# AI-ENABLED FUTURE WIRELESS NETWORKS

## *Challenges, Opportunities, and Open Issues*

Medhat Elsayed and Melike Erol-Kantarci

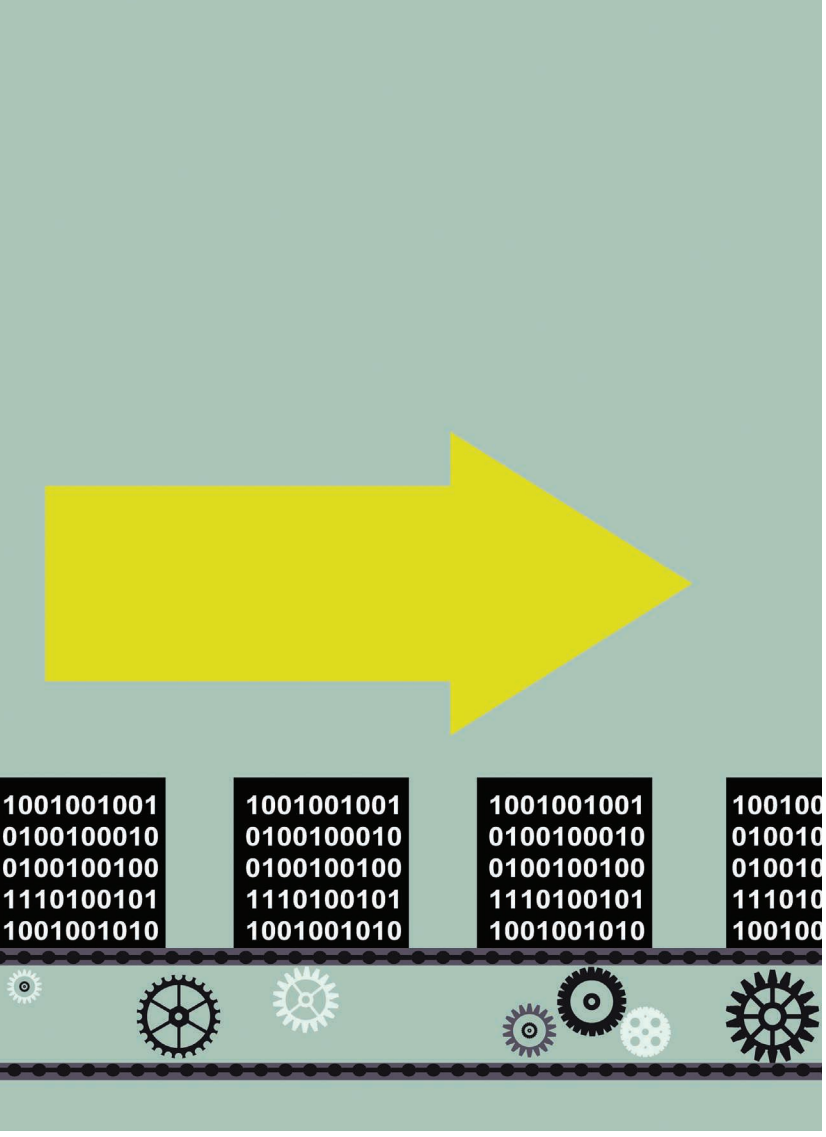

An expected plethora of demanding services and use cases mandates a revolutionary shift in the way future wireless network resources are managed. Indeed, when application requirements for tight quality of service (QoS) are combined with increased network complexity, legacy network-management routines will become untenable in 6G. Artificial intelligence (AI) is emerging as a fundamental enabler to orchestrate network resources from bottom to top. AI-enabled radio access and core will open up new opportunities for automated 6G configurations. At the same time, many challenges in AI-enabled networks need to be addressed. Long convergence times, memory complexity, and the intricate behavior of machine-learning algorithms under uncertainty and the network's highly dynamic channel, traffic, and mobility conditions contribute to the challenges. In this article, we survey state-of-the-art research on using machine-learning techniques to improve the performance of wireless networks. In addition, we identify challenges and open issues to provide a roadmap for researchers.

### Background and Related Work

In the future, wireless networks are expected to support a multitude of services. According to the International Telecommunication Union, 5G network services can be classified into three service scenarios: enhanced mobile broadband (eMBB), ultrareliable and low-latency communications (uRLLC), and massive machine-type communications (mMTC) [1]. Heterogeneous devices with different QoS demands will require an intelligent and flexible allocation of network resources in response to network dynamics. For instance, a highly reliable and low-latency network is necessary to enable the rapid transfer of messages between connected autonomous vehicles.

Concurrently, the same physical infrastructure will be expected to serve users' demand for high-quality video, or even mobile augmented/virtual reality entertainment applications. Next-generation wireless networks, i.e., 5G and the upcoming 6G, are expected to accommodate diverse use cases. In particular, the heterogeneous traffic coming from the mobile, vehicular, smart grid, and tactile domains calls for efficient use of network resources to maintain the QoS requirements of each application. In addition, resource efficiency, reliability, and robustness are becoming more stringent for 5G and beyond networks. To meet these needs, 6G networks must incorporate a paradigm shift for network-resource optimization, in which efficient and intelligent resource-management techniques are employed. AI or, more specifically, machine-learning algorithms are promising tools to intelligently manage networks, such that network efficiency, reliability, and

*Digital Object Identifier 10.1109/MVT.2019.2919236*
*Date of publication: 10 July 2019*



©ISTOCKPHOTO.COM/ALEUTIE

robustness goals are achieved and QoS demands satisfied. The opportunities that arise from learning environmental parameters under varying wireless channel behaviors render AI-enabled 5G and 6G superior to preceding generations of wireless networks. Figure 1 highlights some wireless problems and applications that can leverage the potential of AI.

In the literature, several research efforts have addressed radio-resource allocation, device-to-device communications, access to unlicensed bands, routing, security, and fault management using machine learning [2]. Among these, radio-resource allocation has received significant interest from the research community. In particular, distributed algorithms have been proposed to enable each base station (BS) to learn radio resource-allocation parameters independently, which facilitates readily deploying small BSs without requiring preconfiguration. Among a broad variety of machine-learning algorithms (extensively surveyed in [3]), deep learning has been employed widely in many works. A comprehensive survey of deep-learning algorithms applications for different network layers can be found in [3]. In addition, [4] discusses various machine-learning techniques and provides visionary use cases for their potential application to multiple-input, multiple-output (MIMO), heterogeneous networks, and cognitive radio.

Unlike previous surveys, here we provide state-of-the-art AI-enabled techniques with a focus on resource allocation, spectrum access, BS deployment, and energy efficiency in a wide variety of uses cases, including mobile broadband, tactile Internet, device-to-device communications, and unmanned aerial vehicles. In addition, we present a deep reinforcement-learning-based solution for resource allocation that provides significantly improved delay performance; we also discuss challenges and open issues related to AI-enabled future wireless networks.

## Background on Machine Learning

Over the past decade, the huge growth in data across many different fields has resulted in a "big data" challenge, which amplifies the need for intelligent data-analysis schemes. To cope with this problem, various machine-learning methods, such as deep learning, have been used along with traditional machine-learning methods and adopted in wireless networks. Therefore, in this section, we give a brief overview of widely used techniques.

Machine-learning schemes can be classified into four main categories: supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning. These differ in the way the algorithm is being trained [5].

In supervised learning, the training is initially performed by some labeled data. The labeled data represent a set of inputs with their corresponding outputs, which are known beforehand. Therefore, supervised-learning algorithms are well suited to applications with historical data. Feature extraction and classification have been applied to several signal-processing problems. In classification, the task is to identify to which set of categories a new observation belongs. In contrast, unsupervised-learning algorithms aim to infer features in the data, thus inferring the implied structure. Semisupervised learning algorithms use both labeled and unlabeled data.

Finally, reinforcement learning uses data from the implementation instead of historical data. Its aim is to improve the performance of an agent for a certain task using feedback from the environment. As such, the agent's goal is to predict which next action to take so as to earn the greatest final reward. Reinforcement learning is unsupervised; however, the learning method is different from other unsupervised learning

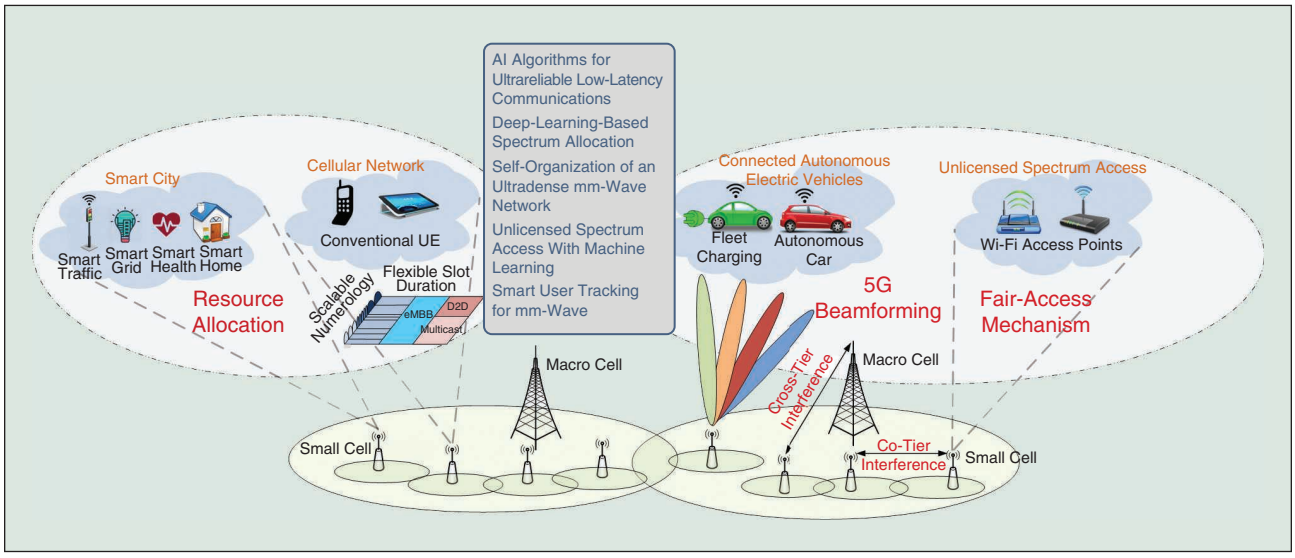



***FIGURE 1*** An AI-enabled future wireless network and associated services. mm-wave: millimeter-wave; D2D: device to device; UE: user equipment.

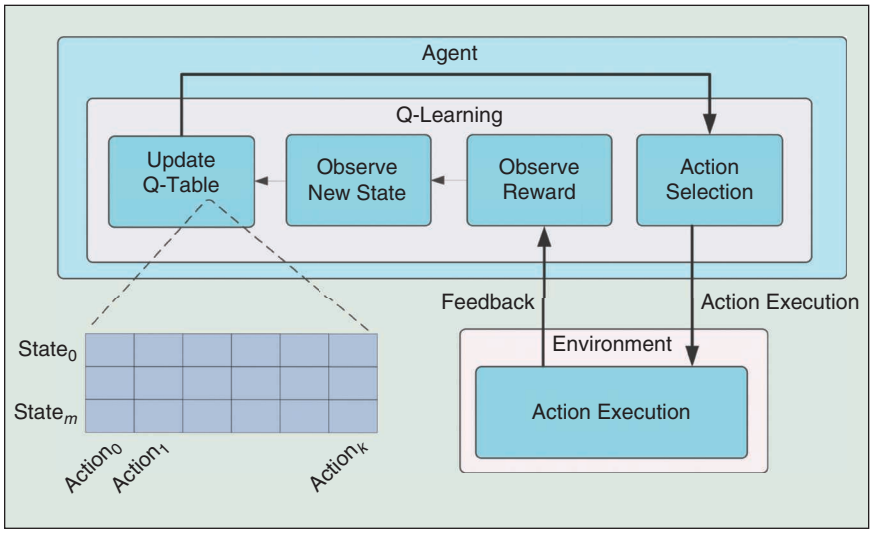



***FIGURE 2*** A conceptual diagram of a Q-learning operation.

techniques. Rather than learning the structure of some data, reinforcement learning tries to explore the best actions in the medium of operation. Hence, the ability to capture the environment through feedback and perform actions makes reinforcement learning suitable for problems involving a series of decisions, i.e., following a policy of actions according to the observed environment's state.

Reinforcement learning can be model-based or model-free. In model-based learning, the agent aims to understand the environment and builds a model to represent it. In contrast, the agent in model-free learning aims to learn a policy that it can follow. Model-free learning is more suitable for wireless networks, because learning from history does not match well with the network dynamics. Q-learning is an example of a model-free reinforcement-learning algorithm that aims to learn a policy informing the agent what action to take within each state. In other words, Q-learning provides the agent with the ability to learn the best actions in each state without knowing the model transition probabilities.

Figure 2 presents a conceptual diagram of Q-learning. The agent starts by selecting an action according to the policy. After the action executes, the environment is influenced, and some feedback is returned to the agent. In wireless networks, feedback can include interference, the queuing state of nodes, path congestion, or learning actions of other nodes, as well as many more factors. Therefore, agents can either compute feedback at their side, such as measuring the signal-to-interference noise ratio, or they can exchange decisions as a form of feedback to one another. The latter poses communication overhead because more network resources must be allocated to facilitate signaling among agents. The feedback will modify the reward, and the agent will observe a new state. The state, action, and reward values obtained characterize the quality of the action taken at the current state, so the agents store this quality value (i.e., Q-value) in a Q-table. By repeatedly exploring different actions in different states, the agent will be able to identify the optimal ones to take. The update of the Q-value is performed iteratively using the Q-learning update equation as [5]

$$Q(s,a) \leftarrow (1-\alpha)\,Q(s,a) + \alpha[R(s,a) + \gamma \max_{a} Q(s,a)], \quad (1)$$

where $\alpha$ is the learning rate, $\gamma$ is the discount factor, $R(s, a)$ is the reward at the state–action pair $(s, a)$, and $Q(s, a)$ is the Q-value of state–action pair $(s, a)$.

Besides Q-learning, a neural network was recently used in state-of-the-art wireless network research. Neural networks are designed to mimic the structure of neurons in the human brain. In particular, a neural network consists of three types of layers: input, output, and hidden. Each layer comprises a set of artificial neurons that perform certain mathematical functions, specifically, neuron activation functions. Neurons in a certain layer are connected to the neurons in the preceding layer, and each connection has a weight. In the training phase, the weights are adjusted according to the training dataset, which provides a set of inputs and the expected outputs. Neural networks can be structured in different forms, such as feedforward, convolutional, or recurrent. A neural network with one hidden layer is a shallow neural network, while one with multiple hidden layers is a deep one. In addition, a deep neural network can take different forms, such as feedforward, convolutional, or recurrent. Figure 3 presents an example of a deep feedforward neural network in which information flows in one direction. In contrast, a deep recurrent neural network incorporates feedback connections among layers. Such feedback allows the deep recurrent neural network to infer relations within long sequential information (which is more efficient for generalizations).

A more recent method is deep Q-learning, which Google DeepMind developed [6]. In contrast to traditional tabular Q-learning, deep Q-learning complements the Q-learning algorithm with a deep convolutional neural network that approximates the Q-value function, avoiding the need to store a huge amount of information. Figure 4 presents the deep reinforcement learning components as proposed by Google DeepMind [6], which incorporates Q-learning, a neural network, and an experience replay memory. The Q-learning algorithm is similar to the one presented in Figure 2 after removing the Q-table. As such, Q-learning performs action selection, reward calculation, and an observation of the new state, as before, which is denoted as Q-learning experience $e$ = old state ($s$), old action ($a$), reward ($rc$) of old state–action, and new state ($s'$). Action selection is performed using Q-learning policy applied to the Q-values estimated by the neural network. The training of the neural network is performed using experiences from the Q-learning algorithm, in which the training samples are drawn from an experience replay memory that stores experiences over many episodes.

Incorporating a deep neural network can enable reinforcement learning to scale to problems that were previously intractable [7]. Those problems have a high-dimensional state–action space that leads to a curse-of-dimensionality problem, which may cause slow convergence behavior. The deep neural network acts as a function approximator, which estimates one Q-value during each iteration, instead of predicting individual actions' Q-values for a given input state with only a single forward pass. As demonstrated in the "Resource Allocation Using Deep Q-Learning for Low-Latency Applications" section, this improves the convergence of Q-learning. Despite its advantage, tailoring a deep neural network to the problem at hand involves choosing several parameters, such as neural network type, number of layers, number of neurons, and so on.

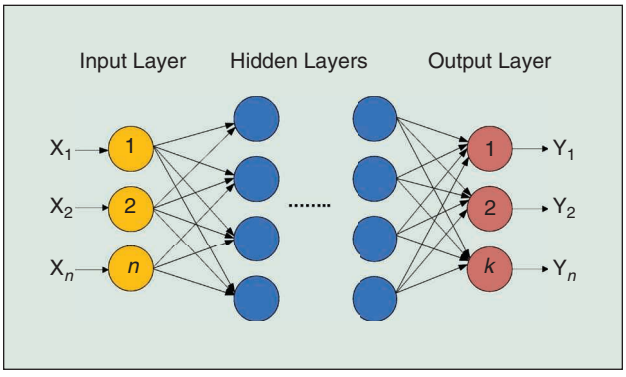



**FIGURE 3** A typical structure of a neural network. (Adapted from [5].)

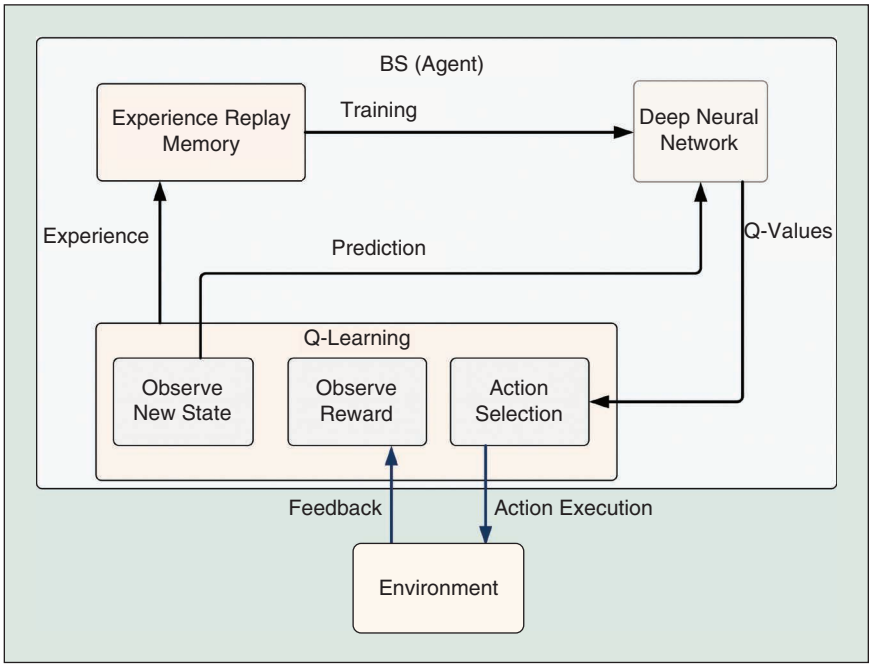



**FIGURE 4** The architecture of deep Q-learning [6].

## Radio-Resource Allocation

Radio-resource allocation is a key function in wireless networks, where using resources efficiently is necessary to improve the quality of both network service and the user experience. Numerous factors impact the performance of radio resource-allocation schemes. For instance, wireless channel quality influences the probability of successful data transmission; hence, it affects both throughput and latency. Meanwhile, interference has an impact on channel quality, which is a highly dynamic parameter due to environment, traffic, and mobility. Because future wireless networks are anticipated to serve a wider variety of users, spectrum scarcity will be more pronounced and, consequently, will make radio-resource allocation more challenging.

Radio resource-allocation approaches can be classified as either centralized or decentralized. The centralized approaches mainly rely on a principal entity that collects information from all users in the network. Resource allocation is then performed according to the global view of the network. The decentralized approaches allow users to make decisions autonomously, with slight cooperation in some cases. The centralized approaches can achieve optimal results; however, they add overhead for the information exchange. Still, the decentralized approaches allow for more flexibility at the cost of suboptimal results. Radio resource-allocation schemes also vary according to their optimization objectives, such as improving throughput, latency, fairness, and energy and spectral efficiency. Finally, radio resource-allocation schemes can also be classified according to the model used, including optimization, heuristic, game theoretic, or machine-learning based. Optimization-based schemes are useful for finding the optimal solution to achieve certain objectives. Such problems are normally complex to solve, as it may be necessary to target the optimization of several parameters to achieve multiple objectives simultaneously in the presence of several constraints.

The heuristic method is an alternative that targets the optimal solution, rather than relaxing the model assumptions and searching for a reasonable solution. However, such relaxation can be loose, and the method is not guaranteed to converge to a good solution. Another alternative is game theory-based methods, in which network nodes are modeled as players interacting and affecting others' decisions. Each player has a set of actions (decisions) to maximize its payoff (utility). An obvious advantage of game-theoretical methods is their flexibility for adapting to network dynamics. Finally, machine-learning schemes are emerging as an alternative solution to the radio resource-allocation problem. In the rest of this article, we discuss machine-learning methods and their state-of-the-art use in wireless networks.

## AI-Enabled Radio-Resource Allocation

Here, we survey the machine learning-based radio resource-allocation techniques and group them according to use cases and application domains.

### *Traditional Mobile Broadband Use Cases*

The research in [8] provides a radio resource-allocation learner architecture as well as a learning-algorithm framework based on reinforcement learning. The architecture is divided into a centralized learner and distributed actors, with the latter being responsible for executing the learned policies and collecting samples of experience from the network.

In [9], the authors adopt a reinforcement-learning approach backed with neural networks to facilitate scheduling decisions so that packet delay and packet-drop rate are improved. The problem is formulated as a multiobjective optimization that aims to select the optimal scheduling rule and resource-block allocation at each transmission-time interval. The performance results are conducted for five reinforcement-learning algorithms for different windowing factors, objectives (i.e., delay or packet-drop rate), and traffic classes.

Radio-resource allocation couples closely with user association. Users associate with BSs that offer better channel quality, and the number of users associated with a BS impacts resource-sharing decisions. Thus, a collaborative neural Q-learning algorithm is used in [10] to perform user-cell association in ultradense small-cell networks. Users strive to improve their rate by selecting the best association with a BS. The proposed deep Q-learning algorithm uses the small-cell BSs (SBSs) selected by neighboring users (as the name *collaborative* implies) along with the local information as an input to its neural network. The output of the neural network is the estimated Q-values, where each Q-value represents an action (i.e., an SBS for the user to select).

### *Device-to-Device Communications and Tactile Internet*

Device-to-device communication is expected to be a significant part of future wireless networks because it provides a way to exchange information among users without a BS. However, this autonomous operation poses further challenges to resource allocation. Bayesian reinforcement learning is used in [11] to form coalitions in device-to-device networks to maximize network throughput subject to power constraints. Devices create coalitions to maximize their long-term rewards. The decision process used in forming a coalition includes selecting the BS, transmission power, transmission channel, and transmission mode (e.g., cellular or device-to-device).

Radio-resource allocation also plays a significant role in the tactile Internet, which is characterized by its ultrareliability and ultralow latency. A Q-learning algorithm was proposed to perform resource-block allocations to

maximize the throughput of data-intensive traffic [12]. A two-tier network of SBSs and an Evolved Node B (eNB) is proposed to carry the traffic of both conventional and data-intensive users. The Q-learning algorithm learns the traffic patterns and channel conditions of the network and then allocates the resource blocks in such a way that data-intensive users experience low latency and high throughput.

### *Unmanned Aerial Vehicle-Assisted Networks*

Unmanned aerial vehicles have recently become a promising approach to augmenting BSs and enhancing the connectivity, capacity, and QoS of wireless networks. The research in [13] considers a network of virtual reality users that communicate using unmanned aerial vehicles. These vehicles behave as relays that receive images on the uplink and forward them on the downlink, using LTE licensed and unlicensed bands, respectively. Furthermore, the quality of images can be adjusted to change the transmitted data size to fit the resources allocated. Therefore, dynamic resource allocation is required to both improve the quality of users' experiences and meet the delay requirements of virtual reality applications. To achieve this, the authors employ a deep echo-state network (ESN) algorithm. The ESN is a type of recurrent neural network with sparsely connected hidden layers. It aims to learn the weights of the output layer only; the weights of hidden layers are fixed and randomly assigned. This, in turn, increases the speed of learning more than in conventional recurrent neural networks.

## AI-Enabled Deployment, Spectrum Access, and Energy-Efficiency Techniques

Other than resource allocation, machine-learning algorithms are finding significant uses in deploying BSs, accessing unlicensed bands, and energy-efficient network design. In this section, we survey the research in domains that are highly relevant to 6G.

### *Deploying Base Stations in Unmanned Aerial Vehicle-Assisted Networks*

The deployment and placement of unmanned aerial vehicles have recently become an active area of research. In [14], the authors address the 3D positioning of an aerial BS to assist ground stations with providing enhanced QoS for mobile users. The network topology gradually changes due to users' mobility, which impacts the QoS they receive. Furthermore, the positioning algorithm will require more time to relearn the network. As such, an agile and fast-learning algorithm is needed. The authors propose using a Q-learning algorithm for determining the efficient placement of aerial stations to maximize network throughput. The difference between the current and previous QoS (i.e., throughput) constitutes the agent's reward, which motivates that agent to improve the decision to gain positive rewards, hence improving the network's throughput. After the training phase of Q-learning, it has been shown that the algorithm can more rapidly adapt to small changes in the network.

### *Spectrum Access in Unlicensed Bands*

Deploying LTE networks in unlicensed bands (i.e., LTE unlicensed bands) offers opportunities to improve cellular performance. However, if they are not deployed properly, LTE unlicensed bands may degrade the performance of wireless local area networks, specifically Wi-Fi. In [15], the authors adopt a proactive resource-allocation algorithm to utilize unlicensed spectrum for LTE small cells while maintaining fairness in terms of existing Wi-Fi networks and other LTE operators. In particular, the proposed algorithm aims to balance the spectrum occupancy to avoid degrading Wi-Fi performance. To achieve this, the allocation is performed proactively by predicting the LTE traffic and serving it either momentarily or shifting part of it to the future. The proposed approach includes reinforcement learning with long short-term memory that performs dynamic channel allocation, carrier aggregation, and fractional spectrum access. The use of long short-term memory offers the ability to predict a sequence of future actions, reinforcing the proactive approach.

### *Energy-Efficiency Techniques*

AI-enabled energy efficiency will be another important pillar of 6G. A reinforcement learning-based, energy-aware resource-allocation technique is introduced in [16]. In particular, the authors use actor–critic reinforcement learning to find the number of users allocated to each BS and the channels and power allocated to each user, as well as to select the energy source for the BSs (e.g., energy from the utility grid or renewable energy resources). Due to the environment's stochastic nature (i.e., wireless channel and renewable energy sources), a model free-learning algorithm is needed. As such, the authors propose to use actor–critic reinforcement learning, which combines both policy- and value-based reinforcement learning. Such integration facilitates learning a continuous action space as well as achieving the convergence to the optimal solution. The actor part is responsible for learning the policy and generating actions, where a Gaussian distribution is used to generate stochastic actions. The critic part evaluates the actor's policy and performs the value-function approximation.

## Resource Allocation Using Deep Q-Learning for Low-Latency Applications

In this section, we present our recent work on utilizing deep Q-learning to improve latency in dense small-cell networks [17]. Here, mission-critical traffic coexists with traditional (i.e., noncritical) LTE traffic in a dense and heterogeneous network scenario, where efficient

resource allocation is needed to achieve ultralow latency for mission-critical nodes. Therefore, our algorithm, delay minimization using deep Q-learning (DMDQ), combines long short-term memory with Q-learning to perform resource-block allocation. This study addresses the uplink-scheduling problem when critical and noncritical traffic coexist in a heterogeneous small-cell network, where the main objective is to minimize the latency of mission-critical devices (MCDs) while maintaining fairness among MCDs and conventional LTE user equipment (UE).

Each BS executes DMDQ in a decentralized manner to minimize the total end-to-end delay of its users, where *end-to-end delay* is defined as transmission and queuing (i.e., scheduling) delays. We define Q-learning tuples as follows.

- *Agents*: These are the SBSs/eNB.
- *States*: We tie Q-learning states to the main objective of delay minimization. Hence, an agent can be either in state 0, where its average delay is less than a target value, or in state 1 otherwise.
- *Actions*: Actions are defined as resource-block allocations to every user the agent covers.
- *Reward*: The reward is defined as the sigmoid function of the total delay.
- *Policy*: An epsilon-greedy policy is used, in which agents perform exploration either by selecting random action or by selecting the action with the maximum Q-value.

We adopt a long short-term memory neural network to estimate Q-values instead of using the Q-table method, as discussed in the "Background on Machine Learning" section.

We compare DMDQ with two algorithms, the tabular Q-learning scheme and a conventional round-robin scheme. Our system-level simulator is based on the MATLAB LTE toolbox. The network consists of a single macro cell network (one eNB) with an 800-m radius and a number of SBSs, each with a 50-m radius. There are five SBSs, 20 UEs, and six UNBs (UEs connected to only eNB). MCDs generate traffic according to the beta distribution as defined by the 3GPP for machine-type traffic, and UEs generate traffic according to the Poisson distribution [17].

Figure 5 shows the average end-to-end delay of all schemes. As can be observed, DMDQ outperforms the Q-learning and round-robin schemes by approximately 28% and 37% delay reductions, respectively. Furthermore, Figure 6 shows the convergence of DMDQ as compared to the Q-learning scheme (calculated as $(1-\Sigma_{t=1}^{T} RC_t/T)$, where $T$ is the subframe number (i.e., x-axis), and $\Sigma_{t=1}^{T} RC_t/T$ represents the average absolute reward of all SBSs [17]. As can be observed, DMDQ converges more rapidly (i.e., after 30 iterations) while Q-learning requires approximately 80 iterations to converge. The results show that the deep Q-learning algorithm can provide a notable delay reduction with a relatively short convergence time.

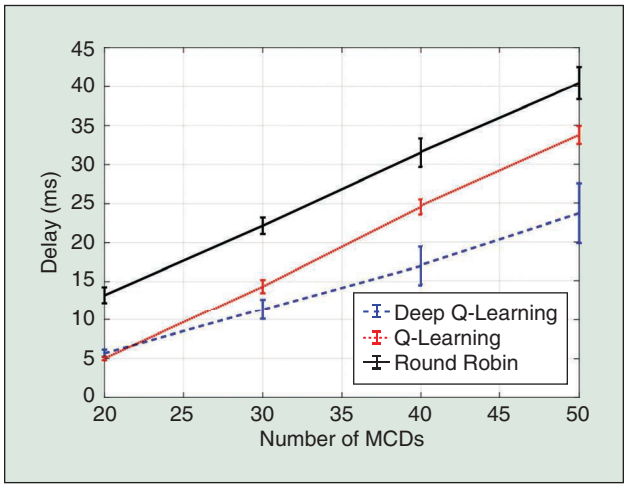



**FIGURE 5** The average end-to-end delay of MCDs.

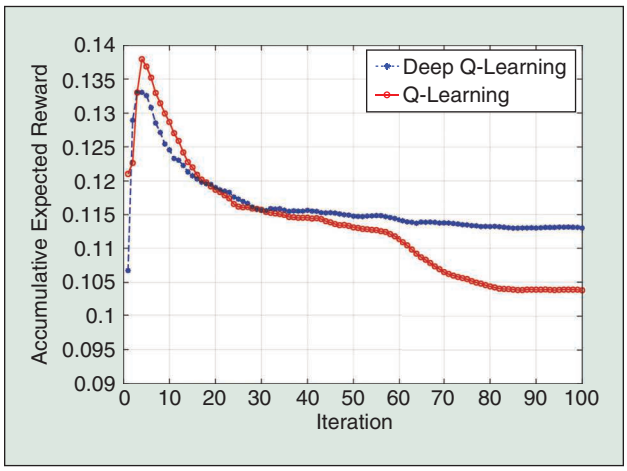



**FIGURE 6** The average discounted reward for both DMDQ and deep Q-learning.

## Open Issues and Future Directions

6G will have greater complexity than the preceding generations of wireless networks, which will introduce the need for intelligent mechanisms to orchestrate the available resources, services, and users. Thus, AI-enabled techniques or, more precisely, the use of machine-learning algorithms may allow future networks to learn from their environment, adapt the changes in an automated fashion, and achieve optimal performance.

Machine-learning algorithms have paved the way for significant agility in network management, yet several challenges are still open for research efforts. Generally, the open issues can be classified into two main pillars: the performance of machine-learning algorithms and the performance of wireless networks.

The relatively long convergence time of machine-learning methods undermines their usefulness in highly dynamic wireless networks. A careful investigation of the convergence problem, as well as the factors that influence the convergence, is needed. Novel machine-learning techniques with faster convergence and online learning capabilities can better benefit wireless networks.

Besides convergence, the uncertainty in the wireless network calls for ongoing updates of the parameters of the machine-learning method or even the method itself. The stochastic nature of the wireless channel may require continuous adaptation. For instance, a network encompassing a large and diverse set of users will have very dynamic operation. In particular, users who join or leave the network may have very different QoS and quality-of-experience requirements. Thus, it is necessary to examine whether a one-size-fits-all approach is feasible in real-world implementations. In addition, the scalability of machine-learning algorithms must be addressed. Machine-learning algorithms can become unfeasible for moderately large data, especially in collaborative-learning approaches. This calls for a scalable learning algorithm to accommodate the dense use cases of future wireless networks.

Furthermore, supervised and unsupervised learning techniques have recently been used for massive MIMO [4]. Further research is needed to investigate whether it is possible to enhance the performance of massive MIMO by using reinforcement learning and deep learning.

Finally, AI-enabled networks also impact e-health applications. For instance, advancing outside-of-clinic operations by using wearable sensors [18] requires harmonizing network resource allocation across several technologies, and machine-learning algorithms can be helpful for such harmonization. Hence, the application-specific use of machine learning must be explored further.

## Acknowledgment

This research was supported by the Natural Sciences and Engineering Research Council of Canada under grant RGPIN-2017-03995.

## Author Information

***Medhat Elsayed*** (melsa034@uottawa.ca) is a Ph.D. degree candidate with the School of Electrical Engineering and Computer Science, University of Ottawa, Canada. His research interests include artificial-intelligence-enabled wireless networks, 5G and beyond, and smart grids.

***Melike Erol-Kantarci*** (melike.erolkantarci@uottawa.ca) is an associate professor with the School of Electrical Engineering and Computer Science, University of Ottawa, Canada; the founding director of the Networked Systems and Communications Research Laboratory; and a courtesy assistant professor with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, New York. Her research interests include artificial intelligence-enabled networks, 5G and beyond wireless networks, smart grid, electric vehicles, and the Internet of Things. She is a Senior Member of the IEEE.

## References

[1] International Telecommunications Union. (2018), "5g overview," in Setting the Scene for 5G: Opportunities and Challenges. ITU. Geneva, Switzerland. [Online]. Available: http://handle.itu.int/11.1002/pub/811d7a5f-en

[2] S. Ayoubi et al., "Machine learning for cognitive network management," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 158–165, Jan. 2018. doi: 10.1109/MCOM.2018.1700560.

[3] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surv. Tut.*, vol. 20, no. 4, pp. 2595–2621, 2018. doi: 10.1109/COMST.2018.2846401.

[4] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017. doi: 10.1109/MWC.2016.1500356WC.

[5] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA: MIT Press, 2014.

[6] V. Mnih et al. "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015. doi: 10.1038/nature14236.

[7] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017. doi: 10.1109/MSP.2017.2743240.

[8] F. D. Calabrese, L. Wang, E. Ghadimi, G. Peters, L. Hanzo, and P. Soldati, "Learning radio resource management in RANs: Framework, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 138–145, Sept. 2018. doi: 10.1109/MCOM.2018.1701031.

[9] I. Comşa et al., "Towards 5G: A reinforcement learning-based scheduling solution for data traffic management," *IEEE Trans. Netw. Service Manag.*, vol. 15, no. 4, pp. 1661–1675, Dec. 2018. doi: 10.1109/TNSM.2018.2863563.

[10] K. Hamidouche, A. T. Z. Kasgari, W. Saad, M. Bennis, and M. Debbah, "Collaborative artificial intelligence (AI) for user-cell association in ultra-dense cellular systems," in *Proc. IEEE Int. Conf. Communications Workshops*, May 2018, pp. 1–6. doi: 10.1109/ICCW.2018.8403664.

[11] A. Asheralieva, "Bayesian reinforcement learning-based coalition formation for distributed resource sharing by device-to-device users in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5016–5032, Aug. 2017. doi: 10.1109/TWC.2017.2705039.

[12] M. Elsayed and M. Erol-Kantarci, "Learning-based resource allocation for data-intensive and immersive tactile applications," in *Proc. 5G World Forum*, 2018, pp. 278–283. doi: 10.1109/5GWF.2018.8517001.

[13] M. Chen, W. Saad, and C. Yin, "Echo state learning for wireless virtual reality resource allocation in UAV-enabled LTE-U networks," in *Proc. IEEE Int. Conf. Communications*, May 2018, pp. 1–6. doi: 10.1109/ICC.2018.8422503.

[14] R. Ghanavi, E. Kalantari, M. Sabbaghian, H. Yanikomeroglu, and A. Yongacoglu, "Efficient 3D aerial base station placement considering users mobility by reinforcement learning," in *Proc. IEEE Wireless Communications and Networking Conf.*, Apr. 2018, pp. 1–6. doi: 10.1109/WCNC.2018.8377340.

[15] U. Challita, L. Dong, and W. Saad, "Proactive resource management for LTE in unlicensed spectrum: A deep learning perspective," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4674–4689, July 2018. doi: 10.1109/TWC.2018.2829773.

[16] Y. Wei, F. R. Yu, M. Song, and Z. Han, "User scheduling and resource allocation in HetNets with hybrid energy supply: An actor-critic reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 680–692, Jan' 2018. doi: 10.1109/TWC.2017.2769644.

[17] M. Elsayed and M. Erol-Kantarci, "Deep reinforcement learning for reducing latency in mission critical services," in *Proc. IEEE GLOBECOM 2018*, pp. 1–6. doi: 10.1109/GLOCOM.2018.8647289.

[18] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 9, no. 21, Apr. 2012. [Online]. Available: https://doi.org/10.1186/1743-0003-9-21

*VT*