IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 December 2023, accepted 6 January 2024, date of publication 10 January 2024,
date of current version 18 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3352032

RESEARCH ARTICLE

SFNAS-DDPG: A Biomass-Based Energy Hub Dynamic Scheduling Approach via Connecting Supervised Federated Neural Architecture Search and Deep Deterministic Policy Gradient

AMIRHOSSEIN DOLATABADI[1], (Graduate Student Member, IEEE), HUSSEIN ABDELTAWAB[2], (Senior Member, IEEE), AND YASSER ABDEL-RADY I. MOHAMED[1], (Fellow, IEEE)

[1]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada
[2]Department of Engineering, Wake Forest University, Winston-Salem, NC 27101, USA

Corresponding author: Amirhossein Dolatabadi (adolatab@ualberta.ca)

This work was supported in part by the Canada First Research Excellence Fund through the University of Alberta's Future Energy Systems Research Initiative, in part by the Alberta Innovates Graduate Student Scholarships for Data-Enabled Innovation (GSS-DEI), and in part by Alberta Innovates.

ABSTRACT The transition to a near-zero-emission power and energy industry for facing up to global warming issues is dominated by the use of renewable energy resources such as bioenergy and solar energy. When these resources are coordinated within an energy hub framework, the system's flexibility is increased and dispatchable energy is provided by enhancing the share of renewable-dominated power. This paper proposes a dynamic scheduling framework for an energy hub with a biomass-solar hybrid renewable system. A hybrid forecasting model based on convolutional neural networks (CNNs) and Gated recurrent units (GRUs) is developed first to capture solar-related uncertainty sensibly, which will provide a great opportunity for the learning-based controller to determine an effective operation strategy in an optimal manner, especially on a cloudy-weather day. Then, a supervised federated neural architecture search (SFNAS) technique has been presented to eliminate the need for manual engineering of deep neural network models and the unnecessary computational burden associated with them. Finally, the deep deterministic policy gradient (DDPG), as an actor-critic deep reinforcement learning (DRL) methodology, enables the biomass-based energy hub to achieve cost-effective dynamic control strategies by addressing the decision-making problem as a highly dynamic continuous state-action model. The major conclusions of the numerical results show the effectiveness of the proposed SFNAS-DDPG method from average operating cost reduction up to 7.31% compared to the conventional DDPG model.

INDEX TERMS Actor-critic deep reinforcement learning, biomass energy, energy hub, federated learning (FL), neural architecture search (NAS).

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

Over the past decade, renewable energy sources (RESs), such as bioenergy and solar, have received unprecedented attention and become more prevalent to meet the ambitious goals of policymakers for reducing greenhouse gas emissions by moving towards a more sustainable energy industry. On the other hand, with the popularity of RESs continually increasing, their inherent volatility and intermittency pose a challenge to the integration and utilization of a high-penetration of RESs, and raise operational concerns for energy system operators. In this regard, the importance of coordinated control of interconnected energy infrastructures has been increasingly recognized for facilitating the development of a low-carbon, economically feasible, and

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

renewable-dominated energy supply. This direction has led to the significant recognition of energy hub systems as resilient, reliable, and economical means of redistributing supply and demand among different energy sectors at the local level [1]. Meanwhile, with biomass energy resources being widely distributed and its potential to support multiple forms of energy demands, such as electricity and thermal, the integrated operation of other RESs and biomass can offer a cost-effective alternative compared to utility grid operations.

B. LITERATURE SURVEY

Several recent representative studies have been reported on the optimization of energy hub systems from different perspectives, including optimal design and expansion planning [1], [2], optimal power dispatch [3], [4], energy trading scheme planning [5], [6], voltage-frequency optimization [7], etc. The literature on the integration of biomass energy resources into the energy hub concept is somewhat limited and scattered. Using climate-independent biomass energy, an energy hub can balance the stochastic output of weather-driven renewable RESs. In both [8] and [9], the planning problem for a biomass-based energy facility was studied and formulated as a two-stage mixed-integer linear programming (MILP) model. While the authors in [8] focused on the optimal planning of solar and biogas energy for reducing the dependency of the energy hub system on battery storage systems in remote regions, the model proposed in [9] centered on the expansion planning of integrated electric power and biogas delivery networks. In terms of the scheduling and operation of biomass-based multi-carrier energy systems, the existing literature studies the use of biomass energy resources as a solution to the increasing demand for diversified and affordable energy services. In [10], a stochastic optimal operation strategy for an integrated solar and biomass system is proposed by considering the uncertainties in market prices and solar irradiation. Reference [4] assesses an optimal operation strategy for an energy hub with a biogas-solar-wind hybrid renewable system. The approach in [4] is extended by providing a distributed stochastic scheduling framework for the coordinated operation of interconnected biogas-solar-wind systems in [11]. Both [12] and [13] propose trading schemes for the biomass-concentrated solar system. Stochastic optimization and information gap decision theory (IGDT) approaches are leveraged to address uncertain factors.

Even though some advances were made, the aforementioned works also had three limitations. Firstly, they need to understand the dynamics of components. Developing a dynamic model that can accurately simulate components' behavior is challenging due to the multitude of factors that influence it. Further, model-based approaches may differ in premises or performance in relation to particular components, so their generalizability may be limited. Secondly, algorithms require explicit knowledge of how uncertainty is represented (for example, RES production's probability distributions). Finally, a third limitation is the lack of support for an online control mechanism, particularly for large-scale solutions. Specifically, these methods need to perform the optimization by selecting the most optimal solution among a set of possible ones, regardless of the problem size.

Alternatively, the disadvantages described above can be overcome through the use of model-free learning-based techniques. Recently, some research has been conducted on applying reinforcement learning (RL) to the energy management problem of multi-carrier energy systems, such as Q-learning (QL) [14], QL-linear programming (LP) [15], and fuzzy QL [16]. The combination of RL and deep neural networks (DNN), as a powerful function approximator, has led to the development of deep reinforcement learning (DRL). Through multiple interpretations, DRL can leverage the growing data collected from a variety of sources, revealing optimum control policies and dealing with uncertainty related to the data. Additionally, DRL is a model-free, self-adaptive method that requires no explicit modeling of probability distributions or constraints on uncertainties. The application of deep Q network (DQN) concepts to energy systems' control and scheduling has been developed for different scenarios involving the optimization of state and action spaces in high dimensions, e.g., energy storage system [17], industrial Internet of Things (IIoT) [18], electric vehicles (EV) [19], cyber uncertainties [20], and building energy management [21]. Although DQN has been regarded as a considerable improvement over conventional QL in approximating Q-value functions, its performance is suboptimal if the action spaces of the environment are continuous. DQN offers considerable advantages over QL, but its performance can be suboptimal regarding continuous action spaces. There are, for example, only five and seven levels of discrete control actions for lithium-ion batteries and EVs in [22] and [23], respectively. The adoption of a continuous action space is therefore of the utmost importance. This being the case, the deep deterministic policy gradient (DDPG) methodology was introduced and implemented on a few decision-making tasks involved in scheduling and controlling microgrids, including electricity market participation [24], distribution networks [25], volt-var control (VVC), and state of charge (SoC) control [26]. Through using both actor and critic networks, it provides both the benefits of value-based and Monte Carlo policy gradient approaches. Comparatively to the Monte Carlo policy gradient, DDPG uses fewer samples to learn action space selection, resulting in lower computational demands. Furthermore, in contrast to the value-based technique, stochastic policies can be used to address RL problems that involve continuous actions.

In the meantime, DNN algorithms have made promising progress in several tasks, including prediction. However, their performance is still heavily influenced by the parameters and architecture of the neural network (the number of layers and the number of nodes, as well as the connectivity between layers). In most cases, it is challenging to evaluate how

parameter settings influence the performance of the available models, even when they have a thorough understanding of both machine learning (ML) and deep learning (DL). Furthermore, recent studies in the field emphasize the use of higher-level models in order to obtain more accurate predictions, increasing the number of parameters included in the model. This has made exploring such large parameter spaces a more challenging undertaking. Neither expert knowledge nor empirical trial and error are always useful in this regard. A neural architecture search (NAS) method finds the optimal network architecture for a given situation by defining a search space, a search strategy, and a performance estimation strategy [27], [28]. NAS-developed network structures have been shown to be superior to those constructed by hand in a variety of fields.

Over the past few years, the field of NAS has become an active area of research and experienced remarkable success [27], [29], [30]. A Bayesian optimization-based NAS has been used to optimize the structure of DNNs. For example, a new kernel captures the relevant parameters in [31] and a joint optimization of the architecture and hyperparameters is performed using Bayesian optimization in [32]. Recent proposals have attempted to turn NAS problems into RL problems. The NAS problem is addressed using Q-learning in [33] and [34]. To improve the search procedure described in [33] and [35] uses a predictor. With shared DNN parameters, [36] increases the speed of search procedures. By developing a differentiable representation of the NAS model, gradient-based NAS techniques seek to optimize the parameters of the NAS. The performance of other neural networks is predicted by a neural network in [37] and [38]. An approach aimed at reducing training time and memory consumption for NAS is described in [39]. A generalization problem in differentiable architecture search (DARTS) models [40] is addressed in [41] by breaking the problem into sub-problems. Recent years have seen significant interest in evolutionary computation approaches to optimize DNN parameters. DNNs are employed as genotypes and phenotypes in a mutation-only Genetic Algorithm (GA) in [42] and [43]. By evolving the cascades of convolutional filters, a GA for classification is presented in [44]. It is noteworthy that evolutionary-based NAS approaches result in neural networks of superior performance, but moderate in size, as shown in [27], [29] and [30]. Additionally, privacy concerns are becoming more prevalent, which has led to a growing interest in machine learning approaches that preserve the privacy of users. As a machine learning paradigm, federated learning (FL) deals with concerns about data privacy, particularly when working with distributed and heterogeneous information [45]. Thus, FL combined with NAS is capable of effectively addressing privacy concerns associated with NAS.

C. CONTRIBUTION AND PAPER STRUCTURE

Considering the limitations of previous studies and to fill the research gaps outlined above, this paper is directed toward presenting a dynamic scheduling framework for a biomass-based energy hub. An improved actor-critic DRL algorithm, DDPG, and the supervised federated neural architecture search (SFNAS) are incorporated and employed to form a novel model-free and self-adaptable energy management algorithm for a RES-based multi-carrier energy supply infrastructure. For a diverse and dynamic RES-based energy hub system, these characteristics can facilitate flexible scheduling and coordinated decision-making. The proposed SFNAS tries to investigate the merits and limitations of various NAS methods for learning the architecture and model parameters for a given task. An innovative method is then presented for engaging multiple NAS approaches and improving their individual performance by utilizing a supervising agent to improve their training losses. As a result of repeated interaction with the environment, the DDPG agent is able to gain experience to provide optimal dynamic control signals without the need to model constrained probability distributions or uncertainty sets. Moreover, a forecasting technique based on convolutional neural networks (CNNs) and gated recurrent units (GRUs) has been incorporated into the DDPG decision-making framework to achieve decent scheduling results. The major contributions of this paper are summarized as follows:

1) Focusing on the ever-increasing role of biomass energy resources worldwide, the impact of several underlying factors on the cost-benefit analysis of a biomass-based energy hub is investigated using a fully data-driven and model-free actor-critic DRL-based decision-making framework taking into consideration the physical characteristics and thermodynamic effect of temperature-sensitive biogas production at the operation stage. To our best knowledge, such method is rarely investigated before.
2) Based on the FL settings, the proposed SFNAS approach can be used directly to search for the optimal architecture by leveraging training loss reshaping by a supervisor. The test results show that the proposed SFNAS approach achieves faster and smoother convergence than the respective baseline methods. In comparison with the baseline model, the proposed SFNAS approach significantly improves both forecasting RMSE and training time by 10.33% and 9.54%, respectively.
3) A hybrid CNN-GRU solar irradiance forecasting model that captures high levels of abstraction from sky images and numerical measurements was developed and optimized by SFNAS to help the DDPG agent choose the most efficient control policy especially when PV power is highly intermittent on cloudy days.
4) Compared with conventional energy hub scheduling methods, such as stochastic programming (SP) or robust optimization (RO), which require a perfect and detailed understanding of the operational model and parameters of the system, the proposed model-free data-driven energy management

framework can overcome potential uncertainties and determine the optimal control decisions by the learning mechanism.

5) The DDPG is a model-free decision-making framework that offers more realistic and cost-effective decision-making strategies than existing conventional decision-making methodologies like SP and RO, which simplify the scheduling problem with linearized models and piecewise linear curve fitting methods. Ignoring the practical limitations and nonlinear constraints like biogas production thermodynamics, dynamic efficiency of fuel cells, valve admission behavior of power generation systems, and CHPs' non-convex operation regions makes the results less robust.

This paper proceeds as follows. Section II presents the proposed SFNAS algorithm and the hybrid DL-based 2-D CNN-GRU prediction approach. Section III describes the highly nonlinear, realistic model of the system. Following that, Section IV presents the DDPG-based dynamic scheduling method for the biomass-based multi-carrier energy system. Section V is dedicated to the case studies and simulation results, and finally, conclusions are provided in Section VI.

II. SUPERVISED FEDERATED NEURAL ARCHITECTURE SEARCH AND HYBRID 2-D CNN-GRU FORECASTING MODEL

In this section, the mathematical representation of the proposed SFNAS, which leverages a supervisor to enhance the performance of multiple NAS agents, is presented. The supervisor (meta-learner) and NAS agents (base learners) are trained using a bi-level optimization strategy. Fig. 1 illustrates the schematic diagram of the proposed SFNAS methodology.

A. NAS AGENTS AS BASE-LEARNERS

1) OBJECTIVE

By learning the architecture, ω_n, and parameters, θ_n, of the model, each base learner i optimizes its own model using *momentum-regularized* NAS technique. (ω_n, θ_n) is learned by each base-learner n using a momentum-based gradient descent algorithm as follows.

$$\begin{aligned} v_n &:= (1-\lambda_n)v_n + \lambda_n \nabla_n J_n(\theta_n, \omega_n) \\ \theta_n &\leftarrow \theta_n - \alpha_\theta v_n, \\ \omega_n &\leftarrow \omega_n - \alpha_\omega v_n \end{aligned} \tag{1}$$

where α_θ and α_ω correspond to the learning rates for θ and ω, respectively, and, v_n represents the momentum from past steps. When the hyperparameter λ_n is set to a lower value in the updates, longer gradient updates are maintained (this can be demonstrated by expanding the recursive updates, which are not shown here due to limited space). According to the NAS algorithm used by each base-learner, the training loss varies. Hence, for ease of explanation, the training loss is denoted by $J_n(\theta_n, \omega_n)$. The following definition of L_2-regularized mean squared error (MSE) will be used when we assume identical training losses $J_n(\cdot,\cdot)$ for all agents.

$$J_n(\theta_n, \omega_n) := \|y - \hat{y}_n\|^2 + L_2(\theta_n) \tag{2}$$

where $\hat{y}_n$ represents the forecasted value, i.e., the outcome of n-th base-learner's model with architecture ω_n.

2) MODEL

The weights λ_n that are proportional to r_n are estimated by training the meta-learner model with the input states (or the embedding of the inputs). For this purpose, a N-head classifier structure is employed, corresponding to λ_n, $n \in \{1, 2, \ldots, N\}$. Unlike a typical classification problem, r_ns change epoch by epoch, so the target values vary.

3) SELECTION PROCEDURE

The proposed SFNAS methodology includes a model selection procedure that is as follows.

$$\min_A \left[\zeta_1.N + \zeta_2.T + \zeta_3.RMSE\right] \tag{3}$$

where N, T, and $RMSE$ are the number of trainable parameters of deep learning model, training time, and the RMSE value for agent A, respectively. A FL application designer can determine the importance of the objectives by adjusting ζ_1, ζ_2, and ζ_3.

B. HYBRID DEEP 2-D CNN-GRU STRUCTURE

Meteorological and atmospheric variables greatly influence solar irradiance in solar power generation forecasting. A hybrid SFNAS-based DL model for solar irradiance prediction is presented to analyze a sequence of sky images and meteorological components by employing CNN and GRU structures, respectively. Several numerical meteorological features are employed in the proposed hybrid forecasting framework to enhance its capability to capture complex solar abstractions, including global horizontal irradiance (GHI), calendar features, temperature, relative humidity, and atmospheric pressure. The mathematical details of the hybrid deep CNN-GRU network can be expressed as follows [46], [47], and [48]:

1) CNN NETWORK

Through the use of learnable kernels and inductive bias, CNN convolutional layers provide better generalization capabilities. The mathematical expression for convolutions and pooling are as follows:

$$z_l^m = f\left(\sum_{n\in N_m}\left(z_{l-1}^n \otimes w_l^{nm}\right) + b_l^m\right) \tag{4}$$

$$z_l^m = f^{down}\left(z_{l-1}^m\right) \tag{5}$$

A max-pooling down-sampling method followed by a fully-connected layer is used in this paper to extract informative features.

$$z_l = g^f\left(\vartheta_l z_{l-1} + b_l\right) \tag{6}$$

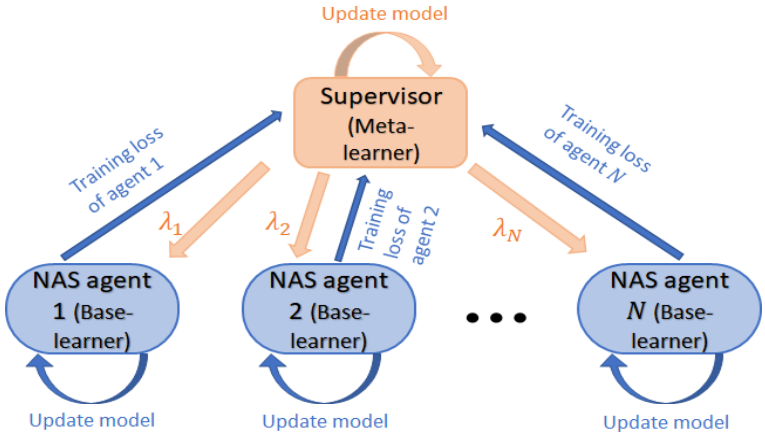

FIGURE 1. Schematic diagram of the proposed SFNAS methodology.

2) BLSTM NETWORK

Each LSTM block consists of three multiplicative units: input, output, and forget. Data is written, read, and reset using these gates. The following equations explain how LSTM forward passes work mathematically [49]:

$$f'_t = \sigma\left(w'^{f'}_{\chi}\chi_t + w'^{f'}_{h}h_{t-1} + b'^{f'}\right) \quad (7)$$

$$z'_t = \sigma\left(w'^{z'}_{\chi}\chi_t + w'^{z'}_{h}h_{t-1} + b'^{z'}\right) \quad (8)$$

$$g'_t = tanh\left(w'^{g'}_{\chi}\chi_t + w'^{g'}_{h}h_{t-1} + b'^{g'}\right) \quad (9)$$

$$o'_t = \sigma\left(w'^{o'}_{\chi}\chi_t + w'^{o'}_{h}h_{t-1} + b'^{o'}\right) \quad (10)$$

$$C'_t = g'_t \odot i'_t + C'_{t-1} \odot f'_t \quad (11)$$

$$h_t = tanh\left(C'_t\right) \odot o'_t \quad (12)$$

Gradient vanishing problems can be handled by LSTMs, but future context cannot be taken into account. As a result, BLSTM networks represent data in a much more sophisticated manner than traditional LSTMs because they utilize all the information of the temporal horizon. Following are the outputs generated by BLSTM networks [50]:

$$h''^{f}_t = tanh\left(w''^{f}_{\chi h''}\chi_t + h''^{f}_{t-1}\right)w''^{f}_{h''h''} + b''^{f}_{h''} \quad (13)$$

$$h''^{f}_t = tanh\left(w''^{f}_{\chi h''}\chi_t + h''^{f}_{t-1}\right)w''^{f}_{h''h''} + b''^{f}_{h''} \quad (14)$$

$$y''_t = h''_t w''_o + b''_o \quad (15)$$

where h''_t is composed of integrating h''^{f}_t and h''^{b}_t.

III. ENVIRONMENT MODEL AND PROBLEM FORMULATION

The purpose of this section is to discuss the main structure of the proposed biomass-based energy hub, taking into account nonlinear characteristics of the facility components such as biogas production thermodynamics, dynamic efficiency of fuel cells, valve admission behavior of power generation systems, and CHPs' non-convex operation regions.

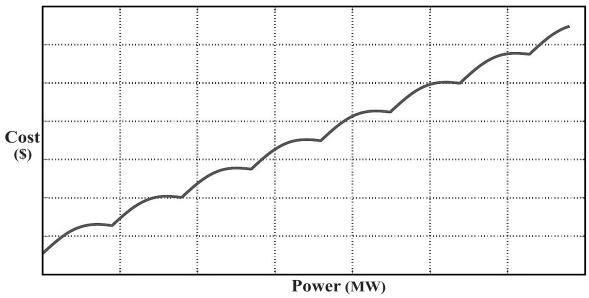

FIGURE 2. Valve-point effects on the fuel cost of a power-only unit.

A. POWER-ONLY UNITS

Generally, power-only generation units are modeled by convex quadratic cost functions. Due to valve admission effects, these models do not take into account ripples in production cost; therefore, they cannot provide a practical solution for real-life applications. In this regard, the quadratic cost function is modified to include an absolute sinusoidal term to capture this phenomenon efficiently. Fig. 2 shows the valve-point effects on the fuel cost of a power-only unit. The total operation cost including valve-point effects can be formulated as follows [51]:

$$C^c_{i,t}\left(E^c_{i,t}\right) = \alpha^{c1}_i.\left(E^c_{i,t}\right)^2 + \alpha^{c2}_i.E^c_{i,t} + \alpha^{c3}_i + V^c_{i,t}\left(E^c_{i,t}\right) \quad (16)$$

$$V^c_{i,t}\left(E^c_{i,t}\right) = \left|\lambda^c_i \sin\left(\rho^c_i\left(E^c_{i,\min} - E^c_{i,t}\right)\right)\right| \quad (17)$$

$$E^c_{i,\min}.I^c_{i,t} \le E^c_{i,t} \le E^c_{i,\max}.I^c_{i,t} \quad (18)$$

where α^{c1}_i, α^{c2}_i, α^{c3}_i, λ^c_i, and ρ^c_i are the cost coefficients of the i^{th} power-only unit. I^c and VPE^c indicate the commitment status and valve-point effects, respectively.

B. THERMAL UNITS

The k^{th} thermal unit's operating cost and limits at time t is defined by (19) and (20).

$$C^b_{k,t}\left(T^b_{k,t}\right) = \alpha^{b1}_k.\left(T^b_{k,t}\right)^2 + \alpha^{b2}_k.T^b_{k,t} + \alpha^{b3}_k \quad (19)$$

$$T^{b}_{k,\min}.I^{b}_{k,t} \le T^{b}_{k,t} \le T^{b}_{k,\max}.I^{b}_{k,t} \tag{20}$$

where $\alpha_i^{b_1}$, $\alpha_i^{b_2}$, $\alpha_i^{b_3}$, and k^{th} represent the cost coefficients and commitment status of the k^{th} heat-only unit.

C. CHP UNITS

Noteworthy is the fact that the electricity and thermal energy produced in CHP units depend on each other, so they cannot be independently adjusted. The cost function of j^{th} CHP is provided in (21). Convex combinations of extreme points represent the power and heat production in CHP units by (22)-(25) [52].

$$C^{chp}_{j,t} = \sum_{s=1}^{S}\sum_{d^s=1}^{D^s_j} \Gamma^{d^s}_{j,t} c^{d^s}_{j} \tag{21}$$

$$E^{chp}_{j,t} = \sum_{s=1}^{S}\left(\sum_{d^s=1}^{D^s_j} \Gamma^{d^s}_{j,t} x^{d^s}_{j}\right) \tag{22}$$

$$T^{chp}_{j,t} = \sum_{s=1}^{S}\left(\sum_{d^s=1}^{D^s_j} \Gamma^{d^s}_{j,t} y^{d^s}_{j}\right) \tag{23}$$

$$\sum_{d^s=1}^{D^s_j} \Gamma^{d^s}_{j,t} = V^{s}_{j,t} \tag{24}$$

$$0 \le \Gamma^{d^s}_{j,t} \le 1 \tag{25}$$

where D^s_j denotes the total number of corner points for the j^{th} CHP in sub-region s and the coefficient $\Gamma^{d^s}_{j,t}$ should satisfy constraints (24) and (25); $\left(x^{d^s}_{chp}, y^{d^s}_{chp}\right)$ represents the electricity and thermal generations related to the corner point d^s; $V^s_{j,t}$ is a binary variable indicating operation in the s^{th} subregion.

D. BIOMASS-BASED BOILER UNITS

The m^{th} Biomass-based boiler's operating cost is modeled in (26). Based on the energy conservation rules, the thermal energy produced by the biomass-based boiler is calculated by (27)-(29), while it is restricted within acceptable limits by constraint (30) [8].

$$C^{bm}_{m,t}\left(T^{bm}_{m,t}\right) = T^{bm}_{m,t}.\alpha^{bm}_{m} \tag{26}$$

$$T^{bm}_{m,t} = \frac{\vartheta^{bm}.\varsigma^{bm}.VOL^{bm}}{T^{bm}_{HR}} \times \left(1 - \frac{\kappa^{bm}}{\left(T^{bm}_{HR}.\pi^{bm}\right) - 1 + \kappa^{bm}}\right) \tag{27}$$

$$\kappa^{bm} = \alpha^{bm}_{1} e^{\alpha^{bm}_{2}.IVS} + \alpha^{bm}_{3} \tag{28}$$

$$T^{bm}_{m,t} = \varpi^{bm}_{m,t}.CV^{bm}.\eta^{bm}_{m} \tag{29}$$

$$T^{bm}_{m,\min}.I^{bm}_{m,t} \le T^{bm}_{m,t} \le T^{bm}_{m,\max}.I^{bm}_{m,t} \tag{30}$$

where α^{bm}_m, ϖ^{bm}_m, CV^{bm}, and η^{bm}_m denote the cost coefficient, mass of injected biomass feedstock, calorific value of biomass feedstock, and efficiency of the m^{th} biomass-based boiler, respectively; κ^{bm} is the stability and process rate kinetic parameter; π^{bm} is the thermophilic and mesophilic digestion's micro-organism growth rate; ϑ^{bm}, ς^{bm}, T^{bm}_{HR} and VOL^{bm} represent biogas biochemical potential, influent volatile solid concentration, hydraulic retention time and digester's volume, respectively. The nonlinear relationship between biogas production and temperature is illustrated in Fig. 3.

E. FUEL CELL UNIT

Fuel cells (FCs) contribute to the energy hub demand by generating electricity, recovering heat, and producing hydrogen. As part of the energy hub consumers' energy requirements, proton exchange membrane FCs are considered to supply some of their demand (since the by-product heat can be captured and used by these FCs, they can also be considered CHP systems). Hydrogen can be produced from the difference between the maximum capacity and the electrical output when the thermal load is low. The energy hub produces electricity by converting hydrogen from FCs into electricity and storing it in hydrogen reservoirs. Note that, the hydrogen reservoirs are used as a backup fuel for the hydrogen engine. The main fuel is hydrogen produced by reforming natural gas.

In the l^{th} FC unit, the amount of recovered thermal power, $T^{fc}_{l,t}$, can be determined as follows:

$$T^{fc}_{l,t} = TER^{fc}_{l,t}.\left(E^{fc,e}_{l,t} + \underline{E}^{fc,h}_{l,t}\right) \tag{31}$$

$$E^{fc,T}_{l,t} = E^{fc,e}_{l,t} + \underline{E}^{fc,h}_{l,t} \tag{32}$$

$$E^{fc}_{l,\min}.I^{fc}_{l,t} \le E^{fc,e}_{l,t} + \underline{E}^{fc,h}_{l,t} \le E^{fc}_{l,\max}.I^{fc}_{l,t} \tag{33}$$

where $\underline{E}^{fc,h}$, $E^{fc,e}$, and TER^{fc} are equivalent electric power for hydrogen production, electrical output power, and the thermal to electrical power ratio, respectively. The amount of stored hydrogen and charged/discharged hydrogen can be formulated as follows:

$$T2T^{fc}_{l,t} = T2T^{fc}_{l,t-1} + \left(\eta^{T2T}_{l,ch}.T2^{in}_{l,t}\right) - \left(T2^{out}_{l,t} \Big/ \eta^{T2T}_{l,dch}\right) \tag{34}$$

$$T2^{in}_{l,t} = \underline{E}^{fc,h}_{l,t}.\psi^{E2T} \tag{35}$$

$$T2^{out}_{l,t} = \overline{E}^{fc,h}_{l,t}.\psi^{E2T} \tag{36}$$

$$T2T^{fc}_{l,\min} \le T2T^{fc}_{l,t} \le T2T^{fc}_{l,\max} \tag{37}$$

$$T2^{in}_{l,\min}.I^{T2T,ch}_{l,t} \le T2^{in}_{l,t} \le T2^{in}_{l,\max}.I^{T2T,ch}_{l,t} \tag{38}$$

$$T2^{out}_{l,\min}.I^{T2T,dch}_{l,t} \le T2^{out}_{l,t} \le T2^{out}_{l,\max}.I^{T2T,dch}_{l,t} \tag{39}$$

$$I^{T2T,ch}_{l,t} + I^{T2T,dch}_{l,t} \le 1 \tag{40}$$

where ψ^{E2T}, $T2^{in/out}$, and $T2T$ indicate the hydrogen (kg) to electric power (kW) ratio, equivalent charged/discharged hydrogen, and stored hydrogen. Most existing studies on FC consider its efficiency to be constant, whereas it can vary according to the amount of electricity produced. Based on Fig. 4, the thermal to electrical power ratio and the efficiency are functions of the part load ratio variable [53].

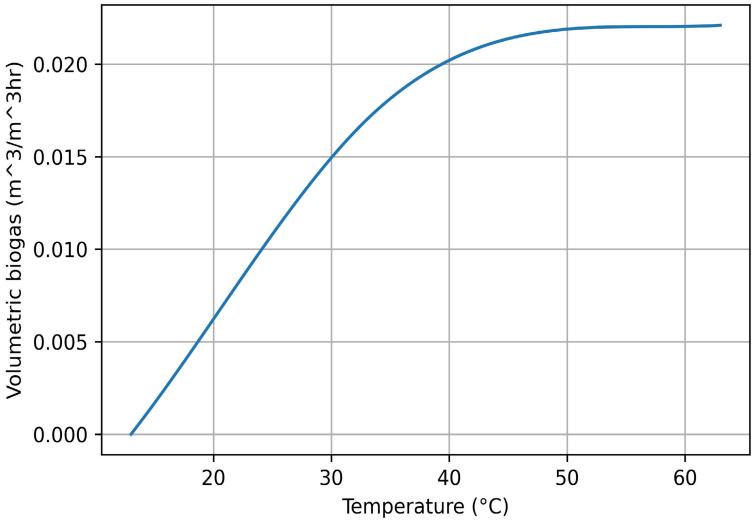

FIGURE 3. Nonlinear relationship between biogas production and temperature.

Equation (41) gives the total operating cost of the FC unit.

$$C^{fc}_{l,t}\left(E^{fc,e}_{l,t}, \underline{e}^{fc,h}_{l,t}\right) = \left(\frac{E^{fc,T}_{l,t}\pi^{ng}_{t}}{\eta^{fc}_{l,t}}\right) + \left(C^{fc,p}_{l}.\underline{E}^{fc,h}_{l,t}\right) + \left(C^{fc,OM}_{l}.E^{fc,T}_{l,t}\right) \tag{41}$$

F. POWER BALANCE

During each time interval t, the amount of electrical and thermal power produced and consumed should be equal, as stated in (42) and (43).

$$\sum_{i\in N_C} E^{c}_{i,t} + \sum_{j\in N_{CHP}} E^{chp}_{j,t} + \sum_{l\in N_{FC}} \left(E^{fc.e}_{l,t} + \bar{E}^{fc,h}_{l,t}\right) + E^{bes}_{t} + E^{pv}_{t} + E^{grid}_{t} = E^{l}_{t} \tag{42}$$

$$\sum_{k\in N_B} T^{b}_{k,t} + \sum_{j\in N_{CHP}} T^{chp}_{j,t} + \sum_{l\in N_{FC}} T^{fc}_{l,t} + \sum_{m\in N_{BM}} T^{bm}_{m,t} + T^{bt}_{t} = T^{l}_{t} + T^{sv}_{t} \tag{43}$$

where E^{grid}, E^{bes}, and E^{l} indicate the amount of power exchange with the upstream network, the amount of power charged/discharged, and the electrical demand, respectively. T^{l}, T^{bt}, and T^{sv} represent the thermal demand, thermal power charged/discharged, and the slack variable [54], respectively. Detailed information on electrical and thermal storage systems is available in [3].

G. OBJECTIVE FUNCTION

In the dynamic scheduling problem for biomass-based energy hub, the objective is to minimize the system operating cost, as stated in (44).

$$OC = \sum_{t\in N_T} \left\{ \begin{array}{l} \sum_{i\in N_C} C^{c}_{i,t} + \sum_{k\in N_B} C^{b}_{k,t} + \sum_{m\in N_{BM}} C^{bm}_{m,t} \\ + \sum_{j\in N_{CHP}} C^{chp}_{j,t} + \sum_{l\in N_{FC}} C^{fc}_{l,t} + C^{grid}_{t} \end{array} \right\} \tag{44}$$

IV. DEEP DETERMINISTIC POLICY GRADIENT ALGORITHM

This section discusses the main structure of the proposed optimal control strategy using the actor-critic DRL algorithm. Since DDPG provides continuous action spaces and does not require discretization, which is computationally expensive, it allows for more effective exploration of the action space. Therefore, obtaining an optimum control decision requires fewer iterations than other techniques regardless of the type of problem encountered. And more importantly, the multi-carrier energy system's infrastructure is controlled more smoothly than when they are regulated at a discrete level.

DDPG assesses the cost-benefit of selecting a control signal over a given period, $a_t \in \mathcal{A}_t$, while making decisions in each state of the system using Q-value, $Q(s_t, a_t)$. According to the Bellman equation, the Q-value of any action, a_t, at state s_t can be calculated based on the optimal action's Q-value at state s_{t+1} as follows.

$$Q(s_t, a_t) = \mathcal{R}(s_t, a_t) + \gamma \mathbb{E}\left[\max_{a_{t+1}\in\mathcal{A}_{t+1}} Q(s_{t+1}, a_{t+1})\right] \tag{45}$$

where $\gamma \in [0, 1]$ and $\mathcal{R}$ represent the discount factor and reward function, respectively.

The energy hub operator, however, does not have any information about either the most effective action at state s_{t+1} or what its value would be over time. The DDPG method addresses this challenge by implementing two

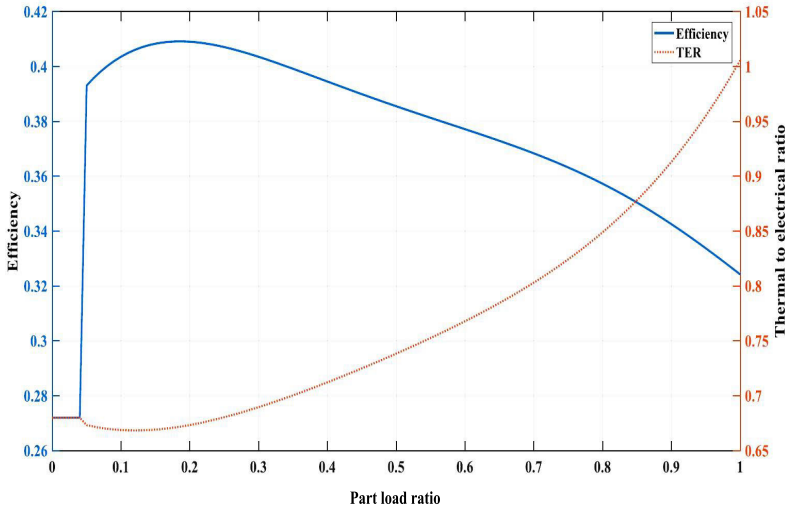

FIGURE 4. An illustration of the FC unit's dynamic performance curve.

separate DNNs. The most beneficial action at state s_{t+1} is estimated by the actor network, $\mu\left(s_t\left|\theta^{\mu}\right.\right)$, while the critic network, $Q\left(s_t, a_t\left|\theta^{Q}\right.\right)$, has the responsibility of estimating the Q-value of action a_t at state s_t. The actor and critic network weight vectors, θ^{μ} and θ^{Q}, are continuously updated based on the observed rewards at each time step during the training process. Using the trained networks, the Q-value can be derived as follows:

$$Q\left(s_t, a_t\right) \approx \mathcal{R}\left(s_t, a_t\right) + \gamma \mathbb{E}\left[Q\left(s_{t+1}, \mu\left(s_{t+1}\left|\theta^{\mu}\right.\right)\right|\theta^{Q}\right)\right] \tag{46}$$

Actor networks respond to sampled states of a system by taking action. Then, the two networks can be simultaneously trained according to the received reward. The s_t and $\max_a Q\left(s_t, a_t\left|\theta^{Q}\right.\right)$ serve as the input and the output for the actor network, while the critic network takes (s_t, a_t), where $a_t = \mu\left(s_t\left|\theta^{\mu}\right.\right)$, and $\mathcal{R}_t$ as the input and the output, respectively. A replay buffer consisting of samples from different time steps, $(s_t, a_t, \mathcal{R}_t, s_{t+1})$, is employed to reduce correlation-induced errors during updating the networks. Copies of the original actor and critic networks, the target networks, are taken and gradually updated to make the training process more stable. Moreover, the actor network is enhanced with a correlated stochastic noise signal to improve exploration during the training process. The most commonly used random noises are Ornstein-Uhlenbeck and Gaussian noises. An Ornstein-Uhlenbeck noise can be used to simulate time-related noise data. The Ornstein-Uhlenbeck process is useful for solving physical control problems that are inertia-driven. Through this perturbation, the DDPG agent is more able to explore and achieve faster convergence as a result. The signal Ω is modeled by employing the Ornstein-Uhlenbeck procedure, as stated in (47).

$$\Omega^{t} = \Omega^{t-1} + \left(\Omega^{mean} - \Omega^{t-1}\right)\kappa T_s + \Omega^{\sigma} n\sqrt{T_s} \tag{47}$$

where Ω_{σ} and Ω_{mean} denote the variance of the noise model and the mean value, respectively. κ and $n \in [0, 1]$ represent the mean attraction constant and a uniformly generated random number, respectively. The critic network is trained using $\mathcal{L}\left(\theta^{Q}\right)$ as the loss function.

$$\mathcal{L}\left(\theta^{Q}\right) = \frac{1}{N}\sum_{i\in N}\left[Q\left(s_i, \mu\left(s_i\left|\theta^{\mu}\right.\right)\right|\theta^{Q}\right) - \mathcal{Y}_i\right]^2 \tag{48}$$

where N is the size of the mini-batch, and $\mathcal{Y}_i$ represents the measured value function as follows.

$$\mathcal{Y}_i = \mathcal{R}\left(s_t, a_t\right) + Q\left(s_{t+1}, \mu\left(s_{t+1}\left|\theta^{\mu'}\right.\right)\right|\theta^{Q'}\right) \tag{49}$$

Meanwhile, the actor network needs to be trained by (50).

$$\nabla_{\theta^{\mu}}\mathcal{L}\left(\theta^{\mu}\right) = \frac{1}{N}\sum_{i\in N}\nabla_a Q\left(s_i, \mu\left(s_i\left|\theta^{\mu}\right.\right)\right|\theta^{Q}\right)\nabla_{\theta^{\mu}}\mu\left(s_i\left|\theta^{\mu}\right.\right) \tag{50}$$

The final step in the process involves gradually updating the target networks with the smoothing factor τ, as stated in (51) and (52).

$$\theta^{\mu'} = (1-\tau)\,\theta^{\mu'} + \tau\theta^{\mu} \tag{51}$$

$$\theta^{Q'} = (1-\tau)\,\theta^{Q'} + \tau\theta^{Q} \tag{52}$$

In the dynamic dispatch problem of the multi-carrier energy system, the state and action vectors can be defined by (53) and (54), respectively.

$$s_t = \left\{E_t^{pv}, \hat{E}_t^{pv}, \pi_t^{e}, \pi_t^{ng}, SoC_t^{bes}, SoC_t^{bt}, E_t^{l}, T_t^{l}\right\} \tag{53}$$

$$a_t = \left\{E_{i,t}^{c}, E_{j,t}^{chp}, T_{k,t}^{b}, T_{m,t}^{bm}, E_{l,t}^{fc,e}, E_{l,t}^{fc,h}, E_t^{bes}, T_t^{bt}\right\} \tag{54}$$

$\Im\left(s_t, a_t, \omega_t\right)$ governs the transition from state s_t to state s_{t+1}. It's not just the decision-making signal a_t that determines transitions, but also environmental uncertainty, ω_t, that influences them. There are several exogenous factors,

including energy consumption habits and solar irradiance, that influence the alignment of probabilistic models within this context. By learning the transition implicitly from collected data samples, a DRL approach can overcome this challenge without the need for statistical models.

A DDPG reward is calculated by the agent by adjusting the parameters as follows:

$$\mathcal{R}(s_t, a_t) = -(OC_t + C_t^{Penalty}) \tag{55}$$

Training with DNN makes DRL-based techniques non-constrained, ignoring the physical limitations of dynamic scheduling. Therefore, the physical constraints cannot be addressed systemically through this process. Nonetheless, energy systems scheduling need to address security and reliability concerns. For a variable $\aleph \in [\aleph^{\min}, \aleph^{\max}]$ a penalty term is added to the reward function to address the constraint violation problem,, as stated in (56) and (57).

$$C^{Penalty} = \sum_{\aleph_i \in N_{\aleph i}} \Im_i . Penalty_{\aleph_i} \tag{56}$$

$$Penalty_{\aleph} = \ln\left(\frac{|\aleph - \aleph^{\min}| + |\aleph - \aleph^{\max}|}{2.\left(\aleph^{\max} - \aleph^{\min}\right)}\right) \tag{57}$$

V. NUMERICAL STUDY

A. DATA AND CONFIGURATION

The biomass-based energy hub shown in Fig. 5 is used as a case study to evaluate the effectiveness of the proposed SFNAS-DDPG dynamic scheduling approach. The proposed SFNAS-DDPG strategy involves scheduling and operating a multi-source multi-product facility that participates in the energy market as a coupled multi-carrier energy supply. The energy loads and prices are adapted from [55] and [56], respectively. Based on a test case presented in [57], the economic data and operational constraints for the generation plants are derived and scaled. The maximum capacities of heat-only, PV, and power-only units are 2.695 MWth, 0.75 MW, and 1.25 MW, respectively. Moreover, the data of the 0.6 MWth biomass-based boiler and the 1 MW FC units are taken from [12] and [53], respectively. Other data corresponding to the available units are listed in Tables 1 and 2. Both the sky images and numerical time series data are obtained from the Solar Radiation Research Laboratory (SRRL) dataset of the National Renewable Energy Laboratory (NREL), located in Colorado [58]. Based on the FL settings, the SFNAS approach is used directly to search for the optimal architecture of the forecasting network using supervisory training loss reshaping. As a representation network, the output of the CNN-GRU model is concatenated with the states of the energy hub before being fed into an actor network. For the critic and actor networks, learning rates of 0.001 and 0.0001 are selected, respectively. The mini-batch of random experience and the discount factor have been selected as 0.95 and 64, respectively. The parameters of the noise model are taken from [26]. Algorithm 1 demonstrates the training process of making control decisions using the proposed SFNAS-DDPG methodology. The training process of the proposed technique is performed by Keras library and TensorFlow as the backend on a workstation with an NVIDIA GeForce GTX 1070 GPU and 32 GB of RAM.

TABLE 1. The characteristics of FC unit.

Characteristics	Value	Characteristics	Value
$E_{max}^{fc,T}$ (MW)	1.0	η_{ch}^{T2T}	0.95
π_t^{ng} (\$/MWh)	40	$C^{fc,p}$ (\$/MWh)	10
$T2_{max}^{in}$ (MW/h)	0.5	$T2T_{max}^{fc}$ (MWh)	2.0
$T2_{max}^{out}$ (MW/h)	0.5	$C^{fc,OM}$ (\$/MWh)	10

TABLE 2. The characteristics of energy storage units.

Characteristics	Value	Characteristics	Value
E^{bes}	0.8	T^{bt}	4
η^{bes}	0.9	η^{bt}	0.6
$E^{bes,ch}$	0.4	$T^{bt,ch}$	1
$E^{bes,dch}$	0.4	$T^{bt,dch}$	1

B. COMPARATIVE RESULTS AND ANALYSIS

In this subsection, the SFNAS algorithm is employed to identify the most accurate architectures for the solar irradiance forecasting task. The CNN and LSTM/BLSTM/GRU networks' hyperparameters significantly affect the model's performance in a complicated manner. However, trying to model the complex interactions between the hyperparameters or manually examining a large number of possible architectures is challenging and time-consuming. To overcome these limitations, the SFNAS algorithm is proposed and implemented to efficiently select the networks' architectures. Algorithm 2 shows the procedure of the proposed SFNAS methodology. The goal is to find optimized models that are capable of better solar irradiance forecasting given the baseline structure of the model discussed in the Section II. For this purpose, we seek to optimize the following parameters of the representation network, shown in Fig. 5:

- Number of convolutional layers in the CNN network (N^C)
- Number of hidden units in the first LSTM/BLSTM/GRU layer (N^{B1})
- Number of hidden units in the second LSTM/BLSTM/GRU layer (N^{B2})
- Number of hidden units in the third LSTM/BLSTM/GRU layer (N^{B3})

Based on our previous work [5], we find that N^{B1} and N^{B2} with fewer than 100 hidden units are likely to exhibit underfitting, while overfitting occurs when the number is greater than 400. To reduce the number of genotype representations from the range 100 to 400, we divide the range by 25 before exploring the two LSTM/BLSTM/GRU layers.

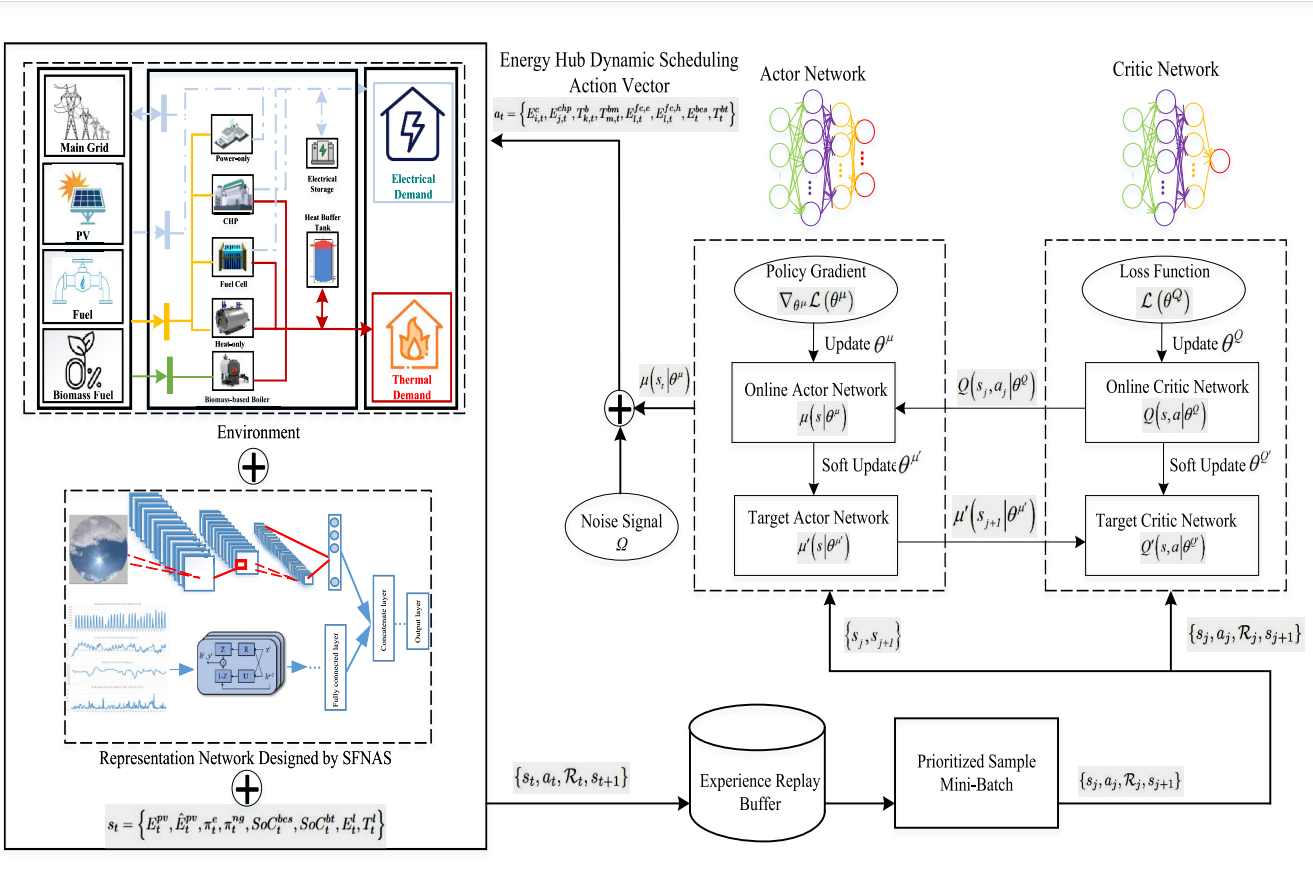

FIGURE 5. The schematic of the proposed SFNAS-DDPG framework for the dynamic scheduling of biomass-based energy hub.

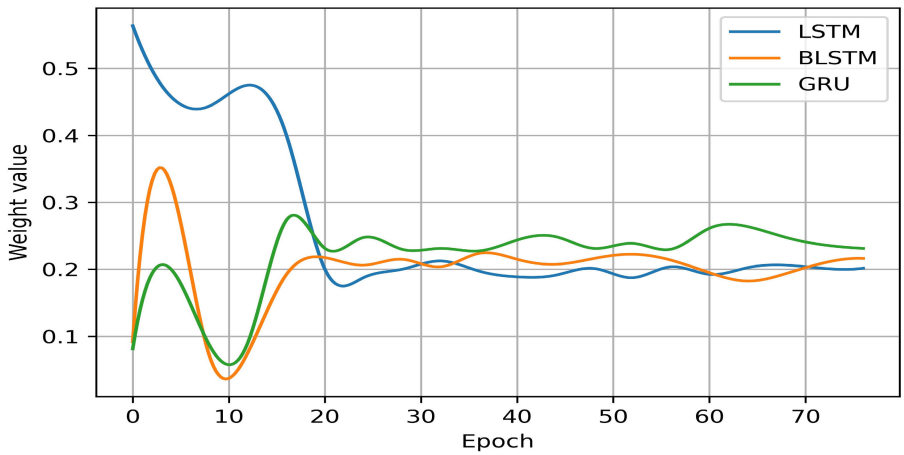

FIGURE 6. The evolution of weight (λ_i) for the gradient term in the momentum update.

In order to translate genotype to phenotype, a multiplier, 25, is assigned to the genotype. According to our previous knowledge, bounds are also set for the other parameters. Our experiments were conducted using the parameters listed in Table 3 as the lower and upper bounds.

From the viewpoint of the EA-based NAS, a genotype then takes the form of a sequence of integers based on the values indicated in Table 3 and phenotypes are the resulting structures. In the proposed stacking model, hierarchical features can be derived from the feature representation of

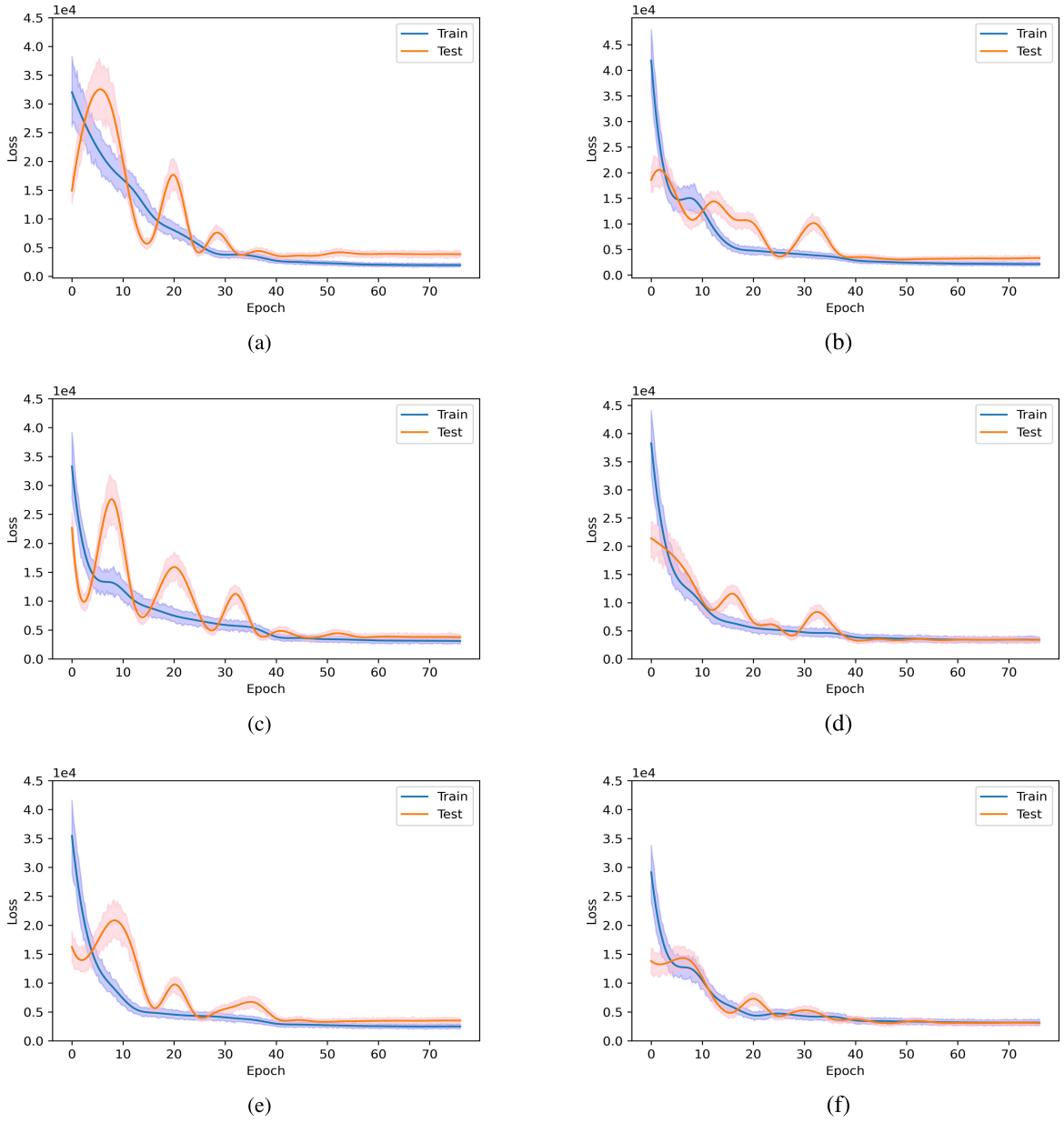

FIGURE 7. The evolution of training and validation losses of the proposed CNN-LSTM/BLSTM/GRU models (a) baseline NAS-based LSTM, (b) SFNAS-based LSTM, (c) baseline NAS-based BLSTM, (d) SFNAS-based BLSTM, (e) baseline NAS-based GRU, and (f) SFNAS-based GRU.

the previous layer and the next layer serves as a layer of abstraction. This hierarchical feature abstraction cannot be achieved when the next layer (i.e., N^{B2}) is greater than the previous layer (i.e., N^{B1}) and results in overfitting. As a result, to ensure that the N^{B2} is equal to or smaller than N^{B1} when converting genotypes to phenotypes, a correction mechanism is applied on N^{B2} (similar correction mechanism is also considered for (N^{B3})) as follows:

$$N^{B_i} = \begin{cases} N^{B_i} & N^{B_i} \leq N^{B_{i-1}} \\ N^{B_{i-1}} & N^{B_i} \geq N^{B_{i-1}} \end{cases} \tag{58}$$

The population is initialized by randomly selecting $N^p - 1$ individuals. The so-called super-fit mechanism is employed to initialize the remaining population based on the parameters of the structure proposed in [5] to begin the evolutionary process with a sufficiently effective individual. An individual's fitness is determined by evaluating the created phenotype for the given genotype based on the root means square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) [5] Both crossover and mutation are used to achieve a reasonable balance between exploration and exploitation. A probability of $P^c = P^m = 0.5$ is applied to each operator independently. Based on

Algorithm 1 Offline Training Process of the DDPG Agent

Initialize: Initialize the weights of the actor and critic networks randomly:
$\mu(s|\theta^{\mu})$: $\theta^{\mu} = \theta^{\mu_0}$
$Q(s, a|\theta^{Q})$: $\theta^{Q} = \theta^{Q_0}$
Initialize Ω as a random process allowing exploration of actions and set up the experience replay buffer M.
for $i \in N^{episodes}$ **do**
  Observe s_t as the state space.
  Concatenate the output of the CNN-GRU network with the state space s_t.
  for $t \in N^{time\ intervals}$ **do**
    Choose the dynamic scheduling action $a_t = \mu(s_t|\theta^{\mu}) + \Omega_t$.
    Solve (44) by taking energy hub control decision a_t and receive R_t as the immediate reward.
    Transfer to s_{t+1} as the new state.
    Store the transition $(s_t,\ a_t,\ R_t,\ s_{t+1})$ into M.
    $k \longleftarrow k + 1$;
    if *batch_size* $\leq |M|$ **then**
      Randomly select $\mathfrak{F} = \{(s_j, a_j, R_j, s_{j+1})\}_{j=1}^{\neq \mathfrak{F}}$ as a mini-batch from M.
      Minimize the loss function (48) to update the critic network.:
      $\theta^{Q} \longleftarrow \eta^{Q} \nabla_{\theta Q} \mathcal{L}(\theta^{Q}) + \theta^{Q}$
      Use (50) and sampled policy gradient to update the actor network:
      $\theta^{\mu} \longleftarrow \eta^{\mu} \nabla_{\theta^{\mu}} \mathcal{L}(\theta^{\mu}) + \theta^{\mu}$
      Use (51) and (52) to Update the target networks:
      $\theta^{Q} \leftarrow \theta^{Q'}$
      $\theta^{\mu} \leftarrow \theta^{\mu'}$
    end if
  end for
end for

the one-point crossover methodology, individuals are firstly ranked according to their fitness, and then the $2n^{th}$ and $(2n+1)^{th}$ ones are selected for crossover ($n \in \left[0, \frac{N^p}{2} - 1\right]$). The population is then uniformly mutated, consisting of individuals generated by crossover and offspring who were not subjected to crossover. Mutation can occur between each architecture parameter (gene) and a value uniformly chosen from its range according to a probability P^{gm}. As a result of this approach, we can conduct searches relatively faster while minimizing disruptive mutations in individuals. The fitness of the parents for each offspring is checked when creating the population of the next generation. Then, to monotonically decrease the population's mean fitness, offspring with better fitness than one of the parents replaces the worst parent.

Algorithm 2 SFNAS Algorithm

for *epoch* $\in$ {*number of epochs*} **do**
  for *agent* $\in$ {*number of agents*} **do**
    Forming the best architecture Z_i^* via agent's NAS
    Compute training loss architecture Z_i^*:
    $J_i(\theta_i, \omega_i) := \|y - \hat{y}_n\|^2 + L_2(\theta_i)$
  end for
  Compute relative contribution of agent:
  $r_i := J_i(\theta_i, \omega_i) / \sum_i J_i(\theta_i, \omega_i)$
  Train the classifier using the relative contribution, r_i, to estimate the weights λ_i
end for
Select the based agent based on:
$\min_A [\omega_1 RMSE + \omega_2 T + \omega_3 N]$

TABLE 3. Hyperparameter bounds for SFNAS algorithm.

Parameter	Lower bound	Upper bound
CNN network (N^C)	1	4
First LSTM/BLSTM/GRU layer (N^{B1})	4	16
Second LSTM/BLSTM/GRU layer (N^{B2})	4	16
Third LSTM/BLSTM/GRU layer (N^{B3})	4	12

Table 4 shows the training time and test RMSE for the best architecture found by each method. According to the results, the proposed SF-NAS technique outperforms the baseline NAS approach in all architectures. Fig 6 illustrates the evolution of the weights (λ) for the gradient term in the momentum update of the solar irradiance forecasting task. As can be seen from the evolution-based figures, the supervisor of the SFNAS provided the agents with guidance that ultimately led to a better optimization process.

Moreover, the evolution of training and validation losses of the proposed CNN-LSTM/BLSTM/GRU networks is depicted in Fig. 7 for both baseline and SFNAS strategies. It can be seen from this figure that networks built with SFNAS exhibit smaller spikes and smoother convergence than those based on baseline NAS.

Table 5 illustrates the average daily operational cost of the biomass-based energy hub using DDPG + state-of-the-art forecasting [6], DDPG + baseline NAS [59], and the proposed SFNAS-DDPG models. As can be seen from this table, the proposed framework, SFNAS-DDPG, has the lowest average daily operating cost when compared with other models. The SFNAS-DDPG method has $5,020 per day average operating costs, which shows a 7.31% improvement over the baseline model (DDPG + State-of-the-art). This improvement is due to the fact that the operator has access

TABLE 4. Forecasting task results for the best architectures found by SFNAS and baseline NAS.

Network's structure	Test RMSE		Average offline training time	
	Baseline NAS [59]	**Proposed SFNAS**	Baseline NAS [59]	**Proposed SFNAS**
CNN-LSTM	66.23	**63.37**	4h 52min	**4h 19min**
CNN-BLSTM	62.88	**57.28**	5h 47min	**5h 11min**
CNN-GRU	62.35	**56.51**	4h 43min	**4h 16min**

TABLE 5. The average daily operating cost for the proposed multi-carrier energy system.

Scheduling Method	Optimization Type	Average Cost ($/day)	Average Computing Time	Profit Change (%)
RO [60]	model-based	7,187	1min 32s	+32.7
SP [61]	model-based	7,001	3min 08s	+29.3
Stochastic/IGDT [3]	model-based	6,516	4min 21s	+20.3
DQN+2D-CNN-BLSTM [5]	model-free	6,081	0.003s	+12.3
DDPG + State-of-the-art model [6]	model-free	5,416	0.008s	0%
DDPG + Baseline NAS [59]	model-free	5,156	0.005s	-4.8%
Proposed SFNAS-DDPG	model-free	**5,020**	**0.005s**	**-7.3%**

to more accurate information about the uncertain parameter (PV generation in future hours), enabling the agent to take more efficient actions.

VI. CONCLUSION

This study presents SFNAS-DDPG as an improved actor-critic DRL framework for the optimal scheduling and operation of a multi-source multi-product facility, taking into account nonlinear characteristics of the facility components such as biogas production thermodynamics, dynamic efficiency of FCs, valve admission behavior of power generation systems, and CHPs' non-convex operation regions. To help the DDPG agent choose the most efficient control policy, especially when PV power is highly intermittent on cloudy days, a hybrid CNN-GRU forecasting model that captures high levels of abstraction from sky images and numerical measurements is developed and employed. The novel SFNAS technique is proposed to learn the architecture and model parameters of the representation network. By utilizing supervisor-led training loss reshaping, the proposed SFNAS approach can be used directly to determine the optimal architecture based on the federated settings. As a result of our case studies, the proposed strategy, by selecting different network structures, is more accurate than the respective baseline forecasting methods. Consequently, The SFNAS-DDPG method has $4,590 per day average operating costs, which shows a 4.79% improvement over the baseline model (DRL+State-of-the-art). While this study was geared toward exploring the effects of the proposed SFNAS methodology on improving actor-critic DRL framework's control strategies, future work could study the effectiveness of replacing the proposed EA-based NAS with heterogeneous NAS methods.

REFERENCES

[1] A. Dolatabadi, B. Mohammadi-Ivatloo, M. Abapour, and S. Tohidi, "Optimal stochastic design of wind integrated energy hub," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2379–2388, Oct. 2017, doi: 10.1109/TII.2017.2664101.

[2] Y. Cheng, N. Zhang, Z. Lu, and C. Kang, "Planning multiple energy systems toward low-carbon society: A decentralized approach," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4859–4869, Sep. 2019, doi: 10.1109/TSG.2018.2870323.

[3] A. Dolatabadi, M. Jadidbonab, and B. Mohammadi-Ivatloo, "Short-term scheduling strategy for wind-based energy hub: A hybrid stochastic/IGDT approach," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 438–448, Jan. 2019, doi: 10.1109/TSTE.2017.2788086.

[4] B. Zhou, D. Xu, C. Li, C. Y. Chung, Y. Cao, K. W. Chan, and Q. Wu, "Optimal scheduling of biogas–solar–wind renewable portfolio for multicarrier energy supplies," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6229–6239, Nov. 2018, doi: 10.1109/TPWRS.2018.2833496.

[5] A. Dolatabadi, H. Abdeltawab, and Y. A. I. Mohamed, "Deep reinforcement learning-based self-scheduling strategy for a CAES-PV system using accurate sky images-based forecasting," *IEEE Trans. Power Syst.*, vol. 38, no. 2, pp. 1608–1618, Mar. 2023, doi: 10.1109/TPWRS.2022.3177704.

[6] A. Dolatabadi, H. Abdeltawab, and Y. A. I. Mohamed, "A novel model-free deep reinforcement learning framework for energy management of a PV integrated energy hub," *IEEE Trans. Power Syst.*, vol. 38, no. 5, pp. 1–13, Oct. 2022, doi: 10.1109/TPWRS.2022.3212938.

[7] P. Zhao, C. Gu, Z. Cao, Z. Hu, X. Zhang, X. Chen, I. Hernando-Gil, and Y. Ding, "Economic-effective multi-energy management considering voltage regulation networked with energy hubs," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2503–2515, May 2021, doi: 10.1109/TPWRS.2020.3025861.

[8] C. Li, H. Yang, M. Shahidehpour, Z. Xu, B. Zhou, Y. Cao, and L. Zeng, "Optimal planning of islanded integrated energy system with solar-biogas energy supply," *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2437–2448, Oct. 2020, doi: 10.1109/TSTE.2019.2958562.

[9] H. Yang, C. Li, M. Shahidehpour, C. Zhang, B. Zhou, Q. Wu, and L. Zhou, "Multistage expansion planning of integrated biogas and electric power delivery system considering the regional availability of biomass," *IEEE Trans. Sustain. Energy*, vol. 12, no. 2, pp. 920–930, Apr. 2021, doi: 10.1109/TSTE.2020.3025831.

[10] Y. Wang, S. Lou, Y. Wu, M. Miao, and S. Wang, "Operation strategy of a hybrid solar and biomass power plant in the electricity markets," *Electr. Power Syst. Res.*, vol. 167, pp. 183–191, Feb. 2019, doi: 10.1016/j.epsr.2018.10.035.

[11] D. Xu, B. Zhou, K. W. Chan, C. Li, Q. Wu, B. Chen, and S. Xia, "Distributed multienergy coordination of multimicrogrids with biogas-solar-wind renewables," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3254–3266, Jun. 2019, doi: 10.1109/TII.2018.2877143.

[12] H. Khaloie, F. Vallée, C. S. Lai, J.-F. Toubeau, and N. D. Hatziargyriou, "Day-ahead and intraday dispatch of an integrated biomass-concentrated solar system: A multi-objective risk-controlling approach," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 701–714, Jan. 2022, doi: 10.1109/TPWRS.2021.3096815.

[13] H. Khaloie, J.-F. Toubeau, F. Vallée, C. S. Lai, and L. L. Lai, "An innovative coalitional trading model for a biomass power plant paired with green energy resources," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 892–904, Apr. 2022, doi: 10.1109/TSTE.2021.3138777.

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

[15] A. Ghadertootoonchi, M. Moeini-Aghtaie, and M. Davoudi, "A hybrid linear programming-reinforcement learning method for optimal energy hub management," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 157–166, Jan. 2023, doi: 10.1109/TSG.2022.3197458.

[16] S. Zhou, Z. Hu, W. Gu, M. Jiang, and X.-P. Zhang, "Artificial intelligence based smart energy community management: A reinforcement learning approach," *CSEE J. Power Energy Syst.*, vol. 5, no. 1, pp. 1–10, Mar. 2019, doi: 10.17775/CSEEJPES.2018.00840.

[17] V.-H. Bui, A. Hussain, and H.-M. Kim, "Double deep *Q*-learning-based distributed operation of battery energy storage system considering uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 457–469, Jan. 2020, doi: 10.1109/TSG.2019.2924025.

[18] J. Wang, C. Jiang, K. Zhang, X. Hou, Y. Ren, and Y. Qian, "Distributed Q-learning aided heterogeneous network association for energy-efficient IIoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2756–2764, Apr. 2020, doi: 10.1109/TII.2019.2954334.

[19] Y. Wang, H. Tan, Y. Wu, and J. Peng, "Hybrid electric vehicle energy management with computer vision and deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 3857–3868, Jun. 2021, doi: 10.1109/TII.2020.3015748.

[20] J. Duan, H. Xu, and W. Liu, "Q-learning-based damping control of wide-area power systems under cyber uncertainties," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6408–6418, Nov. 2018, doi: 10.1109/TSG.2017.2711599.

[21] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg, "On-line building energy optimization using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3698–3708, Jul. 2019, doi: 10.1109/TSG.2018.2834219.

[22] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, and K. Li, "Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4513–4521, Sep. 2020, doi: 10.1109/TSG.2020.2986333.

[23] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time EV charging scheduling based on deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5246–5257, Sep. 2019, doi: 10.1109/TSG.2018.2879572.

[24] Y. Liang, C. Guo, Z. Ding, and H. Hua, "Agent-based modeling in electricity market using deep deterministic policy gradient algorithm," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4180–4192, Nov. 2020, doi: 10.1109/TPWRS.2020.2999536.

[25] X. Sun and J. Qiu, "Two-stage volt/var control in active distribution networks with multi-agent deep reinforcement learning method," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 2903–2912, Jul. 2021, doi: 10.1109/TSG.2021.3052998.

[26] F. Sanchez Gorostiza and F. M. Gonzalez-Longatt, "Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5039–5050, Nov. 2020, doi: 10.1109/TSG.2020.2996274.

[27] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 1997–2017, 2019.

[28] M. Wistuba, A. Rawat, and T. Pedapati, "A survey on neural architecture search," 2019, *arXiv:1905.01392*.

[29] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, X. Chen, and X. Wang, "A comprehensive survey of neural architecture search: Challenges and solutions," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–34, May 2022.

[30] Y. Liu, Y. Sun, B. Xue, M. Zhang, G. G. Yen, and K. C. Tan, "A survey on evolutionary neural architecture search," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 2, pp. 550–570, Feb. 2023, doi: 10.1109/TNNLS.2021.3100554.

[31] K. Swersky, D. Duvenaud, J. Snoek, F. Hutter, and M. A. Osborne, "Raiders of the lost architecture: Kernels for Bayesian optimization in conditional parameter spaces," 2014, *arXiv:1409.4011*.

[32] A. Zela, A. Klein, S. Falkner, and F. Hutter, "Towards automated deep learning: Efficient joint neural architecture and hyperparameter search," 2018, *arXiv:1807.06906*.

[33] B. Baker, O. Gupta, N. Naik, and R. Raskar, "Designing neural network architectures using reinforcement learning," 2016, *arXiv:1611.02167*.

[34] Z. Zhong, J. Yan, W. Wu, J. Shao, and C.-L. Liu, "Practical block-wise neural network architecture generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2423–2432, doi: 10.1109/CVPR.2018.00257.

[35] B. Baker, O. Gupta, R. Raskar, and N. Naik, "Accelerating neural architecture search using performance prediction," 2017, *arXiv:1705.10823*.

[36] H. Pham, M. Guan, B. Zoph, Q. Le, and J. Dean, "Efficient neural architecture search via parameters sharing," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4095–4104.

[37] A. Brock, T. Lim, J. M. Ritchie, and N. Weston, "SMASH: One-shot model architecture search through hypernetworks," 2017, *arXiv:1708.05344*.

[38] G. Bender, P.-J. Kindermans, B. Zoph, V. Vasudevan, and Q. Le, "Understanding and simplifying one-shot architecture search," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 550–559.

[39] H. Cai, L. Zhu, and S. Han, "Proxylessnas: Direct neural architecture search on target task and hardware," 2018, *arXiv:1812.00332*.

[40] H. Liu, K. Simonyan, and Y. Yang, "Darts: Differentiable architecture search," 2018, *arXiv:1806.09055*.

[41] G. Li, G. Qian, I. C. Delgadillo, M. Müller, A. Thabet, and B. Ghanem, "SGAS: Sequential greedy architecture search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1617–1627.

[42] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin, "Large-scale evolution of image classifiers," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2902–2911.

[43] M. Wistuba, "Deep learning architecture search by neuro-cell-based evolution with function-preserving mutations," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2018, pp. 243–258.

[44] D. Połap, "An adaptive genetic algorithm as a supporting mechanism for microscopy image analysis in a cascade of convolution neural networks," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106824.

[45] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*.

[46] A. Dolatabadi, H. Abdeltawab, and Y. A. I. Mohamed, "Hybrid deep learning-based model for wind speed forecasting based on DWPT and bidirectional LSTM network," *IEEE Access*, vol. 8, pp. 229219–229232, 2020.

[47] A. Dolatabadi, H. Abdeltawab, and Y. A. I. Mohamed, "Deep spatial–temporal 2-D CNN-BLSTM model for ultrashort-term LiDAR-assisted wind Turbine's power and fatigue load forecasting," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2342–2353, Apr. 2022, doi: 10.1109/TII.2021.3097716.

[48] H. Li, Z. Ren, Y. Xu, W. Li, and B. Hu, "A multi-data driven hybrid learning method for weekly photovoltaic power scenario forecast," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 91–100, Jan. 2022, doi: 10.1109/TSTE.2021.3104656.

[49] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[50] H. Jahangir, H. Tayarani, S. S. Gougheri, M. A. Golkar, A. Ahmadian, and A. Elkamel, "Deep learning-based forecasting approach in smart grids with microclustering and bidirectional LSTM network," *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8298–8309, Sep. 2021.

[51] M. Moradi-Dalvand, M. Nazari-Heris, B. Mohammadi-Ivatloo, S. Galavani, and A. Rabiee, "A two-stage mathematical programming approach for the solution of combined heat and power economic dispatch," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2873–2881, Jun. 2020, doi: 10.1109/JSYST.2019.2958179.

[52] M. Alipour, K. Zare, H. Zareipour, and H. Seyedi, "Hedging strategies for heat and electricity consumers in the presence of real-time demand response programs," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1262–1270, Jul. 2019, doi: 10.1109/TSTE.2018.2865380.

[53] M. Y. El-Sharkh, M. Tanrioven, A. Rahman, and M. S. Alam, "Economics of hydrogen production and utilization strategies for the optimal operation of a grid-parallel PEM fuel cell power plant," *Int. J. Hydrogen Energy*, vol. 35, no. 16, pp. 8804–8814, Aug. 2010, doi: 10.1016/j.ijhydene.2010.05.108.

[54] J. Cao, C. Crozier, M. McCulloch, and Z. Fan, "Optimal design and operation of a low carbon community based multi-energy systems considering EV integration," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1217–1226, Jul. 2019, doi: 10.1109/TSTE.2018.2864123.
[55] S. Taheri, M. Jooshaki, and M. Moeini-Aghtaie, "Long-term planning of integrated local energy systems using deep learning algorithms," *Int. J. Electr. Power Energy Syst.*, vol. 129, Jul. 2021, Art. no. 106855, doi: 10.1016/j.ijepes.2021.106855.
[56] D. Qiu, Z. Dong, X. Zhang, Y. Wang, and G. Strbac, "Safe reinforcement learning for real-time automatic control in a smart energy-hub," *Appl. Energy*, vol. 309, Mar. 2022, Art. no. 118403, doi: 10.1016/j.apenergy.2021.118403.
[57] M. Basu, "Combined heat and power economic dispatch by using differential evolution," *Electr. Power Compon. Syst.*, vol. 38, no. 8, pp. 996–1004, May 2010, doi: 10.1080/15325000903571574.
[58] T. Stoffel and A. Andreas, "NREL solar radiation research laboratory (SRRL): Baseline measurement system (BMS); golden, Colorado (data)," National Renewable Energy Lab. (NREL), Golden, CO, USA, Tech. Rep. DA-5500-56488, 1981.
[59] H. Mo, L. L. Custode, and G. Iacca, "Evolutionary neural architecture search for remaining useful life prediction," *Appl. Soft Comput.*, vol. 108, Sep. 2021, Art. no. 107474.
[60] P. Zhao, C. Gu, D. Huo, Y. Shen, and I. Hernando-Gil, "Two-stage distributionally robust optimization for energy hub systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3460–3469, May 2020, doi: 10.1109/TII.2019.2938444.
[61] A. Dolatabadi and B. Mohammadi-Ivatloo, "Stochastic risk-constrained optimal sizing for hybrid power system of merchant marine vessels," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5509–5517, Dec. 2018, doi: 10.1109/TII.2018.2824811.

AMIRHOSSEIN DOLATABADI (Graduate Student Member, IEEE) was born in Tabriz, Iran, in 1991. He received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tabriz, Tabriz, in 2014 and 2016, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2023.

From 2016 to 2018, he was a Research Associate with the Smart Energy Systems Laboratory (SES Lab), University of Tabriz. He was also a Research Intern with Hitachi Energy Canada Inc. His research interests include artificial intelligence, machine learning, deep learning, decision-making methods, and their applications to smart girds.

HUSSEIN ABDELTAWAB (Senior Member, IEEE) was born in Bani-Souwaif, Egypt, in April 1987. He received the B.Sc. (Hons.) and M.Sc. degrees in electrical engineering from Cairo University, in 2009 and 2012, respectively, and the Ph.D. degree in electrical engineering from the University of Alberta, Edmonton, AB, Canada, in 2017.

He is currently an Assistant Professor of engineering with Wake Forest University, Winston-Salem, NC, USA. He is also a Licensed Professional Engineer in Saskatchewan, Canada. His research interests include energy management, control system applications in renewable energy, energy storage, and smart distribution systems.

YASSER ABDEL-RADY I. MOHAMED (Fellow, IEEE) was born in Cairo, Egypt, in November 1977. He received the B.Sc. (Hons.) and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, in 2000 and 2004, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2008.

He is a Professor with the Department of Electrical and Computer Engineering, University of Alberta, AB, Canada. His highly cited research focuses on modeling, analysis, stability, control, and optimization of power electronic converters and systems, active distribution systems, and microgrids; grid integration of distributed and renewable energy resources and energy storage; and the development of artificial intelligence technologies for smart grids.

Dr. Mohamed is an Elected Fellow of the Asia-Pacific Artificial Intelligence Association (AAIA). He was an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, IEEE TRANSACTIONS ON SMART GRID, and IEEE POWER ENGINEERING LETTERS, and Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS. He is an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS. He is a registered Professional Engineer in the Province of Alberta, Canada.