

Review

A Review on a Data-Driven Microgrid Management System Integrating an Active Distribution Network: Challenges, Issues, and New Trends

Lilia Tightiz  and Joon Yoo * 

School of Computing, Gachon University, 1342 Seongnamdaero, Seongnam 13120, Republic of Korea

* Correspondence: joon.yoo@gachon.ac.kr; Tel.: +82-31-750-5832

Abstract: The advent of renewable energy sources (RESs) in the power industry has revolutionized the management of these systems due to the necessity of controlling their stochastic nature. Deploying RESs in the microgrid (MG) as a subset of the utility grid is a beneficial way to achieve their countless merits in addition to controlling their random nature. Since a MG contains elements with different characteristics, its management requires multiple applications, such as demand response (DR), outage management, energy management, etc. The MG management can be optimized using machine learning (ML) techniques applied to the applications. This objective first calls for the microgrid management system (MGMS)'s required application recognition and then the optimization of interactions among the applications. Hence, this paper highlights significant research on applying ML techniques in the MGMS according to optimization function requirements. The relevant studies have been classified based on their objectives, methods, and implementation tools to find the best optimization and accurate methodologies. We mainly focus on the deep reinforcement learning (DRL) methods of ML since they satisfy the high-dimensional characteristics of MGs. Therefore, we investigated challenges and new trends in the utilization of DRL in a MGMS, especially as part of the active power distribution network (ADN).

Keywords: active power distribution network; energy management system; microgrid management system; machine learning; deep reinforcement learning; sparse reward



Citation: Tightiz, L.; Yoo, J. A Review on a Data-Driven Microgrid Management System Integrating an Active Distribution Network: Challenges, Issues, and New Trends. *Energies* **2022**, *15*, 8739. <https://doi.org/10.3390/en15228739>

Academic Editors: Xiandong Xu, Hongjie Jia, Yukun Hu, Tao Jiang and Gen Li

Received: 25 September 2022

Accepted: 18 November 2022

Published: 21 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the advent of RESs, the power industry could access infinite independent generators in feasibly electrified remote places [1]. RESs were initially integrated into MGs to supply rural and remote areas and were later deployed to campus and urban locations. A MG is capable of working independently in island mode and interacting with the utility grid in a grid-connected style. The main problem of utilizing RESs in MGs involves their intermittence characteristics, controlled by employing technologies, such as energy storage systems (ESSs), smart inverters, and responsive loads (RLs). Other challenges of MGMSs include the output prediction and load behavior of RES uncertainty. Therefore, it is necessary to coordinate the performances of all elements of MGs together, practiced by energy management systems (EMSs) [2]. EMSs are responsible for the generation and consumer elements of MG coordination by determining each generation unit portion in consumer supply, aside from preserving the stability of the system, such as voltage and frequency regulation [3]. It makes it necessary to provide a multi-objective optimization algorithm for EMSs.

As the main objective of a MGMS is to control energy consumption, prediction, and generation, ML can be used to control and predict in this system, dealing with the complex decision-making process. In the same line of thought, several studies have considered the application of ML in the optimization of MGMS [4–27]. Researchers in these studies considered MGMS implementation by investigating elements, forecasting tools, data and

demand management, standards, and different sights of view in control levels, including communication, energy, and power electronics. The main focus points of these studies on MGMS optimization were on classical mathematics programming, heuristic optimization algorithms, and artificial intelligence (AI) methods, such as fuzzy logic, artificial neural networks (ANN), and game theory. Mosavi et al. in [16] proposed an overview of the ML application in the energy system, mainly focusing on RESs and their output predictions. The authors concluded that the hybridization of ML techniques leads to a better solution for modeling RESs. Han et al. [11], without the definition of the EMS role in the MG, divided MG control levels to primary, secondary, and tertiary levels, and introduced some optimization methods, such as linear programming, genetic algorithm (GA), and particle swarm optimization (PSO) for power management in MG. The authors in [9], after the classification of MGs based on control methodologies, operation modes, and applications, provided a review of the different strategies for EMSs in MGs, including dynamic programming, meta-heuristic approaches, fuzzy logic, and ANN, and showed the limitations of each procedure. The authors of [8] considered EMSs in the operation fashions of MGs, i.e., stand-alone and grid-connected modes. Fuzzy logic and linear programming (LP) were methods introduced to the implementation of EMSs in that study.

Reinforcement learning (RL) and its derivative methods with the combination of deep learning can defeat randomness in the element characteristic MGs; in recent years, trends have risen, with deep reinforcement learning (DRL) being applied in the MGMS. Multi-agent system (MAS) arrangements of power systems, energy management, distribution networks, economic dispatches, cyber security, MGs, demand side management (DSM), DR, and electricity markets are various power system applications that can be optimized by DRL [18,20,21,28]. The robustness of RL in stochastic behavior management of optimization problem elements also drives the use of RL in building energy management systems (BEMSs). Wang et al. [19] and Yu et al. [29] investigated DRL applications in BEMSs. Both studies in the same line of thought arranged the Markov decision process (MDP) of the BEMS issue and classified relevant research based on the DRL hired method, objectives, and implementation environments. Arwa et al. [30] investigated the formation of a MDP for an integrated MG to the utility grid. After presenting a brief introduction of the RL methods, such as Q-learning, batch RL, actor-critic, and DRL, the authors summarized and classified related studies based on their applied techniques and scheduling MG objectives. The authors developed their study in a further attempt [31] to recognize RL method drawbacks and benefits. Another outcome of this paper involved the efficiency of applying MAS methods in the performance optimizations of MGs, in particular, multi-MGs.

Since the general approach of RES integration to the utility grid is through MGs, in a novel approach, in this paper, we will develop on previous studies regarding the application of ML for MG scheduling, mainly focusing on DRL techniques and their pros and cons. Moreover, we clarify which learning algorithms meet the right requirements of each application in MGMS. Since our primary hypothesis focuses an enhanced MG as a service provider of the utility grid, we investigate the neglected constraints of MGs in deploying RL techniques after the classification of existing research. A significant outcome of this study, shown in Table 1, involves the investigation of a recent approach to solving MG EMS problems with a combination of DRL and model-based approaches to overcome its shortcomings, including sparse rewards, implicit rewards, lack of scalability, low convergence speed, and privacy concerns. Therefore, the contributions of this paper are:

- We investigate the MG structure and clarify the role of the EMS in the MGMS.
- We determine EMS requirements, strategies, and tools.
- We explore DRL techniques applied in the literature for the EMS.
- We classify DRL techniques to meet EMS requirements.
- We investigate a new approach in deploying a model-based method to solve weak DRL points in a MGMS arrangement.

The outline of the paper is as follows. Section 2 clarifies the MGMS control levels to recognize its requirements in depth. Additionally, Section 2 classifies MGMS application

requirements. Section 3 determines ML techniques and their applications in an EMS through the literature investigation and classification. After digging into future trends in the MGMS, in Section 4, we specify the roles of RL and DRL techniques in these future trends. Moreover, in Section 4, we classify and reveal the merits and demerits of existing works based on RL and DRL as well as technical gaps in MG contributions to ADN. Ultimately, this paper concludes in Section 5.

Table 1. Related work objectives—a comparison.

Reference	Year	Main Objective
[4]	2014	CCHP based MG modeling, planning and EMS
[5]	2015	Classification of studies on EMS of MG considering cost function
[6]	2016	Control objectives of hierarchical control level of MG
[7]	2016	Categorizing MG EMS optimization algorithms and hired software tools
[8]	2016	Study on Fuzzy logic and linear programming approach in EMS implementation
[9]	2018	Comparative investigation on EMS strategies and communication requirements
[10]	2018	Energy storage system role in EMS
[11]	2018	Focus on different level control of MG optimization with MAS approach in power electronic viewpoint
[12]	2018	Weather prediction methods for EMS
[13]	2019	MG control strategies from power electronic sight of view
[14]	2019	Comparative study on EMS solutions and tools
[15]	2019	Specifying MG generation units cost functions with more concentrating on storage
[16]	2019	RES output prediction
[17]	2020	An extensive classification of trends in EMS research works
[30]	2020	MDP arrangement for EMS problem
[31]	2020	Pros and cons of RL methods
[18]	2020	Application of DRL in power system optimization
[19]	2020	Application of RL in BEMS
[20]	2020	DRL application in power system
[21]	2020	DRL and Multi-agent DRL application in power system
[22]	2021	ESS and MG control strategies
[23]	2021	Key applications in EMS
[24]	2021	EMS in Islanded MG
[32]	2021	DSM role in EMS
[25]	2021	EMS of DC MG
[26]	2021	Protection aspects of MG performance optimization
[27]	2021	EMS of campus MG
[29]	2021	DRL application in BEMS
[33]	2021	EMS in shipboard MG
[34]	2022	Classification of EMS optimization techniques
[35]	2022	bibliometric analysis on EMS studies
[28]	2022	DRL application in power system
Present work	2022	Study on recent trends in DRL techniques combination with model-based method to solve its dedicated drawbacks in Solving EMS

2. Microgrid Management System

2.1. Microgrid Structure and Control Methodology

A MG involves a collection of loads and generators that should handle supplying consumers in grid-connected or islanded modes. MG suppliers include RESs and conventional generators, in which the latter are applied to dominate the intermittency and stochastic characteristics of the former [36]. The ESS is the other element of a MG; it is widely used to make RESs dispatchable, in other words, the ESS supports the grid in the unavailability situation of the RES. Controlling the active power generation and, consequently, the frequency deviation of the MG, facilitates these accessories. An ESS can be a hybrid ESS, e.g., batteries to support the steady power demand or an ultra-capacitor to support the transient power demand [37]. The electric vehicle (EV) is another form of an ESS in a MG [38]. According to the 2003 version of the IEEE 1547 standard, the MG should work in a

grid-connected mode and supply its domain consumers by the utility grid. Then, in case of any failure situation in the utility grid, the MG should disconnect from the grid and supply a whole or a part of the loads autonomously by the ESS. In this scenario, the RES during the grid-connected mode charges ESS [39]. The IEEE 1547 standard, modified in 2018, is used to consider the MG as a member of an ADN, and transacts energy with the utility grid where the RES is equipped with voltage and frequency ride-through capabilities to control abnormal situations in a grid-connected mode [40]. If we consider MG elements, as depicted in Figure 1, they generally include conventional generators, e.g., RESs, ESSs, EVs, and loads. There are two groups for loads in the MG, including emergency loads and responsive loads. While emergency load-serving has priority in both grid-connected and islanded modes of a MG, responsive loads are sheddable and help the grid remain stable [41].

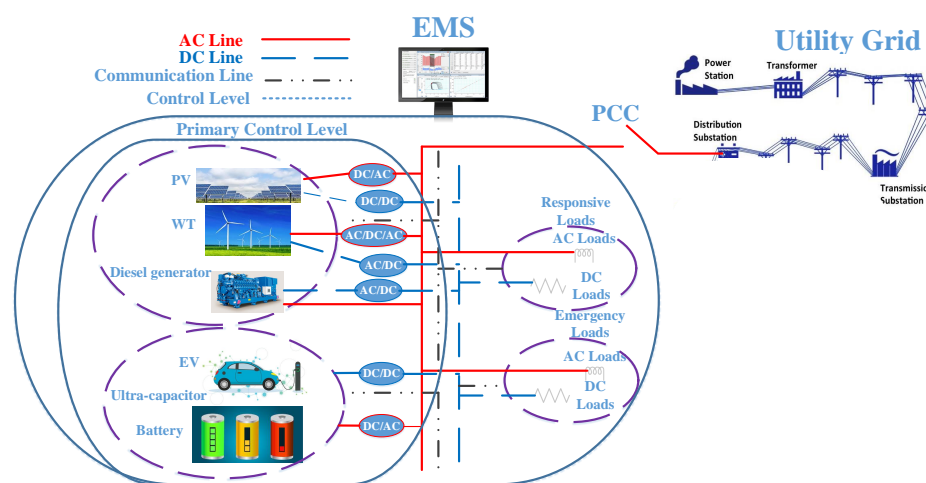


Figure 1. MG structure and elements.

The stability performances of MGs require each element to receive setpoints from the supervisory level to adjust their actions. There are three levels of control for MGs, including primary, secondary, and tertiary levels [42]. The primary level control of a MG is related to the adjustment of output-level resources. The interfaces of the resources and ESSs in the MG involve smart inverters. Smart inverters can follow grid forming, grid feeding, and grid-supporting strategies, as addressed in [43] extensively. Current source inverters (CSIs) connect to RESs with high impedance in parallel to implementing the grid-following strategy. This strategy tracks the maximum power point of a RES. The grid-forming involves a voltage source control inverter (VSI), which connects with a low impedance to the ESS in a “series” arrangement. CSIs and VSIs can control the voltages and currents of RESs through the inner loops and receive references from the primary control levels. The grid-supporting strategy is a droop control that is used in the primary level control and is provided by VSI to utilize dispatchable generators and ESSs in the MGs if grid-forming inverters in the islanded mode cannot maintain the frequency and voltage of the grid. Droop control methods make primary-level control communications “less”, and are reliable by mimicking synchronous generator behaviors; however, due to distances between RESs, this method is not always feasible. Active power-sharing is another method used at the primary control level, which is communication-based. Since primary control specifications require fast responses, communication infrastructure requirement costs would be challenging [44]. The idea of ESS installation near to RESs has been researched because taking source installation space locations has altered the utilization of distributed ESSs [45]. The secondary control level is applied to restore voltage and frequency deviations of MGs to zero. This level of control, which is the MGMS, will determine the power-sharing of resources and ESSs for protecting the stability of the MG. The secondary control level, also known as the energy management system (EMS), interconnects with a supervisory level of the utility grid in the point of common coupling (PCC), which is a distribution management system (DMS).

Secondary control is centralized or decentralized. The centralized controller approach follows the controlling and supervision of resources and ESSs, grid-connected and islanded mode determination, load management, contributions to the market, and forecasting the output power of sources centrally. Because of the multitude of information, decision-maker algorithms are complicated; fast, reliable, and high bandwidth communications in this trend are necessary [42]. Instead of a central control, MG actors in the decentralized method have a local controller. This scenario meets the MAS style requirements. In the MAS, each MG element is an agent, has a decision-maker algorithm in its local controller, and interacts with the neighbor and supervisory agents. Although both methods track hierarchical schemes and require MGMS to communicate with DMS, a centralized controller is commonly used to control distributed generators (DGs) inside MG. Additionally, the decentralized type is used in the grid with a broad horizon, including numbers of MGs incorporating the utility grid [46]. The tertiary control level is a DMS, which coordinates the performance of MG by assigning setpoints to the MGMS.

2.2. Microgrid Management System Requirements and Applications

According to the IEEE 1547 standard, a MG has several stakeholders, including aggregators, maintainers, operators, and distribution system operators (DSO). The MGMS, as a supervisory level of the MG, should be scheduled in a way that satisfies all stakeholders' requirements, as shown in Figure 2. Moreover, Figure 3 reveals the related application of each stakeholder. In this paper, we mainly consider the MG role as an ADN member. This hypothesis is an opportunity to bring merits to whole stakeholders of MGs in several criteria. It facilitates the integration of RESs into the utility grid. MGs aid the role of prosumers for RES owners. Moreover, the power grid receives financial profits by getting rid of the development and refurbishment of transmission lines. MGs aid grid stability by providing services for the grid in the case of a failure in the system, such as frequency stability, voltage regulation, and the black start aid [47]. These characteristics would not be viable without a robust management system and applications coordinating the MG's elemental performances and interactions with the utility grid. The following is a brief of the most notable applications.

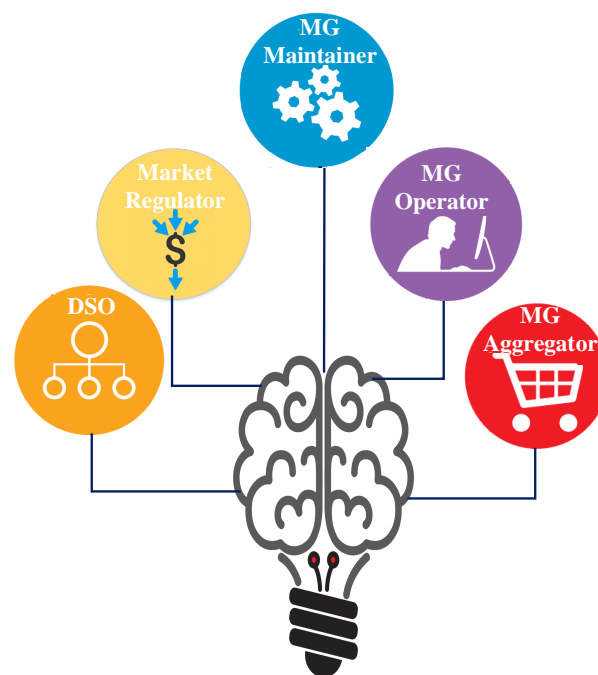


Figure 2. EMS optimization function dependencies.

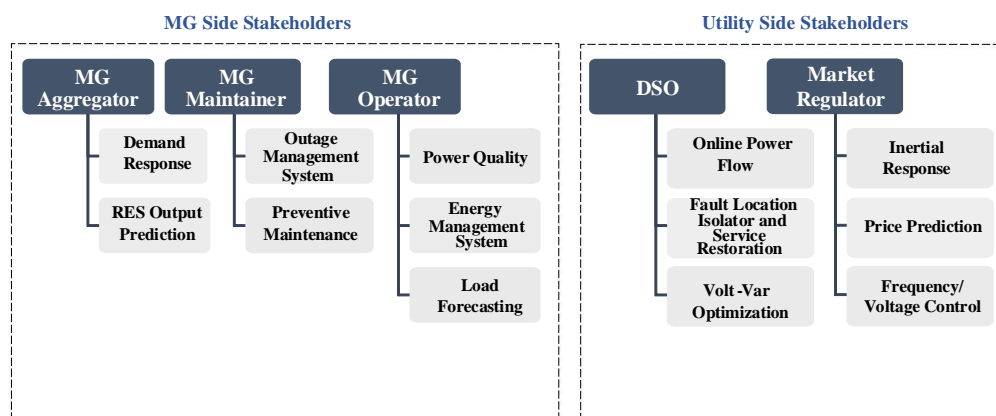


Figure 3. MG application requirements based on stakeholders.

2.2.1. Demand Response

Utilities deploy DR to control energy consumption during peak hours and contingencies through the dynamic contributions of consumers. DR has been applied in MGs to assist EMSs during contingencies in the islanded mode. DR also responds to the utility grid requirements when the MG behaves as an active element of the power system. DR is implemented by offering incentive regulations or time-dependent programs. Consumers decide to cooperate in peak shaving with utilities based on dynamic prices and schedules, such as time of use (TOU), critical peak pricing (CPP), and real-time pricing (RTP) [48].

2.2.2. EMS

The traditional power systems under the control of the generation company (GENCO), TSO, and DSO operated hierarchically. Centralized generation units supply passive loads through transmission and distribution networks in this arrangement. With the addition of smart grid technology and distributed generation, the power system transitioned from a hierarchical to a unidirectional distributed style. DGs are power generation units located at consumer sites and provide dynamic loads. DR programs will strengthen the new role of loads. The distribution system as an adjacent section of the power system to active consumers becomes more flexible, active, and complicated. The dynamic loads are controllable under the supervision of MGs. Both ADNs and MGs can collaborate to satisfy the control and protection requirements. In the ADN concept, one of the responsive elements of the system is the MG. The AS provided by MG EMS includes participation in the remedial action schema of the power system, such as the black start aid, network reconfiguration, frequency, voltage control, power loss reduction, and congestion management. These characteristics require EMS to be able to coordinate the performances of all MG elements. Energy consumption or production of RES, responsive or non-responsive loads, conventional generators (CGs), and ESSs should be determined. This can be done locally or by the ESS unit of the MGMS. Each entity may estimate its situation by its controller independently and send its information to the EMS, or the EMS calculates the portion of entities in the power generation and consumption and issues related commands. The EMS should coordinate its performance with supervisory-level control of the utility grid, which is the DMS [42]. The required input data and output scheduling that the EMS should provide are represented in Figure 4.

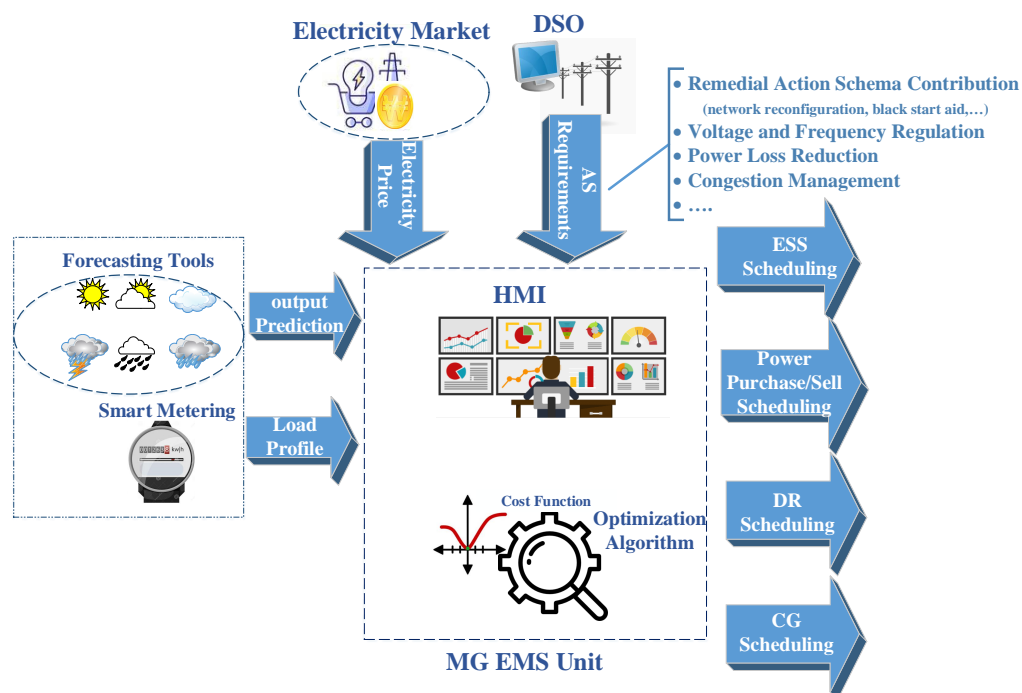


Figure 4. A MG EMS unit under the concept of ADN.

2.2.3. RES Output Prediction

This application is applied in a MG to reveal the amount of available energy from the RESs. They are the main suppliers of the MG and their performances depend on natural sources, such as solar irradiation and wind availability. Hence, forecasting their output assists EMSs in scheduling the MG contribution in market and storage charging and discharging [49].

2.2.4. Preventive Maintenance (PM)

Due to the huge number of elements in the power system, PM aids the grid in controlling risks and reducing the costs of outages and equipment breakdowns. Therefore, it improves the reliability indices of the grid. PM for the power grid in a smart environment is classified as passive and active. The passive class is concerned with the market, while the active class schedules PM without paying attention to the market or prices [50].

2.2.5. Distribution system Operation

The DSO implements the DMS by utilizing different applications, such as online power flow (OPF), fault location isolation–service restoration (FLISR), volt–var optimization (VVO), and smart connection arrangement (SCA). All of these applications help the DSO analyze the status of the power grid and predict its action requirements. The MG under our hypothesis can be an element that supports these applications.

2.2.6. Market regulator

Any violation in the power system stability due to the difference between the scheduled power generation and consumption is known as an area control error (ACE). The electricity market regulator’s traditional task involves the ACE compensation with adjustable generation and consumption units in a market-based manner and determination of the electricity price for end users. The ADN is viable by providing AS services from the MG side. Therefore, in an enhanced power system, the electricity market regulator facilitates the MG participation in the ACE correction as a member of the ADN. The MGs can address the ACE by justifying the amount of generation and consumption according to the signal received from a market-based regulator.

3. Machine Learning Applications in MGMS

3.1. Machine Learning Techniques

The main challenge of obtaining access to the benefits of RESs and the high penetration of MGs is in providing a comprehensive MGMS, which is dependent on the embodiment of the EMS unit as a robust decision center through intelligent ML techniques. ML mimics how humans learn by deploying data and algorithms and improves the solution's accuracy over time, typically in an iterative manner. ML algorithms, with the help of historical data, called data sets, train computers to output values that fall within a predetermined range. A MGMS, with the help of ML techniques, attempts to make the right decision for the control and management of the MG based on existing information and experiences. ML can detect all information patterns inside a MG and predict the behaviors of heterogeneous devices gathered under the supervisory of the MGMS [51]. ML has several sub-models, including supervised learning, unsupervised learning, semi-supervised learning, and RL, as shown in Figure 5.

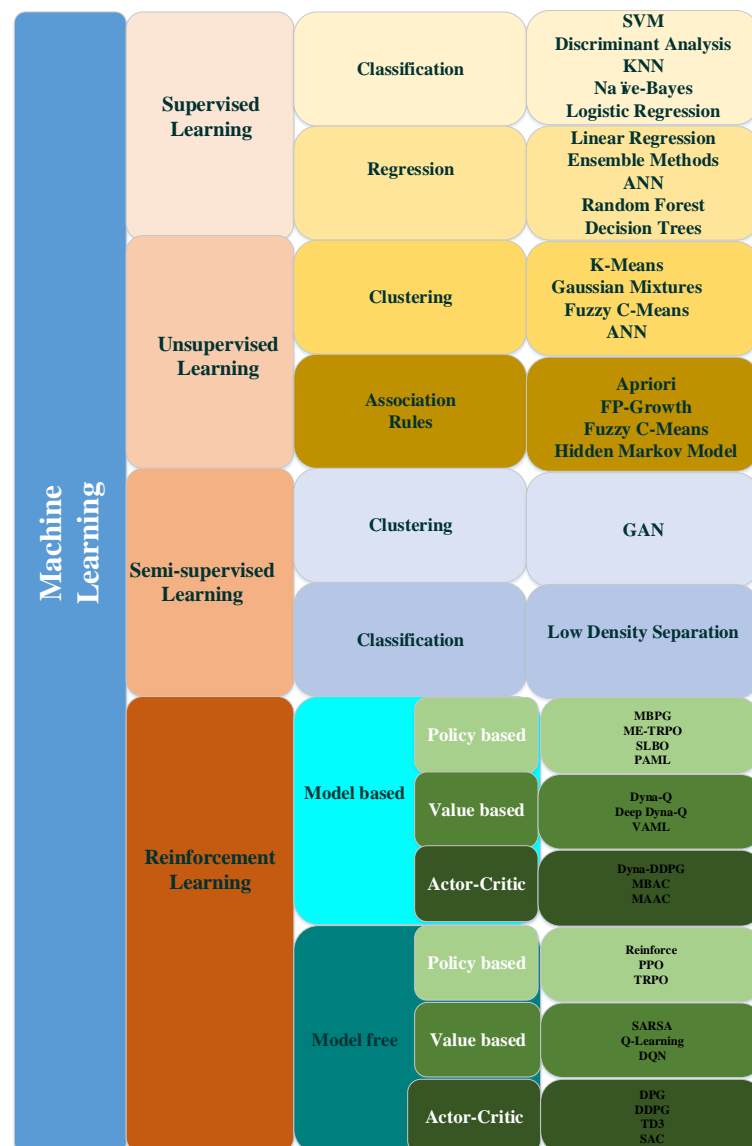


Figure 5. Different types of machine learning.

3.1.1. Supervised Learning

Supervised methods predict output based on existing experiences. This model predicts the output by finding a relationship and dependencies of available input and output data sets. In supervised learning, experts preprocess data sets by labeling input and output data to use them for training and testing the model. The model will classify and forecast previously unknown data sets. Classification and regression algorithms are solutions employed in this method. Classification algorithms can predict discrete output models of new data after feature extraction and label provision for previous data. Support vector machines (SVMs), discriminant analysis, K-nearest neighborhood, naïve Bayes, and logistic regression are a few examples of this algorithm, as shown in Figure 5. The regression algorithms based on the feature and correspondent outputs of previous data predict a continuous output model of new data. Linear and polynomial regression, artificial neural networks (ANNs), random forest, and decision trees are their examples, as can be seen in Figure 5.

3.1.2. Unsupervised Learning

This is a self-organized decision-making method that attempts to find hidden patterns in the system. Despite supervision in the unsupervised approach, there is no access to labels and outputs, and the learning algorithm should predict the model by deriving the feature of input data. Clustering and association rules are methods used to implement unsupervised learning. While clustering extracts features by making different groups for inputs, association rules attempt to do the same task by finding relationships and data set patterns. For clustering, K-means, Gaussian mixtures, and principal component analysis (PCA) and in the case of association rules, Apriori, and FP growth are popular algorithms, as shown in Figure 5.

3.1.3. Semi-Supervised Learning

Semi-supervised learning is a combination of the previous two methods. This technique is efficient when there are a significant amount of inputs and limited access to outputs and a need to expend much on the acquisition of labeling data. Classification and clustering algorithms are used in this method. According to Figure 5, while generative adversarial networks (GAN) is a subset of clustering techniques, low-density separation is an example of classification.

3.1.4. Reinforcement Learning

RL is originally a subset of semi-supervised learning. It works based on four elements, namely agent, environment, reward, and action. Agents take action according to learning from the environment. The environment evaluates each agent's action by positive or negative feedback, which is called reward. The first step in solving a problem with RL is the MDP arrangement. MDP is a set of four members, $\{S, A, P, \text{ and } R\}$, where S is a state space, A is an action space, R is a reward, and P is the state transition probability. The state space consists of the data that assist the agent, which controls the actions accurately. The state space is categorized as fully observed and partially observed. In a fully observed state space, the agent is able to monitor the entire environment. In some cases, the agent can be aware of the environment to some extent. The action space, based on the characteristics of the issue and optimization resolution, can be continuous or discrete. Nonetheless, the continuous action space can be discretized with an acceptable level of resolution to find the trade-off between accuracy and simplicity. The Markovian approach to solving problems specifies a policy to determine the next states according to the current state without requiring knowledge about the previous states of the environment. This policy describes how an agent acts in its current state. The policy can be deterministic or stochastic. In a deterministic style, the same states always lead to the same next states, while with stochastic behavior, there are a set of next states for similar states. Using the Bellman equation, the agent can implement the non-historical dependent policy in choosing

actions for deterministic and stochastic environments, according to (1) and (2), respectively, as follows:

$$V_{s_t} = \max_{a_t} (R(s_t, a_t) + \gamma V(s_{t+1})), \quad (1)$$

$$V_{s_t} = \max_{a_t} (R(s_t, a_t) + \gamma \sum_{s_{t+1}} P(s_t, a_t, s_{t+1}) V(s_{t+1})), \quad (2)$$

where V_{s_t} and $V_{s_{t+1}}$ determine the values of the present and next states, respectively. $R(s_t, a_t)$ is the immediate reward for taking action a in the state s . The agent will choose action a with the maximum reward among the possible actions. γ is the discount factor in the range of (0,1) to consider the effect of the current action on the expected reward in the future. In a stochastic manner, the Bellman equation applies the probability of taking action a with using $P(s_t, a_t, s_{t+1})$.

RL, concerning knowledge of the agent (regarding the environment) has two categories, namely model-based and model-free. In model-based RL, agents attempt to find a model for the environment and capture the transition function by learning. In contrast, the agent attempts to learn the optimal policy to conduct the best action in each state. Since the MG is a collection of elements with stochastic behaviors, model-free RL meets its requirements. There are three model-free RL methods—value-based, policy-based, and actor–critic. Value-based techniques attempt to exploit high-value actions by exploring the environment and finding the policy, which leads the agent to the best action. Unlike value-based methods, policy-based methods directly search for an optimum policy without a value function assistant. Policy-based approaches have better convergence performances for high-dimensional, continuous, and stochastic problems compared to value-based ones. The drawbacks of policy-based methods include converging to local optima rather than global, and inefficient estimation of policies. The third model-free RL category, i.e., actor–critic, is a trade-off between the two previous methods. In the actor–critic model, the critic in each state, according to the value of each action, assigns a reward to the actor. Recently, deep neural network (DNN) was widely applied in RL and offered deep RL (DRL). DRL is a solution used to estimate a model for model-free RL algorithms; it evokes benefits, including dealing with high-dimensional issues in value-based practices and improving variants in policy estimations of policy-based methods. While Q-learning and its promoted derivatives, such as the deep Q-network (DQN) and double DQN (DDQN) are examples of value-based RL, policy gradient and REINFORCE are policy-based ones; utilizing policy and deep learning, actor–critic can be classified to a deterministic gradient policy (DGP) and stochastic gradient policy (SGP). The deep deterministic policy gradient (DDPG), distributed distributional deep deterministic policy gradient (D4PG), twin delayed deep deterministic (TD3), and multi-agent DDPG (MADDPG) are examples of DGP methods. Additionally, the synchronous advantage actor–critic (A3C) algorithm and soft actor–critic (SAC) are SGP ones.

3.2. Classification of Machine Learning Applications for the MGMS

Application requirements of each stakeholder should be satisfied by the MGMS. Easy access to ML modules in different language programming, such as Python-based scikit-Learn [52], MATLAB ML Toolbox [53], and R.Package caret [54], make it popular in optimization, prediction, and training issues in the MGMS. The ML application in the MG can be as follows.

3.2.1. Demand Response

The cooperation of the electricity market, DSO, and consumers leads to DR implementation. Several ML techniques have been applied to propose highly accurate demand prediction. Researchers deployed ML in DR to maximize aggregator profit and predict loads, which can be residential, commercial, and industrial, and assess DR effectiveness. The authors of reference [55], after classifying energy systems to cooling and heating systems, appliances, EVs, DGs, and storage, deployed RL as a solution to the recognition of

consumer demand patterns. This paper also realized the power system able to provide peak shaving by applying RL in the MAS arrangement. The authors of reference [56], using customer incentive price policy and deploying the genetic algorithm as a heuristic procedure, attempted to shift the peak load. In this paper, GA inputs were the consumers' load data and the forecasted day-ahead prices from the utility. The authors developed their model from 5555 customers and 208,000 loads, categorizing into schedulable and non-schedulable, and the K-nearest neighbor was applied for the classification and Gaussian process regressions for learning. They evaluated two models, i.e., (1) the determination effects of schedulable loads on aggregator profit (by selecting them accidentally); (2) the other modeled the effects of daily ahead price effects on the aggregator profit. Their attempts showed—by the classification of schedulable loads—aggregator gain maximum profit. The authors of reference [57] attempted to apply ML in load forecasting of commercial buildings for DR. In this paper, heating and cooling systems participated in DR provision. ANN and SVM with polynomial kernel functions are ML techniques employed to predict the demand of a commercial building at a Dublin university campus in DR and non-DR states. An advanced metering interface (AMI) is one of the applications that provides two-way communication between the end-user and a supervisory system. The leading device in AMI is a smart meter, which should gather information from smart appliances. Many sensors assist smart meters in this case, and ML analyzes this information and uses it for load prediction [58]. Nakabi et al. in [59] reviewed a home energy management system (HEMS), considering that the smart home appliances and home temperatures of ANN and the long short-term memory (LSTM) method could predict an individual end-user's load profile and evaluate his/her proposed method with a mean absolute percentage error (MAPE) and root-mean-square error (RMSE).

3.2.2. RES Output Prediction

The output predictions of RESs are necessary to schedule EMSs of MGs. The predictions can be for ultra-short-term, short-term, medium-term, and long-term forecasts [60]. Research in this area (regarding the application of ML) includes the output prediction of solar power, wind turbines, and geothermal resources [61]. The commonly used method in prediction problems is the statistical method, which is easy to implement and has a high accuracy in short-term prediction. Auto-regressive (AR) and auto-regressive moving average (ARMA) are examples of statistical techniques. ANN is the other method deployed in the field of RES output prediction. ANN is also suitable for short-term prediction. The difference between statistical models and ANN is that the statistical approach determines the model of the element by comparing the predicted results and actual results; ANN will find a relationship between the output and input to create the model [62]. DDPG was deployed in an isolated MG to predict photovoltaic (PV) and wind turbine (WT) output in [63]. Evaluation tools, such as absolute error σ , mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), RMSE and the coefficient of determination R^2 are used to compare forecasting results of DDPG with a traditional multiple linear regression (MLR), auto-regressive-integrated moving average (ARIMA), LSTM, and recurrent deterministic policy gradient (RDPG).

ML techniques can be used for hourly forecasting and are mainly applied in solar prediction [64]. Yangli et al. [65] investigated 68 different ML techniques for short- and long-term irradiation forecasting. They provided a data set involving seven stations from five different climate zones. They found various methods that worked differently in each situation. For instance, tree-based techniques are better in the long term; MLP and support vector regression (SVR) are good in clear sky conditions. Eseye et al. in [66] applied the Wavelet-PSO-SVM model on the actual output of PV and weather prediction data to forecast the output power of PV in four different seasons of the year. After evaluating their method with MAPE and normalizing mean absolute error (NMAE) metrics, the authors benchmarked their approach with back-propagation neural network (BPNN), hybrid GA-SVM (HGNN), hybrid PSO-neural network (HPNN), SVM, hybrid GA-SVM

(HGS), hybrid PSO-SVM (HPS), and (hybrid Hilbert–Huang transform (HHT)-PSO-SVM (HHPS)). Martin et al. [67] predicted solar irradiation by autoregressive, ANN, and fuzzy logic models, and evaluated their results by the root mean squared deviation (rRMSD). Among LSTM, SARIMAX, and their combination, more accurate results based on RMSE, MAPE, and the mean-squared logarithmic error (MSLE) evaluation methods were used to predict the hourly wind turbine output power and load consumption in MG [68]. These predicted values were used to solve the mixed-integer non-linear programming (MINLP) arrangement of the hourly MG economic dispatch. Q-learning improved the accuracy of the LSTM, deep belief network (DBN), and gated recurrent unit (GRU) network weights, employed as wind speed short-term predictors in [69]. This study determined the reward function based on the mean square error (MSE). Q-learning performed better than the meta-heuristic and model-based RL optimization models.

3.2.3. Preventive Maintenance

MG elements during are endangered and subjected to grid unbalance voltages, frequencies, and other parameters during their lifetimes. Therefore, MG elements require maintenance, and ML is one of the tools used in predicting the time of this requirement. This prediction requires a high level of accuracy. Applying ML to analyze historical data helps the power system predict which condition or equipment will disrupt the performance. It also assists the grid in maintenance scheduling. According to Table 2, ML tools promote PM in two criteria—improving reliability and failure prediction. Wu et al. [70] introduced an online ML-based system to prove the efficiency of ML in PM scheduling. Reference [71] focused on the real-time fault diagnosis of a PV system based on gathering information from sensors, such as voltage, current, temperature, and solar irradiance. In this paper, ANN is a ML technique applied to help the MG in pattern recognition to determine faults in the entities, including the battery, PV arrays, loads, and maximum power point tracking (MPPT) system. The authors of [72] optimized the management and maintenance process of RES and controllable generators by combining ANN and Q-learning. The MG joined the power system protection schema in [73] by deploying DDPG and SAC to provide AS in contingencies for the ADN.

Table 2. Related work objective comparisons.

MGMS Application	Problem Statements	References	Methods	Evaluation	Tools
DR	Aggregator profit	[56]	KNN, Gaussian regression	RMSE, NRMSE	MATLAB
	Load forecasting	[57–59,74]	Binary GA, ANN, CNN, Gaussian regression, SVR, LSTM	MAPE, RMSE, CV-RMSE, CV-MAPE	MATLAB
	Assess the effectiveness of DR strategies	[75,76]	Rule-based control algorithm, SARSA, DDQN, PPO, A3C	MAE, RMSE, MBE	MATLAB
RES Output Prediction	wind power and speed forecasting	[69,77,78]	ARMA, ARIMA, Wavelet transform, CNN, Q-learning	MAE, RMSE, MAPE, Average coverage error	MATLAB
	solar power and irradiation forecasting	[65–67]	DBN, FL SVM, KNN, Random forest regression, ARIMA, Hybrid WT-PSO-SVM	MAE, RMSE, BIAS, nRMSE, nMBE, rRMSE, MBE	MATLAB R-language
	Output prediction of different RES with combination of loads	[68,79]	ANN, LSTM, SARIMAX, KNN, K-means	RMSE, MAPE, MSLE	GAMS

Table 2. Cont.

MGMS Application	Problem Statements	Reference	Methods	Evaluation	Tools
EMS	BMS	[80–83]	DQN RL(fitted Q-iteration)	-	-
	EMS based on MG mode	[84,85]	LP, SVM ANN, DQN	-	MATLAB
	EMS with EV presence	[86,87]	DDQN, DDPG	MSE, steady deviation	
	EMS in multi-agent MG	[88–90]	NN-GRU, DDQN	MAPE, RMSE NRMSE	MATLAB Python
PM	Failure prediction	[71,91]	Gaussian process regression, Time aggregation, expert system, ANN	-	-
	Improving reliability	[72,73]	GA, Q-learning SAC, DDPG	NRMSE	Contingency analysis program, MATLAB

3.2.4. Energy Management System

MG in a stand-alone mode should respond to the energy demands of consumers inside a MG, independently. On the other hand, a high MG penetration arises interacting with the utility grid in a bidirectional electricity path to participate in the electricity market and provide profit for the MG stakeholders. The principal challenge in both situations is how the MG can manage EMS. The MGMS should control RES uncertainty characteristics through ESSs and provide coordination between energy consumption and generation in the MG. This coordination depends on the ESS performance, and ML is a tool to implement this scenario. EMSs have diverse aspects, such as building energy management systems (BEMS), EMSs of MGs in grid-connected modes or stand-alone modes, EMSs with the presence of EVs, and EMSs for multi-agent MGs. There are different approaches to the optimization of EMSs in MGs: model-based and model-free. The data in MGs are enormous; thus, the system's complexity–model-free approach will result in more effective results. Q-learning overcomes the lack of availability to the MG's explicit model issue when taking into account its model-free properties.

This model learns from real-time data independent from future rewards or state situations of the system's awareness. Q-learning applied in a BEMS represented the MG in [80,86]. Battery scheduling in [80] was provided based on uncertainty in wind turbine power generation as a power source with the primary purpose of less electricity purchases from the main grid. The authors in this paper declined randomness of load and electricity market prices and considered scheduling for 2 h a day ahead. In this study, MG does not export energy to the utility grid. Conversely, Kim et al. [86] scheduled a real-time smart building storage EMS as a prosumer while the main objective was to reduce the energy cost. Q-learning has been successful in EMS optimization; however, exploiting the capabilities of all MG elements and tackling their uncertainties are high-dimensional tasks. The curse of dimensionality is the hindrance of applying Q-learning to high-dimensional environments. Deep Q-Network (DQN) [92] combines DNN and reinforcement learning, which offer scalability in solving the EMS of a MG. DQN supports high-dimensional problems since it uses ANN for state value approximation rather than a Q-learning tabular representation. Using the same approach, references [81,82] simulated MG as an independent grid without any energy transactions with the utility grid by applying a convolutional neural network (CNN) in DQN. The loads, RESs, and market price uncertainties were considered in [84] when DQN was applied to provide EMS in the MG. This paper evaluates the efficiencies of their method using data gathered from California independent system operators. Despite DQN performing better than Q-learning in a stochastic environment, it still struggles with stability in the training network due to the correlation between the estimated and target values. The double DQN (DDQN) solves this problem by offering separate ANNs for

selecting actions and evaluating them [93]. V. Bui et al. [88] applied DDQN to optimize the battery EMS's community performance of the MG to overcome the overestimation problem of the DQN. The learning process scheduled the ESS to minimize power production costs in the grid-connected mode and serve the critical load in the island mode. This paper examined electricity market prices, loads, and RESs as uncertain elements of the environment and demonstrated that the method presented works well in the multi-MG systems. In contrast to online reinforcement learning, batch reinforcement learning can converge faster in a battery energy management (BMS) in MG, according to the study in [83].

Q-learning, DQN, and DDQN are value-based learning methods. They are adequate for discrete actions, though their slow rates of policy changes make them vulnerable to overestimation. The policy gradient is another subset of RL that uses deep learning. Stochastic and deterministic approaches are two classes of the policy gradient-based method. Stochastic policy gradient (SPG) performs better in addressing the convergence speed and high-dimension action environment issues than the value-based method. However, the possibility of convergence to a local maximum exists. SGP techniques can meet EMS objectives since the environment in the MG is unpredictable. The other subset of RL involves actor-critic methods, which is a compromise between value-based and policy-based methods. In the actor-critic, the actor specifies the action based on the policy while the critic assesses the value of each action.

To construct an EMS that reduces the peak demand and energy costs, Mocanu et al. [87] compared the SPG to DQN. The authors neglected the importance of the ESS as a core component of the EMS in designing the EMS in this system. The MG EMS involved the DR function through peak time-shifting and arbitrage in [75], where loads were divided into price-based and temperature-based categories. In this study, the WT and ESS, in addition to the aforementioned flexible loads, can sell surplus energy generation to the utility grid. The authors offered a proper examination of various DRL algorithms in this study, including state-action-reward-state-action (SARSA), DDQN, actor-critic, proximal policy optimization (PPO), and A3C, in order to address the EMS of the MG problem. They discovered that the learning system convergence would improve when PPO and A3C adopted a semi-deterministic approach through action selection from the replay experience buffer in the exploration. Table 2 summarizes the above-mentioned ML applications in the MGMS and the related research.

4. Future Trends in DRL applications in a Microgrid Management System

Since the EMS is an application in the MGMS, which controls all element performances, other above-mentioned applications are under the supervision of this function of the MGMS. Hence, as a future trend in the MG, we mainly focus on the EMS. MG optimization with traditional ML techniques and initial reinforcement learning techniques require comprehensive knowledge on MG environments and element behaviors or predicting behaviors by forecasting tools. On the other hand, MG high penetration calls for more interaction with the utility grid. In this scenario, the MG acts as a prosumer and requires scheduling in a fully stochastic environment. As it is model-free, we mainly focused on model-free enhanced RL and DRL methods in this section to match the MG environment. Following this idea, we divided the MG, deploying (as a member of an ADN) the challenges into three scopes—the MG interaction with the utility grid, the MDP for the MG, and RL technique candidates.

4.1. Microgrid and The Utility Grid Interaction

As discussed before, there are two trends in the scheduling the EMS of the MG. First, the MG works independently and supplies remote places and rural areas. In the second approach, the MG bides to the utility grid. However, the second approach should support the first situation in the islanded mode. The main difference will be in the cost functions, where the persistence in supplying energy is more important than the cost of energy

production in an independent approach. Scholars considered three styles for the grid-connected method—the MG is just a consumer for the utility grid, the MG is a prosumer, and the MG acts in a multi-agent environment. The MG participates in the electricity market or provides ancillary services for the utility grid in the MG's prosumer fashion. This participation in the electricity market can be directly controlled by the DMS or through time-based prices. Time-based prices are CPP, TOU, or RTP. Researchers considered price as a stochastic parameter in scheduling the MG by deploying TOU and RTP techniques. However, analyzing the direct control of the MG by the DMS and adopting a strategy for deploying this direct control was ignored throughout literature. The main issue in providing cooperation between the MG and the utility grid is the uncertainty of the RES generation impressing on the MG stability and power quality. The new version of IEEE 1547 facilitates the MG integration to the utility grid by a compelling arid via characteristics for inverters to provide the desired levels of the voltage and frequency of the utility grid. RL has been applied to this area [72]. Another area that ML assists the MGMS with is in providing ancillary services of the utility grid [94]. The advent of the MG and shortening the distance of the generation side and end-user's hierarchical power grid structure were altered to a new concept called multi-agent. In this structure, all grid elements are equipped with bidirectional smart grid applications. Hence in the MAS, grid elements based on their capacities participate in electricity generation or consumption while coordinating with their neighbor agents. Each agent has its management system and utilizes RL to optimize and predict its status for the MGMS [94].

4.2. Markov Decision Process of MG

To optimize the EMS performance, first, it is required to investigate formation of the MDP of the MG. As discussed in Section 3.1.4, the state, action, transition function, and reward function, in the form of a four-tuple $\{S, A, P, R\}$, define MDP. The state includes the present conditions of each element, which are further based on the action change to the next state. Figure 6 shows the amount of stored energy in the ESS and EV, the energy consumption of the load, market price, and output power of each generation unit define the state in the MDP model. The traditional methods predict and model the dynamic characteristics of the states. On the contrary, RL (utilizing the previous state's information) determines the system element variables sequentially. The real information databases are usually applied to prove the accuracy of the deployed RL method results. The transition functions in model-free methods are unknown, which we discussed previously in Section 3.1.4. Given the contribution of the MG elements in energy provision, actions are defined. Charging, discharging, and remaining idle are ESS actions, which usually determine how the MG bides the utility grid (as a consumer or generator). The objective function of the MG is the principal factor of the reward. The reward function (to minimize the cost of the system or load shedding directly) applies to the action. One of the challenging elements of the MG is when the load takes part in the MG scheduling through load shedding. The elasticity of the responsive loads in providing a better MG operation, especially in contingencies, was less analyzed and addressed in the literature.

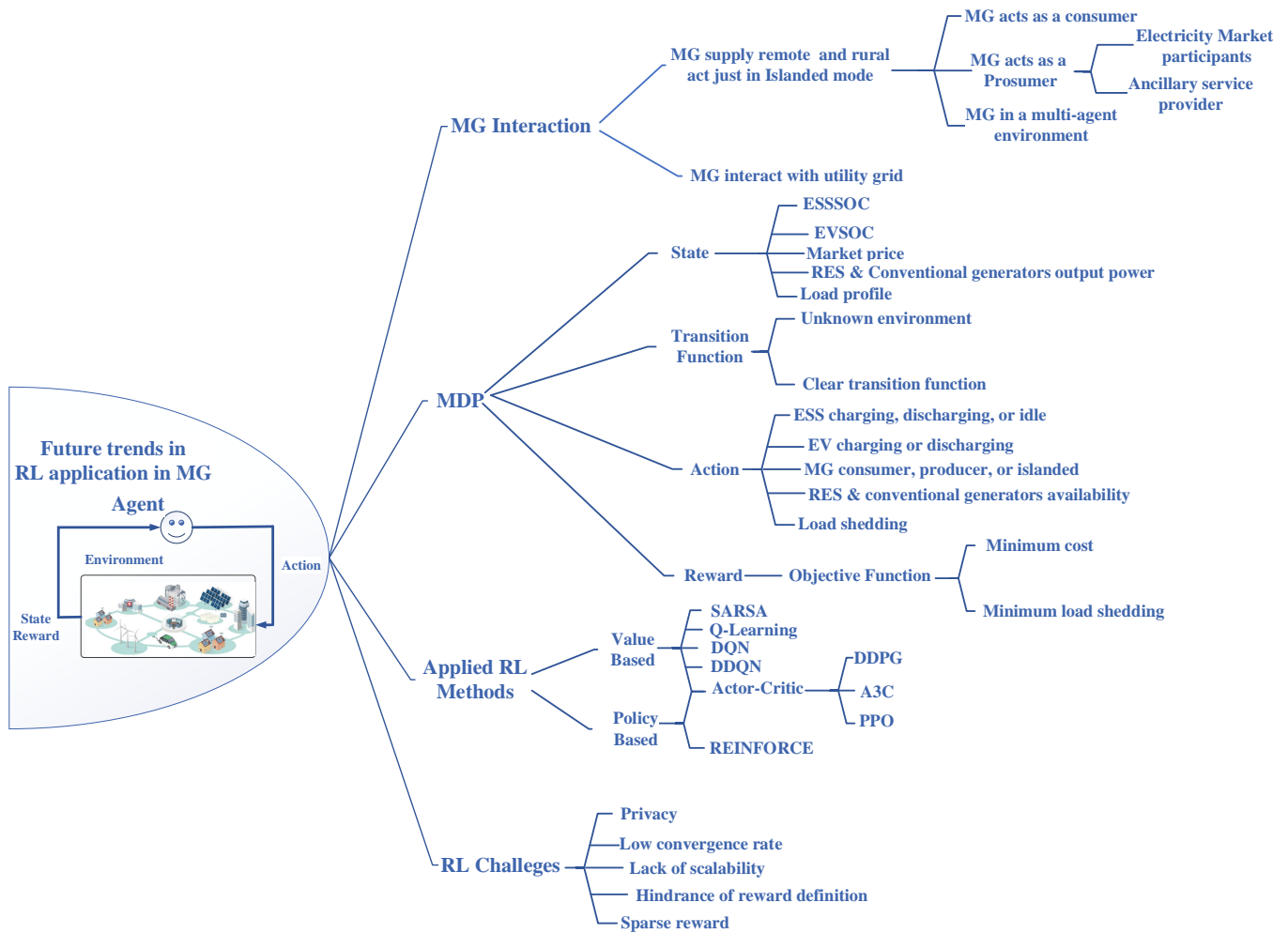


Figure 6. Future trends and requirements of MG scheduling as the utility grid service provider.

4.3. Appropriate Reinforcement Learning Technique

As stated before, value-based RL has some limitations, such as weak convergence for high-dimensional problems. These methods were improved in recent years, such as DQN, which is used to overcome this problem. Rather than creating a table for estimating value functions, the DQN uses a DNN to estimate these values and provides it with i.i.d input through a buffer called replay memory (and containing estimated Q-values). Two separate DNNs are used for Q-value estimations, one working online and the other updating the target network after a predefined time interval. Indeed, value-based methods perform well when dealing with discrete actions; however, the slow pace of policy change makes them vulnerable to overestimation. Efforts to balance value-based and policy-based methods, are achieved through the actor–critic method. Applying deep learning and taking advantage of the aforementioned solutions, the actor–critic method shows better results. Studies show stochastic algorithms follow a semi-deterministic approach in the exploration and exploitation of the environment and perform better. However, there is no comprehensive study that compares the performances of the methods in real situations.

Several challenges are associated with deploying RL as an EMS solution, including privacy concerns, low convergence rates, scalability issues, and the implicit reward definition. Table 3 reveals DRL challenges and research works that propose solutions for these issues.

Table 3. DRL challenges.

Ref	Technique	Main Objective	Advantages to traditional RL	Tools
[95]	Actor-critic DRL + FL	Privacy of users in load scheduling (DR)	Updating NN parameters into a semidefinite programs to deal with non-convex power flow constraints	MATLAB/CVX with MOSEK solver and PYTORCH library in PYTHON
[96]	DDQN+FL	Light weight DRL agents for resource-constrained edge nodes in cluster MG	Improved stability and adaptability to dynamic environments	TensorFlow
[97]	SOCP+ Model based DRL(MuZero)	Residential microgrid EMS	Reduce dimension of action space and consequently convergence time	Python with PyTorch
[98]	MINLP + DDQN	DG and battery scheduling of MG	Reduce dimension of action space space and the complexity of constraints	Python and the PYPOWER 5.1.4 package
[99]	MINLP+ Q-learning	daily and emission costs minimization of a campus MG	Reduce dimension of MINLP arrangement of MG EMS to subproblems with help of RL	DICOPT solver of GAMS
[100]	ADMM+DRL	optimal power flow	Fast convergence	Matpower

4.3.1. Privacy

A MG as an ADN element contributes to the electricity market and offers services to adjacent utility grids. This schema requiring MGs share information with the higher supervisory levels of the power system, such as the DSO/transmission system operator (TSO). Additionally, the consumers and different generation units within the MG require keeping privacy in their interactions with the EMS unit. Researchers follow solutions, such as multi-agent RL and distributed RL with the contribution to the cloud and edge computing arrangements of cost functions to solve this problem. The most recent trend in RL privacy provision involves federated learning RL (FRL), which shows superiority over distributed RL in providing the privacy since, in the multi-agent DRL, the main agent requires collecting data from different agents. FRL distracts the data set into several parties to find a model for each party with a minimum difference from the expected model known as performance loss [101]. FL is solely deployed to implement edge computing in a cluster of MGs and solve load forecasting problems [102]. The FL and SAC combination provides scalability and security in the charging and discharging schedules of EVs [103]. FL also simplifies the non-convex arrangement of load flow constraints in the power system [95]. Therefore, in addition to privacy, the distribution structure of FL resilience respects restrictions in the MG environment, and attention surrounds deploying FRL in solving the EMS. Recently, FRL was hired to optimize the edge computing costs of clustered MGs [96].

4.3.2. Low Convergence Rate

Extreme constraints and high dimensions of action spaces result in inefficient DRL in training the agent and low-speed convergence or even divergence of the learning process. Complex power flow constraints, the non-linear model of BESS, and the equivalency in generation and consumption are examples of evidence that make the learning process unable to converge. The complexity of cost functions and operational constraints can be seen in [97], while the main EMS solver is a model-based DRL. MINLP solves the power flow and applies battery constraints to the MG before deploying DQN to schedule the battery and DG in a MG located in IEEE 10-bus and 69-bus systems [98].

4.3.3. Lack of Scalability

The scalability problem can be looked at from two different views. The first issue is related to the high-dimensional problems, which can already be satisfied by attaching a neural network and providing DRL. The other aspect of the scalability problem is associated with the supporting alteration in the environmental elements and constraints. Since the reward function in DRL is an intricate task and almost impossible to deploy explicit reward, any system variation will result in a lot of effort to justify the reward function. A mathematical model-based algorithm to calculate the reward function can improve RL scalability.

4.3.4. Hindrance of Implicit Reward Definition and Sparse Reward

In DRL, the agent will not learn the best action directly with the policy. It is a reward that guides the agent to the best action. The difficulties related to the implicit reward in DRL, which result in a lack of scalability, are discussed in Section 4.3.3. However, there is another issue related to the reward, which is known as a sparse reward. It is common to implement MG optimization problem constraints using very negative rewards to avoid agents falling into these pitfalls. This approach in the reward definition is also known as a sparse reward and usually results in episode termination before the agent explores and learns valuable states since it stops the agent from exploring the environment accurately. Transfer learning is one of the solutions combined with RL and provides subtasks for the agent to learn constraints and cost functions in different steps. In [99], DRL is associated with alternating the direction method of multipliers (ADMM) to address the sparse reward difficulty of a load flow problem. It is also possible to provide high-density rewards instead of sparse rewards by applying the normalized flow method (NF) [104,105].

5. Conclusions

Regarding power systems, RES penetration development is an egressed strategy for the energy crisis, such as energy resource shortages, environmental pollution, and climate change. Since MG facilitates control and manages the nature-based characteristics of RESs, it is a novel way of utilizing RESs in the grid. ML techniques are tools used for implementing MGMSs considered in this paper, based on MGMS application requirements. A performance optimization study was conducted on a MG to make it an ADN component. After the classification of related research, results showed that the applications of RL and DRL in MG scheduling optimization have risen as they satisfy the MG's stochastic environment. Moreover, our study revealed that the use of a MG as a member of an ADN requires more research in some aspects, including the arrangement of the MG interaction with the utility grid, the establishment of the MDP for the MG environment, and the utilization of deserved RL techniques. Additionally, the use of DRL for MGMSs presents challenges, such as privacy concerns, low convergence rates, scalability issues, and implicit reward definitions. We addressed these hindrances in our study by combining DRL with model-based methods, TL, FL, and NF.

The modeling of MG components and their cost functions will be part of our future efforts regarding the MG role in the ADN. Additionally, we will use the MDP of the MG based on this arrangement to solve EMS issues, highlight the DRL weak points, and use the remedies discussed in this paper.

Author Contributions: Conceptualization, L.T. and J.Y.; methodology, L.T.; software, L.T.; validation, L.T. and J.Y.; formal analysis, L.T.; investigation, L.T.; writing—original draft preparation, L.T. and J.Y.; writing—review and editing, L.T., J.Y.; visualization, L.T.; supervision, J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1F1A1063640).

Data Availability Statement: The data used in this study are reported in the paper's figures and tables.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADMM	alternating direction method of multiplier(s)
ADN	active power distribution network
AI	artificial intelligence
AMI	advanced metering interface
ANN	artificial neural network
AR	autoregressive
ARMA	autoregressive moving average
AS	ancillary services
BEMS	building energy management system(s)
BPNN	back propagation neural network
BMS	battery management system
CSI	current source inverter(s)
CCP	critical peak pricing
CNN	convolutional neural network
DBNN	deep belief neural network
D4PG	distributional deep deterministic policy gradient(s)
DDPG	deep deterministic policy gradient
DDQN	double DQN
DG	Distributed Generators
DGP	deterministic gradient policy
DMS	distribution management system
DNN	deep neural network
DQN	deep Q-network
DR	demand response
DRL	deep reinforcement learning
DSO	distribution system operator
EMS	energy management system
ESS	energy storage system
EV	electric vehicle
FLISR	fault location isolation and service restoration
GA	genetic algorithm
GRU	gated recurrent units
HPNN	hybrid particle swarm optimization-NN
HEMS	home energy management system
HGNN	hybrid GA-NN
LSTM	long short-term memory
MADDPG	multi-agent DDPG
MAE	mean absolute error
MAPE	mean absolute percentage error
MAS	multi-agent system
MDP	Markov decision process
MG	microgrid
MGMS	microgrid management system
MINLP	mixed integer non-linear programming
ML	machine learning
MLR	multiple linear regression
MPPT	maximum power point tracking
MSE	mean square error
OPF	online power flow
PCC	point of common coupling
PM	preventive maintenance
PPO	proximal policy optimization
PV	photovoltaic(s)

RDPG	recurrent deterministic policy gradient
RES	renewable energy resource
RL	reinforcement learning
rRMSD	root mean squared deviation
RTP	real-time pricing
SAC	soft actor-critic
SARIMAX	seasonal auto-regressive integrated moving average
SARSA	state-action—reward-state-action
SCA	smart connection arrangement
SGP	stochastic gradient policy
SPG	stochastic policy gradient
SVM	support vector machine
TD3	twin delayed deep deterministic
TOU	time-of-use
TSO	transmission system operator
VSI	voltage source control inverter
VVO	volt-var optimization
WT	wind turbine

References

1. Tightiz, L.; Yang, H.; Piran, M.J. A survey on enhanced smart micro-grid management system with modern wireless technology contribution. *Energies* **2020**, *13*, 2258. [\[CrossRef\]](#)
2. Tightiz, L.; Yang, H. A comprehensive review on IoT protocols' features in smart grid communication. *Energies* **2020**, *13*, 2762. [\[CrossRef\]](#)
3. Córdova, S.; Cañizares, C.A.; Lorca, Á.; Olivares, D.E. Frequency-Constrained Energy Management System for Isolated Microgrids. *IEEE Trans. Smart Grid* **2022**, *13*, 3394–3407. [\[CrossRef\]](#)
4. Gu, W.; Wu, Z.; Bo, R.; Liu, W.; Zhou, G.; Chen, W.; Wu, Z. Modeling, planning and optimal energy management of combined cooling, heating and power microgrid: A review. *Int. J. Electr. Power Energy Syst.* **2014**, *54*, 26–37. [\[CrossRef\]](#)
5. Minchala-Avila, L.L.; Garza-Castañón, L.E.; Vargas-Martínez, A.; Zhang, Y. A review of optimal control techniques applied to the energy management and control of microgrids. *Procedia Comput. Sci.* **2015**, *52*, 780–787. [\[CrossRef\]](#)
6. Meng, L.; Sanseverino, E.R.; Luna, A.; Dragicevic, T.; Vasquez, J.C.; Guerrero, J.M. Microgrid supervisory controllers and energy management systems: A literature review. *Renew. Sustain. Energy Rev.* **2016**, *60*, 1263–1273. [\[CrossRef\]](#)
7. Khan, A.A.; Naeem, M.; Iqbal, M.; Qaisar, S.; Anpalagan, A. A compendium of optimization objectives, constraints, tools and algorithms for energy management in microgrids. *Renew. Sustain. Energy Rev.* **2016**, *58*, 1664–1683. [\[CrossRef\]](#)
8. Olatomiwa, L.; Mekhilef, S.; Ismail, M.S.; Moghavvemi, M. Energy management strategies in hybrid renewable energy systems: A review. *Renew. Sustain. Energy Rev.* **2016**, *62*, 821–835. [\[CrossRef\]](#)
9. Zia, M.F.; Elbouchikhi, E.; Benbouzid, M. Microgrids energy management systems: A critical review on methods, solutions, and prospects. *Appl. Energy* **2018**, *22*, 1033–1055. [\[CrossRef\]](#)
10. Faisal, M.; Hannan, M.A.; Ker, P.J.; Hussain, A.; Mansor, M.B.; Blaabjerg, F. Review of energy storage system technologies in microgrid applications: Issues and challenges. *IEEE Access* **2018**, *6*, 35143–35164. [\[CrossRef\]](#)
11. Han, Y.; Zhang, K.; Li, H.; Coelho, E.A.A.; Guerrero, J.M. MAS-Based Distributed Coordinated Control and Optimization in Microgrid and Microgrid Clusters: A Comprehensive Overview. *IEEE Trans. Power Electron.* **2018**, *33*, 6488–6508. [\[CrossRef\]](#)
12. Agüera-Pérez, A.; Palomares-Salas, J.C.; De La Rosa, J.J.G.; Florencias-Oliveros, O. Weather forecasts for microgrid energy management: Review, discussion and recommendations. *Appl. Energy* **2018**, *228*, 265–278. [\[CrossRef\]](#)
13. Roslan, M.F.; Hannan, M.A.; Ker, P.J.; Uddin, M.N. Microgrid control methods toward achieving sustainable energy management. *Appl. Energy* **2019**, *240*, 583–607. [\[CrossRef\]](#)
14. García Vera, Y.E.; Dufo-López, R.; Bernal-Agustín, J.L. Energy management in microgrids with renewable energy sources: A literature review. *Appl. Sci.* **2019**, *9*, 3854. [\[CrossRef\]](#)
15. Boqtob, O.; El Moussaoui, H.; El Markhi, H.; Lamhamdi, T. Microgrid energy management system: A state-of-the-art review. *J. Electr. Syst.* **2019**, *15*, 53–67.
16. Mosavi, A.; Salimi, M.; Faizollahzadeh Ardabili, S.; Rabczuk, T.; Shamshirband, S.; Varkonyi-Koczy, A.R. State of the art of machine learning models in energy systems, a systematic review. *Energies* **2019**, *12*, 1301. [\[CrossRef\]](#)
17. Espín-Sarzosa, D.; Palma-Behnke, R.; Núñez-Mata, O. Energy management systems for microgrids: Main existing trends in centralized control architectures. *Energies* **2020**, *13*, 547. [\[CrossRef\]](#)
18. Yang, T.; Zhao, L.; Li, W.; Zomaya, A.Y. Reinforcement learning in sustainable energy and electric systems: A survey. *Annu. Rev. Control* **2020**, *49*, 145–163. [\[CrossRef\]](#)
19. Wang, Z.; Hong, T. Reinforcement learning for building controls: The opportunities and challenges. *Appl. Energy* **2020**, *269*, 115036. [\[CrossRef\]](#)

20. Zhang, Z.; Zhang, D.; Qiu, R.C. Deep reinforcement learning for power system applications: An overview. *CSEE J. Power Energy Syst.* **2020**, *6*, 213–225.
21. Cao, D.; Hu, W.; Zhao, J.; Zhang, G.; Zhang, B.; Liu, Z. Reinforcement learning and its applications in modern power and energy systems: A review. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1029–1042. [[CrossRef](#)]
22. Chaudhary, G.; Lamb, J.J.; Burheim, O.S.; Austbø, B. Review of energy storage and energy management system control strategies in microgrids. *Energies* **2021**, *14*, 4929. [[CrossRef](#)]
23. Battula, A.; Vuddanti, S.; Salkuti, S.R. Review of energy management system approaches in microgrids. *Energies* **2021**, *14*, 5459. [[CrossRef](#)]
24. Raya-Armenta, J.M.; Bazmohammadi, N.; Avina-Cervantes, J.G.; Saez, D.; Vasquez, J.C.; Guerrero, J.M. Energy management system optimization in islanded microgrids: An overview and future trends. *Renew. Sustain. Energy Rev.* **2021**, *149*, 111327. [[CrossRef](#)]
25. Ali, S.; Zheng, Z.; Aillerie, M.; Sawicki, J.P.; Pera, M.C.; Hissel, D. A review of DC Microgrid energy management systems dedicated to residential applications. *Energies* **2021**, *14*, 4308. [[CrossRef](#)]
26. Zahraoui, Y.; Alhamrouni, I.; Mekhilef, S.; Basir Khan, M.R.; Seyedmahmoudian, M.; Stojcevski, A.; Horan, B. Energy management system in microgrids: A Comprehensive review. *Sustainability* **2021**, *13*, 10492. [[CrossRef](#)]
27. Muqeet, H.A.; Munir, H.M.; Javed, H.; Shahzad, M.; Jamil, M.; Guerrero, J.M. An energy management system of campus microgrids: State-of-the-art and future challenges. *Energies* **2021**, *14*, 6525. [[CrossRef](#)]
28. Chen, X.; Qu, G.; Tang, Y.; Low, S.; Li, N. Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. *IEEE Trans. Smart Grid* **2022**, *13*, 2935–2958. [[CrossRef](#)]
29. Yu, L.; Qin, S.; Zhang, M.; Shen, C.; Jiang, T.; Guan, X. A review of deep reinforcement learning for smart building energy management. *IEEE Internet Things J.* **2021**, *8*, 12046–12063. [[CrossRef](#)]
30. Erick, A.O.; Folly, K.A. Reinforcement learning approaches to power management in grid-tied microgrids: A review. In Proceedings of the Clemson University Power Systems Conference (PSC), Clemson, SC, USA, 10–13 March 2020; pp. 1–6.
31. Erick, A.O.; Folly, K.A. Reinforcement learning techniques for optimal power control in grid-connected microgrids: A comprehensive review. *IEEE Access* **2020**, *8*, 208992–209007.
32. Touma, H.J.; Mansor, M.; Rahman, M.S.A.; Kumaran, V.; Mokhlis, H.B.; Ying, Y.J.; Hannan, M.A. Energy management system of microgrid: Control schemes, pricing techniques, and future horizons. *Int. J. Energy Res.* **2021**, *45*, 12728–12739. [[CrossRef](#)]
33. Xie, P.; Guerrero, J.M.; Tan, S.; Bazmohammadi, N.; Vasquez, J.C.; Mehrzadi, M.; Al-Turki, Y. Optimization-based power and energy management system in shipboard microgrid: A review. *IEEE Syst. J.* **2022**, *16*, 578–590. [[CrossRef](#)]
34. Arunkumar, A.P.; Kuppusamy, S.; Muthusamy, S.; Pandiyan, S.; Panchal, H.; Nagaiyan, P. An extensive review on energy management system for microgrids. *Energy Sources Part Recovery Util. Environ. Eff.* **2022**, *44*, 4203–4228. [[CrossRef](#)]
35. Roslan, M.F.; Hannan, M.A.; Ker, P.J.; Mannan, M.; Muttaqi, K.M.; Mahlia, T.I. Microgrid control methods toward achieving sustainable energy management: A bibliometric analysis for future directions. *J. Clean. Prod.* **2022**, *348*, 131340. [[CrossRef](#)]
36. Awan, M.M.A.; Javed, M.Y.; Asghar, A.B.; Ejsmont, K. Economic Integration of Renewable and Conventional Power Sources—A Case Study. *Energies* **2022**, *15*, 2141. [[CrossRef](#)]
37. Punna, S.; Manthathi, U.B. Optimum design and analysis of a dynamic energy management scheme for HESS in renewable power generation applications. *SN Appl. Sci.* **2020**, *2*, 1–13. [[CrossRef](#)]
38. Jan, M.U.; Xin, A.; Rehman, H.U.; Abdelbaky, M.A.; Iqbal, S.; Aurangzeb, M. Frequency regulation of an isolated microgrid with electric vehicles and energy storage system integration using adaptive and model predictive controllers. *IEEE Access* **2021**, *9*, 14958–14970. [[CrossRef](#)]
39. *IEEE Std 1547-2003*; IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems. IEEE: New York, NY, USA, 2003; pp. 1–28.
40. *IEEE Std 1547-2018*; IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces. IEEE: New York, NY, USA, 2018; pp. 1–138.
41. Uz Zaman, M.S.; Haider, R.; Bukhari, S.B.A.; Ashraf, H.M.; Kim, C.-H. Impacts of responsive loads and energy storage system on frequency response of a multi-machine power system. *Machines* **2019**, *7*, 34. [[CrossRef](#)]
42. Tightiz, L.; Yang, H.; Bevrani, H. An Interoperable Communication Framework for Grid Frequency Regulation Support from Microgrids. *Sensors* **2021**, *21*, 4555. [[CrossRef](#)] [[PubMed](#)]
43. Mirafzal, B.; Adib, A. On grid-interactive smart inverters: Features and advancements. *IEEE Access* **2020**, *8*, 160526–160536. [[CrossRef](#)]
44. Arbab-Zavar, B.; Palacios-Garcia, E.J.; Vasquez, J.C.; Guerrero, J.M. Smart inverters for microgrid applications: A review. *Energies* **2019**, *12*, 840. [[CrossRef](#)]
45. Das, C.K.; Bass, O.; Kothapalli, G.; Mahmoud, T.S.; Habibi, D. Overview of energy storage systems in distribution networks: Placement, sizing, operation, and power quality. *Renew. Sustain. Energy Rev.* **2018**, *91*, 1205–1230. [[CrossRef](#)]
46. Li, Q.; Gao, M.; Lin, H.; Chen, Z.; Chen, M. MAS-based distributed control method for multi-microgrids with high-penetration renewable energy. *Energy* **2019**, *171*, 284–295. [[CrossRef](#)]
47. Hirsch, A.; Parag, Y.; Guerrero, J. Microgrids: A review of technologies, key drivers, and outstanding issues. *Renew. Sustain. Energy Rev.* **2018**, *90*, 402–411. [[CrossRef](#)]

48. Hui, H.; Ding, Y.; Shi, Q.; Li, F.; Song, Y.; Yan, J. 5G network-based Internet of Things for demand response in smart grid: A survey on application potential. *Appl. Energy* **2020**, *257*, 113972. [[CrossRef](#)]
49. Spodniak, P.; Ollikka, K.; Honkapuro, S. The impact of wind power and electricity demand on the relevance of different short-term electricity markets: The Nordic case. *Appl. Energy* **2021**, *283*, 116063. [[CrossRef](#)]
50. Mazidi, P.; Bobi, M.A.S. Strategic maintenance scheduling in an islanded microgrid with distributed energy resources. *Electr. Power Syst. Res.* **2017**, *148*, 171–182. [[CrossRef](#)]
51. Cheng, L.; Yu, T. A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems. *Int. J. Energy Res.* **2019**, *43*, 1928–1973. [[CrossRef](#)]
52. Hao, J.; Ho, T.K. Machine learning made easy: A review of scikit-learn package in python programming language. *J. Educ. Behav. Stat.* **2019**, *44*, 348–361. [[CrossRef](#)]
53. Bower, T. *Introduction to Computational Engineering with MATLAB®*, 1st ed.; Chapman & Hall/CRC Numerical Analysis and Scientific Computing Series; CRC: Boca Raton, FL, USA, 2022.
54. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D. Stan: A probabilistic programming language. *J. Stat. Softw.* **2017**, *76*, 1–32. [[CrossRef](#)]
55. Vázquez-Canteli, J.R.; Nagy, Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Appl. Energy* **2019**, *235*, 1072–1089. [[CrossRef](#)]
56. Zheng, Y.; Suryanarayanan, S.; Maciejewski, A.A.; Siegel, H.J.; Hansen, T.M.; Celik, B. An application of machine learning for a smart grid resource allocation problem. In Proceedings of the IEEE Milan PowerTech, Milano, Italy, 23–27 June 2019; pp. 1–6.
57. Kapetanakis, D.S.; Christantoni, D.; Mangina, E.; Finn, D. Evaluation of machine learning algorithms for demand response potential forecasting. In Proceedings of the 15th International Building Performance Simulation Association Conference, San Francisco, CA, USA, 7–9 August 2017.
58. Percy, S.D.; Aldeen, M.; Berry, A. Residential demand forecasting with solar-battery systems: A survey-less approach. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1499–1507. [[CrossRef](#)]
59. Nakabi, T.A.; Toivanen, P. An ANN-based model for learning individual customer behavior in response to electricity prices. *Sustain. Energy Grids Netw.* **2019**, *18*, 100212. [[CrossRef](#)]
60. Wang, X.; Guo, P.; Huang, X. A review of wind power forecasting models. *Energy Procedia* **2011**, *12*, 770–778. [[CrossRef](#)]
61. Dongmei, Z.; Yuchen, Z.; Xu, Z. Research on wind power forecasting in wind farms. In Proceedings of the IEEE Power Engineering and Automation Conference, Wuhan, China, 8–9 September 2011; Volume 1, pp. 175–178.
62. Perera, K.S.; Aung, Z.; Woon, W.L. Machine learning techniques for supporting renewable energy generation and integration: A survey. In Proceedings of the International Workshop on Data Analytics for Renewable Energy Integration, Nancy, France, 19 September 2014; pp. 81–96.
63. Li, Y.; Wang, R.; Yang, Z. Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting. *IEEE Trans. Sustain. Energy* **2021**, *13*, 159–169. [[CrossRef](#)]
64. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [[CrossRef](#)]
65. Yagli, G.M.; Yang, D.; Srinivasan, D. Automatic hourly solar forecasting using machine learning models. *Renew. Sustain. Energy Rev.* **2019**, *105*, 487–498. [[CrossRef](#)]
66. Eseye, A.T.; Zhang, J.; Zheng, D. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. *Renew. Energy* **2018**, *118*, 357–367. [[CrossRef](#)]
67. Martín, L.; Zarzalejo, L.F.; Polo, J.; Navarro, A.; Marchante, R.; Cony, M. Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Sol. Energy* **2010**, *84*, 1772–1781. [[CrossRef](#)]
68. Shirzadi, N.; Nasiri, F.; El-Bayeh, C.; Eicker, U. Optimal dispatching of renewable energy-based urban microgrids using a deep learning approach for electrical load and wind power forecasting. *Int. J. Energy Res.* **2022**, *46*, 3173–3188. [[CrossRef](#)]
69. Yang, R.; Liu, H.; Nikitas, N.; Duan, Z.; Li, Y.; Li, Y. Short-term wind speed forecasting using deep reinforcement learning with improved multiple error correction approach. *Energy* **2022**, *239*, 122128. [[CrossRef](#)]
70. Wu, L.; Kaiser, G.; Rudin, C.; Anderson, R. Data quality assurance and performance measurement of data mining for preventive maintenance of power grid. In Proceedings of the First International Workshop on Data Mining for Service and Maintenance, San Diego, CA, USA, 24–28 August 2011; pp. 28–32.
71. Rao, K.U.; Parvatikar, A.G.; Gokul, S.; Nitish, N.; Rao, P. A novel fault diagnostic strategy for PV micro grid to achieve reliability centered maintenance, In Proceedings of the International Conference on Power Electronics, Intelligent Control and Energy Systems, Delhi, India, 4–6 July 2016; pp. 1–4.
72. Rocchetta, R.; Bellani, L.; Compare, M.; Zio, E.; Patelli, E. A reinforcement learning framework for optimal operation and maintenance of power grids. *Appl. Energy* **2019**, *241*, 291–301. [[CrossRef](#)]
73. Tightiz, L.; Yang, H. Resilience Microgrid as Power System Integrity Protection Scheme Element With Reinforcement Learning Based Management. *IEEE Access* **2021**, *9*, 83963–83975. [[CrossRef](#)]
74. Eseye, A.T.; Lehtonen, M.; Tukia, T.; Uimonen, S.; Millar, R.J. Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems. *IEEE Access* **2019**, *7*, 91463–91475. [[CrossRef](#)]
75. Nakabi, T.A.; Toivanen, P. Deep reinforcement learning for energy management in a microgrid with flexible demand. *Sustain. Energy Grids Netw.* **2021**, *25*, 100413. [[CrossRef](#)]

76. Pallonetto, F.; De Rosa, M.; Milano, F.; Finn, D. P. Demand response algorithms for smart-grid ready residential buildings using machine learning models. *Appl. Energy* **2019**, *239*, 1265–1282. [[CrossRef](#)]
77. Singh, P.K.; Singh, N.; Negi, R. Wind power forecasting using hybrid ARIMA-ANN technique. *Ambient. Commun. Comput. Syst.* **2019**, *907*, 209–220
78. Zaman, U.; Teimourzadeh, H.; Sangani, E.H.; Liang, X.; Chung, C.Y. Wind speed forecasting using ARMA and neural network models. In Proceedings of the IEEE Electrical Power and Energy Conference (EPEC), Toronto, ON, Canada, 21 October 2021; pp. 243–248.
79. Li, K.; Zhang, J.; Chen, X.; Xue, W. Building's hourly electrical load prediction based on data clustering and ensemble learning strategy. *Energy Build.* **2022**, *261*, 111943. [[CrossRef](#)]
80. Kuznetsova, E.; Li, Y.F.; Ruiz, C.; Zio, E.; Ault, G.; Bell, K. Reinforcement learning for microgrid energy management. *Energy* **2013**, *59*, 133–146. [[CrossRef](#)]
81. François-Lavet, V.; Taralla, D.; Ernst, D.; Fonteneau, R. Deep reinforcement learning solutions for energy microgrids management. In Proceedings of the European Workshop on Reinforcement Learning, Barcelona, Spain, 7 December 2016.
82. Domínguez-Barbero, D.; García-González, J.; Sanz-Bobi, M.A.; Sánchez-Úbeda, E.F. Optimising a microgrid system by deep reinforcement learning techniques. *Energies* **2020**, *13*, 2830. [[CrossRef](#)]
83. Mbuwir, B.V.; Ruelens, F.; Spiessens, F.; Deconinck, G. Battery energy management in a microgrid using batch reinforcement learning. *Energies* **2017**, *10*, 1846. [[CrossRef](#)]
84. Ji, Y.; Wang, J.; Xu, J.; Fang, X.; Zhang, H. Real-time energy management of a microgrid using deep reinforcement learning. *Energies* **2019**, *12*, 2291. [[CrossRef](#)]
85. Mahmoud, M.S.; Abouheaf, M.; Sharaf, A. Reinforcement learning control approach for autonomous microgrids. *Int. J. Model. Simul.* **2021**, *41*, 1–10. [[CrossRef](#)]
86. Kim, S.; Lim, H. Reinforcement learning based energy management algorithm for smart energy buildings. *Energies* **2018**, *11*, 2010. [[CrossRef](#)]
87. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M.; Slootweg, J.G. On-line building energy optimization using deep reinforcement learning. *IEEE Trans. Smart Grid* **2018**, *10*, 3698–3708. [[CrossRef](#)]
88. Bui, V.H.; Hussain, A.; Kim, H.M. Double Deep Q-Learning-Based Distributed Operation of Battery Energy Storage System Considering Uncertainties. *IEEE Trans. Smart Grid* **2020**, *11*, 457–469. [[CrossRef](#)]
89. Afrasiabi, M.; Mohammadi, M.; Rastegar, M.; Kargarian, A. Multi-agent microgrid energy management based on deep learning forecaster. *Energy* **2019**, *186*, 115873. [[CrossRef](#)]
90. Areekkara, S.; Kumar, R.; Bansal, R.C. An intelligent multi agent based approach for autonomous energy management in a Microgrid. *Electr. Power Components Syst.* **2021**, *49*, 18–31. [[CrossRef](#)]
91. Shayeghi, H.; Younesi, A. Adaptive and online control of microgrids using multi-agent reinforcement learning. In *Microgrid Architectures, Control and Protection Methods*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 577–602.
92. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602.
93. Van Hasselt, H.; Guez, A.; Silver, D. Deep reinforcement learning with double q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
94. Kezunovic, M.; Pinson, P.; Obradovic, Z.; Grijalva, S.; Hong, T.; Bessa, R. Big data analytics for future electricity grids. *Electr. Power Syst. Res.* **2020**, *189*, 106788. [[CrossRef](#)]
95. Duan, Z.; Qiao, Y.; Chen, S.; Wang, X.; Wu, G.; Wang, X. Lightweight Federated Reinforcement Learning for Independent Request Scheduling in Microgrids. In Proceedings of the IEEE International Conference on Smart Internet of Things (SmartIoT), Suzhou, China, 19–21 August 2022; pp. 208–215.
96. Bahrami, S.; Chen, Y.C.; Wong, V.W.S. Deep Reinforcement Learning for Demand Response in Distribution Networks. *IEEE Trans. Smart Grid* **2021**, *12*, 1496–1506. [[CrossRef](#)]
97. Shuai, H.; He, H. Online Scheduling of a Residential Microgrid via Monte-Carlo Tree Search and a Learned Model. *IEEE Trans. Smart Grid* **2021**, *12*, 1073–1087. [[CrossRef](#)]
98. Shuai, H.; Ai, X.; Fang, J.; Yao, W.; Wen, J. Double Deep Q-learning Based Real-Time Optimization Strategy for Microgrids. *arXiv* **2021**, arXiv:2107.12545.
99. Yoldas, Y.; Goren, S.; Onen, A. Optimal control of microgrids with multi-stage mixed-integer nonlinear programming guided q-learning algorithm. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1151–1159. [[CrossRef](#)]
100. Zeng, S.; Kody, A.; Kim, Y.; Kim, K.; Molzahn, D.K. A reinforcement learning approach to parameter selection for distributed optimal power flow. *Electr. Power Syst. Res.* **2022**, *212*, 108546. [[CrossRef](#)]
101. Bouachir, O.; Aloqaily, M.; Özkasap, Ö.; Ali, F. FederatedGrids: Federated Learning and Blockchain-Assisted P2P Energy Sharing. *IEEE Trans. Green Commun. Netw.* **2022**, *6*, 424–436. [[CrossRef](#)]
102. Firouzi, R.; Rahmani, R.; Kanter, T. Federated Learning for Distributed Reasoning on Edge Computing. *Procedia Comput. Sci.* **2021**, *184*, 419–427. [[CrossRef](#)]
103. Chu, Y.; Wei, Z.; Fang, X.; Chen, S.; Zhou, Y. A Multi-agent Federated Reinforcement Learning Approach for Plug-in Electric Vehicle Fleet Charging Coordination in a Residential Community. *IEEE Access* **2022**, *10*, 98535–98548. [[CrossRef](#)]

-
104. Mazouze, B.; Doan, T.; Durand, A.; Pineau, J.; Hjelm, R. D. Leveraging exploration in off-policy algorithms via normalizing flows. In Proceedings of the Conference on Robot Learning, Virtual Conference, 16–18 May 2020; pp. 430–444.
 105. Ward, P.N.; Smofsky, A.; Bose, A.J. Improving exploration in soft-actor-critic with normalizing flows policies. *arXiv* **2019**, arXiv:1906.02771.