IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 February 2025, accepted 24 February 2025, date of publication 27 February 2025, date of current version 7 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3546578

SURVEY

Reinforcement Learning Solutions for Microgrid Control and Management: A Survey

PEDRO I. N. BARBALHO[1], **ANDERSON L. MORAES**[1], **VINICIUS A. LACERDA**[2], **PEDRO H. A. BARRA**[3], **RICARDO A. S. FERNANDES**[1], **(Senior Member, IEEE), AND DENIS V. COURY**[1]

[1]Department of Electrical and Computing Engineering, São Carlos School of Engineering, University of São Paulo, São Carlos, São Paulo 13566-590, Brazil
[2]Centre d'Innovacio Tecnològica en Convertidors Estatics i Accionaments, Universitat Politècnica de Catalunya (CITCEA-UPC), 08028 Barcelona, Spain
[3]Faculty of Electrical Engineering, Federal University of Uberlândia, Uberlândia 38408-100, Brazil

Corresponding author: Denis V. Coury (coury@sc.usp.br)

This work was supported in part by São Paulo Research Foundation (FAPESP) under Grant 2021/04872-9 and Grant 2023/00182-3; and in part by the National Council for Scientific and Technological Development (CNPq) under Grant 445374/2020-9, Grant 406453/2021-7, Grant 140235/2021-3, and Grant 402334/2023-0.

ABSTRACT A microgrid (MG) is part of a distribution system that comprises loads and distributed energy resources, capable of operating either connected to or islanded from the primary grid. Having an appropriate design, MG controllers improve energy efficiency, playing a vital role in the modern distribution system. Thus, MG management and control has become a broad area of research due to its complex operation. Reinforcement learning (RL) offers adaptive solutions for handling MG complex dynamics and nonlinearity. It is an alternative to traditional algorithms and control methods in tasks, such as load frequency control, resource allocation, and energy management. Due to the relevance of the topic, this survey examined the role of RL in MG control and management, offering a comprehensive update on previous reviews, categorising articles by RL type, control objectives, and MG operational modes. Additionally, hardware implementations and performance assessments across RL-based solutions were evaluated. The present survey identified key research trends and gaps, contributing to understanding the role of RL in MG management and control and guiding future solutions in the field.

INDEX TERMS Distributed energy resources, hierarchical control, microgrid control, reinforcement learning.

I. INTRODUCTION

Distribution system modernisation has been one of the leading research areas recently. In the early stages, studies focused on substation automation, improvement in control, monitoring and measuring equipment, and communication. As the use of distributed energy resources (DER) increased, a new paradigm of electrical grids emerged, called microgrids (MGs). An MG is part of a distribution system that comprises loads and DERs. It can be used in different modes: connected or islanded to the primary grid. An MG enables more efficient energy transfer from generation to loads while reducing greenhouse gas emissions. Moreover, it can improve the distribution system stability by coordinating the DERs with local loads.

MG control has become a broad area of research, as it is challenging to operate such a complex system. Traditionally, classical tools such as proportional-integrative (PI) are used for frequency and voltage control. The work in [1] proposed a decentralized frequency controller using the droop method. Improvements for the droop method to enhance the frequency and voltage controller were proposed in [2]. MG applications also encompass the implementation of optimisation algorithms for cost minimisation and energy, as well as resource management.

Traditional controllers might not operate MGs for all possible scenarios due to their non-linear dynamics. Therefore, adaptive algorithms, such as those based on reinforcement

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

learning (RL), are necessary, as they can perform different tasks, which are: grid-connected virtual synchronous generator control [3]; load frequency control [4]; resource allocation [5]; energy management [6]; and energy arbitrage [7].

Considering RL contributions for control in electrical power systems, some review articles have already discussed this topic. In [8], the increased use of RL for virtual inertia control in islanded MGs was discussed. Some recent works that used RL for electric power system control are shown in [9]. The work in [10] reviewed propositions based on RL for optimal power control in grid-connected MGs. These articles provided relevant insights related to the possible uses of RL algorithms. Finally, studies for grid-connected MGs were presented in [10].

Recently, there has been an in-depth discussion regarding RL algorithms used for MG applications. Model-free RL algorithms for MG control were reviewed in [11]. A comprehensive analysis in [12] explored computational intelligence approaches for MG energy management. Additionally, [13] examined optimisation techniques using metaheuristics for MGs. Finally, [14] provided an update on works regarding AI and MGs.

Given the relevance of MGs in the current scenario of modernisation of distribution systems, this survey discusses RL applications with respect to MG control and management. Over the last five years, the number of studies using RL for MG applications has grown significantly. Therefore, this article updates the RL algorithms found in [9], and adds more research topics compared to [8] and [10]. The present survey considers RL algorithms in a broader sense, not limited to model-free as seen in [11]. Although comprehensive surveys, such as [13] and [14], have extensively explored the applications of AI and computational intelligence in MGs, they did not provide a detailed discussion on MG applications where RL plays a critical role.

Therefore, the present paper adds to the body of knowledge in the field by providing an in-depth analysis of RL-specific solutions for MGs, focusing on their distinct aspects and potential for real-world implementation. This survey divides the articles into categories related to MGs applications. It also compares the solutions of each study in certain aspects, such as the MG type considered; the performance analysis used; the MG operation mode; the type of RL most used, model-based or model-free; as well as shows aspects of the hardware implementation.

The present survey is structured as follows: Section II presents a brief explanation of the MG control, the hierarchical control structure and its characteristics. Section III discusses the basics on RL theory and briefly depicts two fundamentals algorithms used in this area. Section IV shows the evolution of the algorithms used for MG control and management from classical theory to RL, and the main areas studied over the last ten years regarding MGs. Section V compares the article's contributions, presents the research trends and gaps found after reviewing the state-of-the-art articles. Finally, Section VI draws the conclusions of the survey.

II. MICROGRIDS

An MG is an electric grid comprising DERs and local loads, which can operate connected or disconnected from the distribution system. These two operation states are known as grid-connected and islanded modes, respectively. Furthermore, MGs use an energy storage system (ESS) for energy management to improve the power quality and the grid's reliability [15]. Moreover, an MG can connect to neighbouring MGs and exchange energy between areas due to its structure. Fig. 1 illustrates the main elements in an MG.

The hierarchical control of MGs is one of the most used frameworks to coordinate these systems. This control structure is divided into primary, secondary and tertiary levels, differing in function and response speed. Fig. 2 illustrates the MG control structure. The primary level is the fastest and uses local measurements to meet each DER requirement. One of the methods used in this level is the droop control, which changes the DER output power to improve the grid's frequency and the node voltage [15], [16].

The secondary level sends reference signals to the primary controllers to ensure the MG stability and monitors the system to meet power quality requirements. Its speed is slower than the primary one, and it controls the MG frequency and voltage when operating in islanded mode. This level also identifies non-programmable islanding events and is responsible for a smooth transition between the MG operating modes. In addition, it can be centralised or decentralised [15].

The centralised structure receives information from the MG, such as the estimated values of solar irradiation and wind speed, energy cost, power, voltage and current measurements. This secondary level structure is well suited for small MGs to coordinate prosumers with similar goals [17]. Moreover, the communication latency directly influences the controller's stability. The decentralised structure allows prosumers to be more independent in an MG. In this case, controllers act using consensus between the DER local requirements. This structure is commonly used with multi-agents and large MGs, and does not demand a high communication level.

Finally, the tertiary level is the slowest and is responsible for sending reference signals to the secondary one. It changes the MG operating conditions considering the distribution system requirements. Moreover, it is responsible for coordinating interconnected MGs. Finally, when the MG is islanded, only the secondary and primary levels remain operative. Therefore, Section IV provides an overview of the RL algorithm theory and presents how researchers use them for MG control.

III. REINFORCEMENT LEARNING

RL is a machine learning (ML) area widely used for intelligent control and optimisation. This learning is different from supervised or unsupervised learning. RL is based on learning through experience [18]. Using artificial neural

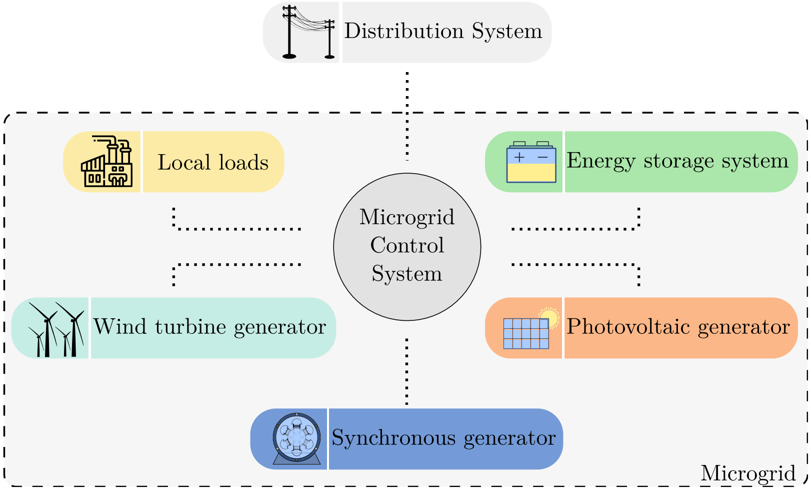


FIGURE 1. MG schematic.

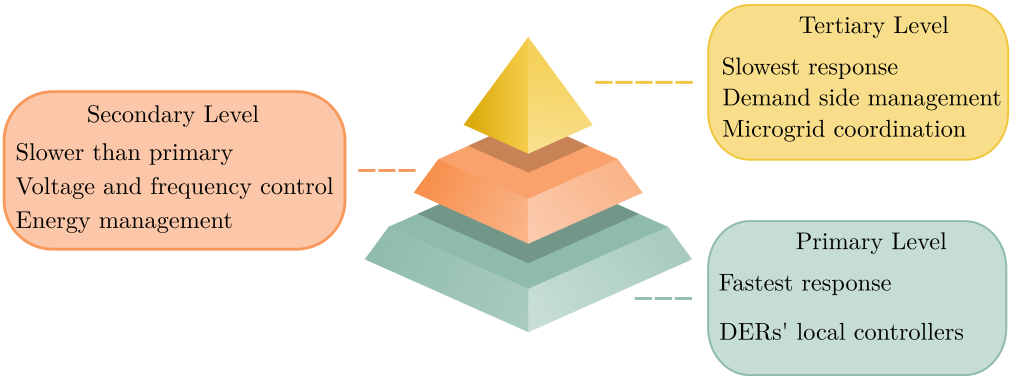


FIGURE 2. Hierarchical control of MGs.

network (ANN) as an example, these types of learning are described as: supervised learning, which adjusts the ANN weights using the error between the network output and the well-known desired value; unsupervised learning, which adjusts the ANN weights to classify clusters in a dataset; and reinforcement learning, which adjusts the ANN weights to maximise the system reward. The better the ANN performance, the better the reward will be.

Reinforcement learning comprises a policy, a reward signal, a value function and, in some cases, the environment model. These elements characterise the agent and how it communicates with the environment. The agent receives the environment states (S) and reward (r) and sends back the computed actions (A) [18], [19]. A schematic of this interaction, considering a Markov decision process (MDP), is illustrated in Fig. 3.

The algorithm awards the agent by evaluating the new state achieved by the environment after the agent's action. The rule that dictates the agent's actions is called policy, which can be stochastic or deterministic [18]. In a stochastic policy, the agent chooses the actions with the highest computed probability to return a good reward. In the second one, the agent does not compute the probability directly.

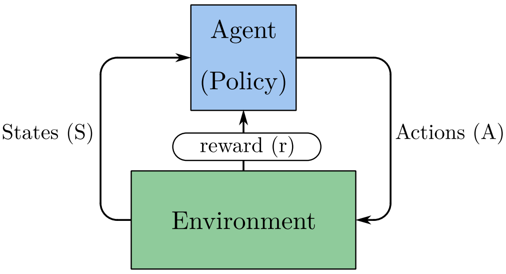


FIGURE 3. MDP diagram.

The main functions and equations underpinning RL theory must be clear to understand how most of the algorithms work. First, consider a finite Markov decision process whose states, rewards, and actions are defined. In this system, the variables S_t and R_t, shown in Fig. 3, have a finite probability distribution and depend only on previous states and actions. The state transition from (1) determines this characteristic [18]:

$$p(s', r|s, a) = Pr\{S_t = s'|S_{t-1} = s, A_{t-1} = a\}. \qquad (1)$$

Equation (1) returns the probability of the agent being rewarded by r in a given state s' reached after transitioning from a previous state s due to action a.

Moreover, the actions and state transitions' set are characterised by equally spaced time intervals. These transitions are finite, and each training episode is limited. The cumulative reward for a given policy, the Return, is given by (2):

$$G_t = R_{t+1} + R_{t+2} + R_{t+3} + R_{t+4} \ldots + R_T, \qquad (2)$$

where T is the episode's final step. The cumulative reward will be finite only if the episode is also limited. In the case of systems with infinite episodes, an adjustment in (2) needs to be defined to limit the final reward. This adjustment, denoted by γ, can be given by discounting the reward of each state transition, as shown by (3):

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} \ldots = \sum_{k=0} \gamma^k R_{t+k+1}. \qquad (3)$$

After formalising some concepts, two fundamental functions can be defined in RL. Using (4) and (5), we have the state value and the action value functions, respectively:

$$v_\pi(s) = E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}|S_t = s], \qquad (4)$$

$$q_\pi(s, a) = E_\pi[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}|S_t = s, A_t = a], \qquad (5)$$

where, function E_π returns the given variable expected value for a given policy π and a specific instant. In (4), the expected value for a given state is calculated, while (5) calculates the expected value of taking a certain action a in a certain state s.

For finite MDPs, the optimal policy returns the maximum value of v_π, which will be the optimal value function, given by (6):

$$v_*(s) = \max_\pi v_\pi(s). \qquad (6)$$

Furthermore, Equation (7) can give optimal state-action value functions:

$$q_*(s, a) = \max_\pi q_\pi(s, a). \qquad (7)$$

The RL algorithms are all based on finding the best policy based on the concept of action-value and the state-value functions. Tabular methods estimate v_* q_* through array or matrix representations. Currently, one of the most relevant tabular methods is Q-learning (QL), based on (8):

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)]. \qquad (8)$$

This algorithm estimates the state-action value function and approximates q_*. This algorithm is considered to be off-policy, which means it learns the optimal policy regardless of the actual policy being followed. As a counterpart, the *sarsa* algorithm is based on the update rule from (9).

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)]. \qquad (9)$$

The *sarsa* algorithm is an on-policy algorithm, which updates the current policy being learned. Note the difference between both update rules, in (9), $Q(S_t, A_t)$ is updated based on the actual estimation of $Q(S_{t+1}, A_{t+1})$, whereas in (8), the Q function is being updated based on the maximum value found among all possible actions $Q(S_{t+1}, a)$.

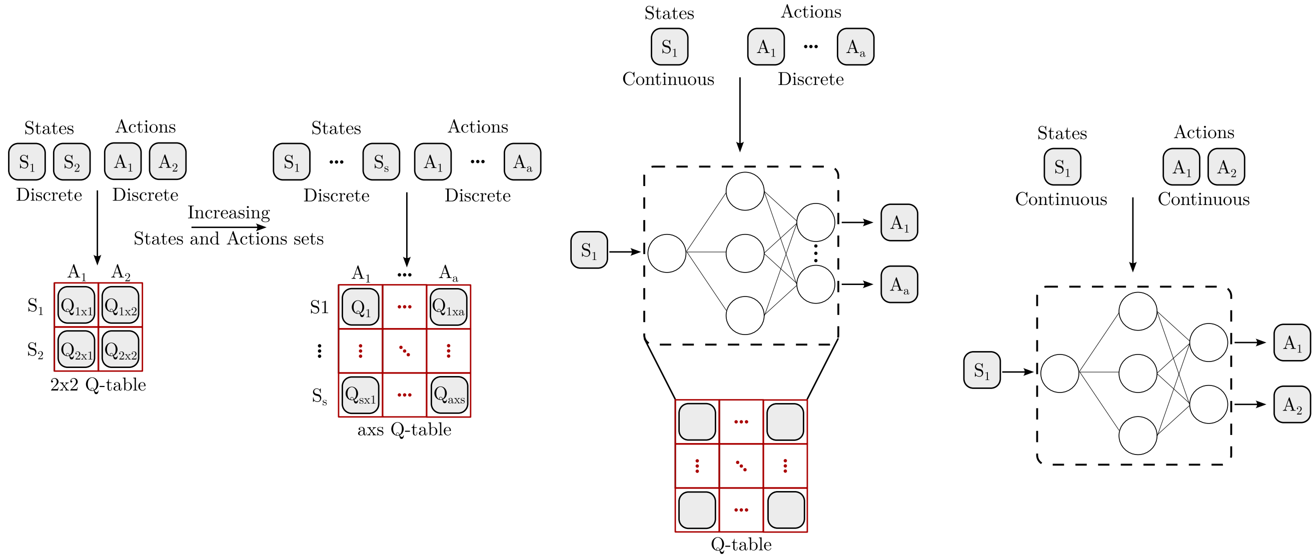
A. CURSE OF DIMENSIONALITY

Although inherently simple, Tabular methods experience an increase in computational burden as the dimensionality of the environment expands [18]. When the number of states and actions is high, the Q-table also grows, causing a possible stack overflow. In contrast, using other methods to estimate the value functions through nonlinear approximations is possible. Among these methods, the ANNs are well-known function estimators and can generalise their training for different conditions. Consequently, a large set of algorithms combine ANNs with Deep Learning (DL), commonly called Deep Reinforcement Learning (DRL) [20]. Fig. 4 illustrates the curse of dimensionality problem by representing different RL agents.

To understand Fig. 4, consider a single-variable control problem where the RL agent has two possible actions. When the QL algorithm is utilised, Fig. 4a shows the Q-table increasing as the variable is discretised into more states and the actions are discretised to allow the agent to process the input with higher resolution, thereby operating the environment with a more significant number of possibilities. As discretisation increases, the Q-table eventually reaches an impractical size for QL implementation. Therefore, the deep Q-network (DQN) was introduced, utilising an ANN to replace the Q-table and approximate the Q-function. This approach enables the processing of continuous state spaces but remains limited to generating discrete actions, as shown in Fig. 4b. To overcome this limitation, other algorithms such as deep deterministic policy gradient (DDPG), twin delayed deep deterministic (TD3), and trust region policy optimisation (TRPO) use a different approach. The DNN trained by these algorithms can process continuous states and compute continuous actions, as illustrated in Fig. 4c.

B. MODEL-FREE VS MODEL-BASED ALGORITHMS

The RL algorithms can also be divided into model-free (MF) and model-based (MB) algorithms. In the MFs, the training is based only on trial-and-error actions, and the agent learns an optimal policy without predicting the state transitions



(a) Relation between Q-table size and the state/action space.
(b) DQN Agent.
(c) DDPG/TD3 Agent.

FIGURE 4. Representation of different RL agents.

TABLE 1. Model-free vs model-based algorithms.

RL Type	Algorithms	Advantages	Disadvantages
Model-free	DQN, DDPG, TD3, A2C, A3C, TRPO, PPO.	● Do not rely directly on the environment model; ● If trained with sufficient scenarios, the algorithms present a robust performance; ● Easier to implement when compared to MB.	● Robustness relies on the representativeness of the training scenarios; ● Sample inefficiency; ● Rely on appropriate design of the reward function, avoiding sparse rewards.
Model-based	Dyna-Q (hybrid), MPC.	● Sample efficiency; ● Ability to plan ahead; ● Robust performance with less samples.	● Rely on the accuracy of the state-transition model and the reward model; ● Computational effort increases with the environment complexity; ● Require an accurate environment model.

of the environment and the reward [19]. The model-based ones use predictions of state transitions and rewards to improve training and find an optimal policy [19]. Examples of MF algorithms are: DQN, DDPG, TD3, proximal policy optimisation (PPO), advantage actor critic (A2C), and asynchronous advantage actor critic (A3C). Examples of MB algorithms are model predictive control (MPC) and Dyna-Q. Table 1 highlights the key differences between the MB and MF algorithms.

As shown in Table 1, MB algorithms are generally more sample-efficient. However, they rely on accurate environment models. The MF algorithms are usually sample-inefficient but show robust performance when trained with many experiences. Therefore, among the main challenges in implementing RL algorithms are the need for novel approaches that address the sample efficiency of MF algorithms and the design of more accurate environment models for MB algorithms.

Recent studies in different areas have been using MB and MF algorithms embedded with information about the dynamics of the controlled or operated physical system. These types of algorithms, known as physics-informed, show a fast convergence rate with robust performance. Moreover, incorporating physics knowledge into MF algorithms enhances the approximation of policy functions, thereby increasing sample efficiency. Physics-informed algorithms have been used in some works regarding MGs control and management, as discussed in Section V.

Section V presents a deeper discussion of RL types in the existing implementation for MG applications, including physics-informed algorithms.

C. REWARD FUNCTION DESIGN

The design of the reward function is critical for RL algorithms, especially for MF. For instance, MF sample efficiency could be improved if a reward function properly translates environment transitions to the training algorithm [21]. Incorporating physics-informed strategies has substantial potential for this subject, as formulating reward structures with environment dynamics could enhance the learning processes of DRL agents [22].

Many works have examined the performance of RL algorithms and the design of the reward function [23]. For each application, there is an effort to define reward structures that optimise the agent's learning. In [24], there is a discussion on reward shaping and how it aids the learning process. In [25], a gradient-ascent-based method is proposed to improve the reward function during training.

Recently, there has been a concern about the design of reward functions with context information. Following this approach, the reward is divided into various components, each playing a role in the agent's learning depending on the scenarios encountered. In [26], the concept of reward machines was proposed, which changes the reward function based on the current state of the environment during learning. Regarding MG control and operation, Section V discusses the role of the reward function design in enabling novel RL solutions to the area suitable for real-world applications.

D. EXPLORATION VS EXPLOITATION

Choosing the trade-off between exploration and exploitation properly during learning helps to train a DRL agent successfully [27]. The agent can adapt to different scenarios and consistently choose actions to prevent the environment from reaching unwanted conditions [18]. Algorithms, such as QL, use a ϵ-greedy approach by choosing a random action during learning with a specified probability. In the case of DDPG, TD3, and others, the learning process adds noise to the agent's actions, allowing it to explore the state-actions space. The exploration phase assists the learning algorithm avoid local optima by encouraging it to experience diverse operating conditions. Over time, the agent transitions from exploration to exploitation, facilitating learning convergence to optimised behaviour when combined with an effective reward structure and proper tuning [18], [27], [28].

E. FINAL REMARKS ON REINFORCEMENT LEARNING

Based on these fundamental RL principles, this technique has been effectively integrated into MG control and management. MG operation is complex and requires innovative strategies to ensure stability and achieve optimal goals. The distinct features and capabilities of RL algorithms make this area of machine learning a powerful tool for addressing the challenges found in MG applications. However, several elements must be considered when developing RL solutions, such as selecting the most suitable RL type, designing a practical reward function, and conducting validation tests to ensure safe implementation in real-world conditions. From this perspective, this survey analysed, in Section IV, the different control solutions used for MGs over the past years and demonstrated how ML algorithms, particularly RL-based methods, have contributed to the evolution of this field. Finally, research gaps and future research directions are described in Section V.

IV. STATE-OF-THE-ART IN MICROGRID MANAGEMENT AND CONTROL USING REINFORCEMENT LEARNING

MG operation is a complex research area that demands interdisciplinary expertise across multiple engineering fields. This section begins with a brief description of some tools historically used for MG control and management, illustrating how RL/DRL algorithms are now being applied to address the limitations of these traditional methods. It then explores key areas where RL is applied to address specific MG challenges, providing a detailed discussion of representative articles within each category.

A. EVOLUTION OF MICROGRID RESEARCH AND TECHNIQUES

Due to their simplicity, researchers have used PI controllers in power systems for decades. One of the most studied applications is the MG frequency and voltage control [29], [30], [31]. These controllers perform reliably if their gains are properly tuned [32]. However, PI controllers experience poor scalability and flexibility [33], which limit their performance. In addition, the droop controllers have been used in voltage and frequency control [34] and generator output power adjustments [35]. However, the droop is affected by transient effects, and its parameters depend on the system configuration [36], which limits the dynamics of the network. Due to the traditional controllers' limitations, researchers have proposed new approaches. Some recent studies rely on intelligent techniques, i.e., ML algorithms. Afterwards, these algorithms are briefly discussed.

In [29] and [37], for example, the MG control parameters were adjusted by a particle swarm optimization (PSO) technique, which is an evolutionary algorithm based on the social behaviour of a flock of birds. In another approach [29], the PSO algorithm was used to set the gains of the PI controller of a voltage source converter, ensuring optimised regulation of the MG voltage, frequency, and power dispatch. Another optimisation tool, known as bacterial foraging optimisation (BFO), was adopted by [30] and [38] in MG control. Inspired by the foraging behaviour of *E. coli* bacteria, the BFO technique in [30] optimised the frequency regulation of an MG comprising different energy resources. In [39], the black widow optimisation (BWO) algorithm was proposed for energy management in MG, achieving optimised solutions with fewer iterations compared to other algorithms.

Fuzzy controllers are another type of intelligent-based approach applied to MG control. In [31], for example, a fuzzy system (FS) was implemented to tune the PI controller, which was responsible for the frequency regulation of the system. A PSO algorithm was used in this methodology to adjust the fuzzy membership functions in real-time to ensure an optimal fit for the PI controller. In [40], FS was combined with PSO optimisation. As another example of FS applications, in [41], the authors developed a fuzzy controller to manage the state of charge (SoC) of battery banks in alternating current (AC) and direct current (DC) MGs, thereby improving their life cycle. Similarly, the methodology presented by [42] used a Fuzzy controller to assist in energy management, minimizing the energy purchased by optimizing the operation of the Sodium Oxide Fuel Cell (SOFC) with a battery energy storage systems (BESS).

In another study, the FS developed in [43] cooperated with the droop and PI controllers to provide system frequency regulation. In this approach, an FS aimed to reduce the dependence of the droop control with the line parameters and introduce an intelligent adjustment of the PI controllers considering load changes in the MG. Furthermore, an adaptive approach using MPC to predict the calculation of optimal control actions was presented in [44]. In this scenario, a fuzzy controller was introduced to the methodology to adjust the MPC and minimise the deviation of the system frequency during MG operation. In [45], an FS-based controller with Lyapunov stability analysis was proposed for DC MGs, enhancing dynamic response and robustness. Simulations and hardware tests show superior performance compared to conventional controllers. Furthermore, a FS-based solution was proposed for the energy management of hybrid electric vehicles in [46].

There are also studies using ANNs to solve optimisation problems. In [47], for example, a radial basis function network-sliding mode (RBFNSM) and a general regression neural network (GRNN) were used to quickly and accurately track the maximum output power of a hybrid power system. In [48], the proposed method combines an Elman neural network (ENN) with the perturb and observe (P&O) technique to determine the maximum power point tracking (MPPT), replacing traditional algorithms used for this task.

In the control role, a primary and secondary level scheme to control the distrubuted generator (DG) of a multi-MG using ANN and adaptive control was proposed in [49]. This approach replaced PI controllers with intelligent ones and used an ANN to ensure voltage and frequency regulation. In [50], a radial basis function (RBF) neural network was used to adjust the switching gains of the sliding mode control in real time. This approach mitigated the chattering phenomenon caused by this controller during the voltage regulation process. Some works used an intelligent hybrid algorithm known as an adaptive neural-fuzzy inference system (ANFIS). Reference [51] used the ANFIS for voltage control of the voltage source converter (VSC) to cooperate in voltage compensation of an MG with DERs and unbalanced loads.

RL overcomes the limitations of supervised and unsupervised learning combined with ANNs in control applications, providing a learning-by-experience scheme. One of the first articles involving RL was proposed in 2012 by [52]. In this study, the RL algorithm minimised electricity costs in an MG. Taking into account a multi-agent environment, in which the units are autonomous and have decision-making power, the proposed method allowed the agents to optimise their actions to reduce the electricity cost and meet the energy balance and generation limit requirements. Table 2 summarises the main characteristics and differences between conventional techniques and RL-based approaches, highlighting scientific challenges. A more in-depth discussion, focused on RL is presented in the following sections.

Given this scenario, the RL algorithm helps to provide the energy balance of each MG in the community, minimising costs involving energy purchase and sale, power generation, and storage. This survey reviewed most journal articles that proposed an RL-based MG controller over the last ten years. This work searched for ''Microgrid'' and ''Reinforcement Learning'' in Scopus. The publications were filtered by year range, 2015 to 2024, and by type, limiting the search to journal articles. Fig. 5 shows the main topics related to RL and MG.

The bibliometric analysis showed how RL is related to different topics of MG research. This work aggregated these topics into four main categories: (i) energy management and power control; (ii) voltage and frequency control; (iii) operation cost minimisation; and (iv) MGs coordination. Each of these categories have significant contributions, with a great variety of works proposing solutions for each area.

B. ENERGY MANAGEMENT AND POWER CONTROL

An MG comprises different energy storage elements, and it is vital to coordinate these elements with the loads and DGs. In a hierarchical structure, the secondary and tertiary levels perform energy management. Thus, the energy management and power control of DERs are the subject of different studies.

An intelligent controller for MG-connected VSC, combining an ANFIS with RL, was proposed in [53]. The ANFIS was the actor and an FS was the critic, which evaluated the actor's performance and updated its parameters. An MG was modelled in MATLAB/Simulink to test the proposed controller, with a further comparison to a PI-based control approach. Moreover, in [54], an optimal active power control of an MG was proposed using an ensemble of ML algorithms.

Reference [55] proposed an intelligent dynamic energy management system (I-DEMS) for MGs. The I-DEMS used action-dependent heuristic dynamic programming (ADHDP), a type of adaptive critic design (ACD). It was implemented an MG comprising a DG, a BESS, and critical and noncritical loads. The management system had to supply the critical loads at any time, maintaining the BESS SoC at an optimal level, maximising the DG utilisation and minimising the energy use provided by the primary grid. This controller allowed the MG to operate in both grid-connected and islanded modes.

In another study, a control scheme was proposed to coordinate ESS in an MG [56]. The controller was developed using QL on a hybrid MG. The main goal of the algorithm was to reduce the power losses on the ESS during run-time. It observed the ESS SoC and the output power of these elements. The controller decides if the ESSs must be turned on or off. In this study, the controller was tested using field data from a real MG, and the ESS power loss and SoC were

TABLE 2. Comparison between techniques from classical to RL-based.

Techniques	Advantages	Disadvantages	Scientific challenges and trends
Classical	• Simplicity of implementation, • High reliability, • Low computational requirements, • Well established in theory and practice.	• Limited performance in nonlinear and complex systems, • Sensitive to uncertainties and disturbances, • Require manual adjustments.	• Combined with advanced methods to improve performance, • Parameter optimisation with ML, • Applicable in low-cost and limited infrastructure requirements.
ML-based	• Ability to handle complex and nonlinear systems, • Adaptability to varying conditions, • Quick pattern recognition.	• Need for large amounts of training data, • Higher computational cost, • Sensitivity to changes outside the training dataset.	• Development of hybrid systems combining ML with classical control, • Real-time use for predictions and optimisation.
RL-based	• Continuous learning in dynamic and non-linear environments, • Real-time optimisation without the need for an explicit model, • Potential for autonomous decision-making.	• High computational cost, • Long training time, • Rely on representative simulations for learning, • Can be unstable if poorly configured (excess exploration/insufficient exploitation).	• Use of DRL for complex MGs, • Application of DRL in MGs with multiple agents, • Design of on optimised reward function, • Hardware-in-the-loop experiments, • Combined with transfer learning and physics information to accelerate implementation in new scenarios.

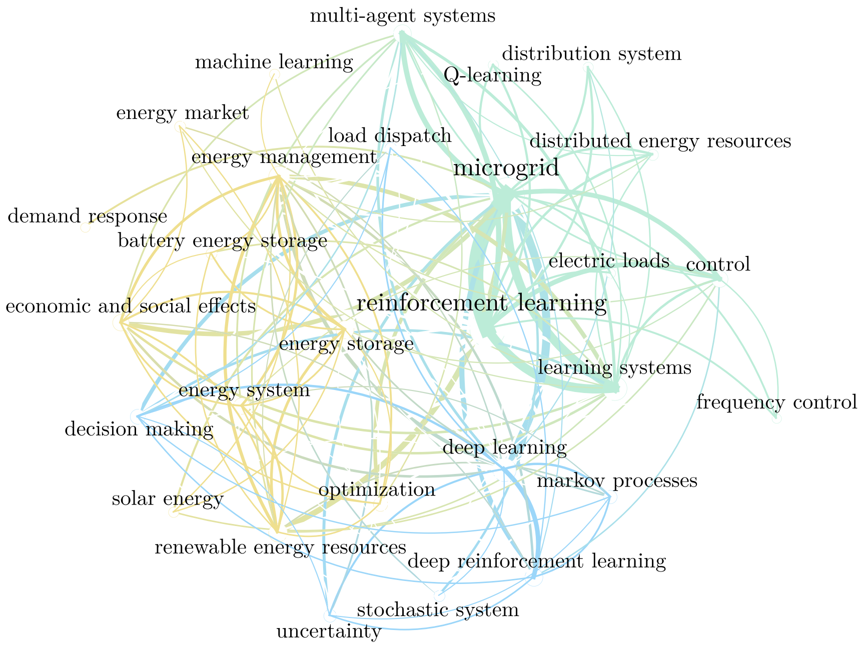


FIGURE 5. Map of topics regarding RL and MGs.

estimated through equations. Reference [57] proposed to use QL to optimise fuzzy-PID controller weights to regulate an islanded MG frequency.

A scheduling algorithm to coordinate a battery using fitted Q-iteration (FQI) was proposed in [6]. This algorithm is a batch RL that optimised the amount of energy an MG uses. The MG comprised a PV, a BESS, the loads and an inverter that connects the DC sources to the AC load. Random trees were used to approximate the Q function. Reference [58] proposed a QL controller to manage the energy of an MG, comprising a WT, a PV, a BESS, a diesel generator, and the loads. Different QL agents operated each MG component. Each of these agents controlled the output power of the DGs, the ESS, and varied the energy consumed by the loads. In addition, each DG agent chose the energy bid based on the availability of resources.

Furthermore, [59] proposed a scheduling algorithm of an electric vehicle (EV) to support an MG. It used QL and decided whether an EV should be charged or discharged, considering the demand and the SoCs of the EV. The proposed algorithm, called mobility-aware vehicle-to-grid control algorithm (MACA), was tested considering the connection of three independent MGs to the same point, each with EVs installed. The simulations considered a simplified model of these MGs, using only their load profile to train and evaluate algorithm performance. The results showed an improvement provided by the MACA in the degree of autonomy of the MG.

In another example, in [60], a multi-agent energy management algorithm using fuzzy Q-learning (FQL) was proposed. It combined an FS with QL to adapt this RL algorithm for continuous applications. Five agents were implemented, one for each DER in the MG, and this control scheme reduced the number of deep discharges of the BESS. In [61], a multi-agent hierarchical control of MGs was proposed to manage the grid's energy, utilising a QL algorithm combined with the evolutionary game theory (GT) to achieve this.

Reference [62] proposed an intelligent power management system in MGs to minimise the exchanged power with the primary grid. Each MG household had a controller based on the Fitted Q-algorithm, and each had its own DG. The transient behaviour of the MG was not considered during disturbances. In [63], an energy management system of an islanded MG based on power pinch analysis was proposed. A QL algorithm was used for adaptive power pinch, improving the energy management system (EMS) performance. Lastly, in [64], an intelligent controller using double deep Q-network (DDQN) to operate the energy storage elements of an islanded MG was proposed. The controller minimised the grid's power loss and was adaptable to the influence of intermittent energy resources. However, the controller was designed for steady-state operation and did not consider transient stability.

C. VOLTAGE AND/OR FREQUENCY CONTROL

The voltage and frequency control are the main functions of the secondary level in an islanded MG. Sometimes, the primary level can perform voltage control to meet local requirements. The design of this control function is essential for the stability of MGs, and recent articles used RL algorithms to improve the system operation.

Reference [65] proposed a load frequency control of an islanded MG using goal representation adaptive dynamic programming (GrADP). This algorithm dampens the deviations in the grid frequency by adding a PID controller output with its own. In this case, the control scheme operates a microturbine and two EVs. The proposed controller performance was compared to two other approaches: a PSO-optimised fuzzy controller and a PID controller without assistance. The results showed fuzzy controller superiority when there was no signal transmission delay. Considering the delay, the GrADP+PID controller oscillated less and was faster than the fuzzy one. However, GrADP+PID exhibited oscillations at a rate similar to that of PID alone, but the transient response of the suggested controller was smoother.

A distribution static compensator (DSTATCOM) RL secondary controller was proposed by [66]. The RL algorithm regulated the DSTATCOM terminal voltage, improving the MG power quality. Moreover, [67] developed the secondary control of an islanded MG using a brain emotional learning-based intelligent controller (BELBIC). The algorithm structure was similar to an RL approach as the agent weights were updated by reading the environment states and the controller performance evaluation. The study used the Lyapunov theorem to evaluate the stability of the MG.

An intelligent DC-DC converter controller for power stabilisation was proposed in [68], using the PPO algorithm, an actor-critic algorithm, for this task. In another case, reference [69] proposed an MG control to ensure the stability of the system after islanding. Two deep reinforcement learning (DRL) algorithms were used: the DQN and the DDPG. The DQN regulated the loads, and the DDPG controlled the DGs. Both algorithms operated without communicating with each other.

Reference [70] proposed an MG secondary voltage controller using multi-agent reinforcement learning (MARL). The algorithm called PowerNet was based on the independent actor-critic, with improvements in the learning stability, the algorithm robustness under uncertainty, and the cooperation of the agents. It was tested the controller in the IEEE 34-bus system with a different number of inverter-based distributed generator (IBDG). In [71], a droop-free control algorithm was proposed to coordinate DGs in islanded DC MGs that regulates the system voltage. An online RL algorithm updates the weights of two neural networks that compute the optimal duty cycle value for each DG.

An RL-based controller for DC MGs to stabilise active loads was proposed in [72], using ANNs to approximate the value function and a non-linear approach to update the agent weights. In this case, a nonlinear analysis was performed using the Lyapunov theorem. In [73], a fuzzy type 2 controller was presented and combined with a DDPG algorithm to regulate an MG frequency. The DDPG improved frequency stabilisation as it can adapt to different disturbances. Finally, in [74], an integral reinforcement algorithm was proposed to optimally control the power buffer in DC MGs. A nonlinear analysis was used to assess the system stability with the controller.

Regarding AC MGs, a DRL-based secondary controller to regulate the voltage and frequency of an islanded MG was proposed in [75]. The DRL agent operated the BESS within the MG, supporting a synchronous generator acting as the slack machine. The DRL-based controller was compared to a classical control approach based on the droop method, demonstrating the transient response under typical operating scenarios for MGs. In addition, a multi-agent approach was proposed in [76] for frequency deviation control, where the DRL algorithm updated the PI gains. In [77], a DRL algorithm trained an ANN, which tuned a PID controller for frequency deviation control with further stability analysis.

D. OPERATION COST MINIMIZATION

Different strategies can optimise the MG operation cost, such as managing the DER generation and storage considering the energy cost; managing available resources to optimise the use of the renewable energy technologies; and optimising combined heat and power systems. In this case, the tertiary and secondary levels were responsible for this control function in the grid-connected mode. In contrast, only the secondary level remains when the MG is in islanded mode.

Reference [52] developed a multi-agent system to minimise the energy cost of a grid-connected MG using an RL algorithm. The RL-MAS minimises the energy cost and satisfies the MG power balance and the DG generation limits. The RL-MAS scheme comprises four types of agents: renewable energy sources; conventional energy sources; loads; and storage elements. Each type of agent has unique actions and can change the DG output power, the energy demanded by the loads, and the storage element output power. This issue requires a high-dimensional matrix using QL, which would require high computational effort to complete the agent's training. Thus, a dynamic hierarchical reinforcement learning (DHRL) algorithm was used to process the rate of change of the state variables. This algorithm uses a divide-and-conquer approach to reduce computational burden.

A bidding strategy to maintain the power balance of an islanded MG using QL was proposed in [78]. Furthermore, in [79], an MG controller was proposed to minimise the operating cost of an MG comprising BESS, hydrogen storage, diesel generator, and photovoltaic (PV) generator was implemented. The DQN was used for this task, which changed the power of BESS, hydrogen storage, and diesel generator.

An MARL algorithm to control an heat, ventilation and air conditioning (HVAC) system was proposed in [80]. The multi-agent framework uses GT to minimise the MG operation cost. Then, the QL was used to reduce the computational burden. In another case, [81] proposed an intelligent control of a battery and a heat pump that reduced the energy cost and the self-consumption. Two RL algorithms were tested for this task, the FQI and the approximate policy iteration with Q-functions (APIQ). A multi-agent structure was used to coordinate the energy resources. The performance of the RL algorithms were similar, and the system achieved close results with both of these algorithms compared to an optimally tuned classical controller. Moreover, the results showed a superiority of the multi-agent framework compared to the single agent.

E. MICROGRID COORDINATION

The last application found in the literature is the MG coordination. In a hierarchical structure, the tertiary level is responsible for the coordination. As an example, [82] proposed a management system for active distribution networks comprising MGs with BESSs and PVs. An ACD algorithm was used for this task, called dual heuristic programming (DHP). The intelligent controller minimises the energy cost and regulates the bus voltages by observing the BESS SoC limits. Moreover, it prevents system overloads, analysing the load over the branches.

In another MG coordination application, an RL-based power management algorithm was proposed in [83]. The proposition uses a multi-agent structure and meets operational constraints. The results showed a better performance of the proposed controller than a DQN-based one. It can be observed that this research area closely relates to operation cost minimisation. However, in this case, multiple MGs were implemented.

V. RESEARCH GAPS AND TRENDS

As mentioned, there is a vast body of literature using RL algorithms for MG control. These algorithms have mainly been used for energy management and power control, voltage and frequency control, operation cost optimisation, and MG coordination. In addition, the studies compared their performances with other ML algorithms and classical controllers. In Table 3, most recent studies are presented on MG control and operation.

Table 3 shows that most studies are related to energy management in MGs or voltage and frequency control. Only two studies are not within the categories presented in the previous section. The method in [5] used delay minimisation Q-learning (DMQ) as a resource allocation algorithm to improve the MG communication latency. In [96], RL was used to identify vulnerable spots in MG against cyberattacks. In terms of RL type, most of the articles used model-free algorithms.

Regarding the performance analysis, the steady-state response is a crucial feature presented in works related to energy management, cost optimisation and MG coordination. Most voltage and frequency control studies show the system transient response. However, only some studies provide stability proof of the MG using non-linear analysis. The MG is a complex system with non-linear dynamics. As discussed, regarding voltage and frequency control, the classical controllers are not designed for plants that operate too far from the linearised model. The RL algorithms can overcome these limitations with proper training and as long as the studies provide sufficient analysis that ensures MG stability. Table 3 shows the recent studies which used non-linear methods to assess the stability.

Every MG control that uses RL performed better than or equal to the classical approach. The RL algorithms usually require more time to train/tune than the traditional frequency controllers, for example. However, the operation of this intelligent algorithm should not be computationally heavy after training. Furthermore, in Table 3, few articles implemented the proposed algorithm in hardware, which demands more information on the feasibility of RL algorithms in real-world MG applications.

The QL is one of the most used RL algorithms. This algorithm is suitable for energy management and operational cost optimisation. In recent studies, the deep learning version of the QL, the DQN, was used for high-dimensional problems, as proposed by [79], for example. Moreover, an important feature of deep neural networks combined with RL is their ability to control continuous variables, as proposed by [73]. In addition, some studies combine RL with other optimisation or ML techniques. The works

TABLE 3. Summary of studies addressing reinforcement learning in MG control problems.

Reference	Algorithm	Control Fun.	RL Type		Perf. Analysis			MG Type			Op. Mode			Hardware
			MF	MB	SSP	TR	NL	DC	AC	HB	CN	IS	MT	
[52]	DHRL	OP		●		●			●		●			
[55]	ADHDP	EMP	●			●			●		●	●		
[6]	FQI	EMP	●			●			●		●			
[60]	FQL	EMP	●			●			●			●		
[67]	BELBIC	VF	●			●			●			●	●	
[79]	DQN	OP	●			●			●			●		
[82]	ACD/DHP	MC		●		●			●		●			●
[5]	DMQ	OF	●			●			●					●
[68]	PPO	VF	●		●			●				●		●
[80]	MARL	OP	●			●			●		●			
[61]	QL/GT	EMP		●		●			●		●			
[81]	FQI	EMP	●			●			●		●			
[62]	FQI	EMP	●			●			●		●			
[63]	QL	EMP	●				●			●		●		
[84]	DHP	VF	●			●			●			●		
[70]	MARL	VF	●			●			●			●		
[71]	RL/ADP	VF		●	●			●				●		●
[72]	RL/ADP	VF		●	●			●				●		
[85]	VA2C	OP	●			●			●		●			
[86]	DDQN	EMP/OP	●			●			●		●			
[64]	DDQN	EMP	●				●			●		●		
[87]	GAN/RL	OP/MC	●			●			●		●			
[88]	GAN/RL	OP/VF	●			●			●		●			
[89]	PPO	OP		●		●			●			●		
[73]	DDPG	VF	●				●			●		●		●
[83]	SMASPL	MC		●		●			●		●			
[90]	DDPG	VF	●		●			●				●		●
[3]	HDP	VF	●			●			●				●	●
[74]	IRL	VF		●	●			●				●		●
[91]	A2C	MC	●			●			●		●			
[7]	RDQN	OP	●				●			●	●			
[92]	DDPG	OP	●		●			●				●		●
[93]	A3C	VF	●			●			●			●		
[94]	MADDPG	MC	●		●			●			●			
[95]	MADDPG	MC	●		●			●			●			
[96]	DDPG/DQN	OF	●		●			●				●		●
[97]	PPO	VF	●		●			●			●			
[98]	MADDPG	OP/VF	●			●			●			●		
[99]	MA3C	EMP	●			●			●		●			
[100]	DDPG	EMP	●			●			●			●		
[101]	TD3	EMP	●			●			●		●			
[102]	DQN	MC	●			●				●	●			
[103]	MATD3	MC	●			●				●	●			
[104]	IOLS-MA4DPG	VF		●		●			●			●		
[105]	TD3	VF	●			●			●			●		●
[106]	DDPG/SAC/TD3	VF	●			●			●			●		
[107]	Phy-Info TD3	EMP	●			●			●		●	●		●
[108]	SDDPG	VF	●		●	●			●			●		
[109]	DDPG	VF	●		●	●			●			●		
[110]	SDDPG	VF	●		●	●			●			●		
[111]	MA-FQL	EMP	●		●			●	●		●			
[112]	PER-SAC	VF	●		●	●			●			●		
[113]	TD3	VF	●		●	●			●			●		
[114]	QL	OP	●		●	●			●			●		
[115]	RMADDPG	EMP/OP	●		●				●		●			
[116]	SAC	MC/OP	●		●				●		●			
[117]	TCSAC	EMP/OP	●		●				●		●			
[22]	Phy-info F-MADRL	MC/OP	●		●				●		●			
[118]	Phy-info MPC	VF		●	●	●			●			●		

Legend:
Control Fun.: EMP = Energy management and power control; VF = voltage and/or frequency control; OP = Operation optimisation;
MC = MG coordination; OF = other functions
RL Type: MF = model-free; MB = model-based
Perf. Analysis: SSP = Steady-state performance; TR = Transient response; NL = non-linear analysis
MG Type: HB = hybrid
Op. Mode: CN = connected; IS = islanded; MT = Mode transition
Algorithm: ADP = adaptive dynamic programming; VA2C = vectorised-A2C;
SMASPL = supervised multi-agent safe policy learning; IRL = integral reinforcement learning
RDQN = Rainbow-DQN; and MA3C = memory-A3C

in [7], [60], and [80] are examples of that approach. Moreover, other studies proposed novel RL algorithms which are more suitable for the desired application, such as [70].

Furthermore, older studies lack information about the number of MG nodes, the load level, and the DER models used in the simulations. The data is provided in other sources, but the MG size is insufficient for broader applicability. In this case, the MG, by definition, is part of the distribution system that can operate in connected or islanded modes, coordinating different DERs and loads. Therefore, it is not ideal for modelling an MG as a two-bus system with a couple of DERs. Although this problem was more visible in earlier publications, the most recent ones could provide a database with the simulations performed. This information allows for the replicability of the work. The most recent articles on RL algorithms in MG control and management presented trends that could be summarised into four main topics. In each case, the research gaps and possible solutions are highlighted.

TABLE 4. Novel concepts combined with DRL to develop novel MGs control and management. solutions.

Concept	Applications and contributions
Physics-informed	• The performance of the agent is more robust, from RL solutions to design new ANN architectures [107], [121], • Integration in different parts of the DRL framework, such as: Reward design and shaping [22], Generates more accurate environment models, contributing to MB algorithms, Improve MF algorithms in better sample efficiency and convergence [107].
Imitation-learning	• Improve sample efficiency and robustness [104], • Implemented to overcome dimensionality issues [120], • The agent learns the policy through demonstration from an expert [120], • The concept can be combined with different ML tools, such as FS or pre-trained DRL agents.
MARL	• Implemented for multiple MGs integration [22], [94], • Suitable for training agents to operate multiple DERs.
HRL	• Combined with MADRL, • Brakes the problem in multiple tasks, • Valuable when used for problems with different goals.

A. NOVEL ALGORITHMS AND LEARNING METHODS

Tabular RL algorithms are not suitable for continuous problems in MGs. These algorithms require state and action quantisation. However, most control problems regarding MG operation have many states and actions, leading to the curse of dimensionality. Recent works minimise this limitation by using deep RL algorithms such as DDPG, TD3, DQN, A2C, and A3C. Lastly, model-free RL algorithms are the most popular, but novel studies use model-based ones, such as MPC, to improve sample efficiency and reduce computational burden [118].

Some studies evaluate different learning methods, such as Imitation Learning. Reference [119] compared the imitation learning algorithm with the QL. In [120], an imitation learning algorithm that performed better than tabular RL algorithms was proposed. In the literature, some articles use hierarchical RL (HRL) algorithms. In this case, the control function is divided into several tasks, improving the convergence and learning process [61]. The combination of the multi-agent framework with RL algorithms has also been explored, as demonstrated in [95] and [98]. Table 4 summarises the novel concepts integrated into recent DRL solutions for MG control and management.

Developing novel training frameworks for ANNs is also valid for improving the performance of the DRL agent. For instance, the generative adversarial networks (GANs) framework, based on training two ANNs to compete, allows the primary network to distinguish relevant data from noise, communication errors, or possible cyberattacks. In [88], the GAN structure was used for improved MG coordination considering communication failure. Inspired by adversarial learning principles, an adversarial DRL-based strategy was proposed in [122] to improve the robustness of inverter-based microgrids against cyberattacks.

In addition to novel training frameworks, different ANN architectures can be used. In [118], a long short-term memory (LSTM) neural network was implemented in the hidden layer of the DDPG agent, allowing it to take actions based on previous information. The recurrent neural network (RNN) architecture, besides LSTMs, is a powerful tool which can be explored more when integrating DRL and MGs. Moreover, convolutional neural networks (CNNs) can potentially contribute to different solutions [121]. When combined with DRL agents, the transformer architecture is also promising for energy management in MGs [123], [124]. Transformer networks have gained popularity in generative artificial intelligence since it was first proposed in [125] as a tool capable of processing information, incorporating the context and correlation between inputs.

B. REWARD FUNCTION DESIGN

The design of the reward function has been little explored in the area of RL combined with MG and this subject is key for a more efficient learning process. Developing structures that optimise agent training considering specific requirements in each area of MG control and management is essential. A proper reward design increases training efficiency, allowing the agent to learn in less time and later showing more robust performance during implementation. In [108], the reward function was based on stability indices, combining an approach proposed in [126] to improve agent learning and performance for the frequency control of AC MGs. In [114], the reward function combined economic and power requirements. The reward structure considered several operational constraints, thus enhancing agent learning when the MG operated within a specified range to optimise cost reductions. In [22], the reward function was based on the physics-informed approach, allowing the agent to coordinate multiple MGs, reducing operational costs.

Adding context information to the reward function is a potential approach to improve the agent's learning. Novel articles could consider adding the reward machine concept for their solutions. Moreover, considering the idea of inferring the reward function is another valuable strategy studied in artificial intelligence, known as inverse RL [127]. Different tools can be used for that task, such as FS or Transformer networks, which could also increase agent learning and produce more autonomous solutions capable of training and operating with minimal human supervision.

C. PERFORMANCE ANALYSIS

Performance analysis is crucial for validating an RL-based solution and allowing future comparison with novel approaches. The selection of each index is determined by the control function and the specific type of analysis under consideration, such as steady-state, transient, or nonlinear. Regarding nonlinear analysis, they are found in VF control studies, using the Lyapunov stability theorems [72], [128]. Although Lyapunov theorems are key for demonstrating the stability of nonlinear systems, the complexity involved in this analysis increases as the MG integrates additional DERs, loads, and operating conditions.

An alternative to nonlinear analysis for VF control is performing a thorough study of transient behaviour, assessing critical operating conditions and observing metrics such as rate of change of frequency (ROCOF) and frequency or voltage deviations. Moreover, evaluating steady-state metrics is also valuable for comparing different solutions, such as mean absolute error (MAE), integral of absolute error (IAE), integral of squared error (ISE) and integral of time absolute error (ITAE). In [109], operational bounds were presented, illustrating the maximum frequency sweep observed during events in MGs.

Regarding EMP control, methods such as Monte Carlo can help analyse the agent performance, adding statistical depth to the proposition. Recent works have focused on coordinating the DERs to reduce operational costs, perform fast restoration, optimise energy storage, preserve the SoC, and reduce the net load. The EMP, OP, and MC usually have similar objectives. For OP, the solutions mainly focus on reducing power losses, increasing efficiency, and observing steady-state stability indices. In the context of MC, there has been an emphasis on lowering the net load, optimising the energy storage operations according to the SoC, and maximising the overall profit among MGs. Table 5 summarises the primary metrics for performance analysis regarding MG operation with RL.

TABLE 5. Performance metrics found in different articles regarding MG operation and control with RL.

Control Function	Performance Metrics
EMP	DERs power and load demand [99]–[101], [117], Computational time for service restoration [100], Energy storage SoC [99], Operational cost [117].
VF	Frequency deviation [108], [109], [118], MAE [108], [109], IAE, ISE, ITAE [108], [113], Peak of the deviation [109], Operational bounds [109], ROCOF [106].
OP	Power losses [92], [117], efficiency [92], Branch stability index [117], Operational cost [114].
MC	Profit (mean, max,min) [102], Net load [103], Local trading [103], Power consumption [115], Power loss [116], Energy storage SoC of each MG [116].

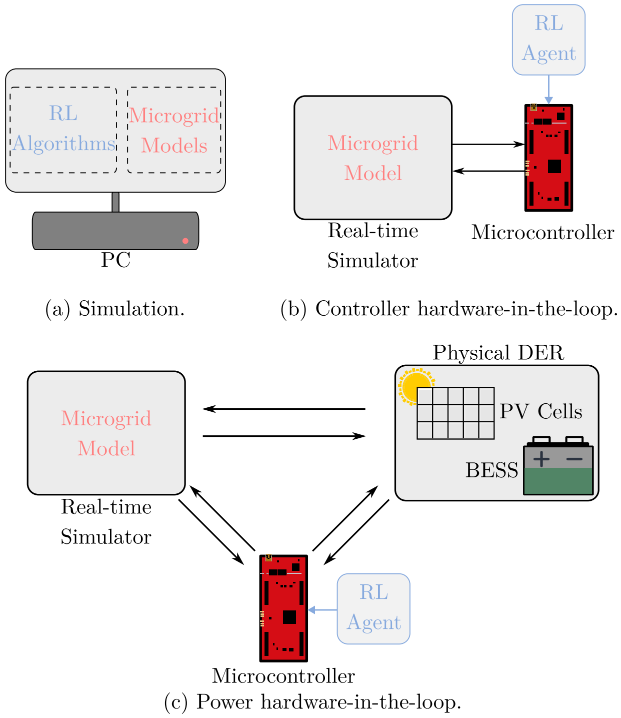
D. SIMULATION AND HARDWARE IMPLEMENTATION

The evolution of DRL implementations is progressively moving toward solutions that are practical for real-world conditions. The highest level of the validation process of these agents can be divided into simulation, controller hardware-in-the-loop (HIL), and power HIL. The simulation phase involves training the agent using simplified models of the environment. The agent model and the learning algorithm can be implemented in programming languages such as Python or MATLAB. The environment models depend on the control function, with VF control solutions using electromagnetic transient (EMT) software such as Simulink or PSCAD for model development. In contrast, EMP, OP, and MC simulations allow for the implementation of steady-state models in tools such as OpenDSS or custom Python-based models.

These simulation tests do not assess the algorithm's performance in real-life applications. Thus, recent works have started showing results for controller HIL implementation. The controller HIL stage involves embedding the RL agent in a microcontroller while communicating with the environment model running in real-time. The environment is modelled on a real-time simulator such as RTDS, OPAL-RT, or Typhoon HIL. This stage is key to capturing limitations not found in simulations with the simplified environment models. Moreover, the experiment considers communication delays between the MG and the RL agent and the impact of noise on system behaviour. These factors ensure the robustness and reliability of the RL agent in real-world applications where delays and noise are present.

Additionally, the setup allows one to evaluate whether the trained RL agent can be embedded on lower-cost microcontrollers with limited memory and processing power or if it requires more expensive hardware with greater computational resources. This experiment assesses the feasibility and cost-effectiveness of implementing the agent in practical scenarios. Furthermore, through controller HIL experiments, the RL agent's performance can be validated, providing an additional layer of security certification. This step ensures that the agent operates safely and reliably under various conditions, allowing it to perform the next testing phase, i.e., the power HIL experiments where the agent is integrated with physical equipment for more advanced validation.

The power HIL differs from controller HIL, using at least one physical DER or element of the MG operating in parallel with the real-time simulator and the microcontroller. The RL agent controls the physical equipment and the other elements of the MG running in the real-time simulator. This stage considers different real-world conditions, such as communication delay and noise, and the complete dynamics of DERs and loads (the physical ones). Following these three steps is essential for validating RL-based solutions and



(a) Simulation.

(b) Controller hardware-in-the-loop.

(c) Power hardware-in-the-loop.

FIGURE 6. **Validation steps of RL agents.**

ensuring their scalability for different implementations. Fig. 6 shows the three main steps in validating RL-based solutions.

The controller HIL has already been used in [129] to validate an RL agent for voltage and frequency control of islanded MGs. In [130], this experiment was used to validate the RL-based voltage controller of DC MGs. Recent works in the field are validating their proposed solutions through power HIL experiments, such as [105] and [107], demonstrating the feasibility of RL-based agents in real-world applications.

VI. CONCLUSION

This article addresses the main challenges in developing MG control and how researchers use RL algorithms in this area. There is also a discussion regarding the drawbacks of classical controllers to ensure system stability, considering their non-linear and complex dynamics. In addition, the present survey shows how these algorithms can contribute to different aspects of MG control and operation, such as operation cost minimisation, energy storage management, power control, and coordination of interconnected MGs.

This survey briefly discusses the limitations of published works that used RL for MG control. Some aspects require improvement, such as a better description of the RL algorithms training data; and a better description of the tested MG, the DERs, and load models. Furthermore, recent research should focus on evaluating whether the solution keeps the MG operational under adverse conditions. In addition, recent solutions should demonstrate superior performance when compared to classical algorithms/controllers. Some studies have already used Lyapunov analysis to demonstrate MG stability. For the other control issues, such as cost optimisation and energy management, it would be possible to use statistical methods, such as Monte Carlo. Statistical analysis allows a better comparison between the RL algorithms and traditional optimisation techniques. Finally, recent works tackled the transient analysis of MGs operated by DRL-based controllers using HIL experiments. Within the class of HIL experiments, Power HIL stands out as one of the most promising methods to validate intelligent controllers.

ACKNOWLEDGMENT

The authors would like to thank the São Carlos School of Engineering, University of São Paulo (Universidade De São Paulo), São Carlos, Brazil for the facilities provided.

REFERENCES

[1] F. Doost Mohammadi, H. Keshtkar Vanashi, and A. Feliachi, "State-space modeling, analysis, and distributed secondary frequency control of isolated microgrids," *IEEE Trans. Energy Convers.*, vol. 33, no. 1, pp. 155–165, Mar. 2018.

[2] J. W. Simpson-Porco, Q. Shafiee, F. Dörfler, J. C. Vasquez, J. M. Guerrero, and F. Bullo, "Secondary frequency and voltage control of islanded microgrids via distributed averaging," *IEEE Trans. Ind. Electron.*, vol. 62, no. 11, pp. 7025–7038, Nov. 2015.

[3] S. Saadatmand, P. Shamsi, and M. Ferdowsi, "Adaptive critic design-based reinforcement learning approach in controlling virtual inertia-based grid-connected inverters," *Int. J. Electr. Power Energy Syst.*, vol. 127, May 2021, Art. no. 106657.

[4] S. Rozada, D. Apostolopoulou, and E. Alonso, "Deep multi-agent reinforcement learning for cost-efficient distributed load frequency control," *IET Energy Syst. Integr.*, vol. 3, no. 3, pp. 327–343, Sep. 2021.

[5] M. Elsayed, M. Erol-Kantarci, B. Kantarci, L. Wu, and J. Li, "Low-latency communications for community resilience microgrids: A reinforcement learning approach," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1091–1099, Mar. 2020.

[6] 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.

[7] D. J. B. Harrold, J. Cao, and Z. Fan, "Data-driven battery operation for energy arbitrage using rainbow deep reinforcement learning," *Energy*, vol. 238, Jan. 2022, Art. no. 121958.

[8] V. Skiparev, R. Machlev, N. R. Chowdhury, Y. Levron, E. Petlenkov, and J. Belikov, "Virtual inertia control methods in islanded microgrids," *Energies*, vol. 14, no. 6, p. 1562, Mar. 2021.

[9] M. Glavic, "(Deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annu. Rev. Control*, vol. 48, pp. 22–35, Jan. 2019.

[10] 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.

[11] B. She, F. Li, H. Cui, J. Zhang, and R. Bo, "Fusion of microgrid control with model-free reinforcement learning: Review and vision," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 3232–3245, Jul. 2023.

[12] M. Bilal, A. A. Algethami, Imdadullah, and S. Hameed, "Review of computational intelligence approaches for microgrid energy management," *IEEE Access*, vol. 12, pp. 123294–123321, 2024.

[13] A. Akter, E. I. Zafir, N. H. Dana, R. Joysoyal, S. K. Sarker, L. Li, S. M. Muyeen, S. K. Das, and I. Kamwa, "A review on microgrid optimization with meta-heuristic techniques: Scopes, trends and recommendation," *Energy Strategy Rev.*, vol. 51, Jan. 2024, Art. no. 101298.

[14] J. Gutiérrez-Escalona, C. Roncero-Clemente, O. Husev, O. Matiushkin, and F. Blaabjerg, "Artificial intelligence in the hierarchical control of AC, DC, and hybrid AC/DC microgrids: A review," *IEEE Access*, vol. 12, pp. 157227–157246, 2024.

[15] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi, C. A. Cañizares, R. Iravani, M. Kazerani, A. H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke, G. A. Jiménez-Estévez, and N. D. Hatziargyriou, "Trends in microgrid control," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1905–1919, Jul. 2014.

[16] A. Bidram and A. Davoudi, "Hierarchical structure of microgrids control system," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1963–1976, Dec. 2012.

[17] L. Meng, E. R. Sanseverino, A. Luna, T. Dragicevic, J. C. Vasquez, and J. M. Guerrero, "Microgrid supervisory controllers and energy management systems: A literature review," *Renew. Sustain. Energy Rev.*, vol. 60, pp. 1263–1273, Jul. 2016.

[18] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[19] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.

[20] N. Locascio and N. Buduma, *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. Sebastopol, CA, USA: O'Reilly Media, 2017.

[21] M. J. Mataric, *Reward Functions for Accelerated Learning*. Amsterdam, The Netherlands: Elsevier, 1994, pp. 181–189.

[22] Y. Li, S. He, Y. Li, Y. Shi, and Z. Zeng, "Federated multiagent deep reinforcement learning approach via physics-informed reward for multimicrogrid energy management," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 5, pp. 5902–5914, May 2024.

[23] J. Eschmann, *Reward Function Design in Reinforcement Learning*. Cham, Switzerland: Springer, 2021, pp. 25–33.

[24] A. Y. Ng, D. Harada, and S. Russell, "Policy invariance under reward transformations: Theory and application to reward shaping," in *Proc. 16th Int. Conf. Mach. Learn.* San Mateo, CA, USA: Morgan Kaufmann, Jun. 1999, pp. 278–287.

[25] J. Sorg, R. L. Lewis, and S. Singh, "Reward design via online gradient ascent," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culott, Eds., Dec. 2010, pp. 2190–2198.

[26] R. Toro Icarte, T. Q. Klassen, R. Valenzano, and S. A. McIlraith, "Reward machines: Exploiting reward function structure in reinforcement learning," *J. Artif. Intell. Res.*, vol. 73, pp. 173–208, Jan. 2022.

[27] T. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. 4th Int. Conf. Learn. Represent. (ICLR)*, Jul. 2016, pp. 1–14.

[28] S. Fujimoto, H. V. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, Jan. 2018, pp. 1582–1591.

[29] W. Al-Saedi, S. W. Lachowicz, D. Habibi, and O. Bass, "Power quality enhancement in autonomous microgrid operation using particle swarm optimization," *Int. J. Electr. Power Energy Syst.*, vol. 42, no. 1, pp. 139–149, Nov. 2012.

[30] S. Mishra, G. Mallesham, and A. N. Jha, "Design of controller and communication for frequency regulation of a smart microgrid," *IET Renew. Power Gener.*, vol. 6, no. 4, pp. 248–258, Jul. 2012.

[31] H. Bevrani, F. Habibi, P. Babahajyani, M. Watanabe, and Y. Mitani, "Intelligent frequency control in an AC microgrid: Online PSO-based fuzzy tuning approach," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1935–1944, Dec. 2012.

[32] M. S. Mahmoud, N. M. Alyazidi, and M. I. Abouheaf, "Adaptive intelligent techniques for microgrid control systems: A survey," *Int. J. Electr. Power Energy Syst.*, vol. 90, pp. 292–305, Sep. 2017.

[33] A. Vasilakis, I. Zafeiratou, D. T. Lagos, and N. D. Hatziargyriou, "The evolution of research in microgrids control," *IEEE Open Access J. Power Energy*, vol. 7, pp. 331–343, 2020.

[34] K. De Brabandere, B. Bolsens, J. Van den Keybus, A. Woyte, J. Driesen, and R. Belmans, "A voltage and frequency droop control method for parallel inverters," *IEEE Trans. Power Electron.*, vol. 22, no. 4, pp. 1107–1115, Jul. 2007.

[35] E. Barklund, N. Pogaku, M. Prodanovic, C. Hernandez-Aramburo, and T. C. Green, "Energy management in autonomous microgrid using stability-constrained droop control of inverters," *IEEE Trans. Power Electron.*, vol. 23, no. 5, pp. 2346–2352, Sep. 2008.

[36] J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. de Vicuna, and M. Castilla, "Hierarchical control of droop-controlled ac and DC microgrids—A general approach toward standardization," *IEEE Trans. Ind. Electron.*, vol. 58, no. 1, pp. 158–172, Jan. 2011.

[37] I. Chung, W. Liu, D. A. Cartes, and S. Moon, "Control parameter optimization for multiple distributed generators in a microgrid using particle swarm optimization," *Eur. Trans. Electr. Power*, vol. 21, no. 2, pp. 1200–1216, Mar. 2011.

[38] K. Roy, K. K. Mandal, A. C. Mandal, and S. N. Patra, "Analysis of energy management in micro grid—A hybrid BFOA and ANN approach," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 4296–4308, Feb. 2018.

[39] M. A. Ayub, U. Hussan, H. Rasheed, Y. Liu, and J. Peng, "Optimal energy management of mg for cost-effective operations and battery scheduling using bwo," *Energy Rep.*, vol. 12, p. 294–304, Dec. 2024.

[40] A. A. Moghaddam, A. Seifi, and T. Niknam, "Multi-operation management of a typical micro-grids using particle swarm optimization: A comparative study," *Renew. Sustain. Energy Rev.*, vol. 16, no. 2, pp. 1268–1281, Feb. 2012.

[41] M. Hosseinzadeh and F. R. Salmasi, "Power management of an isolated hybrid AC/DC micro-grid with fuzzy control of battery banks," *IET Renew. Power Gener.*, vol. 9, no. 5, pp. 484–493, Jul. 2015.

[42] N. Chettibi, A. Mellit, G. Sulligoi, and A. M. Pavan, "Adaptive neural network-based control of a hybrid AC/DC microgrid," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1667–1679, May 2018.

[43] S. Ahmadi, S. Shokoohi, and H. Bevrani, "A fuzzy logic-based droop control for simultaneous voltage and frequency regulation in an AC microgrid," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 148–155, Jan. 2015.

[44] S. Kayalvizhi and D. M. Vinod Kumar, "Load frequency control of an isolated micro grid using fuzzy adaptive model predictive control," *IEEE Access*, vol. 5, pp. 16241–16251, 2017.

[45] U. Hussan, M. A. Majeed, F. Asghar, A. Waleed, A. Khan, and M. R. Javed, "Fuzzy logic-based voltage regulation of hybrid energy storage system in hybrid electric vehicles," *Electr. Eng.*, vol. 104, no. 2, pp. 485–495, May 2021.

[46] U. Hussan, M. Hassan, M. A. Ayub, J. Peng, H. Rasheed, H. Jiang, and F. Asghar, "Smooth and uninterrupted operation of standalone DC microgrid under high and low penetration of RESs," *IEEE Access*, vol. 12, pp. 48620–48629, 2024.

[47] T.-C. Ou and C.-M. Hong, "Dynamic operation and control of microgrid hybrid power systems," *Energy*, vol. 66, pp. 314–323, Mar. 2014.

[48] N. Chettibi and A. Mellit, "Intelligent control strategy for a grid connected PV/SOFC/BESS energy generation system," *Energy*, vol. 147, pp. 239–262, Mar. 2018.

[49] D. O. Amoateng, M. Al Hosani, M. S. Elmoursi, K. Turitsyn, and J. L. Kirtley, "Adaptive voltage and frequency control of islanded multi-microgrids," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4454–4465, Jul. 2018.

[50] G. Lou, W. Gu, X. Lu, Y. Xu, and H. Hong, "Distributed secondary voltage control in islanded microgrids with consideration of communication network and time delays," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 3702–3715, Sep. 2020.

[51] J. Saroha, M. Singh, and D. K. Jain, "ANFIS-based add-on controller for unbalance voltage compensation in a low-voltage microgrid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5338–5345, Dec. 2018.

[52] F.-D. Li, M. Wu, Y. He, and X. Chen, "Optimal control in microgrid using multi-agent reinforcement learning," *ISA Trans.*, vol. 51, no. 6, pp. 743–751, Nov. 2012.

[53] S. S. Khorramabadi and A. Bakhshai, "Intelligent control of grid-connected microgrids: An adaptive critic-based approach," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 3, no. 2, pp. 493–504, Jun. 2015.

[54] X. Zhang, D. Wang, T. Yu, Z. Xu, and Z. Fan, "Ensemble learning for optimal active power control of distributed energy resources and thermostatically controlled loads in an islanded microgrid," *Int. J. Hydrogen Energy*, vol. 43, no. 49, pp. 22474–22486, Dec. 2018.

[55] 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.

[56] X. Qiu, T. A. Nguyen, and M. L. Crow, "Heterogeneous energy storage optimization for microgrids," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1453–1461, May 2016.

[57] M. Esmaeili, H. Shayeghi, H. Mohammad Nejad, and A. Younesi, "Reinforcement learning based PID controller design for LFC in a microgrid," *COMPEL-Int. J. Comput. Math. Electr. Electron. Eng.*, vol. 36, no. 4, pp. 1287–1297, Jul. 2017.

[58] 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.

[59] H. Ko, S. Pack, and V. C. M. Leung, "Mobility-aware vehicle-to-grid control algorithm in microgrids," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2165–2174, Jul. 2018.

[60] P. Kofinas, A. I. Dounis, and G. A. Vouros, "Fuzzy Q-learning for multi-agent decentralized energy management in microgrids," *Appl. Energy*, vol. 219, pp. 53–67, Jun. 2018.

[61] H. Liu, J. Li, and S. Ge, "Research on hierarchical control and optimisation learning method of multi-energy microgrid considering multi-agent game," *IET Smart Grid*, vol. 3, no. 4, pp. 479–489, May 2020.

[62] B. V. Mbuwir, F. Spiessens, and G. Deconinck, "Distributed optimization for scheduling energy flows in community microgrids," *Electric Power Syst. Res.*, vol. 187, Oct. 2020, Art. no. 106479.

[63] B. E. Nyong-Bassey, D. Giaouris, C. Patsios, S. Papadopoulou, A. I. Papadopoulos, S. Walker, S. Voutetakis, P. Seferlis, and S. Gadoue, "Reinforcement learning based adaptive power pinch analysis for energy management of stand-alone hybrid energy storage systems considering uncertainty," *Energy*, vol. 193, Feb. 2020, Art. no. 116622.

[64] Y. Yu, Z. Cai, and Y. Liu, "Double deep Q-learning coordinated control of hybrid energy storage system in island micro-grid," *Int. J. Energy Res.*, vol. 45, no. 2, pp. 3315–3326, Feb. 2021.

[65] Y. Tang, J. Yang, J. Yan, and H. He, "Intelligent load frequency controller using GrADP for island smart grid with electric vehicles and renewable resources," *Neurocomputing*, vol. 170, pp. 406–416, Dec. 2015.

[66] M. Bagheri, V. Nurmanova, O. Abedinia, and M. Salay Naderi, "Enhancing power quality in microgrids with a new online control strategy for DSTATCOM using reinforcement learning algorithm," *IEEE Access*, vol. 6, pp. 38986–38996, 2018.

[67] M. Jafari, A. Ghasemkhani, V. Sarfi, H. Livani, L. Yang, and H. Xu, "Biologically inspired adaptive intelligent secondary control for MGs under cyber imperfections," *IET Cyber-Physical Systems: Theory Appl.*, vol. 4, no. 4, pp. 341–352, Dec. 2019.

[68] M. Hajihosseini, M. Andalibi, M. Gheisarnejad, H. Farsizadeh, and M.-H. Khooban, "DC/DC power converter control-based deep machine learning techniques: Real-time implementation," *IEEE Trans. Power Electron.*, vol. 35, no. 10, pp. 9971–9977, Oct. 2020.

[69] H. Nie, Y. Chen, Y. Xia, S. Huang, and B. Liu, "Optimizing the post-disaster control of islanded microgrid: A multi-agent deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 153455–153469, 2020.

[70] D. Chen, K. Chen, Z. Li, T. Chu, R. Yao, F. Qiu, and K. Lin, "PowerNet: Multi-agent deep reinforcement learning for scalable powergrid control," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1007–1017, Mar. 2022.

[71] A. M. Dissanayake and N. C. Ekneligoda, "Droop-free optimal feedback control of distributed generators in islanded DC microgrids," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 9, no. 2, pp. 1624–1637, Apr. 2021.

[72] A. M. Dissanayake and N. C. Ekneligoda, "Decentralized optimal stabilization of active loads in islanded microgrids," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 932–942, Mar. 2021.

[73] M. H. Khooban and M. Gheisarnejad, "A novel deep reinforcement learning controller based type-II fuzzy system: Frequency regulation in microgrids," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 4, pp. 689–699, Aug. 2021.

[74] P. R. Massenio, D. Naso, F. L. Lewis, and A. Davoudi, "Data-driven sparsity-promoting optimal control of power buffers in DC microgrids," *IEEE Trans. Energy Convers.*, vol. 36, no. 3, pp. 1919–1930, Sep. 2021.

[75] P. I. N. Barbalho, V. A. Lacerda, R. A. S. Fernandes, and D. V. Coury, "Deep reinforcement learning-based secondary control for microgrids in islanded mode," *Electric Power Syst. Res.*, vol. 212, Nov. 2022, Art. no. 108315.

[76] K. Nosrati, A. Tepljakov, E. Petlenkov, Y. Levron, V. Skiparev, and J. Belikov, "Coordinated PI-based frequency deviation control of isolated hybrid microgrid: An online multi-agent tuning approach via reinforcement learning," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Eur. (ISGT-Europe)*, Oct. 2022, pp. 1–5.

[77] K. Nosrati, A. Tepljakov, E. Petlenkov, V. Skiparev, J. Belikov, and Y. Levron, "Constrained intelligent frequency control in an AC microgrid: An online reinforcement learning based PID tuning approach," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2023, pp. 1–5.

[78] Y. Lim and H.-M. Kim, "Strategic bidding using reinforcement learning for load shedding in microgrids," *Comput. Electr. Eng.*, vol. 40, no. 5, pp. 1439–1446, Jul. 2014.

[79] D. Domínguez-Barbero, J. García-González, M. A. Sanz-Bobi, and E. F. Sánchez-Úbeda, "Optimising a microgrid system by deep reinforcement learning techniques," *Energies*, vol. 13, no. 11, p. 2830, Jun. 2020.

[80] J. Hao, D. W. Gao, and J. J. Zhang, "Reinforcement learning for building energy optimization through controlling of central HVAC system," *IEEE Open Access J. Power Energy*, vol. 7, pp. 320–328, 2020.

[81] B. V. Mbuwir, D. Geysen, F. Spiessens, and G. Deconinck, "Reinforcement learning for control of flexibility providers in a residential microgrid," *IET Smart Grid*, vol. 3, no. 1, pp. 98–107, Feb. 2020.

[82] O. Dutta and A. Mohamed, "Reducing the risk of cascading failure in active distribution networks using adaptive critic design," *IET Gener., Transmiss. Distribution*, vol. 14, no. 13, pp. 2592–2601, Jul. 2020.

[83] Q. Zhang, K. Dehghanpour, Z. Wang, F. Qiu, and D. Zhao, "Multi-agent safe policy learning for power management of networked microgrids," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1048–1062, Mar. 2021.

[84] N.-L. Mo, Z.-H. Guan, D.-X. Zhang, X.-M. Cheng, Z.-W. Liu, and T. Li, "Data-driven based optimal distributed frequency control for islanded AC microgrids," *Int. J. Electr. Power Energy Syst.*, vol. 119, Jul. 2020, Art. no. 105904.

[85] Z. Qin, D. Liu, H. Hua, and J. Cao, "Privacy preserving load control of residential microgrid via deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4079–4089, Sep. 2021.

[86] H. Song, Y. Liu, J. Zhao, J. Liu, and G. Wu, "Prioritized replay dueling DDQN based grid-edge control of community energy storage system," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 4950–4961, Nov. 2021.

[87] D. Fang, X. Guan, Y. Peng, H. Chen, T. Ohtsuki, and Z. Han, "Distributed deep reinforcement learning for renewable energy accommodation assessment with communication uncertainty in Internet of Energy," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 8557–8569, May 2021.

[88] L. Yin and B. Zhang, "Time series generative adversarial network controller for long-term smart generation control of microgrids," *Appl. Energy*, vol. 281, Jan. 2021, Art. no. 116069.

[89] S. Totaro, I. Boukas, A. Jonsson, and B. Cornélusse, "Lifelong control of off-grid microgrid with model-based reinforcement learning," *Energy*, vol. 232, Oct. 2021, Art. no. 121035.

[90] M. Gheisarnejad, H. Farsizadeh, and M. H. Khooban, "A novel nonlinear deep reinforcement learning controller for DC–DC power buck converters," *IEEE Trans. Ind. Electron.*, vol. 68, no. 8, pp. 6849–6858, Aug. 2021.

[91] H. Hua, Z. Qin, N. Dong, Y. Qin, M. Ye, Z. Wang, X. Chen, and J. Cao, "Data-driven dynamical control for bottom-up energy Internet system," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 315–327, Jan. 2022.

[92] Y. Tang, W. Hu, B. Zhang, D. Cao, N. Hou, Y. Li, Z. Chen, and F. Blaabjerg, "Deep reinforcement learning-aided efficiency optimized dual active bridge converter for the distributed generation system," *IEEE Trans. Energy Convers.*, vol. 37, no. 2, pp. 1251–1262, Jun. 2022.

[93] P. Chen, S. Liu, B. Chen, and L. Yu, "Multi-agent reinforcement learning for decentralized resilient secondary control of energy storage systems against DoS attacks," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 1739–1750, May 2022.

[94] G. Gao, Y. Wen, and D. Tao, "Distributed energy trading and scheduling among microgrids via multiagent reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10638–10652, Dec. 2022.

[95] D. J. B. Harrold, J. Cao, and Z. Fan, "Renewable energy integration and microgrid energy trading using multi-agent deep reinforcement learning," *Appl. Energy*, vol. 318, Jul. 2022, Art. no. 119151.

[96] A. J. Abianeh, Y. Wan, F. Ferdowsi, N. Mijatovic, and T. Dragicevic, "Vulnerability identification and remediation of FDI attacks in islanded DC microgrids using multiagent reinforcement learning," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 6359–6370, Jun. 2022.

[97] L. Wu and J. Zhang, "Economic operation and management of microgrid system using deep reinforcement learning," *Comput. Electr. Eng.*, vol. 100, May 2022, Art. no. 107879.

[98] J. Li, S. Yang, and T. Yu, "Data-driven cooperative load frequency control method for microgrids using effective exploration-distributed multi-agent deep reinforcement learning," *IET Renew. Power Gener.*, vol. 16, no. 4, pp. 655–670, Mar. 2022.

[99] J. Sang, H. Sun, and L. Kou, "Deep reinforcement learning microgrid optimization strategy considering priority flexible demand side," *Sensors*, vol. 22, no. 6, p. 2256, Mar. 2022.

[100] Y. Du and D. Wu, ''Deep reinforcement learning from demonstrations to assist service restoration in islanded microgrids,'' *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 1062–1072, Apr. 2022.

[101] D. Domínguez-Barbero, J. García-González, and M. Á. Sanz-Bobi, ''Twin-delayed deep deterministic policy gradient algorithm for the energy management of microgrids,'' *Eng. Appl. Artif. Intell.*, vol. 125, Oct. 2023, Art. no. 106693.

[102] H. Xiao, X. Pu, W. Pei, L. Ma, and T. Ma, ''A novel energy management method for networked multi-energy microgrids based on improved DQN,'' *IEEE Trans. Smart Grid*, vol. 14, no. 6, pp. 4912–4926, Nov. 2023.

[103] D. Qiu, T. Chen, G. Strbac, and S. Bu, ''Coordination for multienergy microgrids using multiagent reinforcement learning,'' *IEEE Trans. Ind. Inform.*, vol. 19, no. 4, pp. 5689–5700, Apr. 2023.

[104] J. Li and T. Zhou, ''Prior knowledge incorporated large-scale multiagent deep reinforcement learning for load frequency control of isolated microgrid considering multi-structure coordination,'' *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 3923–3934, Mar. 2024.

[105] O. Oboreh-Snapps, B. She, S. Fahad, H. Chen, J. Kimball, F. Li, H. Cui, and R. Bo, ''Virtual synchronous generator control using twin delayed deep deterministic policy gradient method,'' *IEEE Trans. Energy Convers.*, vol. 39, no. 1, pp. 214–228, Mar. 2024.

[106] Z. Benhmidouch, S. Moufid, A. Ait-Omar, A. Abbou, H. Laabassi, M. Kang, C. Chatri, I. Hammou Ou Ali, H. Bouzekri, and J. Baek, ''A novel reinforcement learning policy optimization based adaptive VSG control technique for improved frequency stabilization in AC microgrids,'' *Electr. Power Syst. Res.*, vol. 230, May 2024, Art. no. 110269.

[107] B. She, F. Li, H. Cui, H. Shuai, O. Oboreh-Snapps, R. Bo, N. Praisuwanna, J. Wang, and L. M. Tolbert, ''Inverter PQ control with trajectory tracking capability for microgrids based on physics-informed reinforcement learning,'' *IEEE Trans. Smart Grid*, vol. 15, no. 2, pp. 99–112, Jan. 2024.

[108] 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.

[109] X. Chen, M. Zhang, Z. Wu, L. Wu, and X. Guan, ''Model-free load frequency control of nonlinear power systems based on deep reinforcement learning,'' *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 6825–6833, Apr. 2024.

[110] 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.

[111] A. Salari, M. Zeinali, and M. Marzband, ''Model-free reinforcement learning-based energy management for plug-in electric vehicles in a cooperative multi-agent home microgrid with consideration of travel behavior,'' *Energy*, vol. 288, Feb. 2024, Art. no. 129725.

[112] Z. Wu, Z. Lv, X. Huang, and Z. Li, ''Data driven frequency control of isolated microgrids based on priority experience replay soft deep reinforcement learning algorithm,'' *Energy Rep.*, vol. 11, pp. 2484–2492, Jun. 2024.

[113] T. Tabassum, S. Lim, and M. R. Khalghani, ''Artificial intelligence-based detection and mitigation of cyber disruptions in microgrid control,'' *Electric Power Syst. Res.*, vol. 226, Jan. 2024, Art. no. 109925.

[114] S. Zheng, Z. Wu, L. Song, W. Gu, W. Liu, J. Zhao, Z. Xu, and T. Hong, ''Distributed economic control strategy based on reinforcement pinning control for microgrids,'' *Electric Power Syst. Res.*, vol. 237, Dec. 2024, Art. no. 111006.

[115] J. Gong, N. Yu, F. Han, B. Tang, H. Wu, and Y. Ge, ''Energy scheduling optimization for microgrids based on partially observable Markov game,'' *IEEE Trans. Artif. Intell.*, vol. 5, no. 11, pp. 5371–5380, Jul. 2024.

[116] Y. Xia, Y. Xu, and X. Feng, ''Hierarchical coordination of networked-microgrids toward decentralized operation: A safe deep reinforcement learning method,'' *IEEE Trans. Sustain. Energy*, vol. 15, no. 3, pp. 1981–1993, Jul. 2024.

[117] C. Mu, Y. Shi, N. Xu, X. Wang, Z. Tang, H. Jia, and H. Geng, ''Multi-objective interval optimization dispatch of microgrid via deep reinforcement learning,'' *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 2957–2970, May 2024.

[118] A. Nandakumar, Y. Li, Z. Xu, and D. Huang, ''Enhancing transient dynamics stabilization in islanded microgrids through adaptive and hierarchical data-driven predictive droop control,'' *IEEE Trans. Smart Grid*, vol. 16, no. 1, pp. 396–410, Jan. 2025.

[119] N. Piovesan, D. López-Pérez, M. Miozzo, and P. Dini, ''Joint load control and energy sharing for renewable powered small base stations: A machine learning approach,'' *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 1, pp. 512–525, Mar. 2021.

[120] S. Gao, C. Xiang, M. Yu, K. T. Tan, and T. H. Lee, ''Online optimal power scheduling of a microgrid via imitation learning,'' *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 861–876, Mar. 2022.

[121] X. Ge and J. Khazaei, ''Physics-informed convolutional neural network for microgrid economic dispatch,'' *Sustain. Energy, Grids Netw.*, vol. 40, Dec. 2024, Art. no. 101525.

[122] Y. Wang and B. Pal, ''Destabilizing attack and robust defense for inverter-based microgrids by adversarial deep reinforcement learning,'' *IEEE Trans. Smart Grid*, vol. 14, no. 6, pp. 4839–4850, Nov. 2023.

[123] N. Uthayansuthi and P. Vateekul, ''Optimization of peer-to-peer energy trading with a model-based deep reinforcement learning in a non-sharing information scenario,'' *IEEE Access*, vol. 12, pp. 111021–111034, 2024.

[124] D. S. Kushwaha, Z. Abdollahi Biron, and M. Hu, ''Transformer neural network-based transfer learning for economic dispatch of microgrids,'' in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2024, pp. 1–5.

[125] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, ''Attention is all you need,'' 2017, *arXiv:1706.03762*.

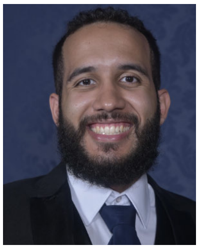
[126] V. Skiparev, K. Nosrati, J. Belikov, A. Tepljakov, and E. Petlenkov, ''An enhanced NN-based load frequency control design of MGs: A fractional order modeling method,'' in *Proc. IEEE 17th Int. Conf. Compat., Power Electron. Power Eng. (CPE-POWERENG)*, Jun. 2023, pp. 1–6.

[127] S. Arora and P. Doshi, ''A survey of inverse reinforcement learning: Challenges, methods and progress,'' *Artif. Intell.*, vol. 297, Aug. 2021, Art. no. 103500.

[128] R. Sepehrzad, A. S. G. Langeroudi, A. Khodadadi, S. Adinehpour, A. Al-Durra, and A. Anvari-Moghaddam, ''An applied deep reinforcement learning approach to control active networked microgrids in smart cities with multi-level participation of battery energy storage system and electric vehicles,'' *Sustain. Cities Soc.*, vol. 107, Jul. 2024, Art. no. 105352.

[129] P. I. N. Barbalho, D. V. Coury, V. A. Lacerda, and R. A. S. Fernandes, ''Hardware-in-the-loop testing of a deep deterministic policy gradient algorithm as a microgrid secondary controller,'' in *Proc. IEEE PES Innov. Smart Grid Technol. Eur. (ISGT EUROPE)*, Oct. 2023, pp. 1–5.

[130] H. Negahdar, A. Karimi, Y. Khayat, and S. Golestan, ''Reinforcement learning-based event-triggered secondary control of DC microgrids,'' *Energy Rep.*, vol. 11, pp. 2818–2831, Jun. 2024.



PEDRO I. N. BARBALHO received the B.Sc. and M.Sc. degrees in electrical engineering from EESC, University of São Paulo (USP), Brazil, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree. In 2023, he was a Visiting Researcher with the University of Strathclyde. His research interests include machine learning algorithms used in power system applications and microgrid control.

ANDERSON L. MORAES was born in Descalvado, Brazil, in 1992. He received the B.Sc. degree in electrical engineering from Central Paulista University Center, São Carlos, Brazil, in 2015, and the M.Sc. degree in electrical engineering from the Federal University of São Carlos, São Carlos, in 2021. He is currently pursuing the Ph.D. degree in electrical engineering with the University of São Paulo, São Carlos. His research interests include the application of machine learning for power quality and fault location in the context of smart grids and microgrids.

VINICIUS A. LACERDA received the B.Sc. and Ph.D. degrees in electrical engineering from the University of São Paulo, Brazil, in 2015 and 2021, respectively. From 2018 to 2019, he was a Visiting Researcher with the University of Strathclyde, Glasgow, U.K. He joined the Centre d'Innovacio Tecnològica en Convertidors Estatics i Accionaments, Universitat Politècnica de Catalunya (CITCEA-UPC), Barcelona, Spain, in 2021. He is currently a Lecturer with CITCEA-UPC. He is a Senior Power Systems Engineer with eRoots Analytics, focusing on EMT and phasor simulation of modern power grids and control and design of grid-forming converters.

PEDRO H. A. BARRA received the B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of Uberlândia, Brazil, in 2015 and 2017, respectively, and the Ph.D. degree in electrical engineering from the University of São Paulo, São Carlos, Brazil, in 2022. He is currently an Adjunct Professor with the Federal University of Uberlândia. His research interests include power system protection, microgrids, distribution systems, and smart grids.

RICARDO A. S. FERNANDES (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Educational Foundation of Barretos, Barretos, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of São Paulo, São Carlos, Brazil, in 2009 and 2011, respectively. From 2015 to 2017, he was a Visiting Professor with the Polytechnic Institute of Porto. He was an Associate Professor with the Federal University of São Carlos, Brazil. He is currently an Assistant Professor with the University of São Paulo. His research interests include embedded systems, signal processing, machine learning, and smart grids.

DENIS V. COURY received the B.Sc. degree in electrical engineering from the Federal University of Uberlândia, Brazil, in 1983, the M.Sc. degree from EESC, University of São Paulo, Brazil, in 1986, and the Ph.D. degree from Bath University, England, in 1992. He joined the Department of Electrical and Computer Engineering, University of São Paulo, São Carlos, Brazil, in 1986. He is currently a Full Professor with the Power Systems Group, Department of Electrical and Computer Engineering, University of São Paulo. He spent his Sabbatical with Cornell University, USA, from 1999 to 2000. His research interests include power system protection, expert systems, and Smart Grids.