IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 September 2022, accepted 26 October 2022, date of publication 1 November 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3218675

RESEARCH ARTICLE

A *Q*-Learning Approach for Real-Time NOMA Scheduling of Medical Data in UAV-Aided WBANs

ZEINAB ASKARI[1], JAMSHID ABOUEI[1], (Senior Member, IEEE), MUHAMMAD JASEEMUDDIN[2], (Member, IEEE), ALAGAN ANPALAGAN[2], (Senior Member, IEEE), AND KONSTANTINOS N. PLATANIOTIS[3], (Fellow, IEEE)

[1]WINEL Research Group, Department of Electrical Engineering, Yazd University, Yazd 89195-741, Iran
[2]Department of Electrical, Computer and Biomedical Engineering, Toronto Metropolitan University (formerly Ryerson University), Toronto, ON M5B 2K3, Canada
[3]Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada

Corresponding author: Jamshid Abouei (abouei@yazd.ac.ir)

The research is supported by NSERC through Discovery Grant.

ABSTRACT Unmanned Aerial Vehicles (UAVs) have emerged as a flexible and cost-effective solution for remote monitoring of the vital signs of patients in large-scale Internet of Medical Things (IoMT) Wireless Body Area Networks (WBANs). This paper deals with the problem of using UAVs for real-time scheduling of the transmission of vital signs in delay-sensitive IoMT WBANs. The main challenge for such a network is to timely and reliably transmit the vital signs of patients to the remote monitoring center without interrupting their daily lifestyles. To achieve this goal, we propose a *Q*-learning-based algorithm to optimize the trajectory of each UAV, as the mobile Base Station (BS), to harvest vital signs of patients in outdoor applications, especially in unreachable areas. In this algorithm, UAVs learn to reach the best 3D position by discovering the network environment step-by-step. It stands for the position in which the covered patients by each UAV have the highest transmission rate, the least delay and energy consumption. Moreover, we employ the Non-Orthogonal Multiple Access (NOMA) technique to simultaneously schedule multiple transmissions by accepting a degree of interference between them in order to enhance the spectrum efficiency of the network. Eventually, the performance of our proposed scheme is evaluated via extensive simulations in terms of throughput, energy consumption, and delay. The simulation results show that our proposed scheme iteratively converges to the benchmark value of the mentioned factors by increasing the information of cluster environment through episodes.

INDEX TERMS IoMT WBAN, UAV, latency, trajectory, NOMA, *Q*-learning.

I. INTRODUCTION

A. BACKGROUND AND RELATED WORK

Expanding Internet of Things (IoT) devices everywhere enables the real-time monitoring of vital signals of patients in indoor/outdoor environments without interrupting their daily lifestyles. In particular, in recent years, remotely tracking the vital signs of patients has been gaining a lot of research interests in terms of 5G and 6G wireless networks and beyond. These devices including smart phones and smart watches should have the ability to connect to the internet via Base Stations (BSs) or cloud servers and transmit vital signs of patients to the remote health-care center with high reliability. The above characteristics have recently expanded the traditional Wireless Body Area Networks (WBANs) to new emerging Internet of Medical Things (IoMT)-based WBANs in emergency situations for outdoor environments [1]. In particular, COVID-19 pandemic that has jeopardized the health and safety of elderly resulted in significant dependence on IoMT-based WBANs. Different from traditional healthcare applications where controlling the vital signs of patients were only possible inside the hospitals and by wired equipment, in recent IoMT-based WBANs, there is no need for patients to monitor their vital signs only by hospitalizing in a medical

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

center. Indeed, in the state-of-the-art IoMT-based WBANs, the vital signs of patients are collected via their smartphones or smartwatches and then transmitted to the remote monitoring center, on a regular basis, without interrupting their everyday life. In other words, these IoMT devices provide access to the Internet everywhere for remotely monitoring of vital signs. In these kinds of networks, a massive volume of traffic load will be invariably generated by several patients. In this regard, the most challenging issue in such an ultra-dense E-health care network is aggregating data packets of these patients in a timely and reliable manner. Moreover, this problem will become far and away crucial when compulsive emergency conditions occur to the patients located in unreachable areas where static Base Stations (BSs) will not be able to provide efficient services. In these delay-sensitive applications, using Unmanned Aerial Vehicles (UAVs), as mobile base stations, will be a viable solution for real-time remote monitoring of vital signs. As demonstrated in the literature [2], [3], [4], UAVs provide an efficient and reliable data harvesting system for wireless networks by hovering over outdoor areas. In addition, UAVs can easily access data packets of patients in unreachable areas by providing air-to-ground communication links and enhance throughput and energy efficiency of WBANs by adapting their horizontal locations and altitudes.

Furthermore, optimizing the horizontal and vertical trajectory of UAVs leads to a remarkable increase in the network coverage area compared to employing static BSs on the ground which further enhances the throughput, e.g., the sum capacity, and quality of service of the network. It is worth mentioning that the authors in [5] state that the use of UAVs as BSs in future wireless communication networks is currently gaining significant attention for its ability to yield ultra-flexible deployments, in use cases like disaster recovery scenarios. Moreover, in [6] it is asserted that future mobile networks aim to realize larger coverage, support more devices, and achieve higher throughput to meet the explosive increasing demand for data. As a result, there has been growing interest in hybrid cellular networks assisted by UAVs as mobile BSs, due to their mobility and flexibility. In [7], furthermore, UAVs are claimed to be capable of providing wireless connectivity even without network infrastructure or complementing the conventional BSs, whose coverage may suffer from severe blockage due to tall buildings or the damage caused by natural disasters.

Although UAVs considerably expand the coverage area of the network, efficiently designing their trajectory is a major challenge in IoMT-based WBANs, as we should reduce the delay and the energy consumption along with increasing the transmission rate of the network. In this regard, several research attentions turned recently to employ Reinforcement Learning (RL)-based algorithms to optimally find the best location of UAVs in different wireless networks such as wireless sensor networks, cellular systems, and vehicular networks [7], [8], [9], [10], [11], [12], even though, to the best of our knowledge, there is not any RL-based algorithm in UAV-assisted WBANs. Generally, the relevant RL-based algorithms are classified into two main categories *i*) model-based, and *ii*) model-free RL schemes. In the model-based class, the agent computes the transition probability distribution and reward function of all possible state-action pairs and then uses this model to optimize the policy by predicting the best actions that lead to higher rewards through interacting with the environment. In contrast, in model-free algorithms like Q-Learning (QL), the agent does not employ the transition probability to predict the best action, instead, it optimizes the policy by making direct decisions via a trial and error mechanism. Recently, QL algorithms have attracted remarkable research attentions in finding an optimal trajectory of UAVs. This algorithm is among the model-free RL category and is based on the Markov Decision Process (MDP) to sequentially find the best position of UAVs in each state in order to achieve the highest reward in terms of optimizing the aforementioned factors. In addition, in the QL class, the agent has no prior knowledge about the environment and the reward of each state-action transition. It is worth mentioning that due to a great deal of time needed by complex machine learning algorithms such as DQN, DDQN, and Rainbow schemes, for collecting enormous datasets as a reply buffer and training procedure of neural networks, they are sort of slow in terms of converging to the optimal value. Taking this issue into account, we employ an efficient Q-learning scheme with the modified state and action space to address the real-time transmission of delay-sensitive data packets in the proposed test-bed.

Although the problem of utilizing UAVs has been studied in various wireless networks, to the best of our knowledge, there exists a few research works on the applications of UAVs in IoMT-WBANs [13], [14], [15], [16], [17], [18], [19], [20]. Authors in [13] and [14] identify open research issues and challenges in UAV-assisted health-care intelligent systems and explain some of the practical attempts that have been made for employing UAVs in emergency medical services. References [15] and [16] study the security issues in outdoor health monitoring systems with the help of UAVs. Authors in [17], [18], and [19] aim to improve the procedure of collecting data from bio-sensors using UAVs. Finally, the authors in [20] focus on optimizing the UAV placement over a serving area where UAV is considered as a fog node to serve the IoMT devices on the ground. In this regard, they propose a particle swarm optimization-based algorithm to improve the communication coverage, energy consumption, exploration area, and optimal number of UAVs.

The Non-Orthogonal Multiple Access (NOMA) scheme is another promising technique in reducing delay and increasing the transmission rate of vital signals to the remote monitoring center. Using this scheme, multiple patients can simultaneously transmit their vital signals satisfying the limited interference at the receiver side. NOMA achieves this goal by utilizing Superposition Coding (SC) at the transmitter and Successive Interference Cancelation (SIC) at the receiver. Thus, the NOMA scheduling scheme can considerably

outperform conventional Orthogonal Multiple Access (OMA) schemes in delay-sensitive and reliable applications.

This paper aims to address the problem of timely and reliable transmission of vital signs of patients to the remote monitoring center in outdoor applications, especially in unreachable and emergency situations. This will be accomplished by simultaneously employing multiple UAVs as mobile BSs and NOMA scheduling technique. To this end, each UAV finds its best trajectory to harvest the vital signs of covered patients in such a way that it increases the throughput and reduces the delay and energy consumption. In this regard, we propose a Q-learning-based algorithm that discovers the environment to reach the best position step-by-step. Our proposed algorithm reduces the computational complexity in UAVs because it does not require all information of entire environment. Additionally, to further increase the transmission rate and reduce the delay, the NOMA technique is employed to schedule multiple transmissions at the same time slot to each UAV.

B. MAIN CONTRIBUTIONS

The key contributions of this paper are briefly explained as follows:

- This paper investigates, for the first time, the NOMA scheduling of vital signs of patients in UAV-enabled IoMT-WBAN for outdoor applications. The patients are equipped with IoMT devices for sensing and gathering the different vital signs and UAVs are responsible for gathering the data packets of patients distributed in the city area which includes unreachable locations.
- Our proposed scheduling algorithm comprises of two levels. In the first level, we schedule the transmissions of data packets of bio-sensors belonging to each patient, to the corresponding hub. In order to eliminate the interference between asynchronous transmissions of bio-sensors, the Walsh Hadamard (WH) coding scheme is employed in which the sensed data packets of each bio-sensor are multiplied by one of the orthogonal codes extracted from WH matrix. Using this scheme, transmissions of all bio-sensors of a patient can be scheduled in the same time slot without occurring any collision among them.
- In level two, the city area is partitioned into multiple clusters and one UAV is assigned to each cluster. All patients belonging to individual cluster are served by the corresponding UAV. Under the circumstances, the transmissions of hubs to UAVs are scheduled using NOMA technique. This technique schedules multiple hubs in the same time slot by accepting a degree of interference between them and satisfying the rate requirements of all these hubs.
- Moreover, we optimize the 3D trajectory of UAVs by jointly considering the transmission rate, energy consumption, and delay. To achieve this goal, our proposed algorithm numerically solves a multi objective problem by Q-Learning method. Using this method, we train each UAV individually for finding the best 3D location where it can achieve the highest sum rate along with the least energy consumption and delay.
- Different priority and emergency levels of vital signs are other challenging issues in IoMT-WBANs. Because of the existing wide variety of chronic diseases, the vital signs of different patients have various delay sensitivity. Furthermore, in disaster situations, unexpected emergency conditions may occur to some patients where the vital signs of them need to be timely transmitted to the monitoring center. In this regard, our proposed algorithm takes the combined effect of data priority, patient priority, and emergency conditions into account in determining the total delay.

TABLE 1. List of notations used in this paper.

Notation	Definition
$\mathbb{U}$	Set of UAVs
$\mathbb{P}$	Set of patients
$\mathbb{C}$	Set of clusters
$\mathbb{S}$	Total state space
$\mathbb{A}$	Total action space
η	Path loss exponent
$\Gamma_{\mathcal{H}_i,u_j}$	Received SINR of hub $\mathcal{H}_i$ at UAV u_j
$\mathcal{R}_{\mathcal{H}_i,u_j}$	Data rate of $\mathcal{H}_i$ for transmitting data to UAV u_j
$\mathcal{E}_{\mathcal{H}_i,u_j}$	Energy consumption of $\mathcal{H}_i$ for transmitting data to UAV u_j
$\mathcal{I}_{em}^{\mathcal{H}_i}$	Emergency index of $\mathcal{H}_i$ for transmitting data to UAV u_j
λ	Learning rate
γ	Discount factor
$\Re(st, ac)$	Instantaneous reward at state st by taking action ac
$\mathcal{U}_{RED}$	Utility function
ep_i	The ith episode
N_0	AWGN power
$d_{\mathcal{H}_i,u_j}$	Euclidean distance between $\mathcal{H}_i$ and UAV u_j
$\mathscr{C}_{\mathcal{H}_i,u_j}$	The fading channel coefficient between $\mathcal{H}_i$ and UAV u_j

The rest of the paper is organized as follows. The preliminaries of our work is introduced in Section II including the system model, channel propagation model, and NOMA-based uplink transmission model. In Section III, we comprehensively express the procedure of our proposed clustering and Q-learning-based trajectory optimization algorithms. In Section V, the simulation results are shown to verify the performance of our proposed scheme. Finally, the results of our proposed algorithm are concluded in Section VI.

Notation: In this paper, scalars are denoted by italic letters. Boldface lower-case letters denote vectors. $\mathbb{R}^{M\times 1}$ denotes the space of M-dimensional real-valued vector. For a vector x, x^T denotes its transpose and $||x||$ represents the Euclidean norm. Table 1 summarizes the notations that will be used in this article.

II. HEALTH CARE SYSTEM MODEL

In this work, we consider a large-scale UAV-aided WBAN consisting of M UAVs, indexed by $\mathbb{U} = \{u_1, \ldots, u_M\}$, as flying BSs that form an adaptive multi-hop network to serve a set of $\mathbb{P} = \{p_1, \ldots, p_K\}$ patients randomly distributed in a large geographical area of size $(\mathcal{A} \times \mathcal{A})$ m^2. Each patient $p_i \in \mathbb{P}$ is equipped with the set of N bio-sensors, indexed by the set $\mathbb{B} = \{b_{i1}, \ldots, b_{iN}\}$, for sensing and transmitting the vital signals of patient's body and one IoMT device as a hub $\mathcal{H}_i$ to gather and transmit data packets of its corresponding

bio-sensors to the monitoring center. We suppose that patients can move in all directions independently with the velocity of $v^t_{\mathcal{H}_i}$, where t represents the current time slot. It should be noted that the velocity of patients is much lower than the velocity of UAVs. Throughout this paper, we occasionally use the term hub $\mathcal{H}_i$ instead of the corresponding patient p_i. It is assumed that each UAV covers a cluster of terrestrial hubs that satisfy the SINR threshold of the UAV. As it will be fully described in Section III, we classify all patients in the network area into M groups based on their locations and horizontal coordinations by proposing a modified version of one unsupervised machine learning-based method, namely the Fast Global K-Means (FGKM) algorithm. Let $\mathcal{CP}_{u_j}$ be the set of all hubs forming a cluster that is overlaid by UAV $u_j \in \mathbb{U}$ and $\mathbb{C} = \{\mathcal{CP}_{u_1}, \ldots, \mathcal{CP}_{u_M}\}$ be the set of all clusters in the network, where $\mathcal{CP}_{u_m} \cap \mathcal{CP}_{u_n} = \emptyset, \forall m \neq n$. In this situation, because of the mobility of patients in the network area, the location and the members of clusters are varied by time, thus, the topology of the backbone adaptively changes by flying the UAVs to the location of new clusters.

To get more insight into the aforementioned system model, Fig. 1 illustrates a practical IoMT-based WBAN architecture proposed by research communities for timely and reliable aggregating data packets of patients distributed in a smart city. As illustrated in this figure, at first, the vital signs are sensed by sticking different bio-sensors on the patients' bodies or using in-body implant devices, which communicate to a hub for further processing via a star topology. This hub can be an IoMT-based device like smart phones or smart watches embedded in the human body. In this layer, since the communication range is less than a meter, the path loss is just considered the loss at the referenced distance which is far and away less than the path loss of long-range communication model WLAN and cellular networks and it can be neglected. Thus, the channel propagation model of Tier I in the proposed network model is totally different from that of WLANs and cellular networks. Since patients are distributed in a smart city area, some of them, located in unreachable areas, cannot be covered by static access points. Therefore, they are not able to transmit their data packets to the monitoring center. This turns into a major when an unexpected emergency condition occurs to the patients in hot spots. UAVs have remarkable performance in terms of online and reliable tracking of the vital signs of patients in outdoor applications. After aggregating the vital signs at each hub, they are transmitted to the corresponding UAV or any Access Point (AP) in proximity. In this level, the city area is divided into different clusters and a single UAV covers the patients located in each cluster. This structure is really useful in unreachable areas where cannot be covered by stationary base stations. Note that the channel propagation model of Tier II is approximately similar to the channel distribution of UAV-assisted cellular networks. To be more precise, similar to cellular networks, Rician distribution has been employed to model the channel between transmitters and receivers, which comprises LoS, NLoS, and shadowing effect terms. However, LoS communication is the dominant term in contrast to WLANs where the transmissions mostly occur in indoor environments. Under the circumstances, thanks to existing walls, doors, and furniture in indoor areas, NLoS links are commonly conducted among transmitters and receivers in WLANs. Eventually, the UAVs transmit data packets to a remote cloud server where medical experts can access them, directly and track the vital signs of patients in a real-time manner. In Tier III of the proposed system model, the collected vital signs are forwarded to the cloud by UAVs and then the cloud server estimates the best path for each data packet in line with its emergency condition and bandwidth requirement. In this tier, transmitting data packets occur via wired links, where the attenuation is relevant to the thermal noise junctions, material, and electromagnetic field of the link which is completely different from the transmission links in WLANs and cellular networks.

A. BIO-SENSORS' SIGNALING MODEL

To enhance the spectral efficiency, we suppose that all bio-sensors of one patient's body use the same bandwidth to transmit their sensed vital signs to the corresponding hub. To mitigate the co-channel interference due to the simultaneous transmissions of different types of bio-sensors of a patient, we assign a code from the Walsh Hadamard (WH) code space to a bio-sensor that guarantees the use of orthogonal codes for transmission. To this end, $\boldsymbol{\mathcal{WH}}_e = [\boldsymbol{rw}_1, \cdots, \boldsymbol{rw}_{k+1}]^T \in \mathbb{Z}^{(k+1)\times 2^k}$ is employed as the matrix of the pre-specified rows extracted from the original WH matrix:

$$\mathbf{WH}_{2^k} = \begin{bmatrix} \mathbf{WH}_{2^{k-1}} & \mathbf{WH}_{2^{k-1}} \\ \mathbf{WH}_{2^{k-1}} & -\mathbf{WH}_{2^{k-1}} \end{bmatrix}, \quad k = 1, 2, \ldots \tag{1}$$

with $\mathbf{WH}_1 = [1]$. It is shown in [21] that the rows of $\boldsymbol{\mathcal{WH}}_e$ represent the codewords which are two-by-two orthogonal in every phase shift $\psi = 0, \cdots, 2^k - 1$. The orthogonality of each pair of rows $\boldsymbol{rw}_i = [c_{i1}, \cdots, c_{i2^k}]$ and $\boldsymbol{rw}_j = [c_{j1}, \cdots, c_{j2^k}]$, $i \neq j$, is calculated by the following cross-correlation expression in the phase shift ψ:

$$\Phi_{\boldsymbol{rw}_i, \boldsymbol{rw}_j}(\psi) = \sum_{\ell=1}^{2^k} c^{\psi}_{i\ell} \times c^{\psi}_{j\ell} = 0, \quad \forall \psi = 0, \cdots, 2^k - 1. \tag{2}$$

Taking this property into account, by the product of signal of each bio-sensor b_{in} to one of the rows $\boldsymbol{rw}_i$, we can guarantee the collision-free transmissions of different signals from bio-sensors in one patient, i.e.,

$$\boldsymbol{rw}^{\psi}_i \perp \boldsymbol{rw}^{\psi'}_j \Rightarrow \boldsymbol{s}_{b_{in}}.\boldsymbol{rw}^{\psi}_i \perp \boldsymbol{s}_{b_{im}}.\boldsymbol{rw}^{\psi'}_j, \quad n \neq m, \tag{3}$$

for all $\psi, \psi' = 0, \cdots, 2^k - 1$, where $\boldsymbol{s}_{b_{in}}$ and $\boldsymbol{s}_{b_{im}}$ are the signal vectors of bio-sensors b_{in} and b_{im}, respectively.

B. UAV CHANNEL PROPAGATION MODEL

In order to model the channel characteristics between each UAV $u_j \in \mathbb{U}$ and a terrestrial hub $\mathcal{H}_i$ belonging to cluster

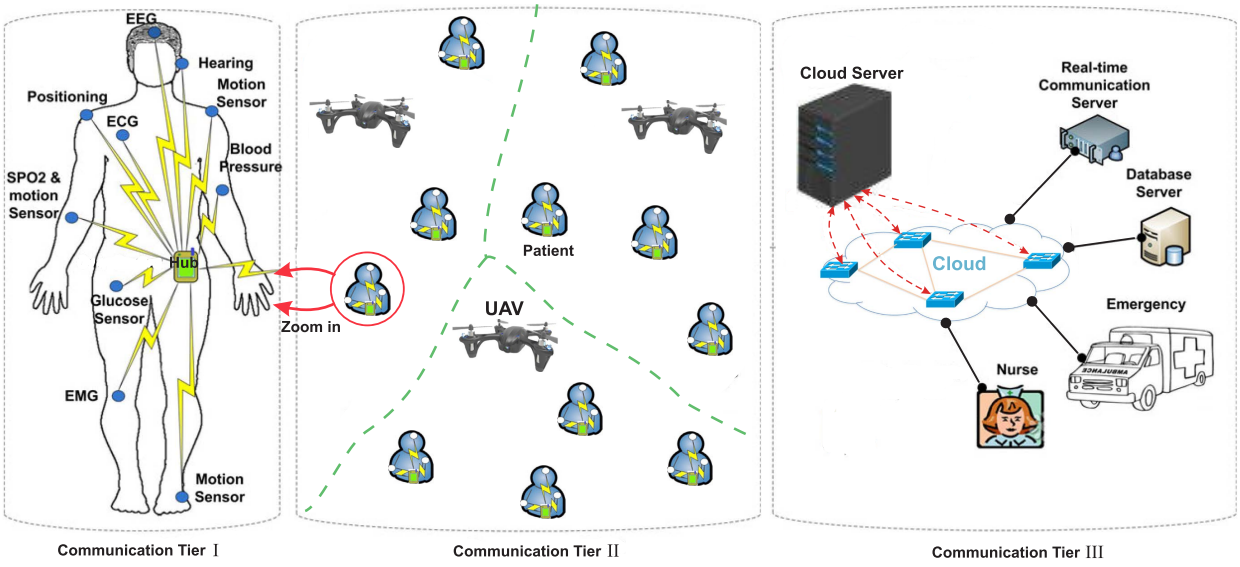


FIGURE 1. Typical IoMT-based WBAN architecture equipped with UAV technology.

$\mathcal{CP}_{u_j}$, we compute the 3D Euclidean distance between u_j and $\mathcal{H}_i$. In this regard, the positions of hub $\mathcal{H}_i$ and UAV u_j are represented by $\mathbf{p}_{\mathcal{H}_i} = \left[x_{\mathcal{H}_i}, y_{\mathcal{H}_i}\right]^T \in \mathbb{R}^{2\times 1}$ and $\mathbf{p}_{u_j} = \left[x_{u_j}, y_{u_j}, a_{u_j}\right]^T \in \mathbb{R}^{3\times 1}$, respectively, where (x_{u_j}, y_{u_j}) denotes the coordinates of UAV u_j on the horizontal plane and a_{u_j} indicates its altitude. Accordingly, the 3D distance between hub $\mathcal{H}_i$ and UAV u_j is calculated by:

$$d_{\mathcal{H}_i,u_j} = \sqrt{(x_{\mathcal{H}_i} - x_{u_j})^2 + (y_{\mathcal{H}_i} - y_{u_j})^2 + (a_{u_j})^2}. \tag{4}$$

We also consider a well-known Air-To-Ground (ATG) channel model [22], in which there are two main propagation groups. Depending on the altitude of UAVs and the elevation angle between the hub $\mathcal{H}_i$ and UAV u_j, denoted by $\theta_{\mathcal{H}_i,u_j}$, these groups are categorized by Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) communication links. In this regard, the probability of LoS communication between UAV u_j at the altitude a_{u_j} and hub $\mathcal{H}_i$ with the Euclidean distance between them $d_{\mathcal{H}_i,u_j}$, is formulated by the sigmoid function as follows:

$$P^{LoS}_{\mathcal{H}_i,u_j} = \frac{1}{1 + \alpha \exp\left(\beta\left(\frac{180}{\pi}\theta_{\mathcal{H}_i,u_j} - \alpha\right)\right)}, \tag{5}$$

where α and β are the environment constants that represent the ratio of built-up area to the total land area multiplied by the mean number of building per unit area, and buildings' height distribution, respectively. Furthermore, the elevation angle $\theta_{\mathcal{H}_i,u_j}$ is defined as

$$\theta_{\mathcal{H}_i,u_j} = \tan^{-1}\left(\frac{a_{u_j}}{r_{\mathcal{H}_i,u_j}}\right), \tag{6}$$

where $r_{\mathcal{H}_i,u_j} = \sqrt{\left(x_{\mathcal{H}_i} - x_{u_j}\right)^2 + \left(y_{\mathcal{H}_i} - y_{u_j}\right)^2}$. On the other hand, the probability of NLoS propagation communication is obtained as

$$P^{NLoS}_{\mathcal{H}_i,u_j} = 1 - P^{LoS}_{\mathcal{H}_i,u_j}. \tag{7}$$

The collected vital signals by the hubs propagate through the network area where they experience shadowing and scattering phenomena imposed by the sky scrapers and huge obstacles, then, they enter the free space and reach the corresponding UAVs. The shadowing effect, represented by $\mathcal{X}_{\mathcal{H}_i,u_j}$, imposes excessive loss to the ATG link which has a Gaussian distribution. Under the circumstances, the LoS and NLoS path loss expressions for given hub $\mathcal{H}_i$ and UAV u_j are determined as

$$\mathcal{L}^{LoS}_{\mathcal{H}_i,u_j}\,[dB] = \mathcal{L}_0 + 10\eta^{LoS}\log\left(d_{\mathcal{H}_i,u_j}\right) + \mathcal{X}^{LoS}_{\mathcal{H}_i,u_j}, \tag{8}$$

$$\mathcal{L}^{NLoS}_{\mathcal{H}_i,u_j}\,[dB] = \mathcal{L}_0 + 10\eta^{NLoS}\log\left(d_{\mathcal{H}_i,u_j}\right) + \mathcal{X}^{NLoS}_{\mathcal{H}_i,u_j}, \tag{9}$$

where $\mathcal{L}_0 = 20\log\left(\frac{4\pi f_c d_0}{c}\right)$ represents the path loss at a reference distance d_0, f_c, and c denote the carrier frequency and the speed of light, respectively. Moreover, η^{LoS} and η^{NLoS} are the path loss exponents for LoS and NLoS links. Based on the aforementioned definitions, the average path loss between hub $\mathcal{H}_i$ and UAV u_j can be calculated as follows:

$$\bar{\mathcal{L}}_{\mathcal{H}_i,u_j} = P^{LoS}_{\mathcal{H}_i,u_j}\mathcal{L}^{LoS}_{\mathcal{H}_i,u_j} + P^{NLoS}_{\mathcal{H}_i,u_j}\mathcal{L}^{NLoS}_{\mathcal{H}_i,u_j}. \tag{10}$$

In addition, the small scale channel fading coefficient for transmitting a symbol from hub $\mathcal{H}_i$ to UAV u_j, denoted by $\tilde{\mathscr{C}}_{\mathcal{H}_i,u_j}$, is represented by a complex Gaussian random variable with the non-zero expected value and variance σ^2. Taking the above definitions into account, the instantaneous channel coefficient between hub $\mathcal{H}_i$ and UAV u_j is represented as

$$\mathscr{C}_{\mathcal{H}_i,u_j} = \frac{\tilde{\mathscr{C}}_{\mathcal{H}_i,u_j}}{\sqrt{\bar{\mathcal{L}}_{\mathcal{H}_i,u_j}}}. \tag{11}$$

Thus, the average channel gain between hub $\mathcal{H}_i$ and UAV u_j, given $\mathbb{E}\left[\left|\tilde{\mathscr{C}}_{\mathcal{H}_i,u_j}\right|^2\right]=1$, is computed as

$$\bar{g}_{\mathcal{H}_i,u_j}=\mathbb{E}\left[\left|\frac{\tilde{\mathscr{C}}_{\mathcal{H}_i,u_j}}{\sqrt{\bar{\mathcal{L}}_{\mathcal{H}_i,u_j}}}\right|^2\right]=\frac{\mathbb{E}\left[\left|\tilde{\mathscr{C}}_{\mathcal{H}_i,u_j}\right|^2\right]}{\bar{\mathcal{L}}_{\mathcal{H}_i,u_j}}=\left(\bar{\mathcal{L}}_{\mathcal{H}_i,u_j}\right)^{-1}. \tag{12}$$

C. NOMA-BASED UPLINK TRANSMISSION MODEL

As previously mentioned, one of the responsibilities of hub $\mathcal{H}_i$ is to transmit the collected vital signs of patient p_i to the corresponding UAV u_j. In the situation of partitioning hubs into different clusters, all the overlaid hubs by UAV u_j, i.e., $\forall\mathcal{H}_i\in\mathcal{CP}_{u_j}$, employ the NOMA signaling technique to simultaneously transmit their data packets to UAV u_j. In fact, the signals of different $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ are superposed with different transmission powers and then those are sent to UAV u_j. The transmission power of each hub $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ is a fraction of the total transmission power $\mathscr{P}_{total}$ supported by UAV u_j. Thus, the average channel gain is computed for all the members of $\mathcal{CP}_{u_j}$ and then the members are indexed in descending order, such as:

$$\bar{g}_{\mathcal{H}_1,u_j}\geq\cdots\geq\bar{g}_{\mathcal{H}_{|\mathcal{CP}_{u_j}|},u_j},\quad\forall\mathcal{H}_i\in\mathcal{CP}_{u_j}, \tag{13}$$

which implies that the first and the last hubs in cluster $\mathcal{CP}_{u_j}$ have the strongest and weakest condition, respectively. Based on the criterion in (13), the NOMA technique assigns a fraction of $\mathscr{P}_{total}$ to each $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ by employing power coefficients $\zeta_1\leq\cdots\leq\zeta_{|\mathcal{CP}_{u_j}|}$, where $\sum_{i=1}^{|\mathcal{CP}_{u_j}|}\zeta_i=1$. For more clarification, the maximum coefficient is assigned to the hub with the weakest channel condition to form its power $\zeta_{|\mathcal{CP}_{u_j}|}\mathscr{P}_{total}=\mathscr{P}_{\mathcal{H}_{|\mathcal{CP}_{u_j}|}}$, while the minimum coefficient is allocated to the hub with the strongest channel condition. Accordingly, the transmission power of all $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ is arranged in the ascending order as follows:

$$\mathscr{P}_{\mathcal{H}_1}\leq\cdots\leq\mathscr{P}_{\mathcal{H}_{|\mathcal{CP}_{u_j}|}},\quad\forall\mathcal{H}_i\in\mathcal{CP}_{u_j}. \tag{14}$$

Based on the above power allocation technique, the transmitted signal by each $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ is expressed as follows:

$$x_{\mathcal{H}_i}=\sqrt{\mathscr{P}_{\mathcal{H}_i}}s_{\mathcal{H}_i}=\sqrt{\zeta_{|\mathcal{CP}_{u_j}|}\mathscr{P}_{total}}s_{\mathcal{H}_i}, \tag{15}$$

where $s_{\mathcal{H}_i}$ indicates the information of hub $\mathcal{H}_i$ with $\mathbb{E}\left[\left|s_{\mathcal{H}_i}\right|^2\right]=1$. In addition, the received signal at UAV u_j can be expressed as

$$r_{u_j}=\sum_{i=1}^{|\mathcal{CP}_{u_j}|}\mathscr{C}_{\mathcal{H}_i,u_j}x_{\mathcal{H}_i}+n_{u_j}=\sum_{i=1}^{|\mathcal{CP}_{u_j}|}\sqrt{\frac{\zeta_i\mathscr{P}_{total}}{\mathcal{L}_{\mathcal{H}_i,u_j}}}\bar{\mathscr{C}}_{\mathcal{H}_i,u_j}s_{\mathcal{H}_i}+n_{u_j}, \tag{16}$$

where n_{u_j} denotes the zero-mean additive white Gaussian noise with power N_0 at UAV u_j. On the other hand, the NOMA scheduling technique employs the SIC scheme at the receiver UAV u_j to detect the massage signal of each $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ without co-channel interference. To this end, the SIC scheme detects the signal of a hub with the strongest channel condition and treats the signals of other poorer hubs as interference. Subsequently, this signal is subtracted from the total signal. This procedure is repeated until the signals of all $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ are detected separately. Under this circumstance, the received SINR of hub $\mathcal{H}_i$ at UAV u_j, represented by $\gamma_{\mathcal{H}_i,u_j}$, is obtained as

$$\Gamma_{\mathcal{H}_i,u_j}=\frac{\zeta_i\mathscr{P}_{total}\bar{g}_{\mathcal{H}_i,u_j}}{\sum_{k=i+1}^{|\mathcal{CP}_{u_j}|}\zeta_k\mathscr{P}_{total}\bar{g}_{\mathcal{H}_k,u_j}+N_0},\quad\forall\mathcal{H}_i\in\mathcal{CP}_{u_j}. \tag{17}$$

The data rate of $\mathcal{H}_i\in\mathcal{CP}_{u_j}$ is computed as follows:

$$\mathcal{R}_{\mathcal{H}_i,u_j}=W\log_2\left(1+\Gamma_{\mathcal{H}_i,u_j}\right), \tag{18}$$

where W is the total bandwidth. In order to satisfy the reliability of the received information, the transmitted SINR of hub $\mathcal{H}_i$ at UAV u_i should satisfy the following constraint:

$$\Gamma_{\mathcal{H}_i,u_j}\geq\Gamma_{th}, \tag{19}$$

where Γ_{th} is the minimum SINR of hub $\mathcal{H}_i$ that is required to achieve a satisfactory Bit Error Rate (BER) for all UAVs.

D. ENERGY CONSUMPTION AND DELAY MODEL

The total energy consumption of each hub $\mathcal{H}_i$ for transmitting L_b data bits to UAV u_i consists of two following components:

$$\mathcal{E}_{\mathcal{H}_i,u_j}=\mathcal{E}^{elec}_{\mathcal{H}_i}+\mathcal{E}^{Tx}_{\mathcal{H}_i,u_j}, \tag{20}$$

where $\mathcal{E}^{elec}_{\mathcal{H}_i}$ and $\mathcal{E}^{Tx}_{\mathcal{H}_i,u_j}$ represent the energy consumption of electronic circuits of hub $\mathcal{H}_i$ and the energy consumption of transmitting data packets on the channel, respectively. The amount of $\mathcal{E}^{Tx}_{\mathcal{H}_i,u_j}$ depends on the channel condition, distance, transmission power, and length of packets, and it can be calculated by:

$$\mathcal{E}^{Tx}_{\mathcal{H}_i,u_j}=\frac{\mathscr{P}_{\mathcal{H}_i}L_b}{\mathcal{R}_{\mathcal{H}_i,u_j}}=\frac{\zeta_i\mathscr{P}_{total}L_b}{W}\log_2^{-1}\left(1+\Gamma_{\mathcal{H}_i,u_j}\right). \tag{21}$$

Since the altitude of each UAV is optimized to maximize the throughput of the network, the distance between each UAV and the corresponding hubs fluctuates between short and long ranges. For the case when the distance between hubs and UAVs is sufficiently large, $\mathcal{E}^{elec}_{\mathcal{H}_i}$ is considerably smaller than $\mathcal{E}^{Tx}_{\mathcal{H}_i,u_j}$, thus, we ignore this term in determining $\mathcal{E}_{\mathcal{H}_i,u_j}$. Moreover, since UAVs are battery-powered with limited energy capacity, thus by reducing the remaining energy of UAVs below a specified value, the vital signs will not be forwarded in time. In this regard, parameter $\mathcal{E}_{th}$ is defined as the remaining energy threshold of UAVs. In our proposed

algorithm, the remaining energy of all UAVs must be higher than $\mathcal{E}_{th}$ to satisfy the real-time and reliable transmission of vital signs.

The total experienced delay by data packets of hub $\mathcal{H}_i$ until receiving by UAV u_j is obtained by the sum of two terms as follows:

$$\mathcal{D}_{\mathcal{H}_i,u_j} = \mathcal{D}^{acc}_{\mathcal{H}_i} + \mathcal{D}^{Tx}_{\mathcal{H}_i,u_j}, \tag{22}$$

where $\mathcal{D}^{acc}_{\mathcal{H}_i}$ and $\mathcal{D}^{Tx}_{\mathcal{H}_i,u_j}$ denote the requirement time for hub $\mathcal{H}_i$ to access the channel and the time of transmitting one data packet with the size L_b on the channel, respectively. In this regard, $\mathcal{D}^{Tx}_{\mathcal{H}_i,u_j}$ is obtained as

$$\mathcal{D}^{Tx}_{\mathcal{H}_i,u_j} = \frac{L_b}{\mathcal{R}_{\mathcal{H}_i,u_j}} = \frac{L_b}{W}\log_2^{-1}\left(1+\Gamma_{\mathcal{H}_i,u_j}\right). \tag{23}$$

E. 3D TRAJECTORY MODEL OF UAV

We assume that the initial and final locations of each UAV u_j are represented by $\mathbf{p}^{(0)}_{u_j} = \left[x^{(0)}_{u_j}, y^{(0)}_{u_j}, a^{(0)}_{u_j}\right]^T$ and $\mathbf{p}^{(F)}_{u_j} = \left[x^{(F)}_{u_j}, y^{(F)}_{u_j}, a^{(F)}_{u_j}\right]^T$, respectively. In order to show the time-spatial changes of UAVs, we divide the time horizon T into multiple equal-length time slots represented by the set $\mathbb{T} = \{t_0, \cdots, t_Z\}$. The length of time slots is obtained by $\delta = \frac{T}{Z}$. The maximum horizontal (in $x-y$ plane) and vertical speeds of each UAV u_j are indicated by $v^{u_j}_{xy}$ and $v^{u_j}_{a}$ where it is supposed each UAV can independently control its horizontal and vertical speeds [23]. Accordingly, the maximum horizontal and vertical distances UAV u_j spans during each time slot are calculated by $\mathcal{S}^{u_j}_{xy} = v^{u_j}_{xy}\delta$ and $\mathcal{S}^{u_j}_{a} = v^{u_j}_{a}\delta$, respectively. In this regard, considering $\mathbf{q}_{u_j} = \left[x_{u_j}, y_{u_j}\right]^T$ as the horizontal coordinates of UAV u_j, the distance between two consecutive coordinates of UAV u_j should satisfy the following constraints:

$$\left|\left|\mathbf{q}^{(t_{\ell+1})}_{u_j} - \mathbf{q}^{(t_\ell)}_{u_j}\right|\right| \le \mathcal{S}^{u_j}_{xy}, \quad \left|a^{(t_{\ell+1})}_{u_j} - a^{(t_\ell)}_{u_j}\right| \le \mathcal{S}^{u_j}_{a}. \tag{24}$$

Moreover, to avoid the obstacles like buildings, each UAV should fly above the minimum altitude denoted by a_{min}, i.e., $a_{u_j} \ge a_{min}$.

F. PROBLEM FORMULATION

Based on the aforementioned definitions, the main objective of our proposed algorithm is to jointly maximize the total effective throughput and minimize the total energy consumption and total delay in each time slot. To this end, these factors are defined as the sum of rates, the sum of energy consumption, and the sum of delay of all hubs belonging to cluster $\mathcal{CP}_{u_j}$ which allows connecting to the UAV u_j, respectively. These factors are determined by:

$$\mathcal{R}^{(t_\ell)}_{u_j} = \sum_{\forall \mathcal{H}_i \in \mathbb{H}^{(t_\ell)}} \mathcal{R}_{\mathcal{H}_i,u_j}, \tag{25}$$

$$\mathcal{E}^{(t_\ell)}_{u_j} = \sum_{\forall \mathcal{H}_i \in \mathbb{H}^{(t_\ell)}} \mathcal{E}_{\mathcal{H}_i,u_j}, \tag{26}$$

$$\mathcal{D}^{(t_\ell)}_{u_j} = \sum_{\forall \mathcal{H}_i \in \mathbb{H}^{(t_\ell)}} \mathcal{D}_{\mathcal{H}_i,u_j}, \tag{27}$$

where $\mathbb{H}^{(t_\ell)}$ is the set of all $\mathcal{H}_i$s connected to the UAV u_j in time slot t_ℓ based on the real time NOMA scheduling technique. Accordingly, we aim to optimize these three metrics, simultaneously, as mathematically expressed below:

$$\max_{\left\{\mathbf{q}^{(\ell)}_{u_j}\right\}} \left\{\mathcal{R}^{(t_\ell)}_{u_j}, \frac{1}{\mathcal{E}^{(t_\ell)}_{u_j}}, \frac{1}{\mathcal{D}^{(t_\ell)}_{u_j}}\right\}. \tag{28}$$

It is worth mentioning that there exists a correlation among sum rates, sum energy consumption, and sum delays as the main optimization objectives in the proposed IoMT-based WBAN. Thus, if these objectives are separately optimized, achieving the local optimum value of one variable leads to inefficient values of the others. Accordingly, there is a tradeoff between optimizing rate and delay values along with a tradeoff between delay and energy consumption. For more clarification, consider a situation in which an unexpected emergency condition would occur to a patient, but the corresponding communications suffer from poor channel conditions. In this case, the network experiences a delay-sensitive situation, but at the cost of a lower transmission rate (according to the Shannon capacity formula). Under the circumstances, separately optimizing the delay value leads to an inefficient transmission rate and vice versa. For another example, consider two different transmissions to be scheduled in the current time slot, and assuming that the first transmission has higher delay sensitivity and energy consumption than the second one. In this situation, if the delay is optimized locally, the first transmission is selected to access the channel, in contrast, if the energy consumption value is only considered in the optimization problem, the second transmission is chosen by that metric. Moreover, based on (21), there is a reverse relationship between energy consumption and transmission rate. Consequently, by selecting the transmissions which have the highest data rate in each time slot, the energy consumption will be decreased to the minimum value. Taking the above considerations into account, our objective is simultaneously optimizing all of these factors to design an energy-efficient scheduling algorithm for timely and reliably transmitting the vital signs to the monitoring center. As will be shown in subsection III-A, to fairly combine the impacts of these metrics in the aforementioned multi-objective optimization problem, the computed value of each metric would be normalized to its maximum value. Using this method, the values of all metrics are mapped to the interval of [0, 1].

III. THE PROPOSED Q-REDTO ALGORITHM

In this section, we propose a two-layer scheduling algorithm for optimizing the trajectory of UAVs, namely Q-Learning-based Rate- Energy- and Delay-aware Trajectory Optimizer (Q-REDTO). The first layer is related to scheduling the transmission of vital signs sensed by bio-sensors of one patient to the corresponding hub, while the second layer schedules the transmission of collected signals by each hub to the

monitoring center via UAVs. The procedure of our proposed Q-REDTO algorithm is explained in the following four stages:

Stage 1 (Interference Avoidance Scheduling in Tier I): In the first stage, the vital signals of different bio-sensors are timely scheduled to transmit to the corresponding hub. Since all bio-sensors of a typical patient share the same bandwidth along with their asynchronous duty cycling mechanism, the collision is not avoidable between transmissions of these bio-sensors, without using any scheduling technique. To solve this problem, we employ the WH coding scheme in which sensed data of all bio-sensors are simultaneously delivered to the hub by multiplying them to the cyclic orthogonal WH codes extracted from (1). This process mitigates the interference between concurrent transmissions because of the orthogonality of these codes in every phase shift. For more clarification, consider $\boldsymbol{s}_{b_{in}} = \left[s_{b_{in}}(1), \cdots, s_{b_{in}}(X)\right]^T$ and $\boldsymbol{s}_{b_{im}} = \left[s_{b_{im}}(1), \cdots, s_{b_{im}}(Y)\right]^T$ as the vital signs of bio-sensors b_{in} and b_{im}, $n \neq m$. As described in the previous section, since the cross-correlation of two different cyclic orthogonal codes $\boldsymbol{rw}_i^{\psi}$ and $\boldsymbol{rw}_j^{\psi'}$, $i \neq j$, is equal to zero, therefore, it can be proved that by multiplying $\boldsymbol{s}_{b_{in}}$ to $\boldsymbol{rw}_i^{\psi}$ and $\boldsymbol{s}_{b_{im}}$ to $\boldsymbol{rw}_j^{\psi'}$, and denoting $\tilde{\boldsymbol{s}}_{b_{in}} = \boldsymbol{s}_{b_{in}}.\boldsymbol{rw}_i^{\psi}$ and $\tilde{\boldsymbol{s}}_{b_{im}} = \boldsymbol{s}_{b_{im}}.\boldsymbol{rw}_j^{\psi'}$, the cross-correlation of these two signals will be zero, i.e.,

$$\begin{aligned}\Phi_{\tilde{\boldsymbol{s}}_{b_{in}},\tilde{\boldsymbol{s}}_{b_{im}}}\left(\psi,\psi'\right) = \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{\kappa=1}^{2^k} s_{b_{in}}(x)c_{i\kappa}^{\psi}s_{b_{im}}(y)c_{j\kappa}^{\psi'} = 0,\\ \forall\psi,\psi' = 0,\cdots,2^k-1,\\ n\neq m,\ i\neq j.\end{aligned} \tag{29}$$

To highlight the benefit of the above interference avoidance scheduling in practical WBAN, let consider four different medical bio-sensors consisting of two MAX30100 chips used for ECG and SpO2 signals, the MPS20N0040D sensor for sensing the systolic blood pressure, MAX30205MAT employed for measuring body temperature, and the X2M200 chip exploited for sensing respiration rate. The different sampling frequencies of these bio-sensors are shown in Fig. 2. As it can be realized from the figure, because of the asynchronous sampling frequencies of bio-sensors, interference is inevitable between the transmissions. To avoid the interference, the cyclic orthogonal WH coding scheme is employed which multiplies the vital signs to a set of codes that are orthogonal in every phase shift. Besides, this simultaneous transmission of vital signs results in reducing the transmission delay of the proposed algorithm which satisfies the real-time monitoring of medical in delay-sensitive WBAN applications.

Stage 2 (Patients Clustering Method): After transmitting vital signs of bio-sensors to their corresponding hubs, the hubs communicate to UAVs to deliver data packets to the monitoring center. As depicted in Fig. 3, because of employing multiple UAVs in the network, we classify the hubs into different clusters using an unsupervised machine learning-based scheme namely Fast Global K-Means (FGKM) [24]. This algorithm establishes an incremental deterministic global optimization method that optimally adds one new cluster at each step until convergence. Assuming a set of hubs $\mathcal{H}_i \in \mathbb{P}$ distributed in the geographical area, the FGKM algorithm partitions these hubs into M disjoint clusters to optimize a specific criterion. In the first iteration, the hub that optimizes the criterion is selected as the optimal cluster center cc_1 for the $m = 1$ clustering problem. Consequently, in the second iteration, i.e., $m = 2$, one new optimal cluster center cc_2 is added by assuming that the previous cc_1 is the first optimal cluster center in the current iteration. This procedure is repeated until the objective function converges to its minimum value. We define this objective function as the Mean Square Error (MSE) of the summation of distances between each $\mathcal{H}_i$ and the corresponding cluster center cc_m, i.e.,

$$\epsilon_M(cc_1,\cdots,cc_M) = \frac{1}{K}\sum_{i=1}^{K}\min_{m=1,\cdots,M}||cc_m - \mathcal{H}_i||^2. \tag{30}$$

In each iteration, the FGKM algorithm adds the cluster that minimizes the upper bound of MSE in the current iteration, i.e., $\epsilon_{m-1} - \Omega_m$, where ϵ_{m-1} denotes the upper bound of MSE in the previous iteration and

$$\Omega_m = \arg\max_{\mathcal{H}_i'\in\mathbb{P}} \Omega_{\mathcal{H}_i'}, \tag{31}$$

in which,

$$\Omega_{\mathcal{H}_i'} = \frac{1}{K}\sum_{i=1}^{K}\max\left(||cc_{m-1} - \mathcal{H}_i||^2 - \left|\left|\mathcal{H}_i' - \mathcal{H}_i\right|\right|^2, 0\right). \tag{32}$$

For the above equations, $\mathcal{H}_i'$ denotes the candidate hub for being the new cluster center, and $||cc_{m-1} - \mathcal{H}_i||^2$ represents the squared distance between hub $\mathcal{H}_i$ and its previous closest cluster center cc_{m-1}. Based on (32), if the squared distance between hub $\mathcal{H}_i$ and $\mathcal{H}_i'$ in each iteration is smaller than the distance of $\mathcal{H}_i$ to cc_{m-1}, it is added to the set of cluster members of $\mathcal{H}_i'$. Thus, regarding (31), this metric selects the hub that has the most number of other hubs in its proximity and its adjacent hubs are in the furthest location from $cc_1,\cdots,cc_{m-1}$, as the new cluster center cc_m. In this situation, it can be easily shown that $\epsilon_m \leq \epsilon_{m-1} - \Omega_m$. Accordingly, this metric reduces the computational complexity of the FGKM scheme and speeds up the convergence of the objective function. We summarize the aforementioned procedure of the FGKM scheme as pseudocode in Algorithm 1.

Stage 3 (Q-Learning-Based Solution for Trajectory Optimization and Vital Signals Scheduling): After partitioning the area into different clusters, all hubs belonging to the same cluster are served by the same UAV. The optimal number of UAVs is determined by the FGKM algorithm. By dividing

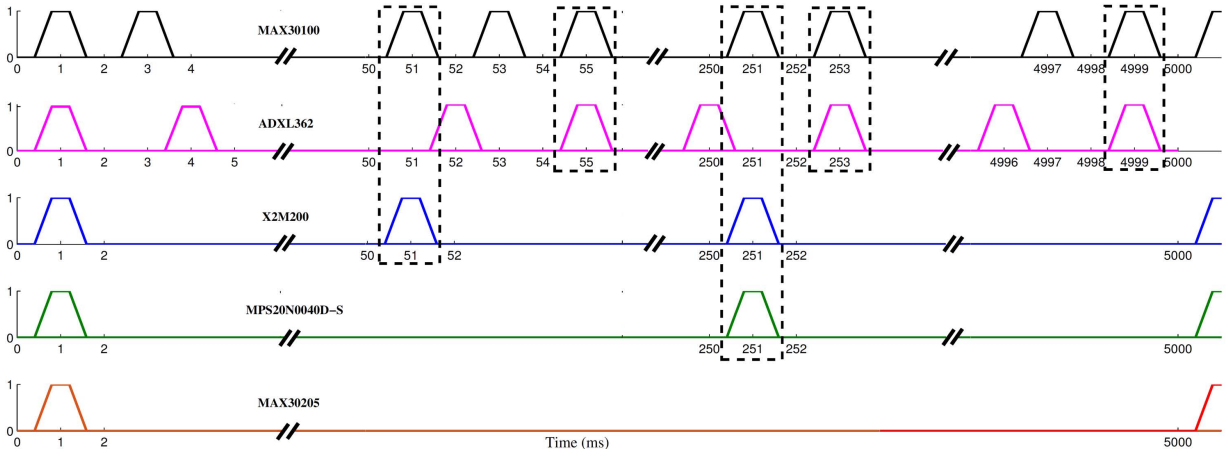


FIGURE 2. Typical duty cycle of four practical medical bio-sensors MAX30100 chip, MPS20N0040D sensor, MAX30205MAT, and X2M200 chip.

Algorithm 1 Pseudocode of the Proposed FGKM Algorithm

Input: Positions of all $\mathcal{H}_i \in \mathbb{P}$ and number of UAVs M.
Output: $\{cc_1, \cdots, cc_m\}$ and $\mathbb{C} = \{\mathcal{CP}_{u_1}, \ldots, \mathcal{CP}_{u_M}\}$.
for $m = 1, \cdots, M$ **do**
  if $m == 1$ **then**
    for each $\mathcal{H}'_i \in \mathbb{P}$ **do**
      Calculate ϵ_1 from (30).
    end for
    $cc_1 \leftarrow \mathcal{H}'_i \Leftrightarrow \arg\min \epsilon_1$ over all $\mathcal{H}'_i \in \mathbb{P}$.
  else
    Suppose $cc_1, \cdots, cc_{m-1}$ as the existed optimal cluster centers.
    for each $\mathcal{H}'_i \in \mathbb{P}$ **do**
      Calculate $\Omega_{\mathcal{H}'_i}$ according to (32).
    end for
    Compute Ω_m using (31)
    $cc_m \leftarrow \mathcal{H}'_i \Leftrightarrow \arg\min(\epsilon_{m-1} - \Omega_m)$ over $\mathcal{H}'_i \in \mathbb{P}$.
  end if
end for

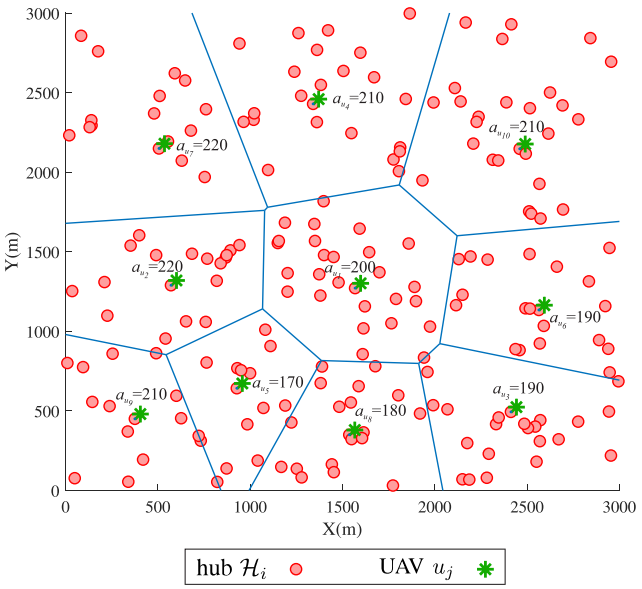


FIGURE 3. A typical city area partitioned into 10 clusters using FGKM where the green star shows the UAV with the optimum value of altitude serving patients in its corresponding cluster.

the total time frame into multiple time slots, we design the Q-REDTO algorithm in such a way that the best vertical and horizontal position of each UAV is determined in each time slot. The objective function given in subsection II-F, which refers to the best position in which each UAV can achieve the highest sum rate along with the least energy consumption and delay. The Q-REDTO algorithm employs the QL-based framework to solve the multi-objective trajectory optimization problem of UAVs to aggregate data packets of their corresponding hubs in each cluster. The Q-REDTO algorithm consists of four core elements, i.e., states, actions, reward, and Q-values which are described as follows:

• **States:** Since the WBAN environment is generally a city area, obviously, we encounter infinite states in our 3D trajectory optimization problem which is not tractable because of infinite decision space. In order to map the continuous environment with a finite number of states, we discretize the WBAN environment into the set of equal-space tiles denoted by $\mathbb{S} = \{st_1, \cdots, st_I\}$. These tiles represent the available states of the environment in which UAVs can only fly between the central points of tiles.

• **Actions:** The procedure of the state transition of UAVs is implemented by taking different actions. In this work, all directions in which UAVs can fly indicate the action space. Based on the aforementioned discrete state-space model, our action space is limited to eight directions. In other words, each state is surrounded by eight other states that each UAV can fly to. In this situation, the set $\mathbb{A} = \{$"*north*","*east*","*south*","*west*", "*north-east*", "*north-west*","*south-east*","*south-west*"$\}$ denotes the action space of a typical UAV.

• **Rewards:** The consequence of each action is defined as a reward. By selecting an action from the set $\mathbb{A}$ and transition to the new state, each UAV can observe a new part of its cluster environment and can achieve a new amount of sum rate, energy consumption, and delay requirements. The combination of these factors is considered as a reward gained by UAV for observing the new state.

• **Q-values:** These values are calculated from the following Q-function that is responsible for the convergence of our proposed Q-REDTO algorithm:

$$Q_{new}(st, ac) \longleftarrow (1-\lambda)Q_{old}(st, ac) + \lambda\left(\Re(st, ac) + \gamma \max_{ac' \in \mathbb{A}} Q_{old}(st', ac')\right), \tag{33}$$

where λ, γ, and st' denote the learning rate, discount factor, and the new observing state, respectively. Moreover, $\Re(st, ac)$ and $Q_{old}(st', ac')$ represent the reward value of the current state-action policy and the expected reward of the new state-action policy. Each UAV keeps Q-values of all possible state-actions in a Q-table in which its rows and columns stand for actions and states, respectively. By observing the new state, the Q-value of the current state is updated using (33).

A. UTILITY MAXIMIZATION

After describing the preliminaries of the QL-based method, the procedure of the proposed Q-REDTO algorithm is presented to find the best position of UAV in each time slot. As mentioned before, the Q-REDTO scheme aims at optimizing both vertical and horizontal trajectories of UAVs to gain the best performance in terms of increasing the total sum rate and reducing the energy consumption and delay. To this end, we propose the following utility function to simultaneously consider the effects of the aforementioned factors:

$$\mathcal{U}_{RED} = \omega_r \mathcal{R}_{ns} - \omega_e \mathcal{E}_{ns} + \omega_i \mathcal{I}_{ns}, \tag{34}$$

where $\mathcal{R}_{ns}$, $\mathcal{E}_{ns}$, and $\mathcal{I}_{ns}$ represent the normalized sum rate, sum of the energy consumption, and sum of the emergency index of all hubs covered by each UAV in the specific position, respectively. In addition, ω_r, ω_e, and ω_i denote the corresponding weighting coefficients. To this end, the proposed optimization problem in (28) is converted to the following optimization problem:

$$\mathcal{P}\mathbf{1}) \quad \max_{\left\{\mathbf{q}_{u_j}^{(\ell)}\right\}} \ \mathcal{U}_{RED} \tag{35}$$

$$\text{s.t. } \mathbf{C1}: \ \mathbf{q}_{u_j}^{(0)} = \left[x_{u_j}^{(0)}, y_{u_j}^{(0)}\right]^T = \mathbf{q}_{u_j}^{(t_0)}, \tag{36}$$

$$\mathbf{C2}: \ \mathbf{q}_{u_j}^{(F)} = \left[x_{u_j}^{(F)}, y_{u_j}^{(F)}\right]^T = \mathbf{q}_{u_j}^{(t_Z)}, \tag{37}$$

$$\mathbf{C3}: \ \gamma_{\mathcal{H}_i,u_j} > \gamma_{th}, \quad \forall \mathcal{H}_i \in \mathbb{H}^{(t_\ell)}, \tag{38}$$

$$\mathbf{C4}: \ \left\|\mathbf{q}_{u_j}^{(t_{\ell+1})} - \mathbf{q}_{u_j}^{(t_\ell)}\right\| \leq \mathcal{S}_{xy}^{u_j}, \quad \forall t_z \in \mathbb{T}, \tag{39}$$

$$\mathbf{C5}: \ a_{u_j} \geq a_{min}. \tag{40}$$

TABLE 2. Data priority mapping in IEEE 802.15.6 standard.

DP index ($\mathcal{I}_{dp}$)	Data packet types
0	Background (BK)
1	Best effort (BE)
2	Excellent effort (EE)
3	Video (VI)
4	Voice (VO)
5	Low-priority medical data (e.g., body TMP)
6	High-priority medical data (e.g., body SBP)
7	Emergency medical data (e.g., body ECG)

The procedure for computing the normalized terms in (34) is demonstrated in detail as follows.

As mentioned before, our Q-REDTO algorithm employs NOMA technique to simultaneously schedule multiple hubs to connect to the corresponding UAV. In this regard, the total throughput achieved by each UAV is equal to the sum of the rates of all hubs located in the coverage area of each UAV that satisfy the constraint (19). According to this metric definition, the proposed Q-REDTO algorithm computes the sum rate of all hubs covered by each UAV in different states from (18) and subsequently divides it to the largest value of sum rate to calculate the normalized metric $\mathcal{R}_{ns}$. With a similar argument, our proposed algorithm obtains the energy consumption of each hub using (20) and then, adds all the values together to obtain sum of the energy consumption. Finally, to calculate the normalized sum of the energy consumption, i.e., $\mathcal{E}_{ns}$, in each state of the cluster area, the computed sum is divided into the largest value of the sum of energy consumption.

The third effective factor in the utility function $\mathcal{U}_{RED}$ is the normalized sum of the emergency indexes of all hubs covered by each UAV u_j that satisfy the constraint (19), based on the NOMA power assignment. In the context of emergency health care systems like WBANs, some vital signs or patients have precedence over others which require a lower delay and higher data rate. Under the circumstances, we should design an appropriate mechanism to address the timely transmission of life-critical vital signs to the monitoring center. The first factor that is effective to design the emergency index of each hub $\mathcal{H}_i$, is the data priority index denoted by $\mathcal{I}_{dp}$. This index is defined as the priority of different vital signals to each other. For instance, the vital signs of respiration rate bio-sensors have priority over the vital signs of body temperature bio-sensors. Our algorithm uses Table 2, categorized by IEEE 802.15.6 standard [25], to determine $\mathcal{I}_{dp}$ for different bio-sensors. The second influential parameter on the total emergency index is the patient priority index $\mathcal{I}_{pp}$ that represents the precedence of some patients to others. As an example, the vital signs of patients who are under surgery or have chronic illnesses, have precedence over other patients. In this situation, the patient priority is mapped by Table 3 extracted from the IEEE 802.15.6 standard, to distinguish between the life-emergency patients and non-critical patients.

The most effective factor in computing the total emergency index is the data severity index represented by $\mathcal{I}_{ds}$. This parameter stands for checking the occurrence of unexpected

TABLE 3. Patient priority mapping in IEEE 802.15.6 standard.

PP index ($\mathcal{I}_{pp}$)	WBAN services
0	Non-medical services (e.g., sport applications)
1	Low priority medical services (e.g., people in laboratory)
2	General health services (e.g., patient in GHU)
3	Highest priority medical services (e.g., patients in ICU)

emergency conditions. In practical health- care applications, each vital sign has a normal range and if the sensed value exceeds this range, it shows anomaly in the vital sign that may be due to the patient's life-critical condition. For more clarification, consider the case when the value of the blood pressure bio-sensor exceeds its normal range. Under this condition, it has the privilege to ECG vital sign which is in the normal range, although $\mathcal{I}_{dp}$ of ECG is higher than $\mathcal{I}_{dp}$ of the blood pressure sensor. In order to model $\mathcal{I}_{ds}$, consider $\varphi_{b_{ij}}$ as the sensed vital sign by bio-sensor b_{ij} and $[\varphi_{low_j}, \varphi_{up_j}]$ as the pre-assigned value of that bio-sensor. In this regard, let us define $\varphi^{(ij)}_{up,b} \triangleq \varphi_{up_j} - \varphi_{b_{ij}}$ and $\varphi^{(ij)}_{b,low} \triangleq \varphi_{b_{ij}} - \varphi_{low_j}$. In this situation, we define the following indicator function that models the normal and abnormal cases that occurred to each patient:

$$\mathcal{I}^{(ij)}_{n} = \begin{cases} 1, & \text{if } \varphi^{(ij)}_{up,b} > 0 \text{ and } \varphi^{(ij)}_{b,low} > 0 \text{ (Normal)}, \\ 0, & \text{if } \varphi^{(ij)}_{up,b} < 0 \text{ or } \varphi^{(ij)}_{b,low} < 0 \text{ (Abnormal)}. \end{cases} \tag{41}$$

To this end, the data severity index of j^{th} bio-sensor of patient p_i is formulated as follows:

$$\mathcal{I}^{(ij)}_{ds} = \begin{cases} 1, & \text{if } \mathcal{I}^{(ij)}_{n} = 1, \\ 1 + \dfrac{|\varphi^{(ij)}_{up,b}|}{\varphi^{(j)}_{up,low}}, & \text{if } \mathcal{I}^{(ij)}_{n} = 0 \text{ and } \varphi^{(ij)}_{up,b} < 0, \\ 1 + \dfrac{|\varphi^{(ij)}_{b,low}|}{\varphi^{(j)}_{up,low}}, & \text{if } \mathcal{I}^{(ij)}_{n} = 0 \text{ and } \varphi^{(ij)}_{b,low} < 0, \end{cases} \tag{42}$$

where $\varphi^{(j)}_{up,low} \triangleq \varphi_{up_j} - \varphi_{low_j}$. From (42), we can realize that in the abnormal case, i.e., $\mathcal{I}^{(ij)}_{n} = 0$, the normalized deviation of $\varphi_{b_{ij}}$ from the corresponding boundary is increased by 1. According to these factors, the emergency index of each hub $\mathcal{H}_i$, represented by $\mathcal{I}_{em}$, is determined as follows:

$$\mathcal{I}^{\mathcal{H}_i}_{em} = \left(\omega_{dp} \frac{\mathcal{I}^{\mathcal{H}_i}_{dp}}{7} + \omega_{pp} \frac{\mathcal{I}^{\mathcal{H}_i}_{pp}}{3} \right) \mathcal{I}^{\mathcal{H}_i}_{ds}, \tag{43}$$

where ω_{dp} and ω_{pp} represent the weighting coefficients of $\mathcal{I}_{dp}$ and $\mathcal{I}_{pp}$, respectively. It is worth mentioning that $\mathcal{I}^{\mathcal{H}_i}_{em}$ indicates the delay sensitivity level of the hub $\mathcal{H}_i$. Accordingly, the higher value of $\mathcal{I}^{\mathcal{H}_i}_{em}$ shows the lower tolerable access delay of data packets belonging to the hub $\mathcal{H}_i$. Thus, in order to minimize the access delay of each hub, the proposed Q-REDTO algorithm should select the hubs with the maximum value of $\mathcal{I}_{em}$. Then, the Q-REDTO algorithm calculates the sum of $\mathcal{I}_{em}$ of all covered hubs by each UAV u_j in each state. Finally, $\mathcal{I}_{ns}$ of each state is computed by dividing its sum of $\mathcal{I}_{em}$ to the maximum value of that summation between all states.

In order to optimize the vertical position of each UAV u_j, the Q-REDTO algorithm changes the altitude of u_j in ascending order from a_{min} to a_{max}, and then at each specific altitude, it calculates the $\mathcal{U}^{u_j}_{RED}$ metric in (34) for all tails belonging to each cluster. In the next stage, the algorithm computes the average $\overline{\mathcal{U}^{u_j}_{RED}}$ over all the tails and eventually, the altitude that maximizes $\overline{\mathcal{U}^{u_j}_{RED}}$ in each cluster is selected as the optimal vertical position of each UAV u_j. After determining the optimal vertical position of each UAV u_j, we optimize its horizontal trajectory. To this end, the proposed Q-REDTO algorithm employs the QL-based mechanism to train each UAV u_j for finding its best position in the corresponding cluster area. According to the aforementioned definitions of the elements in the QL-based mechanism, Q-REDTO runs enough episodes to update values of the Q-table of each UAV u_j step by step until it converges to the optimum value. In this regard, each episode is started by randomly selecting one state in each cluster as the initial state of the corresponding UAV. Then, each UAV u_j selects an action from its available action space $\mathbb{A}$ to reach the next state. In this regard, our proposed algorithm uses the following two different policies:

• **Reward maximizing action selection policy:** Using this policy, the Q-REDTO algorithm selects an action that maximizes the new Q-value of the current state, i.e.,

$$ac^{u_j}_{sel} \leftarrow \arg\max_{ac \in \mathbb{A}} Q^{u_j}_{new}(st, ac). \tag{44}$$

According to (33), the Q-function of each UAV u_j consists of three factors. The first one is the previous Q-value of the current state and the second one is the reward function which has the most important role in identifying the optimal horizontal trajectory of each UAV u_j. In this regard, if $\mathcal{U}^{u_j}_{RED}(st) < \mathcal{U}^{u_j}_{RED}(st')$, the reward function is designed as $\Re^{u_j}(st, ac) = \mathcal{U}^{u_j}_{RED}(st') - \mathcal{U}^{u_j}_{RED}(st)$, otherwise, it takes zero value. To select an optimal action, the reward function is computed for all available actions of the current state. According to this reward function, the states that cover the hubs with lower energy consumption, higher data rate, and delay sensitivity are more probable candidates to be selected as the next state for the corresponding UAV. The third effective factor in (33) is the future expected reward, denoted by $\max_{ac' \in \mathbb{A}} Q_{old}(st', ac')$, which refers to the maximum Q-value of the future state.

After calculating the above three factors, the corresponding Q-values are obtained from (33) and consequently, the action that has the maximum Q-value is selected as the optimal action for each UAV u_j.

• **Random action selection policy:** We use the random action selection mechanism in early episodes to enable UAVs to experience new actions and states. In this regard, we introduce a policy selection variable Ξ that selects a random num-

ber for each state from the interval [0, 1]. If this number is higher than a pre-specified threshold $ep_{th} = \frac{(ep_i - 1)}{|\mathbb{E}|^{\kappa}}$, where ep_i represents the current episode and $\kappa = |\mathbb{E}|, \mathbb{E} = \{ep_1, \ldots, ep_\kappa\}$ is the total number of episodes, the random selection policy will be selected to reach the next state; otherwise, the reward maximization policy is selected. It is worth mentioning that the threshold differs in a ascending manner in each episode.

We summarize the aforementioned procedure of the proposed Q-REDTO algorithm as pseudocode in Algorithm 2.

Beyond WBAN Transmission Model: After collecting vital signs of patients by UAVs, they must be transmitted to the remote monitoring center by means of a cloud backbone as is shown in Fig. 1. The cloud network consists of a cloud server in addition to several routers forming a mesh network for delivering vital signs to beyond the WBAN. These routers have equivalent transmission ranges and all of them employ the same channel for transmitting data packets. Due to the limited transmission range of each router, we use multiple routers to increase the coverage area of the network. After forwarding vital signs to the cloud by UAVs, the cloud server estimates the best path for each data packet in line with its emergency condition and bandwidth requirement. To this end, the shortest path tree (SPT) algorithm is run by the cloud server to find the shortest path for each packet reaching its destination, and the routing tables of the corresponding routers are updated, simultaneously. Furthermore, a weight representing the amount of traffic load is assigned to each link between the routers, to prevent the congestion in the routes and guarantee the load balancing. In this regard, by selecting each path, the weights of the constituent links are increased resulting in reducing the chance of selecting these links in the future paths.

IV. COMPLEXITY ANALYZES

Proposition 1: The computational complexity of Algorithm 1 is of order $\mathcal{O}(M\,|\mathbb{P}|)$, in which M represents the total number of UAVs needed after the convergence of FGKM and $\mathbb{P}$ denotes the set of all patients in the network, where $|\bullet|$ is the cardinality operator.

Proof: It is realized from the pseudo code of Algorithm 1 that there are one main *"for"* loop (i.e., lines $3-17$) and two inner *"for"* loops (i.e., lines $5-7$ and $11-13$). The computational complexity of the first inner loop is $\mathcal{O}(|\mathbb{P} - \{cc_1\}|)$, and the complexity of second one is $\mathcal{O}(|\mathbb{P} - \{cc_1, \cdots, cc_{m-1}\}|)$. Because the number of cluster centers is much less than the number of patients, it can be neglected in comparison to $|\mathbb{P}|$. Thus, the complexity of these two inner loops is of order $\mathcal{O}(|\mathbb{P}|)$. Moreover, complexity of the other lines and the main loop are $\mathcal{O}(1)$ and $\mathcal{O}(M)$, respectively. Thus, the total complexity of the nested loops is equal to $\mathcal{O}(M\,|\mathbb{P}|)$.

Proposition 2: The computational complexity of Algorithm 2 is of order $\mathcal{O}(|\mathbb{E}|\,M\,|\mathbb{S}|\,|\mathbb{A}|)$, where $|\mathbb{E}|$, $|\mathbb{S}|$, and $|\mathbb{A}|$ represent the number of episodes, available states, and actions in the proposed Q-REDTO algorithm, respectively.

Algorithm 2 Procedure of the Proposed Q-REDTO Algorithm

1: **Input:** $\mathbb{U} = \{u_1, \ldots, u_M\}$, $\mathbb{C} = \{\mathcal{CP}_{u_1}, \ldots, \mathcal{CP}_{u_M}\}$, $\mathbb{S}$, $\mathbb{A}$ ={"*north*","*east*","*south*","*west*","*north-east*","*north-west*","*south-east*","*south-west*"}.
2: **Output:** The optimal trajectory of UAVs.
3: **for** each $\mathcal{CP}_{u_i} \in \mathbb{C}$ **do**
4: Discretize the cluster environment into tails.
5: **end for**
6: **for** each $\mathcal{H}_i$ covered with u_j **do**
7: Calculate $\mathcal{R}_{\mathcal{H}_i,u_j}$ from (18).
8: Calculate $\mathcal{E}_{\mathcal{H}_i,u_j}$ from (20).
9: Calculate $\mathcal{I}_{em}^{\mathcal{H}_i}$ from (43).
10: **end for**
11: **for** each $u_j \in \mathbb{U}$ **do**
12: **for** each $a_{min} < a_{u_j} < a_{max}$ **do**
13: **for** each $st_i \in \mathbb{S}$ **do**
14: Compute $\mathcal{R}_{ns}$, $\mathcal{E}_{ns}$, and $\mathcal{I}_{ns}$.
15: Compute $\mathcal{U}_{RED}^{u_j}(st)$ from (34).
16: **end for**
17: Determine the average value of $\overline{\mathcal{U}_{RED}^{u_j}}$.
18: **end for**
19: $a_{u_j}^{opt} \leftarrow \arg\max_{a_{u_j}} \overline{\mathcal{U}_{RED}^{u_j}}$.
20: **end for**
21: **for** each $ep_i \in \mathbb{E}$ **do**
22: **for** each $u_j \in \mathbb{U}$ **do**
23: Randomly select the initial state of u_j.
24: **for** st_i **do**
25: Select a random number $\Xi \in [0, 1]$.
26: Calculate $ep_{th} = \frac{(ep_i - 1)}{|\mathbb{E}|}$.
27: **if** $\Xi > ep_{th}$ **then**
28: Randomly select $ac_{u_j} \in \mathbb{A}$.
29: **else**
30: **for** each $ac_{u_j} \in \mathbb{A}$ **do**
31: $\varpi \leftarrow \mathcal{U}_{RED}^{u_j}(st) - \mathcal{U}_{RED}^{u_j}(st')$
32: **if** $\varpi < 0$ **then**
33: $\Re^{u_j}(st, ac) \leftarrow -\varpi$.
34: **else**
35: $\Re^{u_j}(st, ac) \leftarrow 0$.
36: **end if**
37: Compute $\mathcal{Q}_{new}^{u_j}(st, ac)$ from (33).
38: **end for**
39: $ac_{sel}^{u_j} \leftarrow \arg\max_{ac \in \mathbb{A}} \mathcal{Q}_{new}^{u_j}(st, ac)$.
40: **end if**
41: Update $\mathcal{Q}_{new}^{u_j}(st, ac)$ using (33).
42: **end for**
43: **end for**
44: **end for**

Proof: Regarding the pseudo code of Algorithm 2, there are four main *"for"* loops in this algorithm. Complexity of the first and second loops (i.e., lines $3-5$ and $6-10$) are $\mathcal{O}(1)$ and $\mathcal{O}(|\mathbb{P}|)$, respectively. The third main *"for"* loop (i.e., lines $11-20$) consists of three nested *"for"* loops. The

computational complexity of the outer loop is $\mathcal{O}(M)$, then the complexity of the first inner loop is of order $\mathcal{O}(1)$ and for the second loop is $\mathcal{O}(|\mathbb{S}|)$. Therefore, the complexity of this part is of order $\mathcal{O}(M|\mathbb{S}|)$. The fourth main loop (i.e., lines $21-44$) comprises of four nested *"for"* loops. The computational complexities of these loops from outer on to inner one are $\mathcal{O}(|\mathbb{E}|)$, $\mathcal{O}(M)$, $\mathcal{O}(|\mathbb{S}|)$, and $\mathcal{O}(|\mathbb{A}|)$, respectively. Thus, the total complexity of this term is of order $\mathcal{O}(|\mathbb{E}|M|\mathbb{S}||\mathbb{A}|)$. Finally, the total complexity of the Q-REDTO algorithm is computed by the sum of the main loops complexities as $\mathcal{O}(|\mathbb{P}| + M|\mathbb{S}| + |\mathbb{E}|M|\mathbb{S}||\mathbb{A}|)$. In this equation, the third term is much larger than the other terms when the number of episodes grows. Therefore, the third term is dominant and the other terms can be ignored. It should be noted that, in deep learning methods based on data sets, the complexity is related to the number of data samples used for training the model. However, in Q-learning method which is based on trial and error, the complexity is determined according to the number of episodes.

It is worth mentioning that all computational tasks in the ESTO algorithm should be executed for all tiles in each cluster area. Accordingly, the complexity of this algorithm is of order $\mathcal{O}\left(M\left|\tilde{\mathbb{S}}\right|\right)$, where $\left|\tilde{\mathbb{S}}\right|$ is the number of all tiles in each cluster which is much larger than $|\mathbb{S}|$ in Q-REDTO. Since $|\mathbb{E}|$ and $|\mathbb{A}|$ are upper bounded by the maximum values of deployed episodes and actions in Q-REDTO, it is concluded that the computational complexity of ESTO is higher than that of the proposed Q-REDTO algorithm especially in a large number of tiles.

V. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed Q-REDTO algorithm in terms of average sum rate, delay, and the energy consumption of hubs. In this regard, we investigate how the Q-REDTO algorithm designs the optimal trajectory of UAVs to maximize the average sum rate besides minimizing the average sum delay and energy consumption. Toward this goal, we consider a vast network in which $|\mathbb{P}| = 300$ patients are randomly distributed in the area with the size of 3000×3000 m^2. The optimal number of existing UAVs in the network is calculated by the FGKM scheme described in Algorithm 1. This algorithm divides the network area into $|\mathbb{C}|$ clusters and dedicates a single UAV to each cluster. For modeling the state space of our proposed algorithm, the cluster area is discretized into different tails with an area of 100×100 m^2. To reduce the complexity of the proposed Q-REDTO algorithm and to guarantee the real-time transmission of vital signs in delay-sensitive healthcare applications, the numerical results of Q-REDTO in state spaces with a different number of tiles are compared. Afterward, the state space with tiles' space 100m $\times$ 100m is selected to manage the trade-off between the best-obtained results and the complexity of the algorithm. In addition, we assume that each UAV u_j can horizontally fly at the altitudes $a_{u_j} \in \{170, 180, 190, 200, 210, 220\}$ meters. Moreover, the horizontal radius of each UAV at each specific altitude is equal to that altitude. To consider the dynamic nature of the WBANs environment, Difference Correlated Random Walk (DCRW) model is assumed for the mobility pattern of hubs. In this regard, the velocity of each hub $\mathcal{H}_i$ is randomly selected from the interval $[0, 4]$ m/s and it is changed in each time slot. Furthermore, to achieve fair results, equal weights are considered for transmission rate, energy consumption, and delay in our proposed utility function in (34). Based on this, ω_r, ω_e, and ω_i are assumed to be 1. According to these assumptions, we simulate our proposed algorithm using the MATLAB simulator V.2018. Table 4 illustrates the list of parameters for our simulations.

TABLE 4. Simulation parameters.

Parameter	Value	Parameter	Value
c	3×10^8 m/s	f_c	2.4 GHZ
d_0	1 m	W	20 MHZ
$\mathscr{P}_{total}$	60 dBm	N_0	36 dBm
η^{LoS}	2.5	η^{NLoS}	3
α	9.61	β	0.16
a_{min}	170 m	a_{max}	220 m
λ	0.1	γ	0.7

To compare the performance of the proposed Q-REDTO algorithm in terms of high transmission rate, energy-efficient, and delay-sensitive transmission of vital signs in WBANs, we consider the following three baseline models:

Particle Swarm Optimization (PSO)-based UAV Placement with OMA: To make a fair comparison, in this scheme, the transmissions belonging to communication tier I are scheduled using the proposed WH coding scheme, in which all bio-sensors of each patient can simultaneously transmit data to the corresponding hub by employing the aforementioned orthogonal WH codes. Afterward, in communication tier II, where the aggregated data in each hub is transmitted to UAVs, the OMA-based PSO algorithm proposed in [20] is employed to optimize the trajectory of UAVs in each cluster. This algorithm finds the best position of each UAV in each time slot in which the UAV can achieve the best throughput along with the least energy consumption. It is worth mentioning that this algorithm uses the OMA scheme to schedule the transmission of hubs to UAVs. Accordingly, in each time slot, only one hub can transmit its data to each UAV. To the best of our knowledge, [20] is the only work that has investigated the problem of trajectory optimization of UAVs in WBANs.

Particle Swarm Optimization (PSO)-based UAV Placement with NOMA: In light of expanding the proposed PSO algorithm in [20], the NOMA-based PSO scheme is designed to support the NOMA scheduling technique. Similar to other algorithms, the cyclic orthogonal WH codes are used for scheduling the transmissions of bio-sensors to the corresponding hubs in tier I of this scheme. Subsequently, in communication tier II, the best trajectory of each UAV over each cluster area is determined by the new proposed PSO algorithm to optimize the value of utility function $\mathcal{U}_{RED}$. In this case, the transmission of multiple hubs that satisfies

constraint (19) can be simultaneously scheduled in the same time slot using the NOMA technique.

Exhaustive Search-based Trajectory Optimization (ESTO) Algorithm: Similar to the PSO-based algorithm, in the communication tier I of this scheme, the transmission of vital signs sensed by bio-sensors is scheduled by employing the cyclic orthogonal WH coding scheme. In the next stage, UAVs aggregate vital signs received by hubs. In the ESTO scheme, each UAV has all information of hubs located in its corresponding cluster. In other words, the utility function in ESTO is obtained for each tail from (34). Then, based on this information, each UAV computes its best trajectory in each time slot. Moreover, ESTO uses the NOMA scheduling technique in which multiple hubs can simultaneously transmit their data packets to its UAV.

Proposed Q-REDTO Algorithm: This scheme is our proposed Q-learning-based UAV trajectory optimization algorithm, where we use the cyclic WH codes in communication tier I. Furthermore, as described in the previous section, we design a Q-learning-based search algorithm to find the best location of each UAV in order to maximize the utility function in (34). Additionally, the same as ESTO, our proposed algorithm employs the NOMA scheduling scheme to deliver data packets of hubs to UAVs.

A. RESULTS AND DISCUSSION

In this subsection, we first investigate the trajectory of UAVs in corresponding cluster areas in the IoMT WBAN. This trajectory is obtained by dividing the whole city area into different clusters using FGKM and assigning one UAV to each cluster. Fig. 4 shows the 3D trajectory of two different UAVs attained from Q-REDTO compared to the ESTO algorithm during ten time slots. In each time slot, Q-REDTO and ESTO find the best 3D position of UAVs which maximizes $\mathcal{U}_{RED}$ according to (34). It should be noted that because of the mobility of patients in the network and the variations of vital signs, the best position of UAVs is changed during the time. As shown in Fig. 4, in the proposed Q-REDTO algorithm, UAVs are trained to reach the benchmark trajectory shown in solid line, by knowing the cluster environment episode by episode. Moreover, we can see that in some time slots, the position of UAVs, selected by Q-REDTO, is different from the benchmark value obtained from the exhaustive search. This occurs as a result of limited knowledge of UAVs from the clusters' area when the Q-REDTO algorithm is employed.

Afterward, we evaluate the performance of the proposed Q-REDTO algorithm through various scenarios in terms of the spectral efficiency, energy consumption, and delay.

Scenario I: We first examine the changes in average spectral efficiency defined as the ratio between average sum rate and the total bandwidth W in different number of patients. In this regard, assuming the number of patients, i.e., $|\mathbb{P}|$, is varied from 100 to 300, Fig. 5 illustrates the performance of our Q-REDTO in terms of increasing the average spectral efficiency in comparison to the other mentioned schemes.

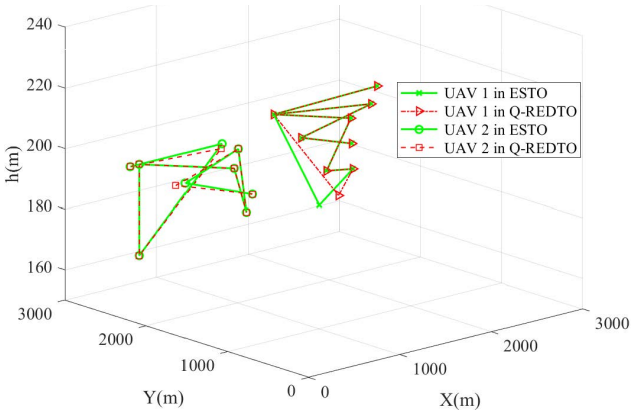


FIGURE 4. 3D trajectory of two UAVs over their cluster areas during ten time slots.

Because of employing NOMA scheduling technique in both ESTO and Q-REDTO algorithms, any increase in the number of patients leads to increasing the number of hubs covered by each UAV. In this situation, based on NOMA, the number of simultaneous transmissions from hubs to each UAV is increased, which results in intensifying the average sum rate. However, by further increasing the number of patients, the interference is highly increased that leads to reducing the rate of raising average spectral efficiency. As it can be realized from this figure, the Q-REDTO algorithm with FGKM achieves considerably better results than Q-REDTO with K-means in terms of increasing the average sum rate. This indicates the poor performance of the K-means algorithm in finding the global optimal solution of the city clustering problem that results in deteriorating the optimized trajectory of UAVs in the network. Furthermore, the results in Fig. 5 show that our proposed Q-REDTO algorithm outperforms the PSO-based algorithm in increasing the average spectral efficiency. This occurs as a result of scheduling the transmissions of multiple hubs in the same time slot using NOMA technique. As demonstrated in the figure, the proposed Q-REDTO algorithm achieves a better performance in finding the best value of the average sum rate compared to the NOMA-based PSO algorithm. The result indicates the lower convergence speed of PSO than the Q-REDTO algorithm. In other words, under very similar assumptions, because the NOMA-based PSO does not converge to the best value of utility function $\mathcal{U}_{RED}$, it obtains the lower average sum rate in comparison to Q-REDTO. Eventually, Fig. 5 demonstrates that the results of our Q-REDTO algorithm are absolutely close to the optimal values of the spectral efficiency obtained by the ESTO algorithm. Thus, we can claim that our Q-REDTO algorithm can achieve near optimal results without requiring all information about the location and channel conditions of hubs, and just by discovering the cluster environment step by step through executing episodes.

Scenario II: This scenario investigates the effect of a different number of patients on the average sum energy consumption at each time slot. To this end, the value of $|\mathbb{P}|$ is

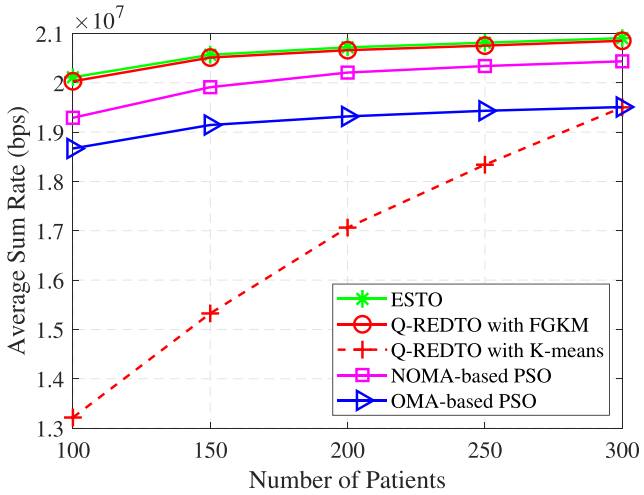


FIGURE 5. The variation of the average spectral efficiency versus different number of patients.

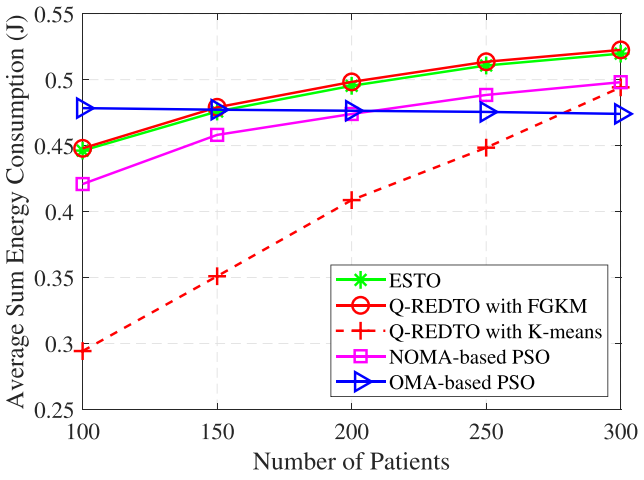


FIGURE 6. The variation of the average sum energy consumption versus different number of patients.

changed between 100 and 300 and the results are shown in Fig. 6. According to this figure, the average sum energy consumption of our proposed Q-REDTO and ESTO algorithms is higher than that for the PSO-based algorithms. This occurs as a consequence of employing the NOMA scheduling technique in Q-REDTO and ESTO algorithms. Regarding this property, multiple adjacent hubs that satisfy condition (19), simultaneously transmit their data packets to the corresponding UAV, while in the OMA-based PSO algorithm, only one hub can transmit its data to each UAV at each time slot. Additionally, the average sum energy consumption of NOMA-based PSO is less than the Q-REDTO algorithm. As a consequence of the lower convergence speed of NOMA-based PSO, the lower number of simultaneous transmissions is scheduled in each time slot. Hence, the sum value of energy consumption of these transmissions is less than that of Q-REDTO. Moreover, from Fig. 6, we can realize that by increasing the number of patients in the network, the average sum energy consumption of Q-REDTO and ESTO is intensified. In these algorithms, by increasing the number of patients in each cluster, the number of hubs covered by each UAV is increased. Thus, by increasing the number of simultaneous transmissions to each UAV, the total energy consumption of the network is increased. However, by further intensifying the number of concurrent transmissions, some of them cannot satisfy condition (19) and they are discarded from the scheduling process. In this situation, the sum energy consumption reduces. In contrast, the curve in blue demonstrates the average sum energy consumption of the OMA-based PSO algorithm where only one hub transmits its data packet to each UAV within each time slot. It should be noted that by increasing the number of patients, the number of hubs transmitting to each UAV will not rise. Taking this problem into account, the energy consumption of OMA-based PSO roughly remains the same in a different number of patients. Taking these features into account, by rising the number of patients in the other algorithms, more hubs will be able to transmit their data packets concurrently which increases the average sum energy consumption of these algorithms and finally results in crossing the blue curve. Finally, Fig. 6 illustrates that the average sum energy consumption value of Q-REDTO with K-means is less than that for Q-REDTO with FGKM. In this case, because of non-globally optimizing the clustering problem in K-means, the number of simultaneous transmissions scheduled in the same time slot is reduced. Consequently, the sum energy consumption value of these transmissions is decreased.

Scenario III: The delay sensitivity of our proposed algorithm is evaluated in this scenario. In this regard, the sum emergency index of different algorithms is compared in a different number of patients. To obtain this metric, the emergency indexes of all covered hubs by all UAVs are added together. This metric shows the delay sensitivity of the algorithms. In other words, a higher amount of computed sum emergency index represents that the hubs with higher emergency conditions can access the channel earlier. This guarantees the timely transmission of vital signals in life-critical situations. The results are shown in Fig. 7 supposing $|\mathbb{P}|$ is varied from 100 to 300. As illustrated in Fig. 7, the sum emergency index of Q-REDTO and ESTO is much higher than the PSO-based algorithms. This is a consequence of scheduling multiple hubs in the same time slot using the NOMA technique in ESTO and Q-REDTO. Accordingly, we can claim that our proposed Q-REDTO algorithm outperforms the PSO-based schemes in terms of delay sensitivity. Furthermore, this figure shows that under similar assumptions, the NOMA-based PSO scheme has worse performance in comparison to the proposed Q-REDTO algorithm. As mentioned before, this occurs because the convergence speed of NOMA-based PSO is less than our proposed algorithm. Therefore, the inefficient of $\mathcal{U}_{RED}$ is selected by the NOMA-based PSO scheme which leads to reducing the sum emergency index value of concurrent transmissions scheduled by the NOMA technique. Moreover, we can realize

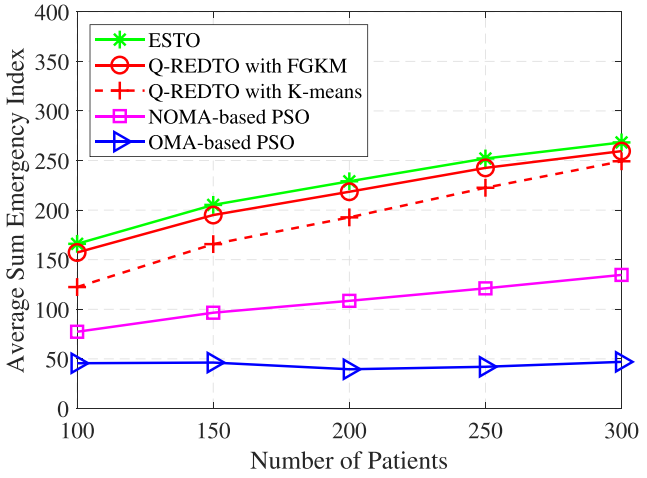


FIGURE 7. The variation of the average sum emergency index versus different number of patients.

that, by increasing the number of patients, the sum emergency index of ESTO and Q-REDTO is increased, whereas it remains constant for the OMA-based PSO algorithm. As it was clarified in the previous scenario, this is a result of increasing the number of simultaneous transmissions to each UAV in ESTO and Q-REDTO. Similar to the argument in Scenario I, our proposed Q-REDTO with FGKM has a better performance in increasing the value of the average sum emergency index in comparison to Q-REDTO with K-means, which is a result of globally finding an optimal solution by the FGKM clustering algorithm.

Scenario IV: In this scenario, we investigate the convergence of our proposed Q-REDTO algorithm throughout episodes and in different learning rates. To this end, the optimal value obtained from the ESTO algorithm in one time slot is assumed to be the benchmark value, and our proposed Q-REDTO is trained to converge to this value step by step during the episodes. In this regard, Fig. 8 illustrates the convergence of our proposed algorithm by increasing the number of episodes in the range [1, 100], in terms of the aforementioned factors for one of the UAVs in different values of the learning rate, i.e., λ. Figs. 8a, 8b, and 8c show respectively, the convergence of sum rate, sum energy consumption, and sum emergency index of all admitted hubs by the NOMA technique in the selected position of the UAV versus different episodes. These figures illustrate that in the earlier episodes, the values obtained from Q-REDTO are really divergent from the benchmark line attained from ESTO. This phenomenon occurs because, in earlier episodes, the UAV has no appropriate cognition from its cluster environment, thus, it randomly selects its positions. However, by further increasing the number of episodes, the environment information of the UAV is enhanced, and eventually, the values of sum rate, energy consumption, and emergency index converge to the corresponding benchmark values. Moreover, we can realize from Fig. 8 that by increasing the value of the learning rate, the amount of the three mentioned factors converges slower. This is because of the unknown nature of the cluster. In this

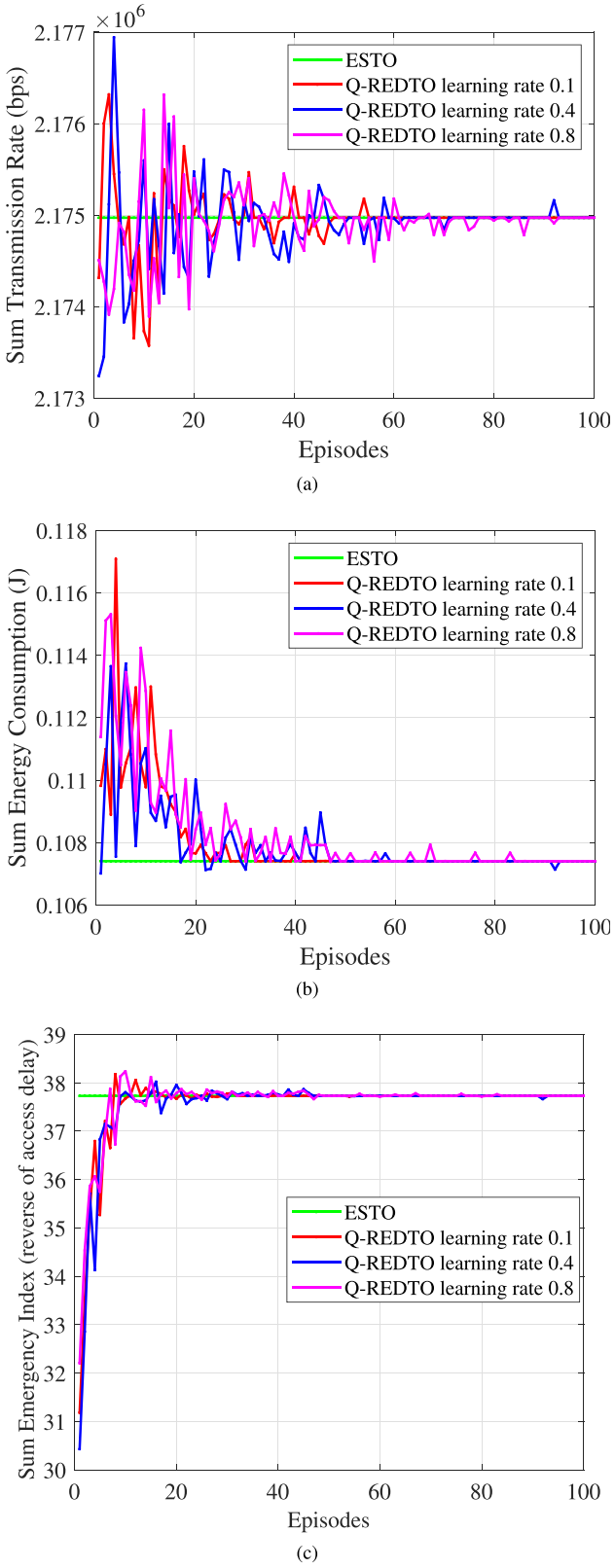


FIGURE 8. Convergence of Q-REDTO in terms of *a*) sum rate, *b*) sum energy consumption, and *c*) sum emergency index.

situation, by increasing the learning rate value, the effect of calculated reward is intensified in updating the Q-value of

selected state-actions according to (33). Therefore, in earlier episodes, due to the lack of information of the environment, selecting some inappropriate positions leads to considerable changes in Q-values of that state's actions. This problem results in a raising number of requisite episodes for convergence of the algorithm. Additionally, Fig. 8c demonstrates the convergence of the sum value of the emergency indexes of all hubs scheduled in the same time slot, which has a reverse relationship with the access delay. Indeed, as explained in Section III-A, the emergency index metric, i.e., $\mathcal{I}_{em}^{\mathcal{H}_i}$, represents the delay sensitivity of each hub $\mathcal{H}_i$. Under the circumstances, the higher value of the sum emergency index in each time slot means that hubs with higher emergency conditions are scheduled in that time slot. This property results in reducing the access delay of life-critical data packets transmitted in the allocated bandwidth.

VI. CONCLUSION AND FUTURE WORK

In this paper, we addressed the Q-learning-based 3D trajectory optimization of UAVs to the timely transmission of vital signs of patients without interrupting their daily lifestyle. To this end, we proposed the Q-REDTO algorithm which efficiently increases the amount of throughput and reduces the energy consumption and delay by training UAVs to achieve the best 3D placement. In this regard, at first, each UAV has no prior cognition of its corresponding cluster area, and it gets to know the environment during the episodes by moving among the states based on their Q-value. It should be noted that the mobility of patients leads to a time-varying topology of the network. In this situation, our proposed Q-REDTO algorithm learns to reach the best 3D position for each UAV in each time slot by updating its Q-table step-by-step. Moreover, our algorithm employs the NOMA technique, which simultaneously schedules the transmission of multiple hubs by considering a degree of interference among them. Under the circumstances, the data rate requirement of all of the simultaneous transmissions should be satisfied using a pre-specified SINR threshold. The simulation results demonstrated that our Q-REDTO scheme can achieve the benchmark value of throughput, energy consumption, and delay without requiring complete information about the environment. One possible future work is to expand this study to use a more sophisticated learning algorithm along with edge computing in the NOMA technique to find the best set of simultaneous transmissions as well as employing federated learning to improve the performance of the proposed trajectory optimization algorithm in absolutely large test-beds.

REFERENCES

[1] A. Ghubaish, T. Salman, M. Zolanvari, D. Unal, A. Al-Ali, and R. Jain, "Recent advances in the Internet-of-medical-things (IoMT) systems security," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 8707–8718, Dec. 2020.

[2] M. T. Mamaghani and Y. Hong, "Intelligent trajectory design for secure Full- duplex MIMO-UAV relaying against active eavesdroppers: A model-free reinforcement learning approach," *IEEE Access*, vol. 9, pp. 4447–4465, 2021.

[3] A. Visintini, T. D. P. Perera, and D. N. K. Jayakody, "3-D trajectory optimization for fixed-wing UAV-enabled wireless network," *IEEE Access*, vol. 9, pp. 35045–35056, 2021.

[4] Y. Du, Z. Chen, J. Hao, and Y. Guo, "Joint optimization of trajectory and communication in multi-UAV assisted backscatter communication networks," *IEEE Access*, vol. 10, pp. 40861–40871, 2022.

[5] O. Esrafilian, R. Gangula, and D. Gesbert, "Learning to communicate in UAV-aided wireless networks: Map-based approaches," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1791–1802, Apr. 2019.

[6] F. Cheng, S. Zhang, Z. Li, Y. Chen, N. Zhao, F. R. Yu, and V. C. M. Leung, "UAV trajectory optimization for data offloading at the edge of multiple cells," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6732–6736, Jul. 2018.

[7] L. Wang, K. Wang, C. Pan, W. Xu, N. Aslam, and L. Hanzo, "Multi-agent deep reinforcement learning based trajectory planning for multi-UAV assisted mobile edge computing," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 1, pp. 73–84, Mar. 2021.

[8] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7957–7969, Aug. 2019.

[9] J. Cui, Z. Ding, Y. Deng, A. Nallanathan, and L. Hanzo, "Adaptive UAV-trajectory optimization under quality of service constraints: A model-free solution," *IEEE Access*, vol. 8, pp. 112253–112265, 2020.

[10] H. Wu, F. Lyu, C. Zhou, J. Chen, L. Wang, and X. Shen, "Optimal UAV caching and trajectory in aerial-assisted vehicular networks: A learning-based approach," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2783–2797, Dec. 2020.

[11] Y.-H. Hsu and R.-H. Gau, "Reinforcement learning-based collision avoidance and optimal trajectory planning in UAV communication networks," *IEEE Trans. Mobile Comput.*, vol. 21, no. 1, pp. 306–320, Jan. 2022.

[12] S. Zhu, L. Gui, N. Cheng, F. Sun, and Q. Zhang, "Joint design of access point selection and path planning for UAV-assisted cellular networks," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 220–233, Jan. 2020.

[13] R. W. Jones and G. Despotou, "Unmanned aerial systems and healthcare: Possibilities and challenges," in *Proc. 14th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2019, pp. 189–194.

[14] S. Ullah, K. Kim, K. H. Kim, M. Imran, P. Khan, E. Tovar, and F. Ali, "UAV-enabled healthcare architecture: Issues and challenges," *Future Gener. Comput. Syst.*, vol. 97, pp. 425–432, Aug. 2019.

[15] R. Gupta, A. Shukla, P. Mehta, P. Bhattacharya, S. Tanwar, S. Tyagi, and N. Kumar, "VAHAK: A blockchain-based outdoor delivery scheme using UAV for healthcare 4.0 services," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 255–260.

[16] A. Islam and S. Y. Shin, "BHMUS: Blockchain based secure outdoor health monitoring scheme using UAV in smart city," in *Proc. 7th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Jul. 2019, pp. 1–6.

[17] A. Mukhopadhyay and D. Ganguly, "FANET based emergency healthcare data dissemination," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 170–175.

[18] A. Kachroo, S. Vishwakarma, J. N. Dixon, H. Abuella, A. Popuri, Q. H. Abbasi, C. F. Bunting, J. D. Jacob, and S. Ekin, "Unmanned aerial vehicle-to-wearables (UAV2W) indoor radio propagation channel measurements and modeling," *IEEE Access*, vol. 7, pp. 73741–73750, 2019.

[19] S. R. Vangimalla and M. El-Sharkawy, "Interoperability enhancement in health care at remote locations using thread protocol in UAVs," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2018, pp. 2821–2826.

[20] C. Tang, C. Zhu, X. Wei, J. J. P. C. Rodrigues, M. Guizani, and W. Jia, "UAV placement optimization for internet of medical things," in *Proc. IEEE Int. Wireless Commun. Mobile Comput.*, Limassol, Cyprus, Jul. 2020, pp. 752–757.

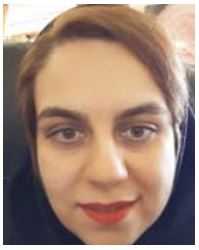
[21] A. Tawfiq, J. Abouei, and K. N. Plataniotis, "Cyclic orthogonal codes in CDMA-based asynchronous wireless body area networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1593–1596.

[22] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.

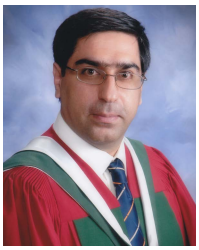
[23] C. You and R. Zhang, "3D trajectory optimization in Rician fading for UAV-enabled data harvesting," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3192–3207, Jun. 2019.

[24] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, p. 451– 461, Feb. 2003.

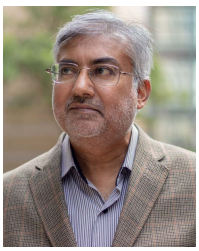
[25] *IEEE Standard for Local and Metropolitan Area Networks—Part 15.6: Wireless Body Area Networks*, IEEE Standard 802.15.6-2012, Feb. 2012, pp. 1–271.



ZEINAB ASKARI received the B.Sc. degree in electronics engineering from Sheikhbahaee University, Iran, in 2012, and the M.Sc. degree in telecommunication systems engineering from Najafabad University, Iran, in 2016. Her research interests include wireless networking specially the Internet of Medical Things (IoMT), wireless body area networks (WBANs), scheduling, reinforcement learning, and resource allocation.



JAMSHID ABOUEI (Senior Member, IEEE) received the B.Sc. degree in electronics engineering and the M.Sc. degree in communication systems engineering from the Isfahan University of Technology, Iran, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Canada, in 2009. He joined as a Lecturer with the Department of Electrical Engineering, Yazd University, Iran, in 1996, and was promoted to an Assistant Professor, in 2010, and an Associate Professor, in 2015. From 2009 to 2010, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Canada. During his sabbatical, he was an Associate Researcher with the Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Canada. His research interests include 5G and wireless sensor networks (WSNs), with a particular emphasis on PHY/MAC layer designs, including the energy efficiency and optimal resource allocation in cognitive cell-free massive MIMO networks, multi-user information theory, mobile edge computing, and femtocaching.



MUHAMMAD JASEEMUDDIN (Member, IEEE) received the B.E. degree from N. E. D. University, Pakistan, the M.S. degree from The University of Texas at Arlington, and the Ph.D. degree from the University of Toronto. He worked with the Advanced IP Group and Wireless Technology Laboratory (WTL), Nortel Networks. He is currently a Professor and the Program Director of Computer Networks Program at Ryerson University. His research interests include network automation, caching in 5G and ICN networks, context-aware mobile middleware and mobile cloud, localization, power-aware MAC and routing for sensor networks, heterogeneous wireless networks, and IP routing and traffic engineering.

ALAGAN ANPALAGAN (Senior Member, IEEE) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto, Canada. He joined with the ELCE Department, Ryerson University, Canada, in 2001, and was promoted to a Full Professor, in 2010. He served the department in administrative positions as an Associate Chair, the Program Director of Electrical Engineering, and the Graduate Program Director. During his sabbatical, he was a Visiting Professor at the Asian Institute of Technology, and a Visiting Researcher at Kyoto University. His industrial experience includes working for three years with Bell Mobility, Nortel Networks, and IBM. He directs a research group working on radio resource management (RRM) and radio access and networking (RAN) areas within the WINCORE Laboratory. He served as an Editor for the IEEE Communications Surveys and Tutorials (2012–2014), IEEE Communications Letters (2010–2013), and *EURASIP Journal of Wireless Communications and Networking* (2004–2009). He also served as the Guest Editor for six special issues published in IEEE, IET, and ACM. He served as the TPC Co-Chair of IEEE VTC Fall 2017, and the TPC Co-Chair of IEEE INFOCOM'16, IEEE Globecom15, and IEEE PIMRC'11. He served as the Vice Chair of IEEE SIG on Green and Sustainable Networking and Computing with Cognition and Cooperation (2015–2018), the IEEE Canada Central Area Chair (2012–2014), the IEEE Toronto Section Chair (2006–2007), the ComSoc Toronto Chapter Chair (2004–2005), and the IEEE Canada Professional Activities Committee Chair (2009–2011). He was a recipient of the IEEE Canada J. M. Ham Outstanding Engineering Educator Award, in 2018, the YSGS Outstanding Contribution to Graduate Education Award, in 2017, the Deans Teaching Award, in 2011, and the Faculty Scholastic, Research and Creativity Award thrice from Ryerson University.

KONSTANTINOS (KOSTAS) N. PLATANIOTIS (Fellow, IEEE) received the B.Eng. degree in computer engineering from the University of Patras, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the Florida Institute of Technology Melbourne, Florida.

He is currently a Professor with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where he directs the Multimedia Laboratory. He has been holding the Bell Canada Endowed Chair in Multimedia, since 2014. His research interests include the areas of image/signal processing, machine learning and adaptive learning systems, visual data analysis, multimedia and knowledge media, and affective computing. He is a fellow of the Engineering Institute of Canada and the Canadian Academy of Engineering, and a Registered Professional Engineer in Ontario. He has served as the Editor-in-Chief of the IEEE Signal Processing Letters. He was the Technical Co-Chair of the IEEE 2013 International Conference in Acoustics, Speech and Signal Processing, and he served as the Inaugural IEEE Signal Processing Society Vice President for Membership (2014–2016) and the General Co-Chair for the 2017 IEEE GLOBALSIP. He served as the General Co-Chair for the 2018 IEEE International Conference on Image Processing (ICIP 2018) and IEEE International Acoustics, Speech and Signal Processing (ICASSP 2021). He is the General Chair for the 2027 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2027), Toronto.