P_{min}^{grid} *and* P_{max}^{grid}. The variable P_{max}^{grid} It is the maximum power that the MG may either purchase from or sell to the utility grid system. If P^{grid} Positive, then electricity is being purchased, and if P^{grid} Negative, then electricity is being sold.

The grid cost $COST^{grid}$ Calculated as follows:

$$COST^{grid} = \begin{cases} B^{grid}(t) \times P^{grid} \cdot \Delta t, & if\ P^{grid} \geq 0, \\ S^{grid}(t) \times P^{grid} \cdot \Delta t, & if\ P^{grid} \leq 0, \end{cases} \tag{3}$$

where $COST^{grid}$ is the biding price of electricity, $B^{grid}\,(t)$ The buying price associated with the active power during a period t.

The second objective function of our projected study is power loss reduction in each branch of the network. At the current time, this step is due to various factors such as inefficiencies in distribution lines, overloads, and system fluctuation. As stated in Section II-A, the calculation of power flow in the MG system is performed using the features supplied by the MATPOWER 7.0 function. Therefore, the power loss in each bus is directly calculated in the modified distribution system using this function. It is essential to minimize these losses to improve energy efficiency and save on expenses.

$$minimize\ \Delta P_{\text{Loss}} = \sum\nolimits_{j=1}^{n} I_j^2 \times R_j \tag{4}$$

where, ΔP_{Loss} It is the change in power loss. n the total number of branches in the distribution system, I_j the current flowing through the branch and R_j The resistance of the branch. We incorporate this function into our approach to improve the efficiency of the MG system.

Both the objectives given in Equations (1) and (4) must be taken into account for efficient allocation of resources and better system planning. The calculation of both the objective functions can be seen below:

$$minimize\ OB = \sum_{t=1}^{M} W_1 cost + W_2 \Delta P_{\text{Loss}} \qquad (5)$$

where, W_1 and W_2 represents the weights, which have equal values, meaning that both objectives are balanced and equally important for EM and grid stability.

2) CONSTRAINTS

The power balance constraint can be considered the most important constraint in the EM field. This constraint applies to all MG connections, regardless of whether they are grid-connected or independent. Below is an illustration of the power balance constraint.

$$\sum_{i}^{D} P_{it}^{DG} + P_{t}^{grid} = P_{t}^{demand} t = 1, 2, \cdots, T, \qquad (6)$$

where, P_t^{demand} It is defined as the total demand.

The system is presumed to have a total of Z distributed generators (DGs). The control variable for DG_i is represented as P_{it}^{DG}.

$$P_{it}^{DG} = P_t^{PV} + P_t^{WT} + P_t^{MT} + P_t^{FC} \qquad (7)$$

where the limits of P_{it}^{DG} It ranges thus: $P_{i,min}^{DG} \leq P_{it}^{DG} \leq P_{i,max}^{DG}$

Optimization should also consider operational constraints such as the voltage limit. This limit ensures the distribution system operates within safe operational boundaries and prevents voltage violations or equipment. Another constraint applies to voltage. Constraints guarantee that the distribution system functions within safe operational parameters and prevents voltage violations or equipment overloads.

$$V_{\min} \leq V_{j,t} \leq V_{\max}, \forall j = 1 \ldots N_{bus}, \forall t = 1 \ldots T \qquad (8)$$

Here, $V_{j,t}$ It is the voltage of the bus j at a time t, and the range of this voltage is 0.9-1.1 (p.u.).

C. MODELLING OF CHARGING STATION

Given the rise in demand for EVs, it becomes necessary to establish charging stations in many locations to alleviate concerns about limited driving range. This, therefore, further necessitates the incorporate of charging stations into the distribution system in order to tackle future planning difficulties. One challenge associated with charging stations is the uncertain amount of electricity they require, which adds to the overall demand for the distribution system alongside the needs of homeowners.

In order to efficiently control energy in the MG system that is connected to charging stations, it is essential to make an accurate prediction of the amount of electrical power that will be required by EVs connecting to stations on a daily basis [41]. This work utilizes a probabilistic approach to ascertain the arrival times of EVs and their State of Charge (SoC).

As suggested in reference [42], the times of arrival and the durations of parking for EVs are determined using two-parameter Weibull distributions. Upon arrival at a station, the State of Charge (SoC) of an EVs is set to 0.95 as the maximum SoC (SOC_{max}), and the desired SoC SOC_{desire}) is set to 0.9, as mentioned in references [43], [44]. Furthermore, the charging rates and battery capacities of FCSs vehicles are derived from the findings reported in [45].

The 24-hour EV load profile, as shown in Figure 3, is generated using vehicle driving patterns from the U.S. Department of Transportation's National Household Travel Survey (NHTS). Vehicles primarily travel during the day and typically charge at public/commercial charging stations between 9:00 a.m. and 5:00 p.m., with peak demand occurring in the early afternoon.

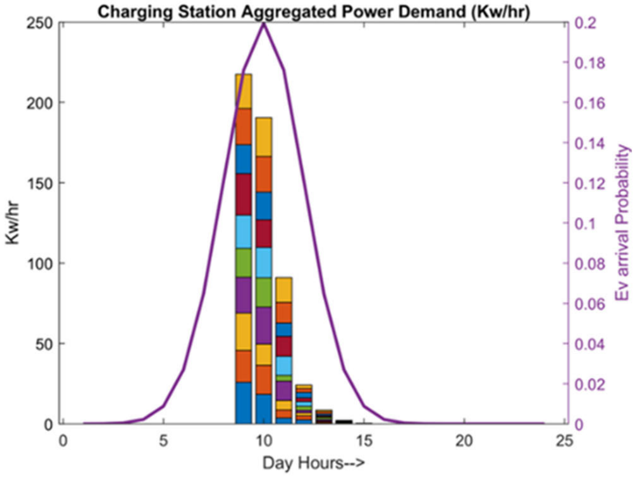

FIGURE 3. Aggregated power demand and EV arrival probability for charging.

The colored bars represent the aggregated power demand across ten charging stations, combining contributions from slow chargers (7.2 kW per EV) and FCs (50 kW per EV). The x-axis represents the 24-hour timeline, while the y-axis (left) indicates the aggregated power demand, capped at 250 kW, which is the maximum load capacity for all ten stations combined. Each charging station can accommodate up to ten EVs simultaneously when fully utilized. The purple curve shows the frequency of EV arrivals per hour, which follows a normal distribution derived from EV charging station statistics and vehicle driving patterns reported in the NHTS [46]. Therefore, we created a demand curve for all ten charging stations over a 24-hour period.

III. PROPOSED METHODOLOGY FOR SOLVING THE OPTIMAL ENERGY MANAGEMENT PROBLEM

This section presents the methodology for the proposed optimal energy management strategy in the MG. The methodology addresses the challenges of uncertainty in fast and slow charging stations, utilizing a model-free approach to handle increasing and unpredictable load demands. An overview of the method is provided in Figure 4.

Based on the MDPs, the proposed model-free method generates optimal decisions for energy management by

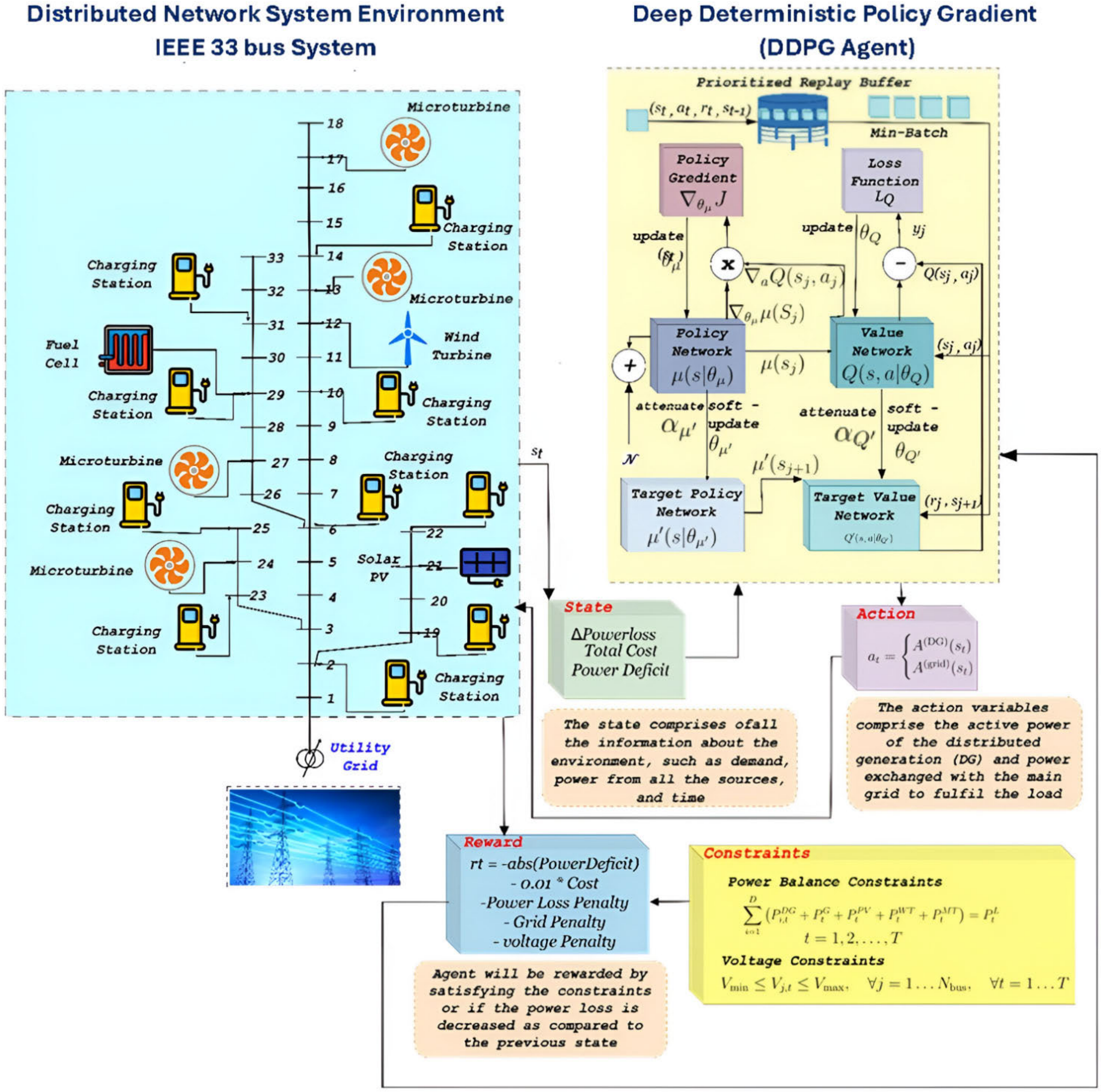

FIGURE 4. Overview of the proposed energy management methodology.

relying solely on the MG's current state. The optimal policy at times $t-1, t-2$, etc., ensures continuous updates to the solution. Conventional methods cannot effectively handle the complexity presented in this study because they often require discretizing the action and state spaces, leading to significant computational challenges. The exponential growth in the action and state spaces, referred to as the "curse of dimensionality," results in increased computation and complexity of storage requirements. These limitations render traditional approaches impractical for addressing the energy management problem in a modern MG environment.

Several heuristic methods, such as those presented in [47] and [48], can solve the optimal energy management problem. However, while these methods partially address the issue, they do not directly overcome the curse of dimensionality. Additionally, their exploration efficiency is significantly hindered by the increasing dimensionality of the search space. The model-free approach proposed in this study overcomes these limitations by leveraging RL techniques to explore the high-dimensional action and state spaces effectively; the pseudocode of the suggested methodology is provided in Table 2, offering a detailed step-by-step process for implementing the model-free strategy.

TABLE 2. Pseudocode of the proposed methodology.

```
Algorithm. 1: Pseudocode of the Proposed Methodology
Input: load IEEE 33 bus data, add charging stations $E_i$, $COST^{DG \in \{PV,WT,FC,MT\}}$, $COST^{Grid}$
Output: $P_t^{DG}$, $P_t^{grid}$
  1.  Generate charging station load profile.
  2.  Run power flow analysis to calculate power loss and system voltage
  3.  If  $total\ generation < P_t^{demand}$   # Check if energy management is required
        a.  Energy management is required
        b.  For energy management: train DDPG with a microgrid environment and define extra sources as action variables $a_t$ As given in eq.
            (10), and define state variable $s_t$ As cost, change in power loss, and power deficit represented as $cost, \Delta P_{\mathrm{Loss}}$, and $D_t$ Respectively.
        c.  For episode =1:M
                i.     Add the action $a_t$ at the $MT_i$, $PV_i$, $WT_i$, $FC_i$
                ii.    Run power flow analysis.
                iii.   Save the new set of observations through power flow analysis. $s_t \in [\Delta P_{\mathrm{Loss}}]$
                iv.    Buy the $D_t$ from the grid
                v.     Calculate cost and update the states as $s_t \in [\Delta P_{\mathrm{Loss}}, cost, D_t]$
                vi.    If load demand is matched as given in eq. (6),
                           1.  Give a reward to the agent $r_t^{P^{Demand}} = 10$
                vii.   else
                           2.  penalize the agent as given in eq. (14)
                viii.  End
                ix.    If the voltage limit is violated $V_{\min} > V_{j,t} > V_{\max}, \forall j = 1 \ldots N_{bus}, \forall t = 1 \ldots T$
                           1.  penalize the agent as given in eq. (14)
                x.     else
                           2.  Give a reward to the agent $r_t^{V_L} = 1$
                xi.    End
                xii.   If the power limit of DGs $P_{i,min}^{DG} \leq P_{it}^{DG} \leq P_{i,max}^{DG}$ as well as grid power limit $P_{min}^{grid} > P_t^{grid} > P_{max}^{grid}$ If it is violated,
                           1.  penalize the agent as given in eq. (14)
                xiii.  else
                           2.  Give a reward to the agent $r_t^{P_{it}^{DG}, P_t^{grid}} = 10$
                xiv.   End
                xv.    If  power loss is greater than the updated loss $P_{loss}^{cs} > P_{loss}^{updated}$
                           1.  Agent gets rewarded as given in eq. (13)
                xvi.   else
                           2.  penalize the agent as given in eq. (14)
                xvii.  End
                xviii. Calculate the reward $r_t$
                xix.   Send feedback to the agent
                xx.    Update the loss function of DDPG agent from eq. (18)
                xxi.   If episodes reach limit
                         Save the results
                xxii.  End
        d.  End
  4.  else
            No energy management is required
  5.  End
```

The methodology begins with loading standardized IEEE 33 bus system data, including the electrical network's topology, line, and bus parameters. It installed RESs $DG \in \{PV, WT, FC, MT\}$ on specific buses to support generation and increase load demand by adding fast and slow charging stations in the particular buses, as detailed in Section II-A. A probabilistic model generates the charging station load profile based on a Gaussian distribution to reflect the stochastic nature of charging demands. The MG is configured with ten EVs charging stations, integrating the generated load profiles. After adding load and generation, we perform power flow analysis using MATPOWER 7.1 to calculate the power losses in the transmission lines and the voltage levels at different buses. We also incorporate the power balance constraint to verify the fulfillment of demand. Step 3 of Algorithm 1 indicates $total\ generation < P_t^{demand}$, such that if the generation fails to fulfill the demand, the EM strategy becomes necessary; otherwise, it is unnecessary. An RL-based DDPG agent is trained to optimize power purchase from RES sources to achieve optimal energy management, as illustrated in Figure 5.

To train the agent, we define the action variables. a_t as given in Equation (10), as well as the state variables $s_t cost$, ΔP_{Loss}, and D_t. Subsequently, the agent fulfills the demand by taking appropriate measures. If demand is still unmet, the cost of purchasing power from the power grid is determined, and the agent is penalized or rewarded based on the constraints presented in Steps vi, xi, xii, and xv of Algorithm 1.

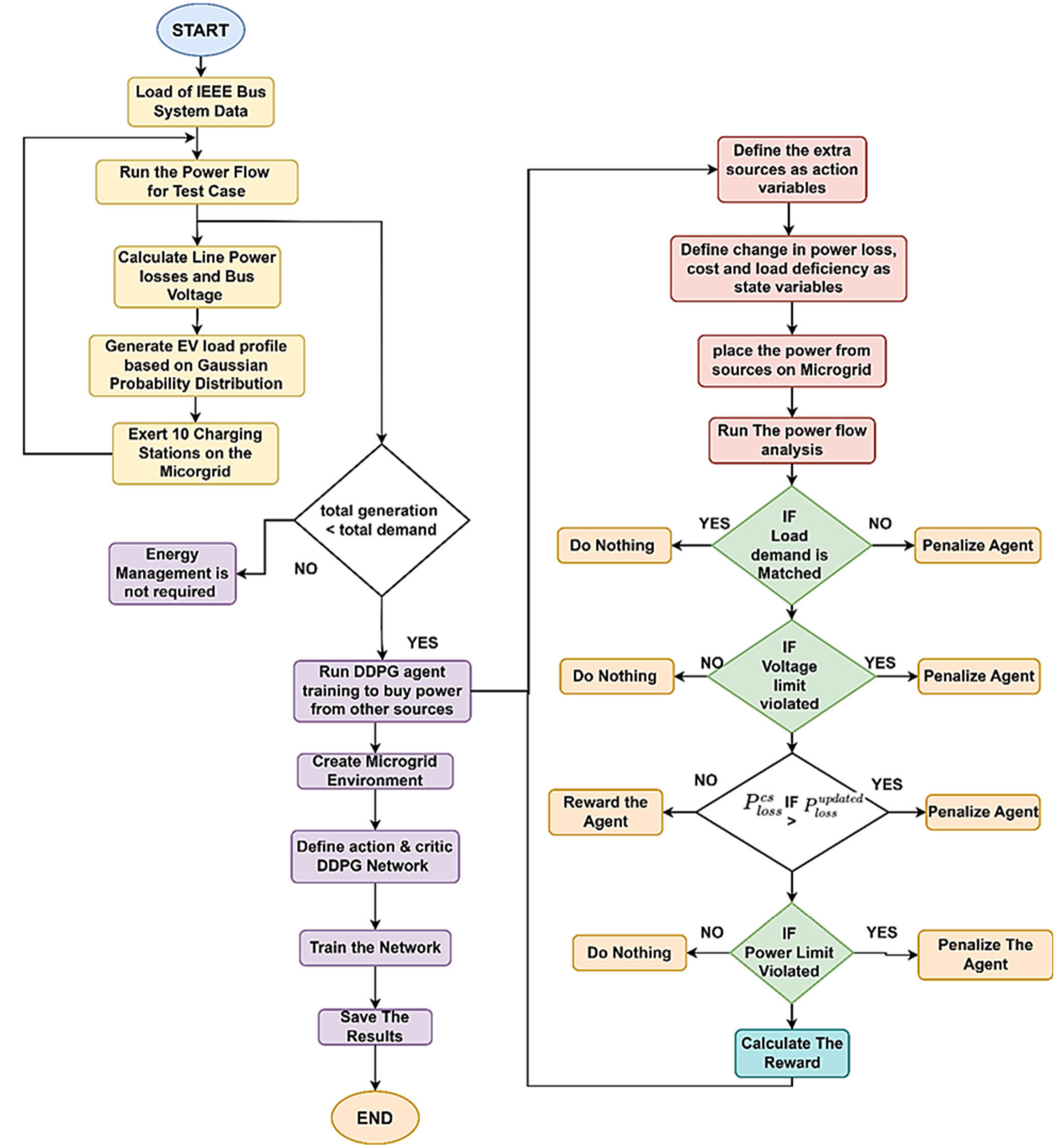
START
Load of IEEE Bus System Data
Run the Power Flow for Test Case
Calculate Line Power losses and Bus Voltage
Generate EV load profile based on Gaussian Probability Distribution
Exert 10 Charging Stations on the Micorgrid
total generation < total demand
NO
Energy Management is not required
YES
Run DDPG agent training to buy power from other sources
Create Microgrid Environment
Define action & critic DDPG Network
Train the Network
Save The Results
END
Define the extra sources as action variables
Define change in power loss, cost and load deficiency as state variables
place the power from sources on Microgrid
Run The power flow analysis
YES
IF Load demand is Matched
NO
Do Nothing
Penalize Agent
NO
IF Voltage limit violated
YES
Do Nothing
Penalize Agent
NO
P^{cs}_{loss} IF $P^{updated}_{loss}$ >
YES
Reward the Agent
Penalize Agent
NO
IF Power Limit Violated
YES
Do Nothing
Penalize The Agent
Calculate The Reward

FIGURE 5. Execution flow of the proposed model.

As a result, rewarding the agent will result in more positive feedback, preserving the solution, and updating the action in the event of a penalty. The process concludes with finalizing the optimal energy management strategy, and the results will be saved. This advanced strategy ensures the agent learns to make cost-effective and efficient decisions, enhancing the overall performance of the MG.

A. TRANSFORMING OPTIMAL EM PROBLEMS INTO REINFORCEMENT LEARNING FRAMEWORKS

In this context, we transform the optimum EM problem into an appropriate format for the RL problem. Within the confines of this investigation, the MG system functions as the environment, which is comprised of several components that generate and use power. Additionally, an agent is utilized to manage the output of DGs and the output from the external grid. An observation of the current state of the system is sent to the agent by the environment at the indicated time t. Both the control policy of the agent and the observable state of the system $s(t) \in$ S are taken into consideration when the agent generates action $u(t)$. In this study, the state $s(t)$ of the MG possesses the MDPs. This implies that the transition probability from $s(t-1)$ to $s(t)$ is solely connected to the action $u(t-1)$ and $s(t-1)$, which is outlined in the Equation (9).

$$P(s^*; s, u) = Ps(t) = s^*|s(t-1) = s, u(t-1) = u)\}, s, s^* \in S, a \in A \tag{9}$$

The MG system state comprises three key metrics: $s_t = [\Delta P_{\text{Loss}}, cost, D_t]$

- Total cost ($cost$) as given in Equation (1): This represents the total cost incurred to fulfill the demand at the current time step. Both the cost of obtaining power from the grid and the cost of acquiring electricity from DG sources are included.
- Change in power loss (ΔP_{Loss}) s given in Equation (4): This represents the amount of power lost within the system during the current time step due to various factors, such as inefficiencies in distribution lines, overloads, and system fluctuations.
- Power Deficit ($\mathrm{D_t}$): This represents the discrepancy between the amount of power required and the available power source at this time step. A positive value indicates a deficit (demand exceeding supply), while a negative value suggests a surplus (supply exceeding demand).

Actions are considered the primary outcomes of an agent's decision-making process, and they can directly impact the agent's interaction with and influence the surrounding environment. The subsequent state of an agent and the benefits or penalties it receives from the environment are contingent upon the activity it takes. This represents the discrepancy between the amount of power required and the power supplied. The action variables are the distributed generation's (DG) active Power and the Power transferred to the main grid to meet the load.

$$A_t = \begin{cases} A^{DG}(s_t) \\ A^{grid}(s_t) \end{cases} \tag{10}$$

where $A^{DG} = P_{it}^{DG} = P_t^{PV} + P_t^{WT} + P_t^{MT} + P_t^{FC}$ and $A^{grid} = P_t^{grid}$, For each time step.

The total reward for a single simulation, starting at time t with state $s(t)$, is computed as

$$R(s(t), t) = \sum\nolimits_{k=0}^{M-t} \gamma^k r(t+k), \tag{11}$$

The attenuation coefficient [0, 1] represents the degree of reduction in value. The reward for transitioning from state $s(t+k-1)$ to $s(t+k)$ with action $a(t+k-1)$is denoted as $r(t+k)$. Then, the value function is defined as:

$$V^\pi(s(t)) = E[R(s(t), t)|s(t) = s] \tag{12}$$

The agent should discover the best regulation strategy. π^* to optimize $V(s(0), 0)$. The agent responsible for managing power distribution is rewarded based on its ability to satisfy specific constraints and reduce power losses compared to the previous state.

$$r_t = -|D_t| - 0.01 \times Cost_t - Penalties \tag{13}$$

$$Penalties = -Power\ Loss\ Penalty_t - Grid\ Penalty_t - Voltage\ Penalty_t \tag{14}$$

In order to encourage effective and consistent power management, the reward function is made to incentivize the RL agent so as to eliminate power deficits, reduce costs, and adhere to grid and voltage constraints. The penalty is defined on the basis of power loss, which is described in Equation (3). The agent will receive a penalty if the power loss increases from the previous state. The grid penalty is defined as the agent incurring the power from the grid within a limit, defined as $P_{min}^{grid} \leq P^{grid} \leq P_{max}^{grid}$. The voltage penalty is defined as the agent's decision to take actions in which the system's voltage should be within its limit. The agent will receive a penalty if the voltage limit is violated, as shown in Figure 3. The Equation and the fundamental limit are shown below:

Therefore, the subsequent connection is estimated as

$$V^\pi(s(0), 0) = E[OB + Penalties] \tag{15}$$

B. POLICY GRADIENT ALGORITHM FOR OPTIMAL DECISION

This methodology optimizes the operation of an MG network using a DDPG agent within a 33-bus distribution network that integrates various distributed energy resources, including MTs, fuel cells, WT, solar PV panels, and multiple charging stations. The network's diverse power generation and storage configuration ensures a robust and flexible energy distribution system, aiming to minimize power losses, costs, and power deficits while maintaining grid stability and adhering to voltage constraints.

The DDPG agent architecture includes a prioritized replay buffer, policy network $\mu(s|\theta_\mu)$, value network $Q(s, a|\theta_Q)$, and target networks, all working together to derive optimal control policies. The agent operates based on states characterized by power losses, total cost, and power deficit, making decisions regarding power generation and grid control actions. Rewards are determined by power deficit, cost, power loss penalties, grid penalties, and voltage penalties, while constraints ensure power balance and voltage levels within specified bounds. The primary objectives are to minimize absolute power deficits, reduce operational costs, penalize excessive power losses and grid imbalances, and maintain voltage levels within acceptable limits to ensure system stability.

Algorithm 2 in Table 3 provides a pseudocode of the DDPG-based approach, outlining the actor and critic networks' initialization, training, and optimization steps for achieving optimal control policies.

TABLE 3. Pseudocode for the DDPG-based policy gradient algorithm.

```
Algorithm 2.
Initialize the main actor $\mu(s|\theta^{\mu})$ and critic $Q(s,a|\theta^{Q})$ network
Set the parameters of the target actor-network $\mu'$and critic network $Q'$
Set up the replay buffer $\mathbb{D}$.
For each episode =1:Z do
  Set up $s_1$
  Set up an arbitrary procedure $\mathbb{N}$ for finding the best action
  For t=1:T do
    If each episode< $K$, then
      Randomly choose the action $a_t$ As given in eq. (10) From the uniform distribution
      Carry out the selected action $a_t$ And observe the reward. $r_t$ As given in eq. (13) and transfer to the next state $s_{t+1}$
      Save the change $(a_t, s_t, r_t, s_{t+1}$ )in reply, buffer $\mathbb{D}$
    Else
      Select action $a_t$using the main actor policy $\mu(s|\theta^{\mu})$ with added exploration noise $\tau * \mathbb{N}_t$
      Implement the $a_t$ and observe $r_t$ As given in eq. (13) and transfer to the next state $s_{t+1}$
      Save the change $(a_t, s_1, r_t, s_{t+1}$ )in reply, buffer $\mathbb{D}$
      Make a batch of the changes called as minibatch $\mathbb{F} = \{(a_j, s_j, r_j, s_{j+1})\}_{j=1}^{\mathbb{F}}$ from reply buffer $\mathbb{D}$
      Calculate the target action value $y_j = r_j + \gamma Q'(s_{j+1}, \mu'(s, a|\theta^{\mu'})\theta^{Q'})$
      Update the main critic by minimizing the loss function as given in eq. (18)
      Update the main actor-network with policy gradient
      Update the target networks
    End if
  End for
End for
```

The following Sections will detail the implementation of the DDPG agent, the training process, and the evaluation metrics used to assess its performance in managing the distributed network system effectively. A detailed overview of the proposed methodology can be found in Figure 4.

Then, the action of the current state s_t is evaluated using the action value function $Q^{\pi}(s, a)$.

$$V^{\pi}(s(t)) = Q^{\pi}(s, a) \tag{16}$$

$$Q^{\pi}(s, a) = E_{\pi}\left[\sum_{k=0}^{K} \gamma^{k} * r_{t+k} \middle| s_t = s; a_t = a\right] \tag{17}$$

In this context, the symbol K represents the number of future time steps taken into account, whereas γ represents a discount factor that determines the balance between immediate and future rewards. Here, the value of γ is set to 1, indicating that DG and grid benefits are considered equally important.

The goal of the RL issue is to determine an optimal policy π^* that maximizes the cumulative rewards obtained. The DDPG approach [49] has been suggested in order to handle the high demand that was caused by FCS charging stations and the voltage profile problem that was associated with high-dimensional and continuous action spaces. This was done by utilizing only low-dimensional data [49]. The DDPG method is a DRL algorithm that utilizes a policy-based approach and employs an actor-critic design. Figure 4 illustrates the process of constructing the DDPG algorithm. The distribution network environment is manipulated by a learning agent in the DDPG in discrete timesteps, much like in traditional RL. Within this framework, the actor and critic components play crucial roles. The actor parameterized as $\mu\,(s\,|\,\theta^{\mu})$, determines actions based on states deterministically, while the critic $Q(s, a)$ evaluates these actions using the Bellman equation akin to Q-learning principles.

The overall performance of the critic is updated in accordance with a loss function that has been explicitly defined. This integration enables the DDPG algorithm to incorporate EVs load profiles and facilitate efficient power purchases from external sources. The update rule for the parameters of the critic is determined by the loss function, as stated in the study [50].

$$L\left(\theta^{Q}\right) = \frac{1}{B}\sum\nolimits_i \left(y_i - Q\left(s_i, a_i\,|\,\theta^{Q}\right)\right)^2 \tag{18}$$

As can be seen in Figure 6, the actor and critic work together to learn the optimal policy for a DDPG algorithm. As observed from the MG environment in Section II-A, the state space is continuous in nature. The optimization is designed for 24 hours, and we need to generate the optimal results for each. In this proposed algorithm, the state s_t= $[\Delta P_{\text{Loss}}, cost, D_t]$ is given to the actor, and the actor has to determine the best action ($A^{DG}(s_t)$ and $A^{grid}(s_t)$) using policy function $\mu\,(s\,|\,\theta^{\mu})$. Power from various MG sources feeds into power flow analysis to evaluate load demand matching, voltage limit violations, and power loss reduction. The agent receives rewards or penalties based on these evaluations.

The actor function $\mu\,(s\,|\,\theta^{\mu})$, parameterized to assign actions deterministically to states, specifies the current policy. Actor value networks update via policy gradients using gradient ascent and decision sampling sequences [50]. For

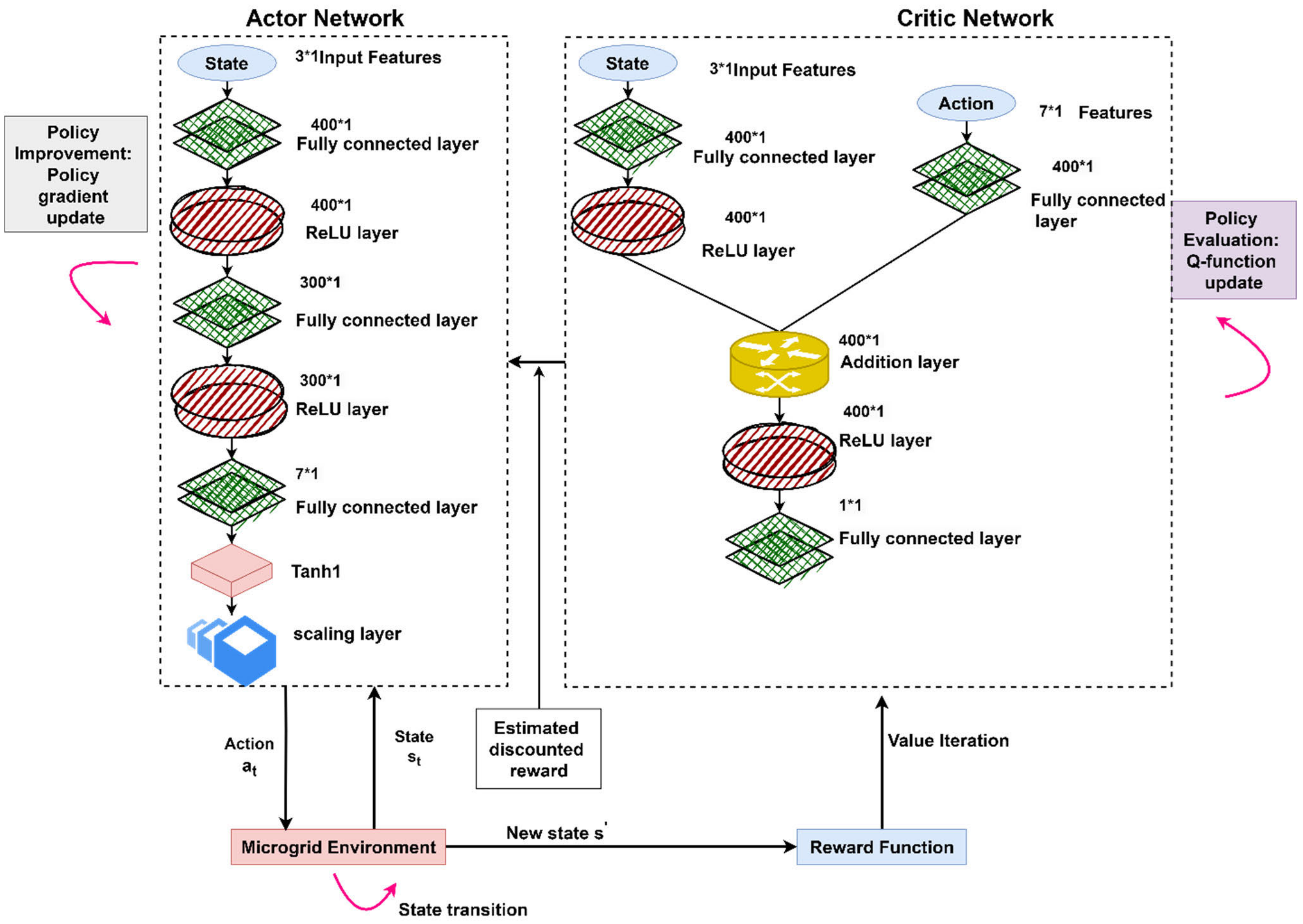

FIGURE 6. The structure of actor- critic network in the DDPG framework.

the training of the actor and critic, the different activation functions are applied in the network architecture, as can be seen in Figure 6.

C. NETWORK ARCHITECTURE

The proposed DDPG model comprises an actor-network and a critic network, which form its architecture. The actor-network determines the power procured from the DGs and the utility grid, referred to as the action.a_t. The critic network evaluates the performance of these actions by producing the action-value $Q^{\pi}(s, a)$, which represents the expected cumulative reward. The overall architecture of the actor and critic networks is illustrated in Figure 6.

1) ACTOR-NETWORK

The actor-network plays a central role in selecting the optimal course of action based on the system's current state. The network architecture comprises an input layer, multiple fully connected layers, activation functions, and a scaling layer. It begins with a feature input layer, designated as the "state," which represents the system's current state. This input layer has dimensions of 3 × 1, capturing three distinct features as described in Section III-A.

The network effectively addresses the energy management (EM) problem by identifying valuable features from the input data, such as the power available from various sources, to meet the load demand. The goal is to optimize the anticipated reward through interactions with the energy management environment.

The first fully connected layer (FC1) processes the input layer with weights of 400 × 3 and bias of 400 × 1, resulting in an output of 400 × 1. A ReLU activation function is applied to introduce non-linearity, enabling the network to capture complex patterns effectively. The second fully connected layer (FC2) further processes the output of the previous layer with weights of 300 × 400 and a bias of 300 × 1, producing an output of 300 × 1. Again, a ReLU activation function is applied to enhance the network's representation capability.

The final fully connected layer (FC3) generates the output with weights of 7 × 300 and a bias of 7 × 1, producing an output of A hyperbolic tangent(tanh) activation function is applied at this stage to constrain the output values within

[−1,1], essential for maintaining the action outputs within the required range. Finally, a scaling layer normalizes the output actions, ensuring compatibility with the energy management environment.

The actor network's detailed architecture is shown in Figure 6 (left side) and summarized in Table 4.

TABLE 4. Description of different layers of actor network.

Name	Type of Layer	Activations	Learnable Sizes
State	Input	3 × 1	-
FC1	Fully connected	400 × 1	weights 400 × 3 bias 400 × 1
ReLU1	ReLU	400 × 1	-
FC2	Fully connected	300 × 1	Weights 300 × 400 Bias 300 × 1
ReLU2	ReLU	300 × 1	-
FC3	Fully connected	7 × 1	Weights 7 × 300 Bias 7 × 1
Tanh1	Hyperbolic tangent	7 × 1	-
Scale	Scaling	7 × 1	-

2) CRITIC NETWORK

The critic network evaluates the actions proposed by the actor-network by predicting the Q-value for each state-action pair. This Q-value represents the expected reward, considering power loss, voltage limit violations, and power limit violations.

The critic network begins with two separate input layers. The state input layer has dimensions of 3 × 1, capturing the current state of the energy system, including variables such as power loss, cost, and power deficit, as detailed in Section III-A. The action input layer has dimensions of 7 × 1, representing the actions generated by the actor-network, such as decisions regarding energy allocation or matching demand.

The fully connected layers begin with FC1, which features 400 × 1 activations and processes the state input with 400×(3+1) weights to account for the state dimensions and biases. The next layer, FC2, also has $400(C) \times 1(B)$ activations and integrates processed state and action information. Its size is $400 \times (400 + 7)$, which combines features from state and action input. The final fully connected layer, FC3, has $1(C) \times 1(B)$ activation, reducing the output to a single scalar value representing the Q-value. A linear activation function is applied at this stage to ensure the Q-value remains continuous.

ReLU layers are employed as activation functions between fully connected layers to introduce non-linearity, as previously discussed in the actor-network. These layers are known for their simplicity and efficiency in handling gradients. Additionally, the element-wise addition layer combines the processed state and action inputs after their respective processing through earlier layers, facilitating effective integration.

The critic network architecture can be seen in Figure 6 (right side) and summarized in Table 5.

TABLE 5. Description of different layers of the critic network.

Name	Type of Layer	Activations	Learnable Sizes
State	Input	3 × 1	-
FC1	Fully connected	400 × 1	weights 400 × 3 bias 400 × 1
ReLU1	ReLU	400 × 1	-
Action	Features input	7 × 1	-
Add	Element-wise Addition (2 inputs)	400 × 1	-
FC2	Fully connected	400 × 1	Weights 400 × 7 Bias 400 × 1
ReLU2	ReLU	400 × 1	
FC3	Fully connected	1 × 1	Weights 1 × 400 Bias 1× 1

This architecture achieves stable convergence during training while avoiding overfitting and unnecessary computational overhead. A balance between performance and computational efficiency guided the design of the actor and critic networks.

IV. RESULTS

This section evaluates the efficacy of the suggested DDPG scheme for the energy management technique for the MG system. The modified IEEE 33 bus test system is used to implement the proposed approach given in [51]. The different DGs are installed on different buses, as shown in Figure 4.

This study used an experimental configuration of 10 charging stations and 10 EVs. The battery capacities range from 18.8 kWh to 98 kWh, while the charging rate varies from 6.6 kW/hr. To 22 kW/hr. The cost calculation is done using the bidding cost of these DGs, which is given in Table 6. This Table also includes the bidding costs for electricity acquired from the main grid.

TABLE 6. Parameters of MG component [11].

Type	Quantity	Maximum power (kW)	Minimum Power (kW)	Bids (€ct/kWh)
MT	4	30	6	0.457
FC	1	30	3	0.294
PV	1	25	0	2.587
WT	1	15	0	1.073
Main grid	1	30	-30	0.83

PV and WT systems are unreliable sources of electricity generation, making them unsuitable for meeting the energy needs of a household. Therefore, it is imperative to procure additional power to fulfill any electricity requirements. Any unfulfilled demand can be fulfilled using MTs and main grid power. All the MTs possess an identical capacity of 30 KW.

Table 7 provides the uncertain data for WT and PV systems. This data is related to the preprocessing steps of the experimental configuration.

The results of its implementation are subsequently examined and deliberated. This study demonstrates the application

TABLE 7. Data on wind velocity, solar irradiance, temperature, and humidity [11].

Hour	Wind Speed (m/s)	Solar Radiation (kW/m²)	Temperature (°F)	Humidity (%)
1	13	0	1.04	88
2	9.3	0	1.4	88
3	11.1	0	1.94	88
4	13	0	-0.04	86
5	9.3	0	-0.04	88
6	11.1	0	-0.58	86
7	9.3	0.0351	0.5	88
8	11.1	0.1275	0.68	87
9	9.3	0.2196	1.04	86
10	11.1	0.2841	1.04	86
11	11.1	0.3169	1.04	86
12	13	0.2826	1.4	87
13	11.1	0.2107	1.94	86
14	14.8	0.1116	2.66	87
15	13	0.0199	3.2	88
16	14.8	0	4.1	89
17	14.8	0	4.82	88
18	13	0	5	88
19	14.8	0	5.9	89
20	14.8	0	6.8	88
21	13	0	7.7	89
22	11.1	0	8.6	89
23	11.1	0	9.32	89
24	9.3	0	9.86	90

of the DDPG for managing energy in MGs, including renewable energy sources. The technical and cost parameters of the MG have been derived from reference [11]. The research did not prioritize the planning of the microgrid, which encompasses aspects such as the design, components, specs, capacity, and size. In contrast, this study explicitly focuses on the optimization of energy management, which entails reducing the desired cost function by generating energy from each individual component.

Figure 7 shows a simulation of the DDPG algorithm used to train the actor-critic network. The attenuation coefficient was set to 0.9 to determine the best possible actions; the proposed approach was trained for 1,000 episodes, with 24 time steps in each episode. Table 8 presents the hyperparameters utilized for the DDPG and actor-critic networks.

At the beginning of training, the DDPG algorithm randomly explored the environment's action space. The agent perceives the state. $s_t = [\Delta P_{\text{Loss}}, cost, D_t]$, earns from the dynamic changes to the MG and evaluates feedback from its actions. Initial oscillations are observed due to the trial-and-error activity and learning process as the agent seeks to identify the optimal solution within the environment. After this initial decrease, there is a substantial increase in rewards after 150 episodes.

As training progresses, the agent learns effective methods for meeting energy needs while avoiding constraint violations. This increase in rewards indicates that the agent has discovered new policies that minimize penalties and, consequently, maximize rewards. The stability observed between 200 and 850 episodes demonstrates that the agent has optimized its policies for effective energy management while adhering to constraints. However, a reduction in rewards is observed after 850 episodes, potentially due to factors such as changes in the exploration-exploitation balance or the dynamics of the environment. Subsequently, the agent recovers and stabilizes, indicating that it either returned to previously learned policies or successfully adapted to new conditions while avoiding constraint violations.

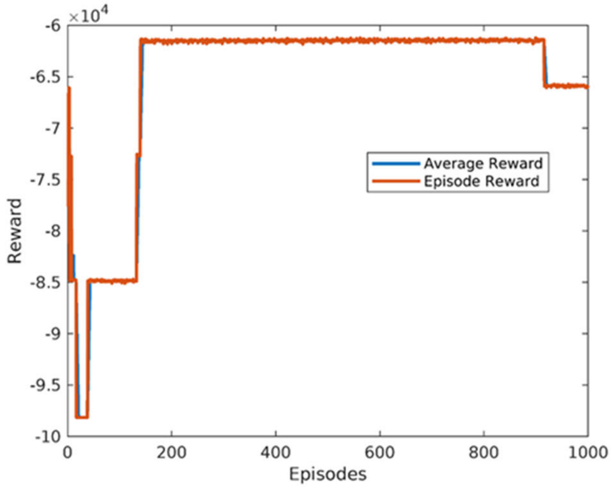

FIGURE 7. Training graph of DDPG.

TABLE 8. The hyperparameters of DDPG.

Hyperparameters	Value
Sample Time	1
Target Smooth Factor	10^{-3}
Discount Factor	0.99
Mini Batch Size	64
Experience Buffer Length	10^{-6}
Actor learning rate	0.01
Critic learning rate	0.01

A. OPTIMAL ENERGY MANAGEMENT RESULTS

In this section, we evaluate the performance of a well-trained DDPG algorithm for optimal energy management. The primary challenge considered in this study is the sudden increase in demand caused by FCSs.

The FCSs are integrated into the IEEE 33 bus system, with the charging stations located on buses. $E_i \in [2, 22, 28, 26, 5, 19, 21, 11, 8, 32]$. As mentioned in Section II-C, vehicles primarily travel on highways and often recharge at public or commercial charging stations between 9:00 a.m. and 5:00 p.m. The combined demand for slow and fast charging stations placed at different buses can be seen in Table 9.

TABLE 9. The load demand of charging stations.

Charging stations placed					Time				
at different buses	9 a.m.	10 a.m.	11 a.m.	12 a.m.	13 a.m.	14 a.m.	15 a.m.	16 a.m.	17 a.m.
CS_3 load demand (kW)	19.5924	22.2011	12.6215	4.4523	1.1916	0	0	0	0
CS_6 load demand (kW)	25.5423	12.1079	2.6053	0.8953	0.4792	0	0	0	0
CS_{10} load demand (kW)	24.3453	23.1985	3.4854	0.7985	0.4274	0	0	0	0
CS_{14} load demand (kW)	24.0109	10.7515	9.4882	2.3955	1.2822	0.5345	0.1157	0.0146	0
CS_{19} load demand (kW)	19.4516	13.2648	4.9289	1.6938	0.9066	0.1998	0.0649	0.0164	0
CS_{22} load demand (kW)	26.6689	25.8315	7.9391	0.7985	0.4274	0.1782	0	0	0
CS_{23} load demand (kW)	26.6689	30.2199	9.8754	2.1293	0.4274	0.1782	0.0578	0	0
CS_{25} load demand (kW)	20.2790	22.9791	15.4381	4.6216	0.4792	0.1998	0	0	0
CS_{29} load demand (kW)	20.4726	18.8101	10.5972	3.7263	0.8548	0.1782	0.0578	0.0146	0
CS_{31} load demand (kW)	24.1517	22.9791	8.4672	2.2261	0.7123	0.2970	0	0	0

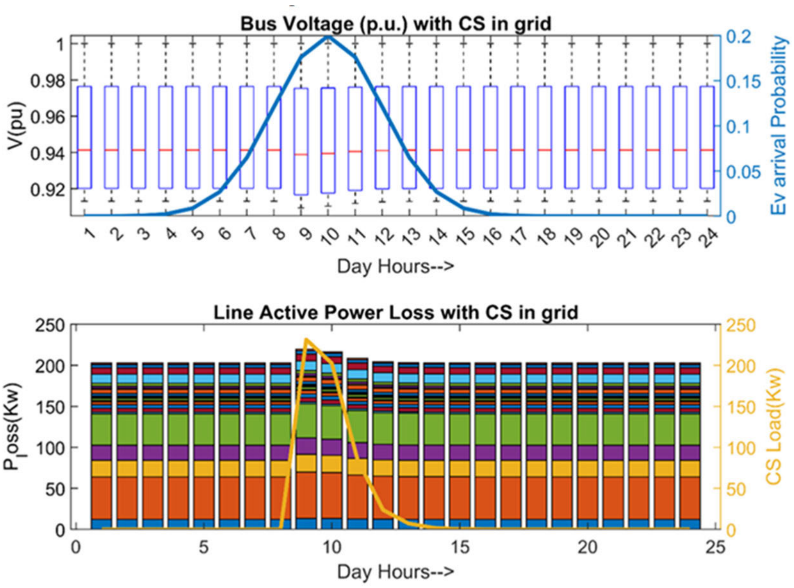

FIGURE 8. Effect of high charging demand on bus voltage and active power.

The highest demand for charging occurs during the early afternoon, as shown in Figure 8. This increased demand must be managed effectively by the actions of the DDPG agent.

The top graph of Figure 8 illustrates the voltage profile for the 33 buses, represented using box plots. Each box plot reflects the voltage range of all buses at a given hour. The red dashed line highlights voltage disruptions during peak hours, while the blue curve represents the EV arrival probability, showing a correlation between increased charging demand and the drop in bus voltage. Between 9:00 a.m. and 11:00 a.m., a noticeable reduction in voltage is observed, caused by the high energy demand at charging stations. The reduction in voltage is primarily observed between 9:00 a.m. and 11:00 a.m., coinciding with a significant increase in power loss, as illustrated in the bottom graph of Figure 8. Each bar in the graph represents the total power demand for the ten charging stations, and the fluctuations in bar heights reflect increased power loss caused by the surge in energy demand during peak hours. The yellow curve highlights the demand distribution, peaking between 9:00 a.m. and 11:00 a.m.

The top graph of Figure 9 illustrates the combined power loss of all the charging stations for 24 hours. The figure below represents the effectiveness of integrating RES in meeting the load demand, significantly reducing power loss within the system. Hence, the DDPG demonstrated significant improvements in the effective energy management in distribution systems, including the continuous action space environment. A justification for the reduction of power loss can be found in Figure 9, and as a result, the voltage profile has been greatly enhanced.

As discussed in Section III, the action space variable includes power from various, such as (PV, WT, FC, and MTs), in addition to the main grid, to fulfill the additional demand. The simulation spans 24 hours, requiring continuous action throughout the day. Policy-based methods, such as DQN, are less stable and exhibit poor convergence compared to value-based methods due to the approximation of the Q-value

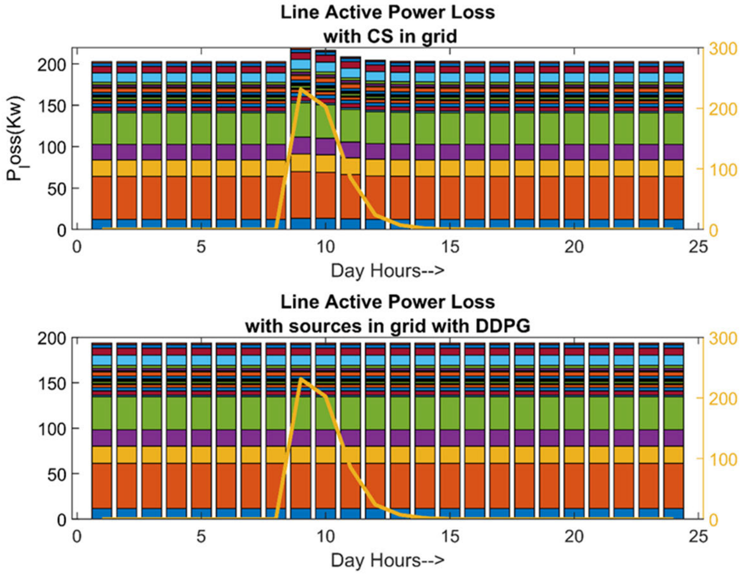

FIGURE 9. Displays the aggregate power loss, comparing the DDPG with base Case.

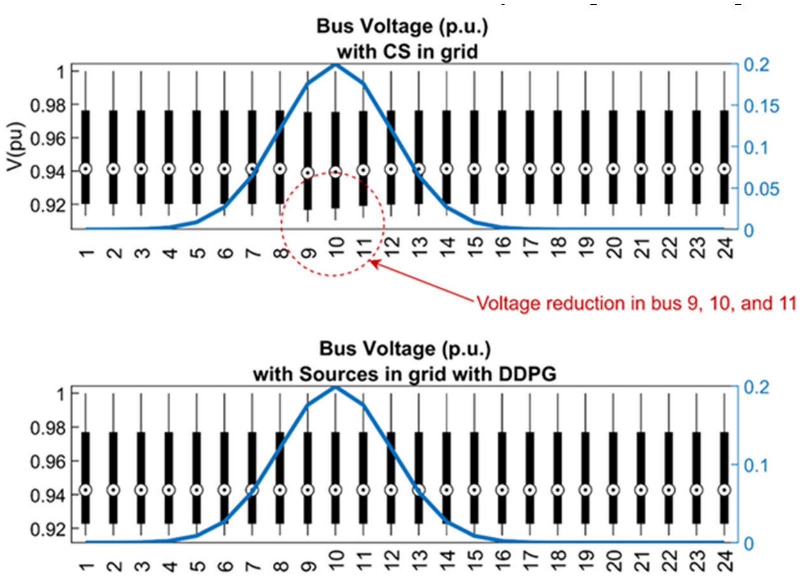

FIGURE 10. Displays the aggregate voltage profile of 33 buses, comparing the DDPG with base case.

function. Q^{π} (s, a) However, value-based methods are prone to suboptimal solutions in high-dimensional action spaces.

To overcome these challenges, this study employs an actor-critic model within the DDPG framework, leveraging the strengths of both approaches. The DDPG agent consists of two deep neural networks: the actor and the critic. The actor generates actions (a_t) based on the current state and policy (π). At the same time, the critic evaluates these actions and provides feedback in the form of a reward (r). The critic evaluates the actor's performance by assigning a numerical score to its actions. Based on this feedback, the actor adjusts its policies to improve future actions [52].

The voltage profile in the IEEE 33-bus system is significantly improved when the load demand caused by the FCSs is fulfilled through the optimal use of available generating sources. The bus voltage for each bus is calculated using power flow analysis, and Figure 10 illustrates these improvements.

The bottom graph demonstrates that, with the integration of RES managed by the DDPG approach, all bus voltages remain within the acceptable range of 0.92–1.0 (p.u.), ensuring system stability and efficiency during peak hours.

The voltage range of 0.92–1.0 (p.u.) represents normal operating conditions within acceptable regulatory limits for the microgrid. The observed voltage dips during peak EV charging hours are due to increased load demand and do not indicate faulty conditions, such as short circuits. These variations highlight the system's dynamic behaviors under-load changes, effectively managed by the proposed DDPG approach.

We have two objective functions: minimizing the total costs incurred to fulfill the demand and reducing power losses in the system. As shown in Figure 9, RES integration significantly reduces power loss, resulting in an enhanced voltage profile. Figure 10 demonstrates that the voltage profile achieved through DDPG control variables exhibits higher voltage magnitudes per unit (p.u.) compared to the base case across all buses. Additionally, the voltage profiles of all nodes remain within the prescribed band limits, ensuring system stability. Maintaining a stable voltage profile in the distribution system effectively minimizes power losses and enhances overall system efficiency.

B. COMPARISON RESULTS

This section presents a detailed comparison of the proposed DDPG algorithm with the DQN and Dueling DQN algorithms. The comparison is performed based on training rewards and cost performance.

1) TRAINING PERFORMANCE

The DQN and Dueling DQN algorithms utilize value-based techniques, where a deep neural network (DNN) approximates the Q-value function to evaluate policy performance. While DQN is well-suited for discrete action spaces, it requires discretization to accommodate continuous action spaces, which can reduce accuracy and increase computational complexity.

Figure 11 illustrates the training progression of DDPG, Dueling DQN, and DQN over 1000 episodes.

As shown in Figure 11, the DQN algorithm converges early but achieves the lowest reward value among the three algorithms. This behavior is attributed to its limited exploration capacity in high-dimensional continuous action spaces.

The Dueling DQN, an enhanced version of DQN, independently estimates the state value $V(s)$ and the advantage function $A(s, a)$, which are combined to obtain the Q-value. This architecture improves the ability to distinguish between the value of states and the relative benefit of actions, leading to faster learning and higher rewards than DQN. As a result, Dueling DQN converges earlier than DDPG but still falls short in reward value.

For additional clarity, DQN operates as a value-based RL method where a deep neural network approximates the Q-value function. DQN is limited by its dependence on discretization, making it less effective in high-dimensional continuous environments. Although Dueling DQN enhances performance through its ability to evaluate state and action benefits separately, both models face determination in handling dynamic MG environments because of their discrete nature. These constraints lead to increased expenses and diminished flexibility relative to the policy-based DDPG structure.

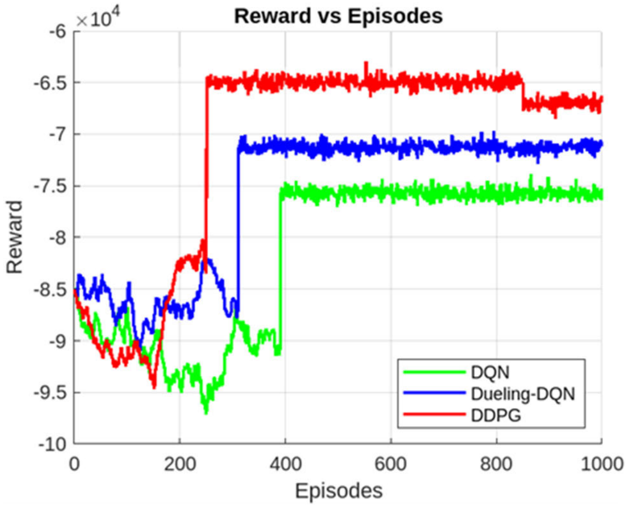

FIGURE 11. Training comparison of DDPG, DQN, and Dueling DQN.

In contrast, DDPG explores a multi-dimensional continuous action space, making it inherently more challenging and time-consuming than Dueling DQN's discrete action exploration. However, as the training progresses, DDPG surpasses DQN and Dueling DQN by achieving the highest rewards. This is due to DDPG's ability to make more accurate and precise decisions through its actor-critic architecture, which leverages experience replay and target networks for improved stability and learning efficiency.

2) COST COMPARISON

For the cost comparison, to further validate the performance, we compare the energy management costs of the three algorithms DDPG, Dueling DQN, and DQN at peak demand periods (10 a.m.). This period is selected based on the charging station demand model described in Section II-C, where maximum energy demand occurs between 9:00 a.m. and 11:00 a.m. Figure 12 presents the cost comparison among the three algorithms.

The DDPG algorithm's median cost is approximately 52 €ct/kWh, with the interquartile range (IQR) extending from 51.8 €ct/kWh to 52.2 €ct/kWh. A single outlier at 50.8 €ct/kWh indicates a lower variability in cost, highlighting DDPG's consistent performance in managing energy demand effectively.

The Dueling DQN algorithm shows a slightly higher median cost of 53.588 €ct/kWh, with an IQR ranging from 52.6 €ct/kWh to 53.4 €ct/kWh. Two outliers, one below the lower whisker and one above the upper whisker, demonstrate its increased cost variability compared to DDPG.

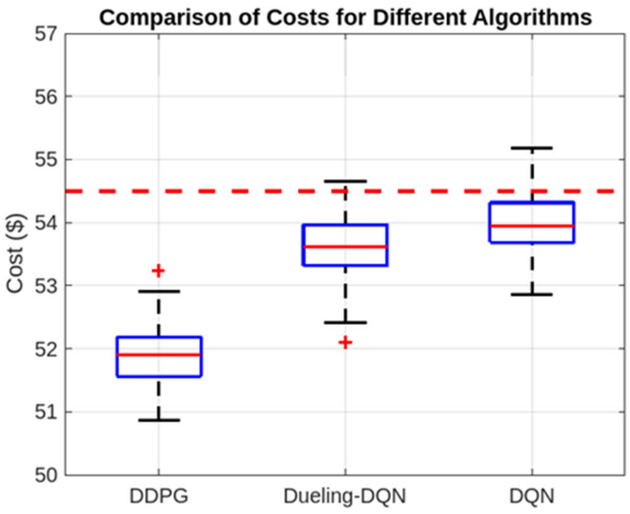

FIGURE 12. Cost comparison among DDPG, DQN, and Dueling DQN.

The DQN algorithm exhibits the highest median cost of 54.0256 €ct/kWh and shows greater variability, as evidenced by its wider IQR and two prominent outliers. This result indicates the limitations of DQN in handling the dynamic and continuous energy management environment.

Table 10 concisely compares runtime, convergence speed, training stability, computational overhead, and cost for the three algorithms.

TABLE 10. Comparison table.

Metric	DQN	Dueling DQN	DDPG
Convergence Speed	Early but suboptimal	Faster than DQN	Slower but optimal
Training Stability	Moderate	Higher than DQN	High
Computational Overhead	High	High	Low

This comparative analysis highlights the strengths and weaknesses of each algorithm in terms of performance metrics, demonstrating the superiority of DDPG for energy management in MGs.

3) PRACTICAL DEPLOYMENT CHALLENGES

Deploying RL-based strategies, such as DDPG, in real-world MG systems presents significant hardware and computational challenges. The computational requirements of the DDPG algorithm, particularly for training and real-time decision-making, can overwhelm the processing capabilities of typical MG controllers. Training DDPG models involves handling large state and action spaces, which necessitates advanced hardware with high computational power and memory capacity. However, the hardware in MG environments is often resource-constrained, making it difficult to execute the algorithm effectively. Moreover, real-time operation demands low-latency inference, which can be challenging for complex models running on limited hardware. Inadequate computational resources can lead to delayed or suboptimal decisions, adversely affecting the stability and efficiency of the MG.

V. CONCLUSION

This study proposed a DRL-based model-free approach for optimal energy management in MGs. Traditional model-based methods have shown limitations in handling uncertainties in MG environments, particularly with the increasing integration of RES and EVs. The model-free approach introduced in this work has demonstrated its ability to manage these uncertainties effectively and deliver optimal performance under challenging and dynamic conditions.

In this study, an IEEE 33 bus system integrated with RES and fast and slow CSs is used to make the system more realistic and challenging. This research optimizes RL agent actions using an actor-critic-based DDPG algorithm. We selected the policy-based function DDPG because it performs better in the continuous action space. The proposed model is composed of an actor-network and a critic network. The actor-network generates the action, which refers to the power supply obtained from various sources to meet the load demand within a specific continuous action area. The critic network evaluates the execution of these actions.

Numerical simulations demonstrate that the suggested strategy reduces electricity costs and power losses, which enhances the system's efficiency by minimizing the surge in demand caused by the FCSs. In addition, the simulation results show that the DDPG has performed better than the DQN and Dueling DQN. As a result, this novel strategy successfully solved the challenges of integrating RESs and managing CSs demand, making power systems more resilient and efficient. The total cost of the DDPG model is 51.8770 €ct/kWh, which is 3.19% less than the Dueling DQN and 4% less than the DQN model. The comparison analysis results with other methods and traditional optimization approaches confirm that the suggested control strategy based on actor-critic DDPG may reduce the overall cost by 15.91% and 15.59%, respectively.

The model-free DDPG algorithm, being a policy-based, end-to-end optimization method, eliminates the need for iterative calculations during online testing. This provides significant computational advantages compared to conventional algorithms. However, further work is needed to enhance the scalability of the proposed DDPG-based strategy for more complex and larger MG systems.

The current study focuses on optimizing the operation of distributed MGs without direct coordination with centralized grid management systems. This lack of integration may limit the approach's applicability to real-world scenarios where microgrids are often part of a larger grid infrastructure. Future work could investigate hybrid models that incorporate centralized control signals alongside decentralized decision-making to ensure seamless grid integration.

REFERENCES

[1] M. F. Zia, E. Elbouchikhi, and M. Benbouzid, "Microgrids energy management systems: A critical review on methods, solutions, and prospects," *Appl. Energy*, vol. 222, pp. 1033–1055, Jul. 2018, doi: 10.1016/j.apenergy.2018.04.103.

[2] L. Ahmethodžic, M. Music, and S. Huseinbegovic, "Microgrid energy management: Classification, review and challenges," *CSEE J. Power Energy Syst.*, vol. 9, no. 4, pp. 1425–1438, Jul. 2023, doi: 10.17775/CSEEJPES.2021.09150.

[3] N. Hatziargyriou, B. Mohammadi-Ivatloo, H. Abdi, and A. Anvari-Moghaddam, *Microgrids Advances in Operation, Control, and Protection*. Berlin, Germany: Springer, 2021, doi: 10.1007/978-3-030-59750-4_5.

[4] C. B. Ndeke, M. Adonis, and A. Almaktoof, "Energy management strategy for a hybrid micro-grid system using renewable energy," *Discover Energy*, vol. 4, no. 1, p. 1, Feb. 2024, doi: 10.1007/s43937-024-00025-9.

[5] V. S. Tabar, M. A. Jirdehi, and R. Hemmati, "Energy management in microgrid based on the multi objective stochastic programming incorporating portable renewable energy resource as demand response option," *Energy*, vol. 118, pp. 827–839, Jan. 2017, doi: 10.1016/j.energy.2016.10.113.

[6] B. N. Silva, M. Khan, and K. Han, "Futuristic sustainable energy management in smart environments: A review of peak load shaving and demand response strategies, challenges, and opportunities," *Sustainability*, vol. 12, no. 14, p. 5561, Jul. 2020, doi: 10.3390/su12145561.

[7] M. Rizvi, B. Pratap, and S. B. Singh, "Optimal energy management in a microgrid under uncertainties using novel hybrid metaheuristic algorithm," *Sustain. Comput., Informat. Syst.*, vol. 36, Dec. 2022, Art. no. 100819, doi: 10.1016/j.suscom.2022.100819.

[8] V. Boglou, C. Karavas, A. Karlis, K. G. Arvanitis, and I. Palaiologou, "An optimal distributed RES sizing strategy in hybrid low voltage networks focused on EVs' integration," *IEEE Access*, vol. 11, pp. 16250–16270, 2023, doi: 10.1109/ACCESS.2023.3245152.

[9] D. Loukatos, V. Arapostathis, C.-S. Karavas, K. G. Arvanitis, and G. Papadakis, "Power consumption analysis of a prototype lightweight autonomous electric cargo robot in agricultural field operation scenarios," *Energies*, vol. 17, no. 5, p. 1244, Mar. 2024, doi: 10.3390/en17051244.

[10] J. J. Makrygiorgou, C.-S. Karavas, C. Dikaiakos, and I. P. Moraitis, "The electricity market in greece: Current status, identified challenges, and arranged reforms," *Sustainability*, vol. 15, no. 4, p. 3767, Feb. 2023, doi: 10.3390/su15043767.

[11] M. Petrollese, L. Valverde, D. Cocco, G. Cau, and J. Guerra, "Real-time integration of optimal generation scheduling with MPC for the energy management of a renewable hydrogen-based microgrid," *Appl. Energy*, vol. 166, pp. 96–106, Mar. 2016, doi: 10.1016/j.apenergy.2016.01.014.

[12] E. Craparo, M. Karatas, and D. I. Singham, "A robust optimization approach to hybrid microgrid operation using ensemble weather forecasts," *Appl. Energy*, vol. 201, pp. 135–147, Sep. 2017, doi: 10.1016/j.apenergy.2017.05.068.

[13] Z. Li, C. Zang, P. Zeng, and H. Yu, "Combined two-stage stochastic programming and receding horizon control strategy for microgrid energy management considering uncertainty," *Energies*, vol. 9, no. 7, p. 499, Jun. 2016, doi: 10.3390/en9070499.

[14] T. Morstyn, B. Hredzak, R. P. Aguilera, and V. G. Agelidis, "Model predictive control for distributed microgrid battery energy storage systems," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 3, pp. 1107–1114, May 2018, doi: 10.1109/TCST.2017.2699159.

[15] G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam, and A. Ahmadi, "Dynamic energy management system for a smart microgrid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1643–1656, Aug. 2016, doi: 10.1109/TNNLS.2016.2514358.

[16] E. Foruzan, L.-K. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5749–5758, Sep. 2018, doi: 10.1109/TPWRS.2018.2823641.

[17] J. Liang and W. Tang, "Sequence generative adversarial networks for wind power scenario generation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 110–118, Jan. 2020, doi: 10.1109/JSAC.2019.2952182.

[18] J. Faraji, A. Ketabi, H. Hashemi-Dezaki, M. Shafie-Khah, and J. P. S. Catalão, "Optimal day-ahead self-scheduling and operation of prosumer microgrids using hybrid machine learning-based weather and load forecasting," *IEEE Access*, vol. 8, pp. 157284–157305, 2020, doi: 10.1109/ACCESS.2020.3019562.

[19] V. Skiparev, K. Nosrati, A. Tepljakov, E. Petlenkov, Y. Levron, J. Belikov, and J. M. Guerrero, "Virtual inertia control of isolated microgrids using an NN-based VFOPID controller," *IEEE Trans. Sustain. Energy*, vol. 14, no. 3, pp. 1558–1568, Jul. 2023, doi: 10.1109/TSTE.2023.3237922.

[20] B. Mbuwir, F. Ruelens, F. Spiessens, and G. Deconinck, "Battery energy management in a microgrid using batch reinforcement learning," *Energies*, vol. 10, no. 11, p. 1846, Nov. 2017, doi: 10.3390/en10111846.

[21] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, no. 8, p. 2010, Aug. 2018, doi: 10.3390/en11082010.

[22] J. Xian, A. Meng, and J. Fu, "Short-term load probability prediction based on conditional generative adversarial network curve generation," *IEEE Access*, vol. 12, pp. 64165–64176, 2024, doi: 10.1109/ACCESS.2024.3395659.

[23] Y. Xia, Y. Xu, Y. Wang, S. Mondal, S. Dasgupta, A. K. Gupta, and G. M. Gupta, "A safe policy learning-based method for decentralized and economic frequency control in isolated networked-microgrid systems," *IEEE Trans. Sustain. Energy*, vol. 13, no. 4, pp. 1982–1993, Oct. 2022, doi: 10.1109/TSTE.2022.3178415.

[24] R. Yan, Y. Wang, Y. Xu, and J. Dai, "A multiagent quantum deep reinforcement learning method for distributed frequency control of islanded microgrids," *IEEE Trans. Control Netw. Syst.*, vol. 9, no. 4, pp. 1622–1632, Dec. 2022, doi: 10.1109/TCNS.2022.3140702.

[25] E. O. Arwa and K. A. Folly, "Reinforcement learning techniques for optimal power control in grid-connected microgrids: A comprehensive review," *IEEE Access*, vol. 8, pp. 208992–209007, 2020, doi: 10.1109/ACCESS.2020.3038735.

[26] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, "Reinforcement learning and its applications in modern power and energy systems: A review," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1029–1042, Nov. 2020, doi: 10.35833/MPCE.2020.000552.

[27] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: A comprehensive review," *Int. J. Energy Res.*, vol. 44, no. 9, pp. 7136–7157, Jul. 2020, doi: 10.1002/er.5331.

[28] T. Yang, L. Zhao, W. Li, and A. Y. Zomaya, "Reinforcement learning in sustainable energy and electric systems: A survey," *Annu. Rev. Control*, vol. 49, pp. 145–163, Jan. 2020, doi: 10.1016/j.arcontrol.2020.03.001.

[29] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE J. Power Energy Syst.*, vol. 7, no. 2, pp. 209–220, Mar. 2021, doi: 10.17775/CSEEJPES.2020.02700.

[30] T. Levent, P. Preux, E. Le Pennec, J. Badosa, G. Henri, and Y. Bonnassieux, "Energy management for microgrids: A reinforcement learning approach," in *Proc. IEEE PES Innov. Smart Grid Technol. Eur. (ISGT-Europe)*, Sep. 2019, pp. 1–5, doi: 10.1109/ISGTEUROPE.2019.8905538.

[31] G. Muriithi and S. Chowdhury, "Optimal energy management of a grid-tied solar PV-battery microgrid: A reinforcement learning approach," *Energies*, vol. 14, no. 9, p. 2700, May 2021, doi: 10.3390/en14092700.

[32] Y. Yoldas, S. Goren, and A. Onen, "Optimal control of microgrids with multi-stage mixed-integer nonlinear programming guided Q-learning algorithm," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1151–1159, Nov. 2020, doi: 10.35833/MPCE.2020.000506.

[33] Y. Yu, Y. Qin, and H. Gong, "A fuzzy Q-learning algorithm for storage optimization in islanding microgrid," *J. Electr. Eng. Technol.*, vol. 16, pp. 2345–2353, pp. 2343–2353, Apr. 2021, doi: 10.1007/s42835-021-00769-7.

[34] K. Nosrati, V. Skiparev, A. Tepljakov, E. Petlenkov, and J. Belikov, "Intelligent frequency control of AC microgrids with communication delay: An online tuning method subject to stabilizing parameters," *Energy AI*, vol. 18, Dec. 2024, Art. no. 100421, doi: 10.1016/j.egyai.2024.100421.

[35] Y. Shang, W. Wu, J. Guo, Z. Ma, W. Sheng, Z. Lv, and C. Fu, "Stochastic dispatch of energy storage in microgrids: An augmented reinforcement learning approach," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114423, doi: 10.1016/j.apenergy.2019.114423.

[36] D. Liu, C. Zang, P. Zeng, W. Li, X. Wang, Y. Liu, and S. Xu, "Deep reinforcement learning for real-time economic energy management of microgrid system considering uncertainties," *Frontiers Energy Res.*, vol. 11, pp. 1–18, Mar. 2023, doi: 10.3389/fenrg.2023.1163053.

[37] C. Yang, Y. Xu, Y. Li, Y. Wang, and X. Zou, "Deep reinforcement learning based optimal energy management of multi-energy microgrids with uncertainties," *CSEE J. Power Energy Syst.*, pp. 1–12, Jan. 2023.

[38] C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning," *Energy*, vol. 238, Jan. 2022, Art. no. 121873, doi: 10.1016/j.energy.2021.121873.

[39] H. Hua, Y. Qin, C. Hao, and J. Cao, ''Optimal energy management strategies for energy Internet via deep reinforcement learning approach,'' *Appl. Energy*, vol. 239, pp. 598–609, Apr. 2019, doi: 10.1016/j.apenergy.2019.01.145.

[40] N. G. Paterakis, O. Erdinç, A. G. Bakirtzis, and J. P. S. Catalão, ''Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies,'' *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1509–1519, Dec. 2015, doi: 10.1109/TII.2015.2438534.

[41] C. Leone, M. Longo, L. M. Fernández-Ramírez, and P. García-Triviño, ''Multi-objective optimization of PV and energy storage systems for ultra-fast charging stations,'' *IEEE Access*, vol. 10, pp. 14208–14224, 2022, doi: 10.1109/ACCESS.2022.3147672.

[42] Y. Kim and S. Kim, ''Forecasting charging demand of electric vehicles using time-series models,'' *Energies*, vol. 14, no. 5, p. 1487, Mar. 2021, doi: 10.3390/en14051487.

[43] D. Fischer, A. Harbrecht, A. Surmann, and R. McKenna, ''Electric vehicles' impacts on residential electric local profiles—A stochastic modelling approach considering socio-economic, behavioural and spatial factors,'' *Appl. Energy*, vols. 233–234, pp. 644–658, Jan. 2019, doi: 10.1016/j.apenergy.2018.10.010.

[44] Å. L. Sørensen, K. B. Lindberg, I. Sartori, and I. Andresen, ''Analysis of residential EV energy flexibility potential based on real-world charging reports and smart meter data,'' *Energy Buildings*, vol. 241, Jun. 2021, Art. no. 110923, doi: 10.1016/j.enbuild.2021.110923.

[45] C. B. Saner, A. Trivedi, and D. Srinivasan, ''A cooperative hierarchical multi-agent system for EV charging scheduling in presence of multiple charging stations,'' *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2218–2233, May 2022, doi: 10.1109/TSG.2022.3140927.

[46] Y. Jiang, T. Ortmeyer, and M. Fan, ''Data-driven fast uncertainty assessment of distribution systems with correlated EV charging demand and renewable generation,'' *IEEE Trans. Sustain. Energy*, vol. 14, no. 3, pp. 1446–1456, Jul. 2023, doi: 10.1109/TSTE.2023.3236446.

[47] B. Papari, C. S. Edrington, T. V. Vu, and F. Diaz-Franco, ''A heuristic method for optimal energy management of DC microgrid,'' in *Proc. IEEE 2nd Int. Conf. DC Microgrids (ICDCM)*, Jun. 2017, pp. 337–343, doi: 10.1109/ICDCM.2017.8001066.

[48] A. Yasmeen, S. Zahoor, and H. Iftikhar, ''Optimal energy management in microgrids using metaheuristic technique,'' in *Microgrids Using Meta-Heuristic*, vol. 1. Cham, Switzerland: Springer, 2018, pp. 303–314, doi: 10.1007/978-3-319-75928-9_27.

[49] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, ''Mastering the game of go with deep neural networks and tree search,'' *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016, doi: 10.1038/nature16961.

[50] T. P. Lillicrap et al., ''Continuous control with deep reinforcement learning,'' in *Proc. 4th Int. Conf. Learn. Represent. (ICLR)*, 2016, doi: 10.48550/arXiv.1509.02971.

[51] M. E. Baran and F. F. Wu, ''Network reconfiguration in distribution systems for loss reduction and load balancing,'' *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr. 1989. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=25627

[52] Y. Wang and V. Friderikos, ''A survey of deep learning for data caching in edge network,'' *Informatics*, vol. 7, no. 4, p. 43, Oct. 2020, doi: 10.3390/informatics7040043.

ODIA A. TALAB (Member, IEEE) was born in Mosul, Iraq. He received the B.Sc. degree in computer engineering from the Technical College, Mosul, and the M.Sc. degree in computer engineering from Erciyes University, Türkiye, in 2012. Since 2003, he has been with the Ministry of Electricity, Iraq. His research interests include microgrids, smart cities, artificial intelligence, wireless communication, and machine learning algorithms.

ISA AVCI received the Ph.D. degree from Istanbul University–Cerrahpasa. He is an Assistant Professor with the Department of Computer Engineering, Karabuk University. He was the Project Manager with Istanbul Gas Distribution Company and Turkish Airlines. His fields of study are project management, business analysis, SCADA, smart grids, smart cities, artificial intelligence, machine learning, artificial neural networks, deep learning, cybersecurity, patch management, AHP method, cyber risks and threats, critical infrastructures, and ICS systems.