Voltage Regulation of DC-DC Buck Converters Feeding CPLs via Deep Reinforcement Learning

Chenggang Cui, *Member, IEEE*, Nan Yan, Baixiang Huangfu, Tianxiao Yang, and Chuanlin Zhang, *Senior Member, IEEE*

***Abstract*—Modeling accuracy of DC-DC converters may deviate largely in the presence of different variation levels of constant power loads (CPLs), hence is well acknowledged as a main hurdle for the design of advanced model-driven control strategies in the literature. Aiming to enhance the bus voltage regulation performance of DC-DC buck converters, a model-free deep reinforcement learning (DRL) control strategy is proposed in this brief. Firstly, a Markov Decision Process (MDP) model and a deep Q network (DQN) algorithm are utilized for the stabilization issue of the converter. Secondly, through a subgoal reward/penalty mechanism, the control objective and prescribed performance of the system are therefore guaranteed. Moreover, a specified action space is designed to match the switch speed of the switching element. As a distinguishable feature, the settling time under the proposed control scheme is significantly reduced in the occurrence of disturbance, resulting from the fast adaption ability of DRL. The simulation comparison results in reference to PI and fuzzy PI controllers demonstrate the efficacy and superiorities under large signal perturbation conditions.**

***Index Terms*—DC-DC converter, constant power load, deep reinforcement learning, large signal stability.**

Nomenclature

α	Learning rate of the loss function
β_1, β_2	Parameters of the positive reward
β_3	Parameter of the penalty reward
ϵ_1, ϵ_2	Subgoals of the expected error
$\frac{de(t)}{dt}$	Time derivative of the tracking error
$\frac{dv_o(t)}{dt}$	Time derivative of the output voltage
γ	Discount factor of MDP
θ	Parameter collection of DNN
θ^-	Parameter collection of the target Q network
ε	Exploration rate
A	Action space of MDP
$a_{t-\mathrm{ran}}$	Random action
a_t	Decision made by the agent
d	Duty ratio of the DC-DC buck converter
e	Tracking error of the output voltage
e_{del}	Delay signal of the tracking error
i_{CPL}	CPL current
$J(\theta)$	Loss function
P	State transition probability matrix of MDP
p	Random value between 0 and 1 about choosing action
P_{CPL}	Power of CPL
Q_π	State-action value of a policy π
R	Current reward of MDP
r_t	Reward/penalty signal
S	Finite state space of MDP
s_{t+1}	Next environment state perceived by the agent
s_t	Environment state perceived by the agent
T	A time-step of the training process
$v_{o-\mathrm{del}}$	Delay signal of the output voltage
V_{ref}	The nominal voltage
y_j	Approximation of a current Q value.

Manuscript received August 19, 2021; accepted August 22, 2021. Date of publication August 25, 2021; date of current version March 15, 2022. This work was supported in part by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning; in part by the National Natural Science Foundation of China under Grant 51607111 and Grant 62173221; and in part by Shanghai Rising-Star Program under Grant 20QA1404000. This brief was recommended by Associate Editor S. Li. *(Corresponding author: Chuanlin Zhang.)*

The authors are with the Intelligent Autonomous Systems Laboratory, College of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China (e-mail: clzhang@shiep.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSII.2021.3107535.

Digital Object Identifier 10.1109/TCSII.2021.3107535

I. Introduction

During the past few decades, the rapid development of power electronics technology has laid a splendid foundation for renewable energy sources (RESs) to be integrated into the grid [1]. Since most distributed power sources are DC in nature, more and more attentions have been concentrated on DC microgrids. It is well acknowledged that CPL shows negative incremental impedance characteristic, which reduces the power supply reliability when cascaded with DC-DC converters [2]. Normally, the conventional PI controller is very difficult to achieve a large signal stability. Thus, advanced control methods for DC-DC converter system feeding CPLs, aiming to enhance the voltage regulation performance, have become a trend in the literature recently, see, e.g., sliding mode control (SMC) [3], NDO based composite control [4] and model predictive control (MPC) [5], etc.

On the other hand, considering the case when the system is affecting large CPL variations, the mathematical model's accuracy may be greatly degraded. Thereafter, the dynamic performance might deteriorate significantly in practices [6]. By acknowledging the fact that a precise modeling procedure for industrial systems could be much difficult or impossible, therefore intelligent control strategies have drawn extensive attention in many fields in recent decades, such as the intelligent control issue for robotics [7]–[9] and power systems [10], [11], etc. Regarding the field of DC-DC converters, a particle swarm optimization (PSO) method provides the optimal fuzzy rules for the controller to generate the duty ratio in [10]. Due to the powerful approximation ability of a fuzzy neural network (FNN), a FNN is designed to imitate the control law of a SMC in [11]. These intelligent controllers refer to the experience of the previous model for supervised learning, in order to optimize the dynamic performance while the system stability is ensured. However, these methods may behave limited learning and adaption capabilities [12]–[14].

Recently, Deep Reinforcement Learning (DRL) has aroused much concern from the scientific society [15]. However, there are very few existing results that can be found in the literature regarding the application of the DRL algorithm into DC-DC converters. Notably, a recent work [16] studies a self adaptive control scheme in which a Proximal Policy Optimization (PPO) based parameter tuner is settled to suppress the destructive caused by CPLs. Later on, a deep deterministic policy gradient (DDPG) algorithm provides a compensation signal for the duty ratio, which can regulate the output voltage of a DC-DC buck-boost converter feeding CPL [17]. However, it can be clearly observed that these DRL control strategies are still partly based on the mathematical model and their effectiveness is still affected by the accuracy of system modeling. DRL algorithm is used to tune parameters of model to regulate controlled object indirectly.

As a pioneer work, this brief considers a total model-free DRL control design issue for a DC-DC converter of buck type, in order to improve the dynamic performance of the system facing a wide range of operating conditions. Firstly, the Markov Decision Process (MDP) is designed while the environment model is established to calculate the reward and state. Secondly, a DQN algorithm is adopted to generate the duty ratio signals, which maintains the bus voltage stability by adjusting the PWM signals. The state-action value function of the action is approximated by a neural network to evaluate the undertaken action. Thirdly, the intelligent agent learns to understand the voltage variation after a period and immediately take action to offset the destabilizing effects imposed by the variation of CPLs. Finally, a simulation is conducted to verify the adaptivity and dynamic performance of the proposed control strategy under various operating conditions.

In reference to existing related results, this brief mainly distinguishes itself in the following three aspects: 1) The proposed controller is able to adaptively regulate the DC bus voltage without any prior knowledge, i.e., the requirement of the system modeling procedure is totally removed. 2) The agent reacts to high frequency switching rapidly, which is derived by the discrete action space with the switching constraint. 3) The proposed controller could accelerate the recovery speed of voltage fluctuation while a large signal stability can be guaranteed.

II. Problem Formulation and Preliminaries

A. Problem Formulation

In this brief, the control objective is to regulate the output voltage at a nominal value for a DC-DC buck converter. As a matter of fact, existing model-based control strategies rely on a small-signal system model or a large-signal model for the converters. However, small-signal model has limited accuracy with the presence of large signal disturbances, and the large-signal system model based controllers may behave certain robustness redundancy in the case of small signal disturbances. In this brief, thanks to the adaptability of the DRL method, even under different working conditions, the voltage recovery rate could be accelerated and an adaptive optimal control result is obtained.

B. DRL Methodology

The DRL framework can be described as the MDP, which gives a tuple as $\{S, A, R, P, \gamma\}$. In each time step, the agent perceives the environment state s_t and makes the decision a_t. The environment is then transformed into a new state s_{t+1} and received a reward/penalty signal r_t that estimates the state transformation. The training objective is to maximize the cumulative reward $E[\sum_{k=0}^{T} \gamma^k r_{t+k}]$ received from interacting with the environment over a long term where T is a time-step at which an episode terminates.

In DRL algorithm, a deep neural network is considered as an approximator to estimate the state-action value function: $Q_\pi(s_t, a_t) \approx Q_\pi(s_t, a_t, \theta)$. DQN is a value-based DRL algorithm, which acts upon the environment according to the maximum Q value. In the DQN framework, the approximation of a current Q value y_j is obtained via a multi-layer neural network. After implementing each decision, the Q value is updated according to the following equation:

$$y_j = \begin{cases} r_j, & \text{if episode terminate at step } j+1; \\ r_j + \gamma \max\limits_a \hat{Q}\big(s_{j+1}, a_{j+1}, \theta^-\big), & \text{otherwise.} \end{cases} \tag{1}$$

The loss function can be expressed as the following relation:

$$J(\theta) = E[(r_{t+1} + \gamma \max_a Q(s_{t+1}, a_{t+1}, \theta^-) - Q(s_t, a_t, \theta))^2]. \tag{2}$$

The network is updated by minimizing the loss function. Generally, the DNN is trained using the gradient descent method through backpropagation [18]. It is worthy of mentioning that in order to disrupt the data correlation, two DNN with the same parameters are initiated in the DQN algorithm. The Q network updates the Q value synchronously, while the target Q network gives the target Q value. In addition, the transitions of each step $\{s_t, a_t, r_t, s_{t+1}\}$ are put into an experience buffer. Then, a mini-batch of experience data is sampled from this buffer to train the DNN, which utilizes data more effectively. To balance the "exploration" and the "exploitation", the following action is adopted:

$$a_t = \begin{cases} \arg\max\limits_a Q(s_t, a_t), & \text{if } p < \varepsilon; \\ a_{t-\text{ran}}, & \text{otherwise.} \end{cases} \tag{3}$$

III. DRL Controller Design

In this section, a DQN algorithm is adopted to adjust the duty ratio by using the superior learning. The undertaken action is applied to create an effective duty ratio in the presence of large disturbances, which is achieved by maximizing the feedback signal r_t. After training, the DQN agent learns how to take action immediately to eliminate the overshoot, resulting from the CPL variation. Moreover, the control performance of the DRL controller largely depends on the designation of state space, action space, reward/penalty function and exploration strategy. The MDP model for the control method of a DC-DC buck converter is defined as follows.

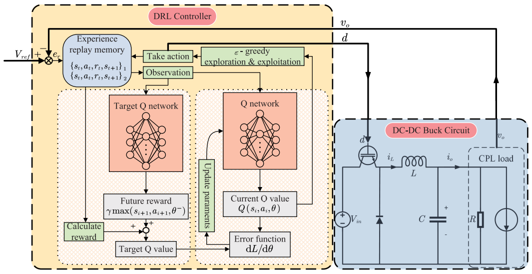


Fig. 1. Control diagram of the converter system under the proposed DRL controller.

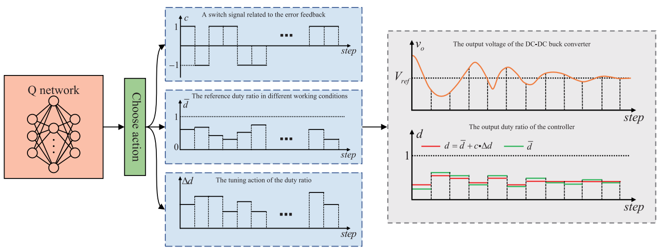


Fig. 2. Action space design in the proposed DRL controller.

A. State Space

Primarily, the output voltage v_o and the tracking error $e(t) = v_o(t) - V_{ref}$ are considered as the underlying signals to determine the system state. Since the inductor and capacitor cannot charge and discharge instantaneously, the delay signal, derivative of output voltage and tracking error are regarded as the supplementary state to characterize the system state more comprehensively. Therefore, the state is defined as:

$$S_t = \left\{ v_o(t), v_{o\_\mathrm{del}}(t), \frac{\mathrm{d}v_o(t)}{\mathrm{d}t}, e(t), e_{\mathrm{del}}(t), \frac{\mathrm{d}e(t)}{\mathrm{d}t} \right\}. \quad (4)$$

B. Action Space

Theoretically, continuous action space provides different values of duty ratio d for various operating conditions, which can enhance the system's dynamic performance. Nevertheless, continuous action space can increase the difficulty of seeking the optimal strategy, while the learning time will be longer. It is necessary to discretize the action space.

For the switching characteristics of switch element, the calculation speed of the traditional DRL algorithm cannot follow the switching speed. Thus, the switch control is chosen as a reference to design a discrete action space, which is described in Fig. 2 concretely.

C. Reward Function with Subgoal

The tracking error $e(t)$ between the current state and control objective is designed as the potential reward function. A sparse reward are given to guide the learning agent by two subgoals ϵ_1 and ϵ_2 of the expected errors. Then, the positive reward β_1 and β_2 are given when the output voltage is close to the nominal voltage. Furthermore, a penalty reward is designed as $\beta_3 e(t)$ according to the subgoals. β_1, β_2 and β_3 are selected according to the specific situations.

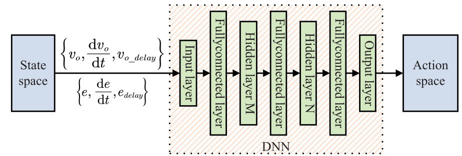


Fig. 3. Internal structure of DNN.

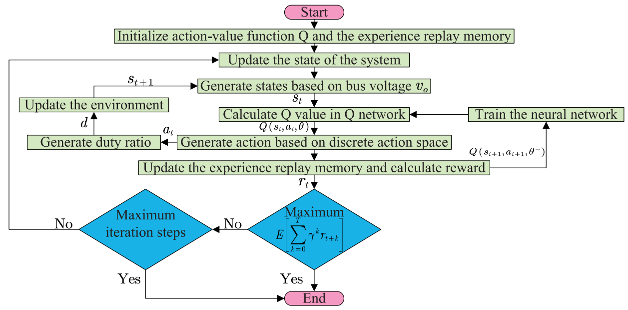


Fig. 4. Flowchart of the DRL methodology.

Therefore, the reward function of the DRL controller gives:

$$r = \begin{cases} \beta_1 - \beta_3 e(t), & \text{if } 0 \le |e(t)| < \epsilon_1; \\ \beta_2 - \beta_3 e(t), & \text{if } \epsilon_1 \le |e(t)| \le \epsilon_2; \\ -\beta_3 e(t), & \text{else.} \end{cases} \quad (5)$$

D. DNN Design

DQN represents the action value function with neural network approximator. The network has 7 layers, including an input layer, three fully-connected layers, two hidden layers and an output layer. The first hidden layer has M neurons and the second hidden layer has N neurons. Relu function is adopted as the activation function of each hidden layer. The fully-connected layers is utilized to combine state space with action space.

Remark 1: The time complexity of the proposed algorithm relies on the size of the DNN input n_l, the number of neurons in the hidden layers, and the dimension of the DNN output m_l [19]. The total time complexity of all DNN layers is $O(\sum_{l=1}^{d} n_l \cdot m_l)$, where l is the index of a layer and d is the depth (number of layers). Thus, the time complexity of the designed DNN is given by $O(D_s \times M + M \times Re + M \times N + N \times Re + N \times D_a)$, where D_s and D_a are the dimension of state space and action space, Re is the dimension of Relu output, respectively.

IV. Simulation Verification

To validate the feasibility and effectiveness of the proposed controller, a DC-DC buck converter model is established in MATLAB/SIMULINK. The classical double-loop PI control method and the fuzzy PI method are selected to compare with the proposed DRL controller.

The parameters of the buck DC-DC converter model are shown in Table I. For a fair comparison, with the optimal parameter configuration method in [6], the PI controller is tuned to have similar dynamic response with the proposed controller. The parameters of the double-loop PI controller and DQN algorithm are shown in Table II. In the fuzzy

TABLE I
CIRCUIT PARAMETERS OF BUCK POWER CONVERTER

Parameter	Definition	Value
V_{in}	Converter input voltage	200V
V_{ref}	Nominal bus voltage	100V
L	Nominal inductance value	2mH
C	Nominal capacitance value	150μF
f	Switching frequency	10kHz

TABLE II
CONTROLLER PARAMETER CONFIGURATION

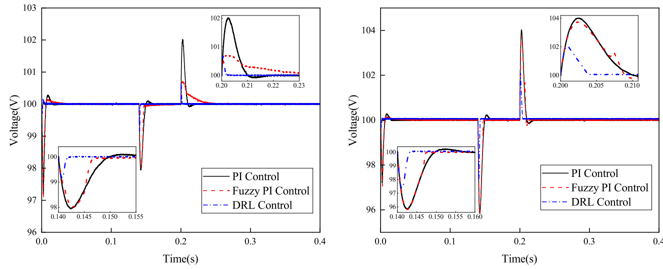
Controller	Parameter	Definition	Value
PI Controller	k_{cp}, k_{ci}	Current Loop Gains	0.02,100
	k_{vp}, k_{vi}	Voltage Loop Gains	1,300
DQN Controller	α	Learning rate	0.001
	γ	Discount factor	0.9
	B	Replay memory capacity	1e-6
	b	Minibatch size	256
	ε	5 greedy exploration	0.1
	u	Duty ratio	0.5
	$\beta_1, \beta_2, \beta_3$	Reward function parameters	10, 1, -10
	ϵ_1, ϵ_2	Subgoals of the reward function	0.1, 1
	Δd	Tuning action of the duty ratio	0, 0.02, 0.04, 0.06
	$\bar{d}$	Reference duty ratio	0.5
	c	Switch signal	± 1
	M	Number of neurons	64
	N	Number of neurons	64

TABLE III
TOEPLITZ TYPE RULE TABLE FOR Δk_p AND Δk_i

e \ ec	NB	NM	NS	ZE	PS	PM	PB
NB	PB NB	PB NB	PM NM	PM NM	PS NS	ZE ZE	ZE ZE
NM	PB NB	PB NB	PM NM	PS NS	PS NS	ZE ZE	NS ZE
NS	PM NB	PM NM	PM NS	PS NS	ZE ZE	NS PS	NS PS
ZE	PM NM	PM NM	PS NS	ZE ZE	NS PS	NS PM	NM PM
PS	PS NM	PS NS	ZE ZE	NS PS	NS PS	NM PM	NM PB
PM	PS ZE	ZE ZE	NS PS	NM PS	NM PM	NM PB	NB PB
PB	ZE ZE	ZE ZE	NM PS	NM PM	NM PM	NB PB	NB PB

PI controller, the adjustments of gain parameters in different segments of the system response curves are displayed in Table III. The two elements in each grid represent Δk_p and Δk_i respectively [20]. To verify the dynamic performance of the proposed DRL controller, the simulations with CPL variation are conducted in the following two cases.

Case I: Initially, a 200W CPL is connected. At 0.14s, the CPL increases from 200W to 500W, and at 0.2s, the CPL steps back to 200W. As can be observed in Fig. 5(a), when the CPL variance occurs, the voltage can be adjusted to reach a stable state within a short time and the voltage deviation is less than 0.2% (about 0.2V) during all simulation time. While the bus voltage change obviously under the control of double loop PI method with the maximum overshoot reaches 2.02V. When the fuzzy PI method is adopted, the settling time of the system is reduced to a certain extent. In the meantime, the settling time of DRL controller is 2ms which is much shorter than the ones of PI and Fuzzy PI controllers. It is shown that the proposed



(a) (b)

Fig. 5. Voltage response with CPL variations: (a) from 200W to 500W, and from 500W to 200W; (b) from 200W to 800W, and from 800W to 200W.

TABLE IV
INFLUENCE OF SUBGOALS ON CONTROL PERFORMANCE

Parameter	Value	Settling time	Overshoot
ϵ_1	0.01	-	-
	0.1	1.79ms	0.75%
	0.5	3.065ms	0.91%
ϵ_2	0.5	2.214ms	0.81%
	1	1.79ms	0.75%
	10	1.689ms	0.76%

TABLE V
INFLUENCE OF POSITIVE/PENALTY REWARD PARAMETERS

Parameter	Value	Settling time	Overshoot
β_1	10	1.79ms	0.75%
	20	3.0ms	1.01%
	50	3.66ms	0.9%
β_2	0.1	-	-
	0.5	1.95ms	0.85%
	2	2.149ms	0.84%
	5	-	-
β_3	1	3.239ms	0.98%
	2	2.255ms	1.12%
	5	3.46ms	0.96%

controller can stabilize the bus voltage within a short time in the presence of the CPL variation.

Case II: The CPL is initialized as 200W. At 0.14s, the CPL increases from 200W to 800W and it steps back to 200W at 0.2s. The comparison results with a double-loop PI controller are presented in Fig. 5(b). It shows that the proposed DRL controller achieves smaller voltage fluctuation and shorter settling time.

A. Analysis on Subgoals of Reward Function

The DRL agent is trained with two subgoals ϵ_1 and ϵ_2, which prescribes the expected tracking error. The control performance for different ϵ is shown in Table IV.

Note that only $\epsilon_1 = 0.01$ causes the training divergent. The other subgoals obtain satisfactory scores. On the other hand, the settling time is less than 4ms and the overshoot is within 1% for the DC-DC buck converter.

B. Positive and Penalty Reward Parameters Analysis

In this section the DRL agent is trained with three reward parameters β_1, β_2 and β_3, which affect the convergence speed and the final performance.

Table V presents the overall training parameters. The settling time is less than 5ms and the overshoot is almost less than 1%. Generally, the larger β_1 and the smaller β_3 are, the worse

TABLE VI
INFLUENCE OF NEURONS ON CONTROL PERFORMANCE

Parameter	Settling time	Overshoot
M=64, N=64	1.79ms	0.75%
M=32, N=32	1.78ms	0.82%
M=16, N=16	2.87ms	0.77%
M=64, N=32	2.62ms	0.83%
M=16, N=64	1.77ms	0.84%

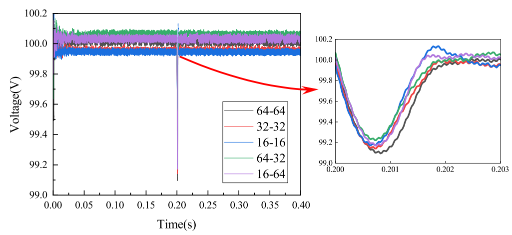


Fig. 6. Voltage response with different number of neurons.

the control performances are after training. β_2 is supposed to limit to a small range to avoid training divergent.

C. Hyper-parameter Sensitivity Analysis

In this part, $M = 64, N = 64$; $M = 32, N = 32$; $M = 16, N = 16$; $M = 32, N = 64$ and $M = 16, N = 64$ are trained. The details of control performance for different M and N are shown in Table VI. Compared with the control performance under adjusted $M = N = 64$, the settling time and overshoot increase slightly as M and N increase or decrease. However, the settling time is less than 3ms and the overshoot is almost within 1%.

After testing different reward function parameters and the number of network layers, it can be concluded that when the reward function parameters and the number of network layers variate in a small range, the control performance of the reinforcement learning controller is rarely affected. Thus, in the case of CPL variations, the DRL controller is not sensitive to the reward function parameters and the number of network layers.

As is well acknowledged, a deep neural network with different neurons can get different training performances. However, in this brief, various neurons bring nearly the same effects to the output voltage. As can be observed from Fig. 6, each controller with different numbers of neurons can track the nominal value with an error less than 0.5V.

V. CONCLUSION

A model-free DRL controller for DC-DC buck converter feeding CPL is proposed in this brief. The simulation results verify that the proposed control strategy shows behaves great adaptivity and dynamic performance under more complete operating conditions, especially when the CPL changes on a large scale. Besides, we have faced a lot of difficulties when trying to transfer algorithm into experiment. The deviation between ideal conditions and real environments, the high sensibility to hyper-parameters leads to the poor transfer result. Future work may focus on designing the innovative and robust DRL controller to solve the problem above. Meanwhile, sim-to-real transferring is also the focus of future research. Later on, one can also try to use other neural networks, such as ANN, CNN and RBFNN methods.

REFERENCES

[1] F. Li *et al.*, "Review of real-time simulation of power electronics," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 4, pp. 796–808, Jul. 2020.

[2] A. Emadi, A. Khaligh, C. H. Rivetta, and G. A. Williamson, "Constant power loads and negative impedance instability in automotive systems: Definition, modeling, stability, and control of power electronic converters and motor drives," *IEEE Trans. Veh. Technol.*, vol. 55, no. 4, pp. 1112–1125, Jul. 2006.

[3] Z. Wang, S. Li, and Q. Li, "Discrete-time fast terminal sliding mode control design for DC–DC buck converters with mismatched disturbances," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1204–1213, Feb. 2020.

[4] C. Zhang, X. Wang, P. Lin, P. X. Liu, Y. Yan, and J. Yang, "Finite-time feedforward decoupling and precise decentralized control for DC microgrids towards large-signal stability," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 391–402, Jan. 2020.

[5] Q. Xu, Y. Yan, C. Zhang, T. Dragicevic, and F. Blaabjerg, "An offset-free composite model predictive control strategy for DC/DC buck converter feeding constant power loads," *IEEE Trans. Power Electron.*, vol. 35, no. 5, pp. 5331–5342, May 2020.

[6] S. Bacha, I. Munteanu, and A. I. Bratcu, "Power electronic converters modeling and control," in *Advanced Textbooks in Control and Signal Processing*, vol. 454. London, U.K.: Springer, 2014, p. 454.

[7] A. T. Khan and S. Li, "Human guided cooperative robotic agents in smart home using beetle antennae search," *Sci. China Inf. Sci.*, to be published.

[8] A. H. Khan, S. Li, D. Chen, and L. Liao, "Tracking control of redundant mobile manipulator: An RNN based metaheuristic approach," *Neurocomputing*, vol. 400, pp. 272–284, Aug. 2020.

[9] A. H. Khan, S. Li, and X. Luo, "Obstacle avoidance and tracking control of redundant robotic manipulator: An RNN-based metaheuristic approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4670–4680, Jul. 2020.

[10] P. Siano and C. Citro, "Designing fuzzy logic controllers for DC–DC converters using multi-objective particle swarm optimization," *Elect. Power Syst. Res.*, vol. 112, pp. 74–83, Jul. 2014.

[11] H. Farsizadeh, M. Gheisarnejad, M. Mosayebi, M. Rafiei, and M. H. Khooban, "An intelligent and fast controller for DC/DC converter feeding CPL in a DC microgrid," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 6, pp. 1104–1108, Jun. 2020.

[12] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," *CSEE J. Power Energy Syst.*, vol. 6, no. 1, pp. 213–225, Mar. 2020.

[13] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 601–614, Feb. 2019.

[14] A. T. Khan, X. Cao, Z. Li, and S. Li, "Enhanced beetle antennae search with zeroing neural network for online solution of constrained optimization," *Neurocomputing*, vol. 447, pp. 294–306, Aug. 2021.

[15] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.

[16] 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.

[17] 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.

[18] H. Lu, L. Jin, X. Luo, B. Liao, D. Guo, and L. Xiao, "RNN for solving perturbed time-varying underdetermined linear system with double bound limits on residual errors and state variables," *IEEE Trans. Ind. Informat.*, vol. 15, no. 11, pp. 5931–5942, Nov. 2019.

[19] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 5353–5360.

[20] K.-S. Tang, K. F. Man, G. Chen, and S. Kwong, "An optimal fuzzy PID controller," *IEEE Trans. Ind. Electron.*, vol. 48, no. 4, pp. 757–765, Aug. 2001.