An Actor-Critic Approach for Resource Allocation in Energy Harvesting-Powered Wireless Body Area Network

Khaled Sabahein,
Mississippi Valley State University,
Itta Bena, MS, USA
Khaled.sabahein@mvsu.edu

Feng Wang
University of Mississippi,
University, MS 38677
fwang@cs.olemiss.edu

Zhonghui Wang
Louisiana State University in Shreveport,
Shreveport, LA,
zhonghui.wang@lsus.edu

Abstract This study introduces an innovative actor-critic deep reinforcement learning approach for optimizing resource allocation in energy-harvesting Wireless Body Area Networks (WBANs). Facing the challenge of limited sensor energy, our method efficiently manages key parameters like transmission mode, relay selection, and energy utilization, significantly enhancing WBAN's energy efficiency and delivery probability. Through simulations, we demonstrate our technique's superior performance over traditional models, showcasing its potential for future WBAN implementations.

1. Introduction:

Wireless Body Area Networks (WBANs), as highlighted by [1], have become pivotal in modern healthcare, providing a platform for continuous monitoring through sensors placed on or within the human body. sensors, crucial for transmitting vital health data, face the significant challenge of limited energy resources, as [2] and [11] emphasize, particularly in the context of energy harvesting from environmental and body-generated sources. The effective allocation of these limited resources, a topic explored by [4] and [10], is key to optimizing WBAN performance. The emerging field of artificial intelligence, particularly reinforcement learning (RL), has shown promise in improving performance in dynamic environments like WBANs, as identified in recent studies [17]. RL, framed as a Markov decision problem (MDP), involves an agent interacting with an environment, garnering rewards, and performing actions to maximize cumulative rewards. This concept has inspired the development of resource allocation techniques to boost energy efficiency in WBANs [16]. While there's limited integration of RL in energy-harvesting WBANs (EH-WBANs), some recent advancements have modeled resource allocation in EH-WBANs as MDPs, applying Q-learning for optimal energy efficiency [16]. However, given the extensive traffic generated by WBANs, Q-learning's effectiveness diminishes due to its discrete state-space, highlighting the need for more sophisticated techniques to optimize resource allocation policies in EH-WBANs. Our study introduces a novel approach using actor-critic deep reinforcement learning to enhance resource allocation efficiency in energy-harvesting WBANs. This method is particularly aligned with the QoS-aware strategies and energy-efficient designs proposed by [6] and [13]. Our approach promises to significantly improve the energy efficiency and operational reliability of WBANs, addressing the urgent need for sustainable and uninterrupted network functionality in critical healthcare applications.

The following are the paper's primary contributions:

- We formulate the energy efficiency as an actor-critic learning DRL framework to learn the resource allocation policy in EH-WBANs.
- The simulation results show that the proposed AC approach can minimize the energy efficiency and speed of convergence and outperforms the traditional Q-learning by efficiently learning the optimal resource allocation policy in EH-WBANs.
- efficiently learning the optimal resource allocation policy in EH-WBANs.

2. SYSTEM MODEL

In the proposed model, we have integrated an intricate actor-critic deep reinforcement learning (DRL) framework within a Wireless Body Area Network (WBAN) system equipped with multiple energy-harvesting (EH) sensors. This network includes a variety of sensors such as EEG (electroencephalogram), ECG (electrocardiogram), motion detectors, glucose monitors, and EMG (electromyogram) sensors, all strategically embedded within the human body. These sensors are tasked with continuously monitoring a wide range of physiological parameters, capturing critical health data.

The data gathered by these sensors is then relayed to a centralized medical server. This transmission occurs via a base station (BS) or a personal digital assistant (PDA), which functions as a crucial gateway in the system. The actor-critic DRL framework, which is meticulously implemented on this server, plays a pivotal role. It intelligently and autonomously learns to optimize resource allocation policies by processing and analyzing various network states. These states include

*This research is partly supported by an NSF I/UCRC Grant (1822104)

diverse parameters such as the lengths of energy queues, data rates from the EH-WBAN sensors, and time slot allocations.

In our proposed system, the actor module of the DRL framework is responsible for executing actions. These actions include dynamically adjusting the allocated time slots, judiciously choosing a relay node for data transmission, and selecting the most efficient transmission mode based on current network conditions. Parallel to this, the critic module undertakes the critical task of evaluating the effectiveness of the actions taken by the actor. This continuous feedback loop allows the actor module to refine and enhance its policy decisions progressively, with the overarching goal of maximizing the energy efficiency of the EH-WBAN system.

Data transmission within this system can occur through two distinct modes: cooperative and direct. In the cooperative mode, the data transmission involves two hops, effectively using intermediate nodes to relay data. In contrast, the direct mode limits the data transmission to a single hop, directly from the sensor to the gateway.

The decision to select between these transmission modes is governed by a binary variable within the system. Additionally, the Medium Access Control (MAC) layer of our system employs the Time Division Multiple Access (TDMA) protocol. This protocol divides the communication channel into multiple time slots, allowing for efficient and orderly data transmission, minimizing interference and maximizing throughput in the network. This advanced, dynamic approach to managing and transmitting data in the WBAN ensures optimal utilization of resources, enhancing the overall performance and reliability of the network in monitoring and managing patient health data. In the case of direct transmission mode $\alpha\, Rn = 1$, Two constraints as in Eq. (1) and (2) are considered; Eq. (1) indicates that the sink can only receive data from one sensor at each time slot, Eq (2) indicates that each sensor assigned at most to a one-time slot to forward the traffic in each time frame, and is represented as [17],

$$\sum_{n=1}^{N} D_{R_N}^{k} \leq 1, \mathrm{k} \in \psi, \qquad (1)$$

$$\sum_{k=1}^{K} D_{R_n}^{k} \leq 1, \ \mathrm{n} \in (1, 2, \ldots, \mathrm{N}), \qquad (2)$$

Where D_{Rn}^{k} represents the data of the n^{th} WBAN sensor forwarded on k^{th} time slot time using a binary variable. We assume that the WBAN can forward the traffic on a single relay, and each relay node can forward the traffic from a single source node at a time, and the constraints can be seen in Eq. (3) and (4) as,

$$\sum_{m=1, m\neq n}^{N} C_{R_n \to s_m}^{k} \leq 1, \sum_{n=1,\ n\neq m}^{N} \delta_{R_n \to s_m}^{k} \leq 1, \quad (3)$$

$$\sum_{n=1, n\neq m}^{N} C_{R_n \to H}^{k} \leq 1, \ \sum_{m=1,\ m\neq n}^{N} \delta_{R_n \to H}^{k} \leq 1 \qquad (4)$$

Where $C_{R_n}^{k}$ represents that the data of n^{th} node can be forwarded on k^{th} time slot of the channel. The transmission rate of the direct mode and cooperative mode, as in Eq. (5) and (6) are used for the transmission of the traffic that can be written according to Shannon's theorem as follows [17],

$$T_n^d = \sum_{k=1}^{K} D_{R_n}^{k} . B. log_2\left(1 + SINR_{n,k}^{d}\right) \qquad (5)$$

$$T_n^{c,\ s\to r} = \sum_{\substack{m=1 \\ m\neq n}}^{N} \sum_{k=1}^{K} C_{R_n \to R_m}^{k} . B. log_2\ (1 + SINR_{n,m,k}^{s\ \to r}) \qquad (6)$$

Where, T_n^d shows the data rate of n^{th} sensor in direct transmission mode and T_n^c is the data rate of the n^{th} body in cooperative transmission approach. The data is stored as packets in the device's buffer with an average rate of λd [18]. We assume the buffer space is finite and follows a FIFO. In timeslot k, IQ_{Rn}^{k} represents the instantaneous queue length at the n^{th} sensor and IQ_{Rn}^{max} denotes the maximum queue length of the device that can be written as follows [18],

$$IQ_{Sn}^{k} = min\left\{IQ_{Tn}^{max}, IQ_{Tn}^{k-1} min\left\{\left\lfloor \frac{C_{Tn}.T_n^d + (1- C_{Tn}).T_n^c}{S_{data}} \right\rfloor, IQ_{Tn}^{k-1}\right\} + A_{Rn}^{k-1}\right\} \qquad (7)$$

In the above equation, traffic packet size is denoted by S_{data}, and $\frac{C_{Tn}.T_n^d + (1- C_{Tn}).T_n^c}{S_{data}}$ is the instantaneous service rate of transmission in the $(k-1)^{th}$ timeslot of the n^{th} sensor, and A_{Sn}^{k-1} is the arriving traffic packet.

The proposed system model utilizes the EH model as in [19], where the energy harvested in the k time slots by the n^{th} WBAN sensor is denoted by $\{EH_{n,1}, EH_{n,2}, \ldots, EH_{n,t}, \ldots, EH_{n,K}\}$ that shows the sequence of energy harvested in a transmission frame. As a result, the instantaneous energy with a queue length can be represented as,

$$IQ_{Tn}^{k} = \min\left\{IQ_{Tn}^{max}, Q_{Tn}^{k-1} - min\left\{\left\lfloor \frac{P_n.k-1}{PS_{energy}} \right\rfloor, IQ_{Tn}^{k-1}\right\} + In, k-1\right\} \qquad (8)$$

where IQ_{Tn}^{k} is represented as instantaneous energy sequence length. IQ_{Tn}^{max} is denoted for the max energy sequence length of body sensors. PS_{energy} is the energy packet size. $P_n.k-1$ denotes the transmission power of the body sensor in the k − 1th time slot. I$n, k-1$ shows the time sequence of energy harvested in a transmission frame at the k-1 time slot.

The objective function (OF), which is the energy efficiency of the n^{th} WBAN in the k time slots for the proposed system, can be mathematically represented as,

$$OF_{Rn}^{k} = \frac{C_{Sn}.T_n^d + (1- C_{Sn}).T_n^c}{P_{n,k}} \ \forall_n \in (1,2, \ldots, N), \forall_n \in \varphi \ (9)$$

We define average efficiency problem as,

$$OF = \frac{1}{N} . \sum_{k=1}^{K} \sum_{N=1}^{KN} OF_{\mathrm{Sn}}^{k} \quad (10)$$

Finally, the proposed energy-efficiency in EH-WBAN can be formulated as,

$$max \quad \mathrm{OF},$$

subject to:

$$\sum_{n=1}^{N} D_{T_N}^{k} \leq 1, \mathrm{k} \in \psi, \quad (10\ \mathrm{a})$$

$$\sum_{k=1}^{K} D_{T_n}^{k} \leq 1, \qquad \mathrm{n} \in (1, 2, \ldots, \mathrm{N}), \quad (10\ \mathrm{b})$$

$$R_{\mathrm{n}}^{d} = \sum_{k=1}^{K} D_{T_n}^{k} . B. log_2\left(1 + SINR_{n,k}^{d}\right) \quad (10\ \mathrm{c})$$

$$T_{\mathrm{n}}^{c,\ s \to r} = \sum_{\substack{m=1 \\ m \neq n}}^{N} \sum_{k=1}^{K} C_{T_n \to T_m}^{k} . B. log_2\left(1 + SINR_{n,m,k}^{s \to r}\right) \quad (10\ \mathrm{d})$$

$$\sum_{k=1}^{K} C_{s_n \to s_m}^{k} \sum_{k=1}^{K} C_{s_n \to H}^{k} \leq 1\ n \neq m \quad (10\ \mathrm{e})$$

$$\sum_{k=1}^{K} C_{s_n \to s_m}^{k} \sum_{k=1}^{K} C_{s_n \to H}^{k} \leq 1\ n \neq m \quad (10\ \mathrm{f})$$

$$\sum_{k=1}^{x} C_{\mathrm{Sn} \to \mathrm{Sm}}^{\mathrm{k}} - \sum_{k=x+1}^{K} C_{\mathrm{Sm} \to \mathrm{H}}^{k} \geq 0 \quad (10\ \mathrm{g})$$

$$\sum_{k=1}^{K} IQ_{\mathrm{Sn}}^{k} - \sum_{k=1}^{K} \left\lfloor \frac{Pn,k-1}{PS_{energy}} \right\rfloor \leq IQ_{\mathrm{Sn}}^{max} \quad (10\ \mathrm{h})$$

3. PROPOSED AC-DRL FRAMEWORK

In our reformulated RL problem for EH-WBAN, we frame it as a Markov Decision Process (MDP) with four key components: action a_t, state-space s_t, reward r_t, and transition probability P_t. Recognizing the limitations of traditional RL methods like Q-learning in extensive state-space scenarios, we've adopted an actor-critic (AC) deep reinforcement learning (DRL) framework.

This framework divides into two parts: the actor, responsible for actions to maximize cumulative rewards (policy improvement), and the critic, which evaluates these actions using a function approximator under policy π (policy evaluation).

The critic's function approximator adaptively refines the actor's policy for optimal resource allocation. We delve into both components, highlighting how they process data from WBAN sensors $w_n^k \in w$ such as data D_n^k and energy queue length $E_n^{k,}$ and how the actor adjusts actions like relay node selection, transmission mode, and time slots for maximum efficiency rewards, subsequently evaluated and refined by the critic.

A. Actor part

The objective of the actor part is to search for the best θ under a given policy $\pi\,\theta$ to maximize the expected reward $J(\pi\,\theta)$. The policy gradient technique is used to update the policy of actor with respect to varying θ as,

$$\theta_{t+1} = \theta_t + a\,\nabla_{\theta t}\, log\pi_{\theta t}\,(s_t, a_t)\delta_t. \quad (11)$$

The expected total reward while following a policy π can be mathematically written as,

$$\nabla_\theta J\,(\theta) = E_{\pi\theta}[\nabla_\theta \log \pi_\theta\ (s, a)\ \delta_t] \quad (12)$$

B. Critic part

The function of the critic component is to approximate the actions taken by the actor part and update the policy π. The state-action value function used for function approximation can be written as,

$$Q^{\pi}(s,a) \sum_{i=1}^{n} \theta i\ \alpha i(s, a) \quad (13)$$

The approximation function used by the critic follows a temporal difference (TD) that is used for updating the value of $Q^{\pi}(s,a)$ and is written as,

$$\delta_t = R_t + \gamma^V\,(V_{t+1}) - (V_t) \quad (14)$$

The problem of EH-WBAN is formulated as an MDP, and its details can be seen as follows:

States: The states from the WBAN sensors D_n^k and E_n^k which show the data and energy queue length of the sensors in the n^{th} body sensor, are generated from the EH-WBAN environment. The states are forwarded from the WBAN environment to the actor-critic framework.

Actions: The action $a_t \in A$ taken by the agent is to vary the resource allocation variables, αR_n is the transmission mode, $\delta k R_n$ shows the relay selection, $p_{n,k}$ is the power allocation and $\beta k R_n$ is the allocation of time slot. The actor component can take the actions to maximize the energy efficiency of the network.

Rewards: The objective of the proposed AC is to maximize the energy efficiency as shown in Eq. (10) of the network.

Algorithm 1

1. **Initialize** the parameters of the AC framework $\theta, \gamma, and\ learning\ rates$
2. **for** *t=1.. T* : **do**
3. Generate action according to $\pi\,\theta(a|\mathrm{s})$
4. Observe the reward r_t and next state s_{t+1}
5. Store the observations in tuple (a_t, s_t, r_t, P_t)

6. Select mini-batch from samples
7. Update parameters of critic

$$\delta_t = R_t + \gamma^V (V_{t+1}) - (V_t)$$

8. Update parameters of actor

$$\nabla_\theta J(\theta) = E_{\pi\theta}[\nabla_\theta \log \pi_\theta (s, a)\, \delta_t]$$

9. **end for**

Algorithm 1 shows the proposed AC framework for resource allocation in EH-WBANs. Initially, the agent in the actor part explores the environment and performs actions randomly, such as relay node, transmission mode, time slot, and transmission power without, considering the queue and data state of WBANs. The learning rate α, weights θ, and discount factor γ of the AC framework are initialized (line 1).

The agent initially takes random action following a policy $\pi\theta(a|s)$ (line 3), and receives a reward value EE in our framework and next-state (line 4),.

The agent's experience with the EH-WBAN framework is stored in a tuple form as in (line 5). After sufficient samples are collected, the AC framework takes a mini-batch of the samples for the training. The critic uses a function approximator and, based on reward, minimizes the error by using the TD as in (line 7).

The critic forwards the updated weights to the actor as in (line 8), and the agent tunes its weight. After training, the agent will try to take those actions (relay node, transmission mode, time slot, and transmission power) that can maximize the EE considering the data and queue state of the sensor in the EH-WBAN.

4. SIMULATION RESULTS

The simulation setup for our actor-critic (AC) framework in training the EH-WBAN is detailed in Table 1. We benchmarked our AC framework against a standard model [17] that utilizes Q-learning RL for resource allocation in EH-powered WBANs.

For evaluating the effectiveness of our proposed scheme against this benchmark, we used two key metrics: energy efficiency and average delivery probability. This comparative analysis aims to demonstrate the enhanced performance of our AC framework in resource allocation efficiency within EH-WBANs.

Parameters	Actor	Critic
Hidden layers	2	2
Nodes	32	32
Activation function (hidden layer)	ReLU	ReLU
Activation function (output layer)	Sigmoid	Linear
Learning rate	0.9	0.9
Batch size	64	64
Discount factor	0.5	0.5
Number of episodes	200	200
Simulator	Python 3.6	
Library	Keras	

Table 1

a) Energy efficiency per episode

Figure 1 illustrates the enhanced performance of our proposed actor-critic (AC) framework compared to benchmark schemes in energy efficiency, particularly as the number of episodes increases. This figure clearly demonstrates that our AC technique effectively explores the WBAN environment and learns the optimal resource allocation policy. It notably achieves a 24% improvement in energy efficiency over the benchmark model. Additionally, the scalability of our AC technique is evaluated by expanding the WBAN network size, reflecting its potential in large-scale intelligent healthcare networks, particularly in the context of IoT-generated healthcare traffic.

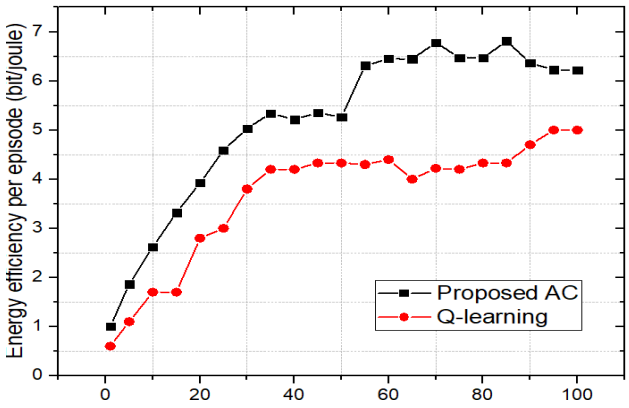

Fig. 1 Energy efficiency comparison with increasing number of episodes.

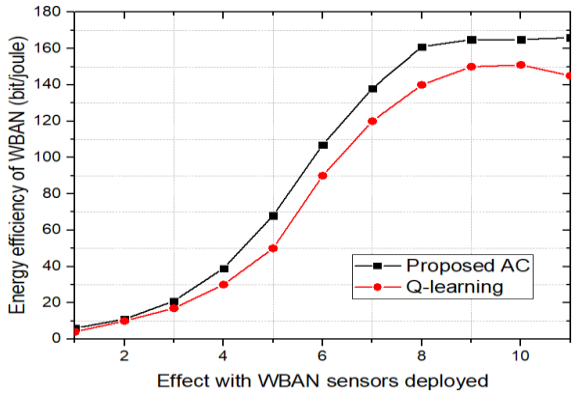

Fig. 2 Energy efficiency comparison with increasing number of WBAN sensors

b) Energy efficiency with varying sensors

Figure 2 demonstrates that our AC framework boosts energy efficiency by 15% compared to the benchmark, especially as the number of nodes rises. The improvement is particularly notable with 11 sensors, illustrating AC's ability to optimize resource allocation in larger networks. This contrasts with the benchmark's Q-learning algorithm, which underperforms in expanded networks.

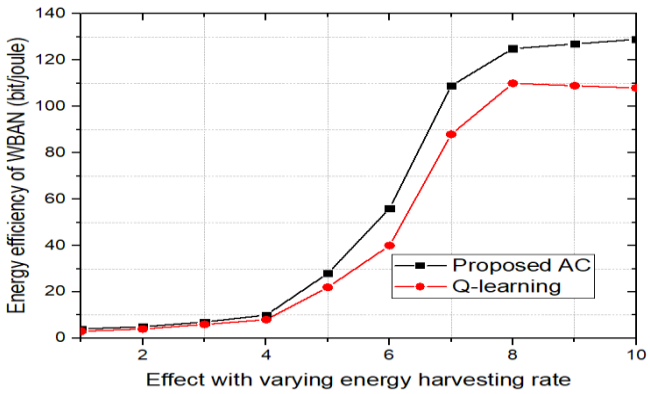

Fig. 3 Energy efficiency comparison with varying energy harvesting rate

c) Energy efficiency with varying harvesting rate

Figure 3 reveals that the proposed AC technique significantly surpasses the traditional Q-learning in energy efficiency, especially when the energy harvesting rate exceeds eight packets per second, achieving a 20% improvement. The AC's strength lies in its ability to discern the interplay between transmission mode, power allocation, and energy harvesting, unlike Q-learning, which struggles with extensive EH-WBAN networks due to its inability to evaluate actions effectively in large state-spaces.

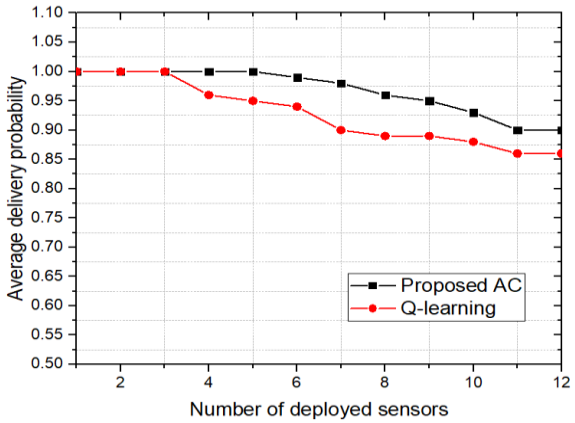

Fig. 5 Average delivery probability

d) Average delivery probability

Figure 4 indicates that the AC technique outperforms Q-learning in delivery probability, especially important given WBANs' diverse quality of service requirements. The AC method not only meets these requirements but also shows superior learning and exploration capabilities for resource allocation, resulting in higher delivery probabilities. While Q-learning initially performs well, it struggles and its effectiveness drops significantly as the network size increases, showcasing the scalability and higher performance of the proposed AC approach.

5. RELATED WORK

Wireless Body Area Networks (WBANs) serve as a key technology in healthcare, utilizing sensors implanted in the human body for continuous monitoring of physiological data. However, a major challenge in WBANs is maintaining consistent network functionality due to energy limitations. To overcome this, Energy Harvesting (EH) techniques have been proposed to extend the network's lifespan by harnessing energy from environmental sources, thus improving the network's sustainability and reliability in healthcare applications. In the field of WBAN, optimization techniques like Particle Swarm Optimization (PSO) have been employed for resource allocation in Energy Harvesting (EH)-based networks, as cited in references [12-15]. However, due to the complex and mobile nature of EH-WBAN systems, these conventional optimization methods struggle in accurately formulating the network's mathematical representation. This limitation points to the need for model-free approaches to effectively tackle resource allocation challenges in EH-based WBANs. Reinforcement Learning (RL) is a model-free approach in which an agent learns an optimal policy through interaction with an environment, receiving rewards for actions taken. In the context of EH-WBAN, researchers have explored RL, specifically Q-learning techniques, to devise resource allocation policies, as noted in reference [17]. This

approach allows for adaptive learning in complex and dynamic network environments. However, the Q-learning technique fails to perform well when the number of state-space such as deployed sensors increases in EH-WBAN. This highlights a notable gap in current literature, underscoring the need for advanced RL methods capable of efficiently managing resource allocation in large and continuously expanding WBAN networks.

6. CONCLUSION

In this study, we have innovatively applied an actor-critic-based Deep Reinforcement Learning (DRL) technique to optimize resource allocation in Energy Harvesting-powered Wireless Body Area Networks (EH-WBAN). Our algorithm stands out for its adaptive learning from the network's dynamic, varied parameters, demonstrating superior energy efficiency optimization over traditional benchmark models. Notably, it excels in handling expanded network state-spaces, showcasing its practical applicability in real-world EH-WBAN scenarios.

Future research directions include exploring federated learning. This approach aims to further improve the generalization capability of our DRL model within EH-WBANs, potentially enhancing its performance across a broader range of network environments. The integration of federated learning is expected to enrich the model's learning process, drawing from diverse data sources while maintaining essential privacy and security in healthcare data handling. Overall, our study lays a foundational step towards more intelligent, efficient, and scalable wireless healthcare monitoring systems.

References:

1. M. Salayma, A. Al-Dubai, I. Romdhani, and Y. Nasser, "Wireless body area network (WBAN): A survey on reliability, fault tolerance, and technologies coexistence," ACM Comput. Surv., vol. 50, no. 1, 2017, Art. no. 3
2. C. Dagdeviren, Z. Li, and Z. L. Wang, "Energy harvesting from the animal/human body for self-powered electronics," Annu. Rev. Biomed. Eng., vol. 19, no. 1, pp. 85–108, 2017.
3. R. Zhang, H. Moungla, J. Yu, and A. Mehaoua, "Medium access for concurrent traffic in wireless body area networks: Protocol design and analysis," IEEE Trans. Veh. Technol., vol. 66, no. 3, pp. 2586–2599, Mar. 2017.
4. M. Razzaque, M. T. Hira, and M. Dira, "QoS in body area networks: A survey," ACM Trans. Sensor Netw., vol. 13, no. 3, 2017, Art. no. 25.
5. Liu, Z., Liu, B., & Chen, C. W. (2017). Transmission-rate-adaption assisted energy-efficient resource allocation with QoS support in WBANs. *IEEE Sensors Journal*, *17*(17), 5767-5780.
6. Ramis-Bibiloni, J., & Carrasco-Martorell, L. (2020). Energy-Efficient and QoS-Aware Link Adaptation With Resource Allocation for Periodical Monitoring Traffic in SmartBANs. *IEEE Access*, *8*, 13476-13488.
7. Askari, Z., Abouei, J., Jaseemuddin, M., & Anpalagan, A. (2021). Energy-Efficient and Real-Time NOMA Scheduling in IoMT-Based Three-Tier WBANs. *IEEE Internet of Things Journal*, *8*(18), 13975-13990.
8. Z. Liu, B. Liu, C. Chen, and C. W. Chen, ''Energy-efficient resource allocation with QoS support in wireless body area networks,'' in Proc. IEEE Global Commun. Conf. (GLOBECOM), San Diego, CA, USA, Dec. 2015, pp. 1–6.
9. B. Liu, Z. Yan, and C. W. Chen, ''Medium access control for wireless body area networks with QoS provisioning and energy efficient design,'' IEEE Trans. Mobile Comput., vol. 16, no. 2, pp. 422–434, Feb. 2017.
10. Liu, Z., Liu, B., & Chen, C. W. (2018). Joint power-rate-slot resource allocation in energy harvesting-powered wireless body area networks. *IEEE Transactions on Vehicular Technology*, *67*(12), 12152-12164.
11. Akhtar, F., & Rehmani, M. H. (2017). Energy harvesting for self-sustainable wireless body area networks. *IT Professional*, *19*(2), 32-40.
12. Huang, C., Zhang, R., & Cui, S. (2014). Optimal power allocation for outage probability minimization in fading channels with energy harvesting constraints. *IEEE Transactions on Wireless Communications*, *13*(2), 1074-1087.
13. Goyal, R., Patel, R. B., Bhaduria, H. S., & Prasad, D. (2021). An energy efficient QoS supported optimized transmission rate technique in WBANs. *Wireless Personal Communications*, *117*(1), 235-260.
14. Panhwar, M. A., Zhong Liang, D., Memon, K. A., Khuhro, S. A., Abbasi, M. A. K., & Ali, Z. (2021). Energy-efficient routing optimization algorithm in WBANs for patient monitoring. *Journal of Ambient Intelligence and Humanized Computing*, *12*(7), 8069-8081.
15. S. Leng and A. Yener, "Resource allocation in body area networks for energy harvesting healthcare monitoring," in Handbook of Large-Scale Distributed Computing in Smart Healthcare, Berlin, Germany: Springer, 2017, pp. 553–587
16. Chen, G., Zhan, Y., Sheng, G., Xiao, L., & Wang, Y. (2018). Reinforcement learning-based sensor access control for WBANs. *IEEE Access*, *7*, 8483-8494.
17. Xu, Y. H., Xie, J. W., Zhang, Y. G., Hua, M., & Zhou, W. (2020). Reinforcement learning (RL)-based energy efficient resource allocation for energy harvesting-powered wireless body area network. *Sensors*, *20*(1), 44.
18. Mitran, P. On optimal online policies in energy harvesting systems for compound poisson energy arrivals. In Proceedings of the IEEE International Symposium on Information Theory, Cambridge, MA, USA, 1–6 July 2012