# 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:*** Wireless body area network (WBAN) is a kind of network that provides continuous monitoring of health parameters. One of the critical constraints for satisfying the quality of service (QoS) in WBAN is the limited energy of the sensors implanted in the human body. Energy harvesting (EH)- powered WBAN is a paradigm that collects the power from surroundings to overcome the energy constraint of the sensors. However, the heterogeneous and time-varying nature of the EH states of the sensors need to be optimized while learning the optimal resource allocation strategy. In this paper, we propose an actor-critic (AC) deep reinforcement learning (DRL) that optimizes the transmission mode, relay selection, transmission power, and time slots with the energy constraint of the sensors to improve the energy efficiency of the EH-enabled WBAN. The simulation results show that the proposed AC technique can efficiently learn the optimal resource allocation policy and achieve a performance improvement in average delivery probability and energy efficiency.

## 1. Introduction:

With the rapid development in wireless network paradigm, the wireless body area networks (WBANs) are a promising approach that continuously monitors the real-time vital signs of the human body and serves different applications (e.g., sports, entertainment, and military) [1-2]. WBAN architecture consists of lightweight and low-power sensors implanted in the human body to monitor the physiological data. The monitored physiological traffic is further forwarded to remote servers using the existing wireless infrastructure to analyze various healthcare and medical applications. The characteristics of the WBAN are highly heterogeneous and dynamic, and some physiological parameters, such as electroencephalogram (EEG) and electrocardiogram (ECG), need continuous data transmission with stringent quality of service (QoS) requirements [3]. Consequently, one of the critical constraints in limiting the data transmission performance in WBAN is the limited energy of the deployed sensors in the human body [4]. To address the energy-efficiency issue in WBANs, the researchers have investigated different energy-efficient schemes in terms of power control, MAC protocol, and cross-layer optimization techniques to prolong the lifetime of the WBAN [5-7]. The researchers in [8] proposed an optimization approach to maximize energy efficiency while considering the transmission power, transmission rate, and QoS as constraints. In another work, the authors proposed a time division multiple access (TDMA) that dynamically adjusts the transmission duration and order of a TDMA slot to minimize the energy consumption [9]. However, one of the challenges of keeping the WBAN network *uninterrupted* in these optimization techniques cannot be ensured due to the depletion of energy of the nodes, which is a critical performance requirement of the WBAN [10].

To address the uninterrupted issue in WBANs, the emerging concept of energy harvesting (EH) has been considered a promising approach to overcome the energy efficiency in WBANs and improve the performance of the communication systems in terms of network lifetime and throughput [11-12]. The EH can collect energy from different ambient sources (e.g., heat, light, and electromagnetic radiations) or the human body and convert it into electric energy to continuously supply the power [11]. Moreover, the WBAN sensors can collect energy from different EH sources and satisfy more stringent performance requirements of the applications.

The EH-powered WBAN resource allocation techniques are divided into offline and online schemes in the literature. In an offline resource allocation strategy, the knowledge of the WBAN network states such as data state, channel state, and energy state is assumed ideally [12-14]. On the other hand, the online EH-powered resource allocation techniques only need the statistical information of channel states, data states, and energy states [15]. The optimization techniques formulated for resource allocation in WBANs require a mathematical representation of an environment. However, WBANs will have highly heterogeneous

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

and dynamic characteristics, and existing optimization techniques formulated for the resource allocation in EH-enabled WBANs cannot perform well. As a result, a novel model-free technique is required for achieving optimal resource allocation techniques in EH-WBANs.

Recently, the novel paradigm of artificial intelligence (AI), known as reinforcement learning (RL), has been proposed, and it has shown performance improvement in heterogeneous and dynamic environments [9]. The RL problem is formulated as a Markov decision problem (MDP), where an agent interacts with an environment and receives rewards, and next state, based on the rewards, the agents perform actions that can maximize the sum of cumulative rewards. Inspired by the RL, the authors proposed a resource allocation technique for optimizing the energy efficiency of WBAN [16]. There is very little work integrating the RL framework in EH-WBANs to achieve energy efficiency. In recent work, the authors modeled the resource allocation issue in EH-powered WBANs as an MDP and proposed a Q-learning technique to learn the optimal energy efficiency [17]. The Q-learning performs well in the scenarios where the network state-space is small. However, the WBANs will generate a massive flux of traffic, and Q-learning fails to learn the optimal resource allocation policy due to discrete state-space in EH-WBANs. The Q-learning technique cannot evaluate the actions taken by the agent to optimize its policy. To address the issue, we propose an actor-critic (AC) based DRL framework known as (AC-DRL) that uses a function approximator to learn the optimal resource allocation policy even in complex scenarios, which is not in the case of Q-learning techniques. In the proposed AC-DRL framework, the actor component performs actions, and the critic component evaluates the agent's action to further improve the resource allocation policy in large and dynamic EH-powered WBANs. Based on the evaluation of the actions taken by the critic component on the rewards, the agent in the critic component will take those actions that can maximize the sum of cumulative rewards that increase the energy efficiency in our framework.

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.

## 2. SYSTEM MODEL

The proposed system model consists of an actor-critic-based DRL framework where multiple EH-enabled sensors are in the WBAN. Different types of sensors are implanted in the human body to record physiological parameters, such as electroencephalography sensor (EEG), motion sensor, electrocardiogram sensor (ECG), glucose sensor, and electromyography sensor (EMG). The traffic from the WBAN is forwarded to the server by using the base station (BS) or personal digital assistant (PDA) as a gateway. The proposed AC DRL framework is implemented in the centralized medical server. The network states such as time slots, energy queue lengths, and the nth body sensor data rate from the EH-WBAN environment are forwarded to the centralized server where the proposed AC framework is implemented to learn the resource allocation policy in EH-WBAN intelligently. The actor module is responsible for taking actions such as varying the time slot, relay node, and transmission mode. As the actor performs the action, the critic model is used for evaluating the performance of the action taken by the agent. Based on the critic's evaluation, it updates the actor policy so the agent can take the actions that can maximize the energy efficiency of the proposed EH-WBANs. The data can be transferred by using cooperative and direct transmission modes. In cooperative mode, the data can be forwarded to only two hops, and in the case of direct transmission, mode traffic can be only forwarded to a single hop. The binary variable is defined to select the transmission modes. We adopted the network model in [17] to validate our approach. The time division multiple access (TDMA) is used in the MAC layer, where the channel is divided into k time slots. 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],

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

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

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,

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

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

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],

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

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

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],

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

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,

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

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. $In, k-1$ shows the time sequence of energy harvested in a transmission frame at the k-1 time slot.

we outline the energy efficiency of the network as the ratio of the transmission rate to the consumed transmission power. The objective function (OF), which is the energy efficiency of the $n^{th}$ sensor in the $k$ time slots for the proposed system, can be mathematically represented as,

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

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,

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

subject to:

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

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

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

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

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

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

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

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

## 3. PROPOSED AC-DRL FRAMEWORK

The RL problem is formulated as an MDP and is based on four components $\{a_t, s_t, r_t, P_t\}$, where $a_t$ is the actions taken by the agent, $s_t$ shows the state-space, $r_t$ represents the rewards, and $P_t$ is the transition probability. The existing RL , such as Q-learning fail to perform well when the network state-space grows exponentially. As a result, we formulated the problem of resource allocation in EH-WBAN as an actor-critic (AC) framework. The AC DRL is divided into two components; actor and critic. The agent's

responsibility in the actor is to take those actions that can maximize the sum of cumulative rewards, and this process is known as policy improvement.

On the other hand, the critic evaluates the action taken by the actor by using a function approximator under a policy π, and this process is called policy evaluation. The function approximator in the critic is used to adaptively update the parameters of the actor component policy to learn the optimal resource allocation in EH-WBAN. Next, we discuss in detail the actor and critic components and their formulation in detail. In a nutshell, the states from the WBAN sensors denoted as $w_n^k \in w$, such as $D_n^k$ and $E_n^k$ denoting the data and sensors' data and energy queue length, are forwarded to the actor-component. The actor part performs actions by varying the relay node, transmission mode, and time slots. As the actor takes action, the reward, which is our framework's energy efficiency, is forwarded to the critic part. Based on the reward function, the critic evaluates the actor's action and updates the actor's parameters to take those actions with the highest reward.

### A. Actor part

The objective of the actor part is to search for the best $\theta$ 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 $\theta$ as,

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

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

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

### 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,

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

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,

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

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, $\alpha R_n$ 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 $\pi\,\theta(a|s)$
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 $\theta$, 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 parameters for the proposed actor-critic framework for the training of the EH-WBAN can be seen in table 1. We compared the proposed AC-framework with the benchmark scheme [17] which uses a Q-learning RL technique to learn the resource allocation policy in EH-powered WBAN. The energy efficiency and average probability are the metrics used to compare for evaluating the effectiveness of the proposed scheme with benchmark paper

| 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 | |

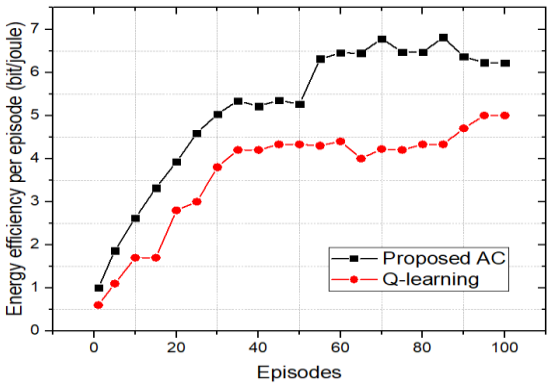


**Fig. 2 Energy efficiency comparison with increasing number of episodes.**

### a) Energy efficiency per episode

Fig 2 shows the performance of the proposed AC with the benchmark schemes in terms of energy efficiency with increasing the number of episodes. It can be clearly seen from Fig. 2 that the proposed AC technique can explore the WBAN environment well and learn the optimal resource allocation policy. The agent can select actions that can maximize the energy efficiency of the EH-WBAN network and achieve an improvement of 24% compared to the benchmark paper.

The performance of the proposed AC technique is evaluated by increasing the size of the WBAN network as intelligent healthcare networks will be based on emerging IoT applications that will generate a massive amount of healthcare traffic for WBAN sensors. So, to evaluate the scalability, we analyzed the proposed AC algorithm in terms of energy efficiency with increasing the WBAN nodes.

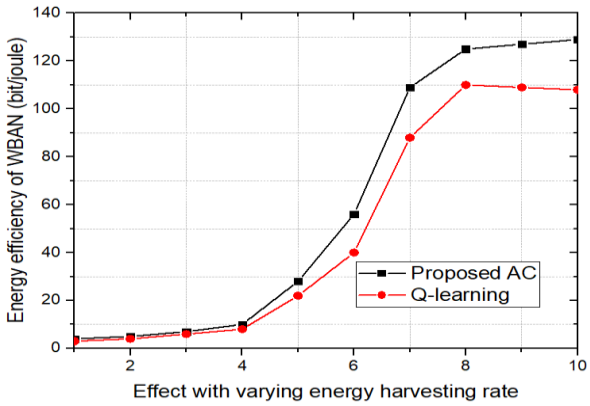


**Fig. 4 Energy efficiency comparison with varying energy harvesting rate**

### b) Energy efficiency with varying harvesting rate

Fig. 4 shows the performance of the proposed AC technique with increasing the effect of the harvesting rate. It can be clearly seen that when the energy harvesting rate is increased beyond eight packets per second, the energy efficiency performance is significantly outperformed compared to the traditional Q-learning technique. The proposed AC framework achieves an improvement of 20% in terms of efficiency than Q-learning. The AC technique can efficiently learn the correlation between the transmission mode, power allocation, transmission mode, and energy harvesting by exploring the EH-WBAN environment. On the other hand, the actions taken by the agent in the Q-learning cannot be evaluated, and it always chooses the action with a higher cumulative reward in the Q-table. As a result, for an extensive EH-WBAN network where the number of network state-space is very high, Q-learning fails to learn the optimal energy efficiency when the network state-space is increased exponentially.

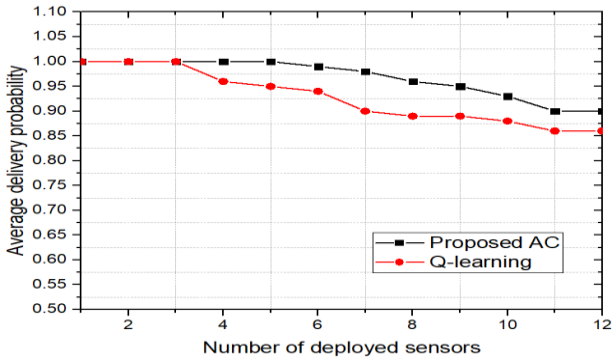


**Fig. 5 Average delivery probability**

### c) Average delivery probability

The WBAN will generate traffic with diverse quality of service requirements, and as a result, the proposed algorithm needs to satisfy the traffic requirements and achieve a higher delivery probability. Fig. 5 shows the performance of the AC technique compared to Q-learning in terms of delivery probability. It can be clearly seen that the AC technique can explore and learn the resource allocation optimally and achieves a higher delivery probability than Q-learning. The Q-learning performs initially well; however, when the size of the network is increased, it fails, and its performance is reduced drastically. As a result, the proposed approach is scalable and can achieve higher performance.

## 6. CONCLUSION

In this paper, we proposed an actor-critic based DRL technique to address the resource allocation issue in EH-powered WBAN. The proposed algorithm is able to learn the dynamic and heterogeneous network parameters of the EH-WBAN and outperforms the benchmark scheme in terms of optimizing the energy-efficiency. Moreover, the AC is also able to perform well when the network state-space is increased. Thus, proposed algorithm can be deployed in practical EH-WBAN systems. In future work, we will try to investigate the federated learning to improve the generalization issue in EH-WBANs.

## 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