ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

Pareto Set Learning for Multi-Objective Reinforcement Learning Applied to Autoadaptive Duty Cycle in Medical Wearable Technology

H. A. Ramasombohitra[1]; H. N. Ramanantsihoarana[1]

[1]SE-I-MSDE, STII Doctoral School, University of Antananarivo, Madagascar

Publication Date: 2025/12/27

Abstract: Portable medical devices designed for continuous monitoring of physiological parameters face a critical trade-off between energy autonomy and monitoring quality. This study presents an adaptive energy management approach based on the Pareto Set Learning - Multi-Objective Reinforcement Learning (PSL-MORL) algorithm for a system powered by photovoltaic solar energy recovery. The proposed approach dynamically generates Duty Cycle (DC) modulated according to the wearer's National Early Warning Score 2 (NEWS2), simultaneously targeting 24-hour energy neutrality operational (ENO) and compliance with clinical recommendations. A comparative study using MATLAB simulation evaluates three strategies: the PSL-MORL algorithm, an intensive monitoring policy (D = 14.3%), and an energy-saving policy (DC = 5.26%). The results demonstrate that the proposed approach guarantees 24-hours autonomy (+42.9% vs. intensive policy), improves the average NEWS2 compliance reward by 58.4% compared to the energy-saving policy, and optimizes the allocation of 142 measurements over clinically critical periods. These performances validate the effectiveness of the PSL-MORL algorithm in reconciling the energy and clinical constraints of autonomous portable medical devices.

Keywords: *Adaptive Power Management, Energy-Aware Scheduling, IoWT, MORL, Pareto Set Learning, Physiological Monitoring.*

How to Cite: H. A. Ramasombohitra; H. N. Ramanantsihoarana (2025) Pareto Set Learning for Multi-Objective Reinforcement Learning Applied to Autoadaptive Duty Cycle in Medical Wearable Technology. *International Journal of Innovative Science and Research Technology*,10(12), 1783-1793. https://doi.org/10.38124/ijisrt/25dec1251

I. INTRODUCTION

Wearable technologies coupled with artificial intelligence (AI) are playing an increasingly important role in the field of connected health, enabling early diagnosis and continuous remote monitoring of patients' health status [1], [2], [3]. However, the requirements associated with the acquisition, processing, and continuous transmission of physiological data demand increased computing power, resulting in significantly higher energy consumption [4]. Connected wearable devices, integrated into the Internet of Wearable Things (IoWT) ecosystem, are generally battery-powered, which limits their operational autonomy [5]. In this context, optimizing energy efficiency is one of the major challenges in ensuring the viability and widespread adoption of IoWT systems in clinical and ambulatory settings [6]. Xcxx.

II. SYSTEM MODEL

The proposed system is based on an architecture consisting of four main components, illustrated in Fig 1: (i) a photovoltaic solar energy recovery module, (ii) an electrochemical storage element (battery), (iii) a set of physiological sensors constituting the system load, and (iv) a power management unit (PMU) responsible for optimizing the device's energy consumption [7]. The battery has a maximum capacity of. We consider that the sensor node has a variable operating cycle. The power management unit uses information on the current battery charge level, light intensity, and sensor data to establish the device's operating cycle. We also assume that a higher duty cycle results in increased energy consumption and higher performance of the sensor node [8].

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

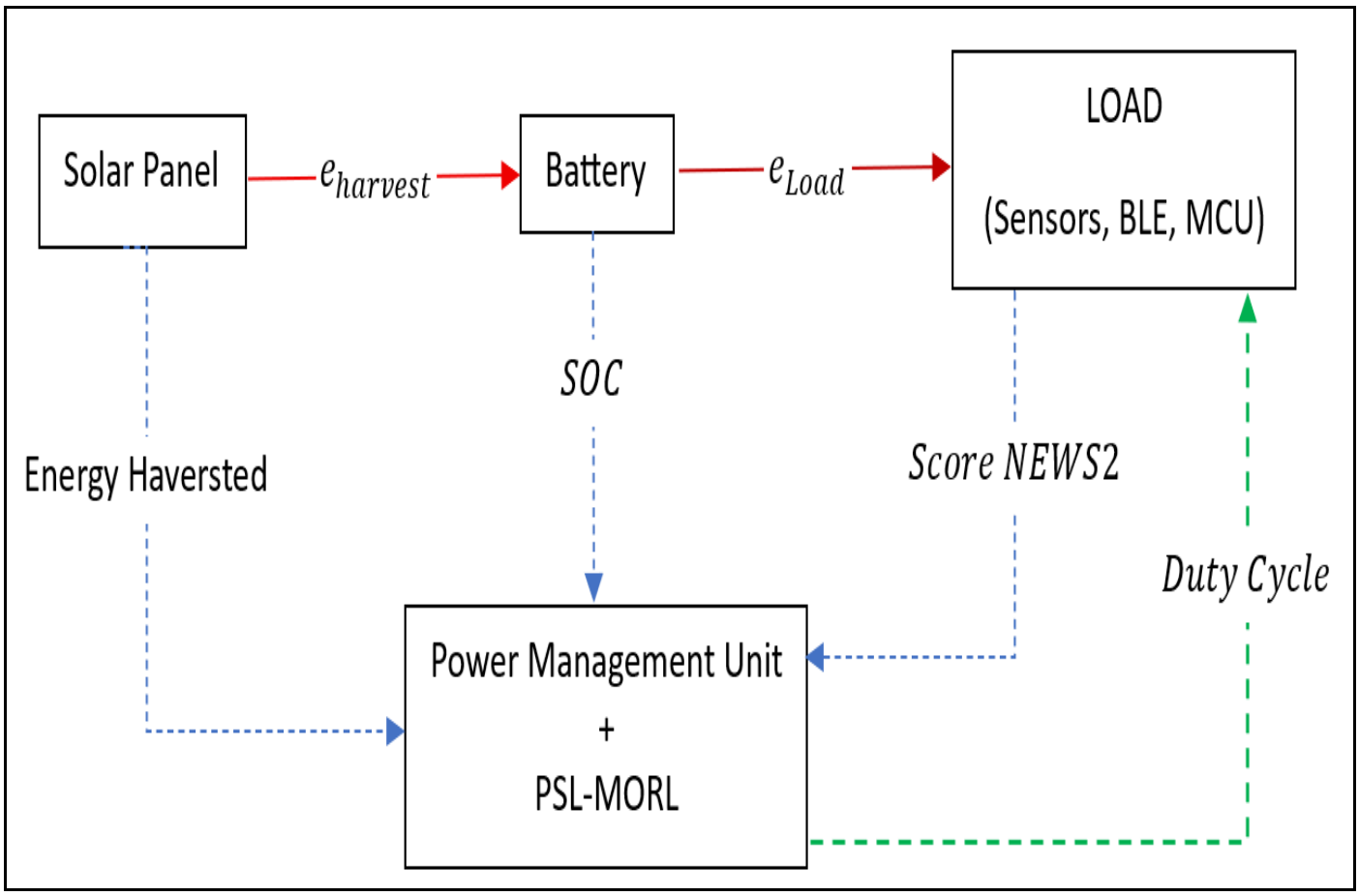

Fig 1 System Model

➢ *Solar Energy Harvester*

The energy harvester (EH) circuit is connected to a small 1W solar photovoltaic cell measuring 80×100 mm, with an energy conversion efficiency of 35% [9]. The monocrystalline panel has an open-circuit voltage of 8.2V, with a voltage and current of 6.4V and 170mA, respectively, at maximum power point. The EH circuit is based on the LT3652 monolithic buck converter, which operates with a minimum input voltage of 4.95 V. The LT3652 optimizes energy transfer from the photovoltaic panel to the battery. Once charging is complete, the LT3652 automatically enters a low-power standby mode, in which supply bias currents are minimized to 85 µA [10].

➢ *Load Components*

- *Microcontroller with BLE Module*

The NINA-B302 module includes a small, self-contained Bluetooth Low Energy 5 circuit paired with the nRF52840 microcontroller [11]. It also incorporates a planar inverted-F antenna, which is tuned using a passive network consisting of inductors and capacitors, creating a perfectly matched transceiver system. The nRF52840 microcontroller includes an Arm Cortex-M4 processor, 1 MB of flash memory, and 256 KB of RAM. It also includes a CryptoCell CC310 cryptographic security unit and a fast 32 MHz SPI interface. This chip operates with a supply voltage range of 1.7 to 5.5 V and has a current consumption of 0.6 µA in standby mode and 4.8 mA in TX mode at 0 dBm.

- *Heart Rate and SpO2-PPG Sensor*

The MAX30102, a clinical-grade sensor, integrates a pulse oximetry module and a heart rate module. The sensor incorporates internal LEDs, light detectors, optical elements, and low-noise electronic components that filter ambient light. This sensor is software-disabled, enabling low-power operation with high sampling rates and good signal-to-noise ratio.

- *Temperature Sensor*

The MAX30205 is a clinical-grade temperature sensor offering 16-bit resolution, ±0.1°C accuracy, an operating range of 0 to +50°C, and a supply voltage ranging from 2.7V to 3.3V. It consumes 600µA during normal operation and 1.65µA in standby mode. This sensor complies with ASTM E1112 for digital electronic thermometers.

➢ *Sleep and Wake-up Algorithm*

The algorithm for sleep and wake functions of the portable device is shown in Fig 2. The system functions in two states: dormant and operational. Initially, the sensor device enters sleep mode.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

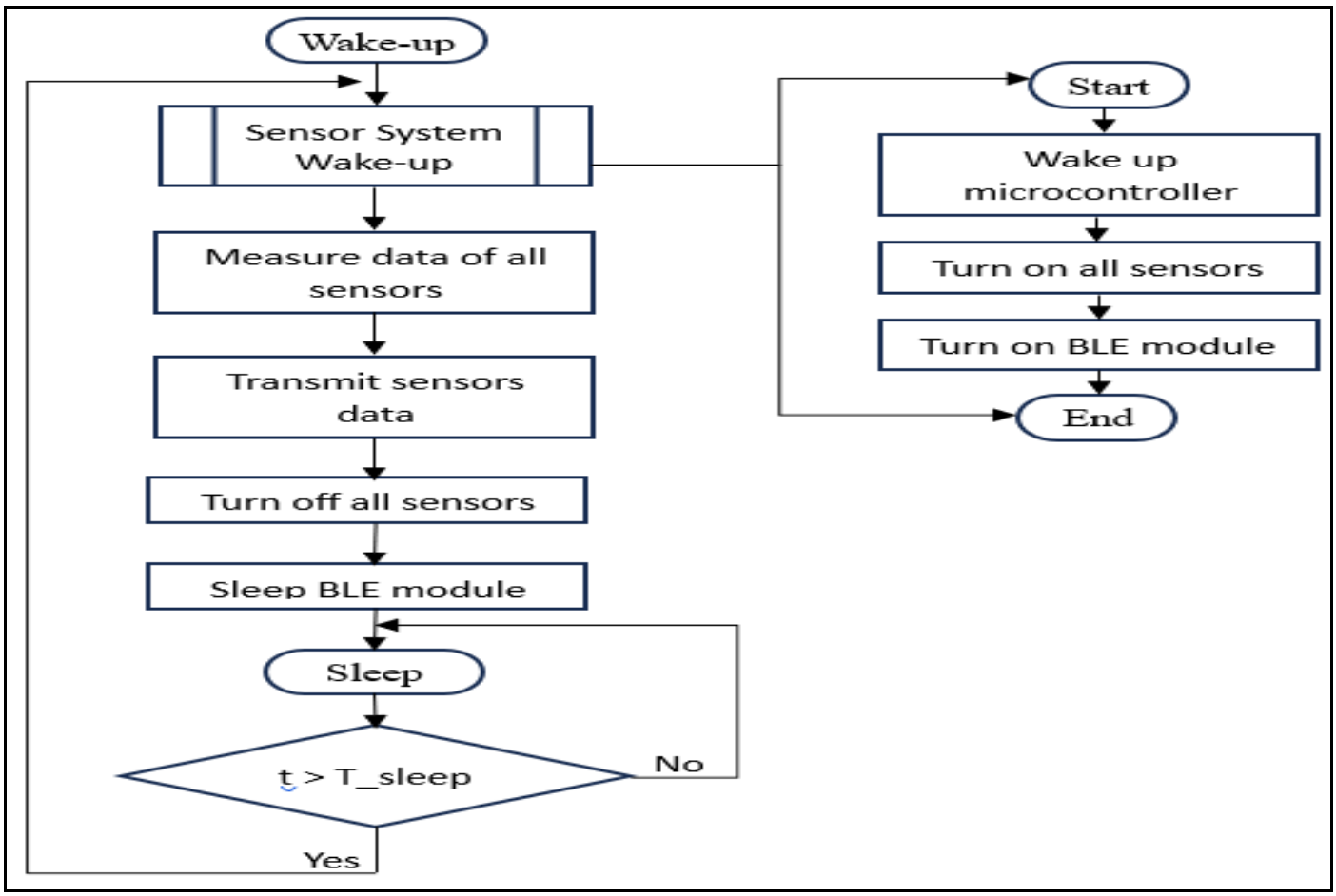

Fig 2 Sleep-Wake Scheduling Algorithm

Following that, two sensors measure the essential signs "temperature, heart rate, SpO2". This crucial data is subsequently transmitted to the server through the BLE module. The gathering and relay of crucial information occur during the T_{ON} phase. Third, the sensors are disabled, after which the BLE module goes into sleep mode. Fourth, the medical device's microcontroller goes into sleep mode during T_{OFF} to save energy. Next, the microcontroller reactivates to enable all sensors and the BLE module during T_{ON}, and this cycle continues. Ultimately, the phases of this algorithm are carried out in a repetitive cycle [12].

➢ *Model Mathematique*

- *Solar Energy Source*

The I–V curve of a single-diode photovoltaic model is described by equation 1. The photogenerated current varies according to irradiance and temperature. Its value is given by Eq 1. The saturation current of the diode depends on temperature T and is expressed by Eq 2.

$$I = I_{ph}(G,T) - I_s(T)\left(e^{\frac{V+IR_s}{nV_t}} - 1\right) - \frac{V+IR_s}{R_{sh}} \quad (1)$$

$$I_{ph}(G,T) = [I_{ph,\text{ref}} + \alpha(T - T_{\text{ref}})] \cdot \frac{G}{G_{\text{ref}}} \quad (2)$$

$$I_s(T) = I_{s,\text{ref}}\left(\frac{T}{T_{\text{ref}}}\right)^3 \cdot \exp\left[\frac{qE_g}{nk}\left(\frac{1}{T_{\text{ref}}} - \frac{1}{T}\right)\right] \quad (3)$$

To represent the captured solar energy, we use the energy function, e(t), which is equivalent to the total amount of energy accumulated up to time t. It is represented by Eq 4 [13].

$$e_{harvest}(t) = \int_0^t P_{harvest}(u)du \quad (4)$$

- *DC-DC Converter*

To ensure energy conversion, the system incorporates a buck converter. The conversion efficiency η of this converter is expressed by Eq 5 [14]. The energy produced is stored in a battery with capacity C. In each interval $t \in T$, the battery receives energy from the harvesting sources and a charging source, if charging is enabled. The harvested energy in an interval ξ_t^H is a random variable with uncertainty because environmental conditions can affect the harvested energy. We have a random variable for the energy harvested for each interval in T. Furthermore, the target application will draw energy from the battery to execute the required tasks. These dynamics of the battery can be captured in Eq 6 [15].

$$\eta = \frac{V_{out}I_{out}}{V_{in}I_{in}+P_{loss}} \quad (5)$$

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

$$E_{t+1}^{B} = E_t^B + \eta \xi_t^H + E_t^I B(t) - E_t^c, \quad t \in T \tag{6}$$

Where E_t^B and E_{t+1}^B are energy at the beginning of the current and next interval respectively, η is the energy harveter efficiency, $\xi_t^H \in \Xi_t^H$ is the realization of harvested energy at interval $t \in T$, $B(t)$ is the battery charging flag, E_t^I is the charging energy, and E_t^c is the energy consumed in the interval t.

- *Power Consumption*

To model the energy required for the device to operate, we use the data line, E_{DEV}, which represents the total energy required to measure, process, and transmit all information related to physical quantities [13].

$$E_{DEV}(t) = \int_0^t P_{DEV}(u)du \tag{7}$$

$$E_{DEV}(t) = P_{avg} * t \tag{8}$$

Most health monitoring applications adopt a similar operating model: data is collected by sensors, analyzed in a processing unit, and then transmitted via a wireless module. This process is repeated over time, and the function of its duty cycle is essential in energy consumption: the shorter the duty cycle, which can be achieved by reducing the active time t_{ON} or extending the inactive phases t_{OFF}, the lower the average power consumption. In fact, the expression of power P_{avg} as a function of duty cycle D is given by Eq 9, and the duty cycle D as a function of t_{ON} is expressed in Eq 10. In most cases, power $P_{OFF} \ll P_{ON}$.

$$P_{avg} = DP_{ON} + (1 - D)P_{OFF} \tag{9}$$

$$D = \frac{t_{ON}}{t_{ON} + t_{OFF}} \tag{10}$$

For a multi-state system, the average power consumption is expressed by Eq 11. For which the sum of the duty cycles of D_i follows Eq 12.

$$P_{avg} = \sum_{i=1}^{k} D_i P_i \tag{11}$$

$$\sum_{i=1}^{k} D_i = 1 \tag{12}$$

Subsequently, the energy required to operate a wireless sensor can be broken down into three essential blocks: for data detection or acquisition P_{ACQ}, for information processing P_{PRC}, and for data communication or networking P_{NET}. In addition, a small portion of the available energy is devoted to system management functions, such as the operation of a real-time operating system (RTOS) or regular system wake-up. The requirements associated with these management activities are grouped together in this contribution P_{SYS}. Together, these elements constitute the overall expression of the system's power consumption P_{DEV}, which is formulated by Eq 13 [16]. Thus, we obtain the average power from Eq 14 by combining Eq 11 with Eq 13.

$$P_{DEV} = P_{SYS} + P_{ACQ} + P_{PRC} + P_{NET} \tag{13}$$

$$P_{avg} = D_{SYS}P_{SYS} + D_{ACQ}P_{ACQ} + D_{PRC}P_{PRC} + D_{NET}P_{NET} \tag{14}$$

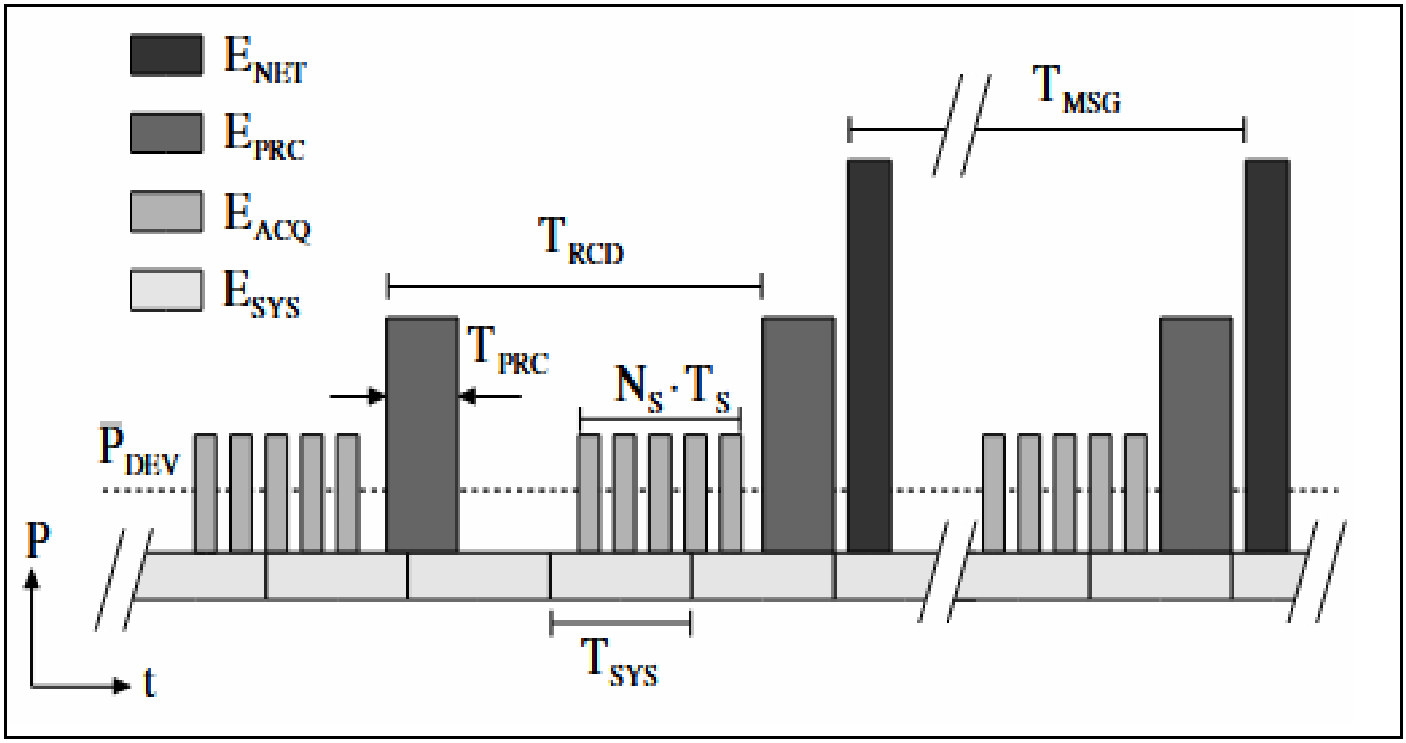

Fig 3 Power Consumption Profile

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

The power P_{ACQ} is formulated in Eq 15, since a sensor node consists of an active sensor, an op-amp, and an analog-to-digital converter (ADC).

$$P_{ACQ} = V_{CC} I_{ACQ\_all} \quad (15)$$

$$I_{ACQ\_all} = I_{sens} + I_{AOp} + I_{ADC} \quad (16)$$

The power P_{PRC} is formulated in Eq 17. In this model, it is a function of the average packet throughput per second λ.

$$P_{PCR} = \lambda V_{DD} I_{PRC} t_{PRC} \quad (17)$$

P_{NET} is the power consumed by the BLE radio module to transmit (Tx) and receive (Rx) packets. It is expressed by Eq 18. This model depends on the data rate D and the RF power P_{out}.

$$P_{NET}(D, P_{out}) = V_{DD} \frac{D}{8L_{data}} [I_{wake} t_{wake} + I_{Rx} t_{Rx} + I_{Tx}(P_{out}) t_{Tx}] \quad (18)$$

III. ADAPTIVE DUTY CYCLE

➢ *Multi-Objective RL Problem*

The energy management problem can be modelled as a Multi-Objective Markov Decision Process (MOMDP) [17], represented by a tuple $\langle S, A, T, \gamma, \mu, R \rangle$, where S is the state space, A is the action space, $T : S \times A \times S \rightarrow [0,1]$ is a probabilistic transition function, $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_m] \in [0,1]^m$ is a discount factor vector, $\mu : S \rightarrow [0,1]$ is a probability distribution over initial states, and $R = [r1, r2, \ldots, rm]^T : S \times A \times S \rightarrow \mathbb{R}^m$ is a vector-valued reward function, specifying the immediate reward for each of the considered $m\ (m \geq 2)$ objectives [18]. In MOMDPs, an agent behaves according to a policy $\pi_\theta \in \Pi$, where Π is the set of all possible policies. A policy is a mapping $\pi_\theta : S \rightarrow A$, which selects an action according to a certain probability distribution for any given state [19]. For brevity, we use π to denote π_θ . The performance of the policy is measured by the associated vector of expected returns, i.e., $\mathbb{F}(\pi) = J^\pi = [J_1^\pi, J_2^\pi, \ldots, J_m^\pi]^T$ with :

$$J_i^\pi = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma_i^k r_i(s_k, a_k, s_{k+1}) \mid \pi, \mu\right] \quad (19)$$

The MORL problem is then formulated in Eq 20.

$$\max_\pi \mathbb{F}(\pi) = \max_\pi [J_1^\pi, J_2^\pi, \ldots, J_m^\pi]^T \quad (20)$$

In MORL, there exists no single optimal policy that can maximize all the objectives simultaneously, and we need the following Pareto concepts [20], [21].

- Definition 1 (Pareto Dominance). For $\pi, \pi' \in \Pi$, π is said to weakly dominate $\pi'(\pi \succeq \pi')$ if and only $if\ \forall i \in \{1, \ldots, m\}$, $J_i^\pi \geq J_i^{\pi'}$; π is said to dominate $\pi'(\pi \succeq \pi')$ if and only if $\pi \succeq \pi'$ and $\exists j \in \{1, \ldots, m\}$, $J_i^\pi > J_i^{\pi'}$.

- Definition 2 (Pareto Optimality). A solution $\pi^* \in \Pi$ is Pareto optimal $if\ \nexists \pi \in \Pi$ such that $\pi \succ \pi^*$.

- Definition 3 (Pareto Set/Front). The set of all Pareto optimal solutions is called the Pareto set, and the image of the Pareto set in the objective space is called the Pareto front.

Each Pareto solution represents an optimal trade-off among the objectives, and it is impossible to further improve one of the objectives without deteriorating any other objectives [21].

➢ *PSL-MORL for ENO-Sensing Priority*

E. Liu et al. propose a novel MORL method called PSL-MORL, which, to the best of their knowledge, is the first MORL method that covers all preferences over multiple objectives and outputs a personalized policy network for each preference [19]. PSL-MORL is a general framework that can be integrated with any single-objective RL algorithm. As part of this research, this model was used to solve the energy management problem of a portable medical device, with the aim of ensuring energy autonomy while meeting the wearer's requirements. With this in mind, Figure 1 presents the basic PSL-MORL algorithm.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

Algorithm 1 : PSL-MORL

Input: Preference distribution Λ, environment $\mathcal{E}$, number E of episodes, number K of weights per episode, replay buffer D, batch size N
Output: Hypernetwork parameters ϕ and primary policy network parameters θ_1

```
Initialize the hypernetwork parameters ϕ and policy network parameters θ1;
Initialize replay buffer D;
for episode e = 1 to E do
    for each step in the episode do
        // Parallel child processes for k ∈ {1, 2, ..., K}
        ω_k ~ SamplePreference(Λ);
        Generate parameters of the policy network by hypernetwork, i.e.,
        θ2 = ϕ(ω_k);
        Obtain the mixed θ using the Parameter Fusion method on θ1 and
        θ2;
        Observe state s; select action a by the policy π_θ;
        Execute action a in environment E, and observe reward r and next
        state s';
        Store transition (s, a, r, s', d, ω_k) in D;
        // Main process
        Sample N transitions from D;
        Update ϕ and θ1 by conducting a single-objective RL algorithm to
        maximize the scalarized return by weight ω_i, i ∈ {1,2,3, ..., N}
    end for
end for
```

Fig 4 PSL-MORL Algorithm

- *State's Definition*

The state of the system at instant t_k is expressed ($S_{batt}(t_k)$, $S_{illu}(t_k)$, $S_{sens}(t_k)$,$S_{dist}(t_k)$) ∈ S.

$S_{batt}(t_k)$ is the state of charge of battery

Table 1 Battery State of Charge Classification Levels

Battery's SOC	$S_{batt}(t_k)$
SOC < 15%	Critical
SOC < 40%	Low
SOC < 70%	Medium
SOC < 90%	Good
SOC > 90%	Full

$S_{illu}(t_k)$ is illuminance (0 to 1000W/m2)

Table 2 Solar Irradiance Classification Levels

Power	**Illumination**
G < 50 W/m2	Dark
G < 200 W/m2	Low
G < 500 W/m2	Medium
G < 800 W/m2	Strong
G > 8000 W/m2	Very Strong

$S_{sens}(t_k)$ is the priority of measurements, classified according to the NEWS standard (0 to 3).

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

Table 3 NEWS Score Clinical Priority Classification

NEWS Score	Priority
0	Monitoring
1	Normal
2	Urgent
3	Critical

$S_{dist}(t_k)$ is the distance from the initial battery level.

Table 4 Energy Distance Classification Levels

Edist Range	Priority
Edist < -30%	Severe_deficit
Edist < -15	Moderate_deficit
Edist < -5	Slight_deficit
Edist < 5	Neutral
Edist < 15	Slight_excess
Edist < 30	Slight_moderate
Edist > 30	Slight_significant

➢ *Action's Space*

The set of actions, A, determines the agent's possible choices. In this study, the scope of action is defined by a set of five Duty cycle, denoted by A= {5.26%,6.25%,7.7%,10%,14.3%}. To save energy, the smallest value of A must be chosen. But in our case, we are talking about a medical device. Thus, the ideal value of A is governed by the NEWS2 standard. These two concepts are therefore paradoxical. Consequently, the objective is not to identify an ideal solution, but a set of relevant compromises. In this sense, we have chosen to apply the multi-objective approach with Pareto Front to evaluate the adequacy of action A with state $S_{sens}(t_k)$ [13].

➢ *Adaptive Composite Reward Function*

To solve the bi-objective optimization problem inherent in the energy management of portable devices under clinical constraints, we propose an adaptive weighted scalarization function based on the patient's condition. This approach falls within the framework of Multi-Objective Reinforcement Learning (MORL) with state-dependent preferences [22]. The composite reward function is defined by the Eq 21.

$$R(s,a) = \alpha\left[w_1.Q_{surv}(s,a) + w_2.Q_{eff}(s,a)\right] + \beta \quad (21)$$

where $s \in \mathcal{S} = \{0,1,2,3\}$ represents the patient's NEWS2 score, , $a \in \mathcal{A} = [0, T_{\max}]$ denotes the sensor's wake time, α and β are the scaling factor and centering term, respectively. The adaptive weights $w_1(s)$ and $w_2(s)$ satisfy the normalization constraint $w_1(s) + w_2(s) = 1$ and are dynamically adjusted according to clinical severity [23], [24]. For non-critical states (s=0), the system prioritizes energy efficiency with $(w_1, w_2) = (0.3,0.7)$, while for critical states (s=3), priority is given to monitoring with $(w_1, w_2) = (0.9,0.1)$. This state-dependent scalarization allows the Pareto front to be explored efficiently while respecting the clinical constraints imposed by the NEWS2 protocol [17], [25]. The parameters α=20 and β=-10 were chosen to center the reward function around zero, thus facilitating the convergence of the reinforcement learning algorithm [26].

- *The Two Conflicting Objectives are Formalized as Follows:*

✓ *Objective 1 - Clinical Monitoring Quality: T*

he function $Q_{\text{surv}}(s,a)$ evaluates the compliance of the measurement frequency with the NEWS2 recommendations, defined in pieces according to the criticality of the condition:

$$Q_{surv}(s,a) = \begin{cases} \min\left(\frac{a}{T_{min}}, 1\right) & si\ s = 0 (Low\ risk) \\ 1 - \frac{|a - T_{opt}(s)|}{T_{opt}(s)} & si\ s \in \{1,2\} (Medium\ risk) \\ \max\left(0,1 - \frac{a}{T_{crit}}\right) & si\ s = 3\ (Severe\ risk) \end{cases} \quad (22)$$

✓ *Objective 2 - Energy Efficiency:*

The $Q_{\text{eff}}(a)$ quantifies energy savings by promoting extended standby times.

$$Q_{eff}(a) = \frac{a}{T_{max}} \quad (23)$$

IV. RESULTS

To validate the effectiveness of the PSL-MORL algorithm in the adaptive management of sleep cycles for a wearable physiological monitoring system, a comparative study was conducted with two reference policies. This evaluation aims to demonstrate the ability of the proposed approach to simultaneously satisfy two fundamental objectives: ensuring device energy neutrality over a 24-hour cycle and maintaining an acquisition frequency compliant with NEWS2 score recommendations.

The three evaluated strategies differ in their management philosophy. The policy (D = 14.3%) adopts an intensive monitoring approach independent of the patient's clinical status, there by maximizing measurement frequency

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

at the expense of energy efficiency. Conversely, the policy (D = 5.26%) prioritizes energy savings by spacing out acquisitions, potentially compromising the quality of physiological monitoring. The PSL-MORL algorithm, for its part, implements a self-adaptive duty cycle dynamically modulated according to the NEWS2 score, aiming for an optimal trade-off between clinical monitoring and energy autonomy.

The simulation protocol was executed in MATLAB over a 26-hours duration, with a solar irradiance of 500 W/m² maintained for 6 hours, an initial state of charge B_0 = 40%, and a time constant T_{on} = 50 s. The clinical data used correspond to the NEWS scores of a patient in a pathological state, thus constituting a test scenario representative of real-world operating conditions.

➢ *Comparison of Energy Profiles*

ENO is characterized by $Edist(t) = B_0 - B(t)$

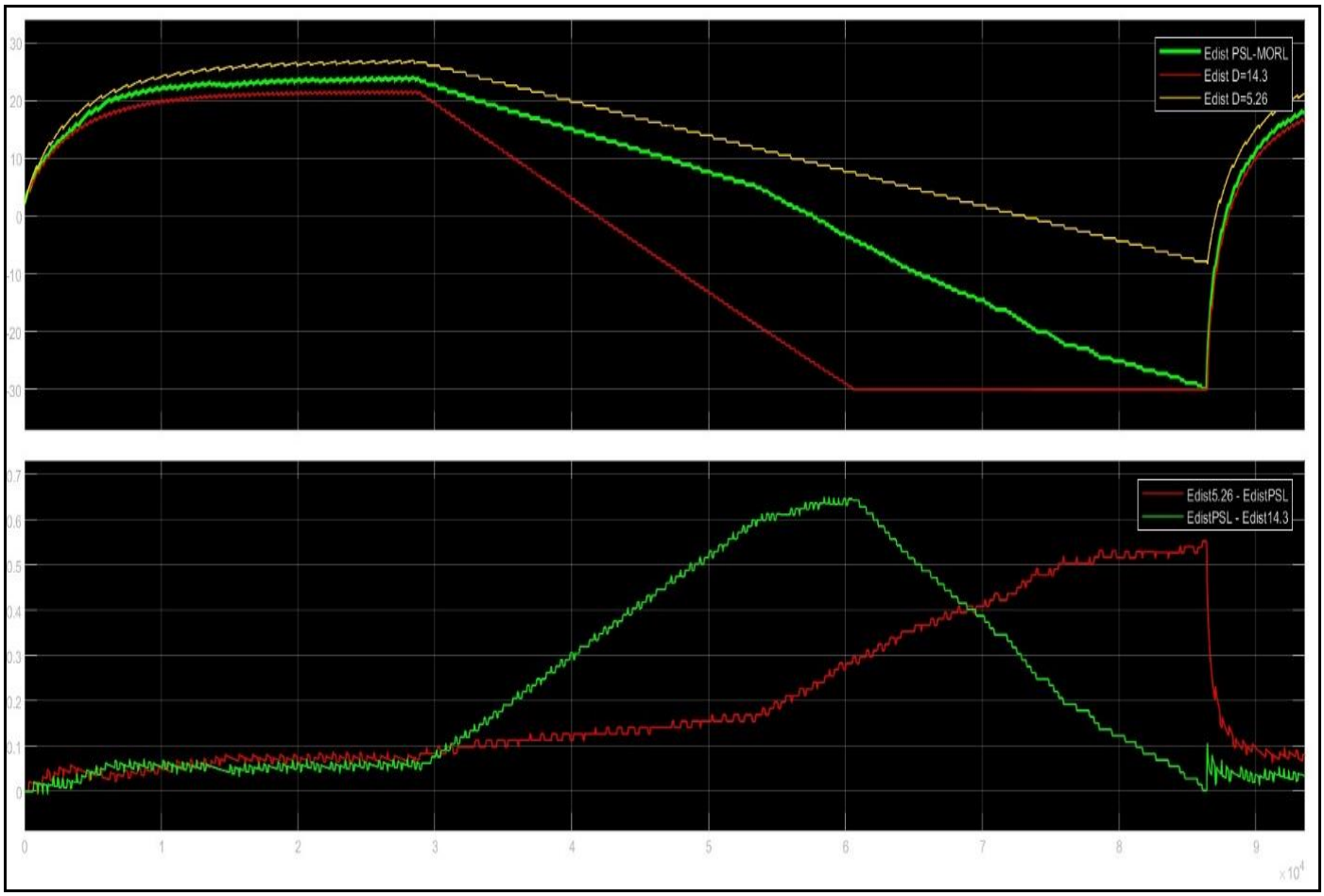

Fig 5 State of Charge Variation Profiles for Different Energy Management Policies.

Table 5 Comparative Evaluation of Energy Autonomy Metrics

Metric	π_1 (D = 14.3%)	π_2 (D=5.26%)	PSL-MORL	Gain vs π_1	Gain vs π_2
Autonomy (h)	16.8	24.0	24.0	+42.9%	0%
ΔSoC Range (%)	[-30%,+21%]	[-9%,+27%]	[-30%,23%]	+64%	-56%

The comparative evaluation of energy profiles demonstrates the superior performance of the PSL-MORL algorithm. In terms of autonomy, the proposed approach guarantees continuous operation for 24 hours, thereby satisfying the objective of energy neutrality over a daily cycle. This result, equivalent to that obtained by policy P2, represents a 42.9% improvement compared to policy P1, whose 16.8-hour autonomy proves insufficient to ensure uninterrupted monitoring.

The analysis of state of charge (SoC) variation relative to the initial value B_0 confirms the effectiveness of the proposed strategy. The PSL-MORL algorithm exhibits a fluctuation range of [-30%, +23%], demonstrating judicious utilization of storage capacity comparable to that of P1 [-30%, +21%].

In contrast, policy P2 displays a more restricted range [-9%, +27%], revealing underutilization of energy resources in favor of excessively conservative management. These results indicate that the PSL-MORL algorithm succeeds in reconciling complete energy autonomy with dynamic battery utilization, thereby validating its ability to simultaneously optimize the system's energy and operational constraint.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

➢ *Comparison of NEWS2 Compliance*

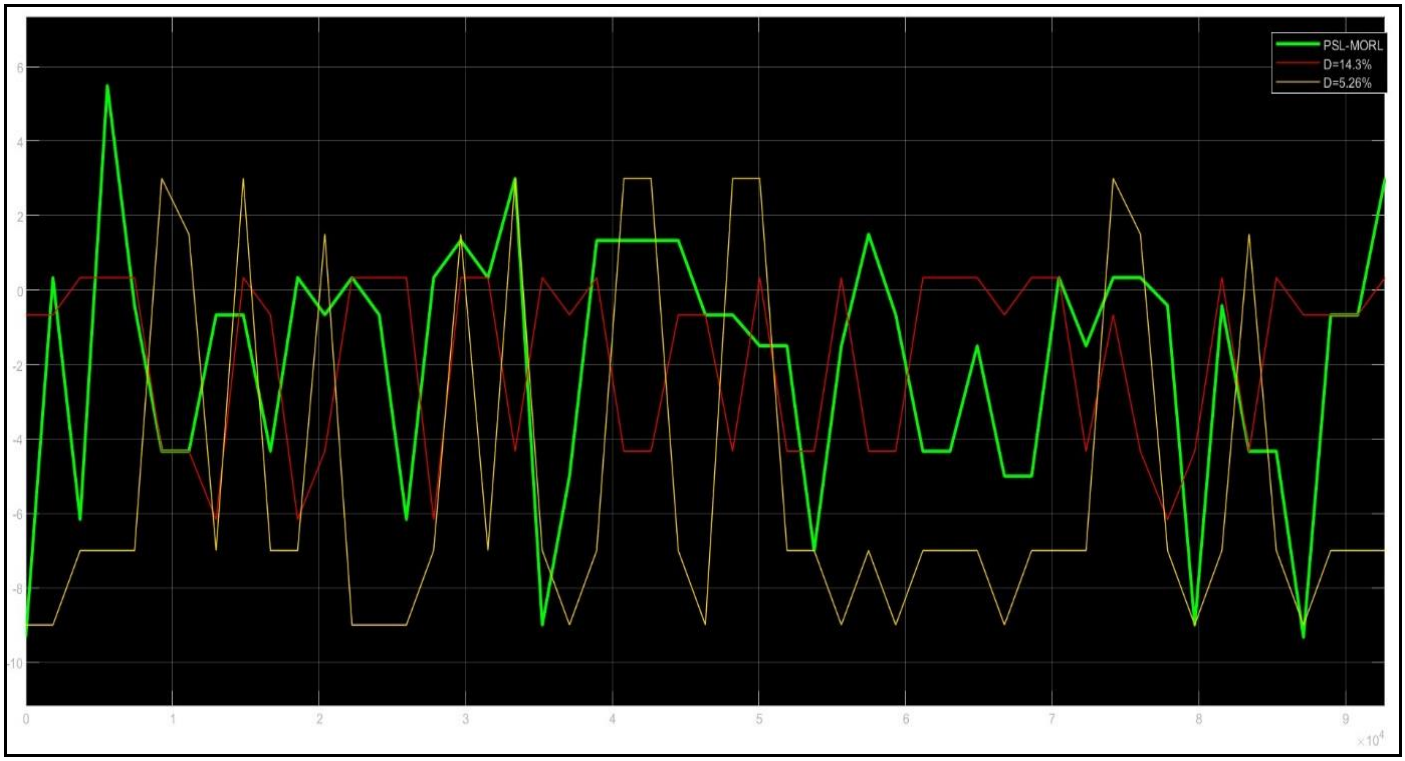

Fig 6 Reward Evolution Comparison Across Energy Management Policies.

Table 6 Comparative Evaluation of NEWS2 Compliance Metrics

Metric	π_1 (D = 14.3%)	π_2 (D=5.26%)	PSL-MORL	Gain vs π_1	Gain vs π_2
Reward Range	[**-6.17 , 0.33**]	[-9.0 , 3.0]	[-9.33 , 5.50]	----------	----------
Average Reward	**-1.54**	-4.45	-1.85	-20%	+58.4%
Number of Measurements	173	91	142	-17.9%	+56%

The evaluation of NEWS2 protocol compliance reveals the effectiveness of the PSL-MORL algorithm in clinical monitoring. With a reward range of [-9.33, +5.50], the proposed approach presents the highest maximum value among the three strategies, reflecting better adaptation to critical phases requiring intensive monitoring.

The average reward obtained by PSL-MORL (-1.85) represents a 58.4% improvement compared to π_2 (-4.45), whose spaced measurement strategy proves inadequate for clinical requirements. The 20% gap with π_1 (-1.54) remains acceptable given the energy gains achieved. In terms of acquisitions, the PSL-MORL algorithm performs 142 measurements, 56% more than π_2 while reducing the number required by π_1 by 17.9%. This distribution demonstrates optimized resource allocation, concentrating measurements during clinically relevant periods while preserving the system's energy balance.

V. CONCLUSION

This study presented an adaptive energy management approach based on the PSL-MORL algorithm for autonomous wearable medical devices. The objective was to address the inherent trade-off between energy efficiency and clinical monitoring quality in the context of IoWT systems powered by solar energy harvesting. Experimental results demonstrate that the PSL-MORL algorithm achieves 24-hour autonomy, ensuring system energy neutrality over a complete daily cycle, representing a 42.9% improvement compared to the intensive monitoring policy. Concurrently, the proposed approach maintains satisfactory compliance with the NEWS2 protocol, with a 58.4% improvement in average reward compared to the energy-saving strategy, while reducing the number of measurements by 17.9% relative to the intensive policy.

The originality of this contribution lies in the integration of an adaptive composite reward function, whose weights are dynamically modulated according to the patient's NEWS2 score. This state-dependent scalarization enables efficient exploration of the Pareto front, concentrating measurement resources during clinically critical periods while preserving overall energy balance.

However, this study has some limitations. The model relies on constant solar irradiance conditions during sunlight phases, without accounting for real-world weather variations. Furthermore, validation was performed exclusively through numerical simulation in MATLAB, without experimentation on a physical prototype. Finally,

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

the discrete action space limited to five duty cycle values may restrict the system's adaptation granularity.

Nevertheless, this work opens promising perspectives for the large-scale deployment of autonomous physiological monitoring devices in ambulatory settings. Future research will focus on extending the model to hybrid energy harvesting sources, integrating variable irradiance profiles, and conducting experimental validation under real clinical conditions.

REFERENCES

[1]. A. Esteva et al., « A guide to deep learning in healthcare », Nat. Med., vol. 25, no 1, p. 24-29, janv. 2019, doi: 10.1038/s41591-018-0316-z.

[2]. E. J. Topol, « High-performance medicine: the convergence of human and artificial intelligence », Nat. Med., vol. 25, no 1, p. 44-56, janv. 2019, doi: 10.1038/s41591-018-0300-7.

[3]. N. Mehta et A. Pandit, « Concurrence of big data analytics and healthcare: A systematic review », Int. J. Med. Inf., vol. 114, p. 57-65, juin 2018, doi: 10.1016/j.ijmedinf.2018.03.013.

[4]. P. Szántó, T. Kiss, et K. J. Sipos, « Energy-efficient AI at the Edge », in 2022 11th Mediterranean Conference on Embedded Computing (MECO), juin 2022, p. 1-6. doi: 10.1109/MECO55406.2022.9797178.

[5]. A. Ometov et al., « A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges », Comput. Netw., vol. 193, p. 108074, juill. 2021, doi: 10.1016/j.comnet.2021.108074.

[6]. W. B. Qaim et al., « Towards Energy Efficiency in the Internet of Wearable Things: A Systematic Review », IEEE Access, vol. 8, p. 175412-175435, 2020, doi: 10.1109/ACCESS.2020.3025270.

[7]. F. K. Shaikh et S. Zeadally, « Energy harvesting in wireless sensor networks: A comprehensive review », Renew. Sustain. Energy Rev., vol. 55, p. 1041-1054, mars 2016, doi: 10.1016/j.rser.2015.11.010.

[8]. S. Shresthamali, M. Kondo, et H. Nakamura, « Adaptive Power Management in Solar Energy Harvesting Sensor Node Using Reinforcement Learning », ACM Trans. Embed. Comput. Syst., vol. 16, no 5s, p. 181:1-181:21, sept. 2017, doi: 10.1145/3126495.

[9]. O. Veligorskyi, M. Khomenko, R. Chakirov, et Y. Vagapov, « Performance analysis of a wearable photovoltaic system », in 2018 IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES), janv. 2018, p. 376-381. doi: 10.1109/IESES.2018.8349905.

[10]. J. Ortegón-Aguilar et al., « Multimodal Power Management Based on Decision Tree for Internet of Wearable Things Systems », Appl. Sci., vol. 13, no 7, Art. no 7, janv. 2023, doi: 10.3390/app13074351.

[11]. « nRF52840 ». Consulté le: 22 novembre 2025. [En ligne]. Disponible sur: https://www.nordicsemi.com/-/media/Software-and-other-downloads/Product-Briefs/nRF52840-Dongle-product-brief.pdf

[12]. S. Mohsen, A. Zekry, K. Youssef, et M. Abouelatta, « A Self-powered Wearable Wireless Sensor System Powered by a Hybrid Energy Harvester for Healthcare Applications », Wirel. Pers. Commun., vol. 116, no 4, p. 3143-3164, févr. 2021, doi: 10.1007/s11277-020-07840-y.

[13]. O. B. Akan, O. Cetinkaya, C. Koca, et M. Ozger, « Internet of Hybrid Energy Harvesting Things », IEEE Internet Things J., vol. 5, no 2, p. 736-746, avr. 2018, doi: 10.1109/JIOT.2017.2742663.

[14]. R. D. Middlebrook et S. Cuk, « A general unified approach to modelling switching-converter power stages », in 1976 IEEE Power Electronics Specialists Conference, juin 1976, p. 18-34. doi: 10.1109/PESC.1976.7072895.

[15]. D. Hussein, G. Bhat, et J. R. Doppa, « Adaptive Energy Management for Self-Sustainable Wearables in Mobile Health », Proc. AAAI Conf. Artif. Intell., vol. 36, no 11, Art. no 11, juin 2022, doi: 10.1609/aaai.v36i11.21451.

[16]. B. Martinez, M. Monton, I. Vilajosana, et J. D. Prades, « The Power of Models: Modeling Power Consumption for IoT Devices », IEEE Sens. J., vol. 15, no 10, p. 5777-5789, oct. 2015, doi: 10.1109/JSEN.2015.2445094.

[17]. C. F. Hayes et al., « A practical guide to multi-objective reinforcement learning and planning », Auton. Agents Multi-Agent Syst., vol. 36, no 1, p. 26, avr. 2022, doi: 10.1007/s10458-022-09552-y.

[18]. Y. Rioual, Y. Le Moullec, J. Laurent, M. Khan, et J.-P. Diguet, Design and Comparison of Reward Functions in Reinforcement Learning for Energy Management of Sensor Nodes. 2021. doi: 10.48550/arXiv.2106.01114.

[19]. E. Liu et al., « Pareto set learning for multi-objective reinforcement learning », in Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, in AAAI'25/IAAI'25/EAAI'25, vol. 39. AAAI Press, févr. 2025, p. 18789-18797. doi: 10.1609/aaai.v39i18.34068.

[20]. C. Qian, Y. Yu, et Z.-H. Zhou, « On constrained boolean pareto optimization », in Proceedings of the 24th International Conference on Artificial Intelligence, in IJCAI'15. Buenos Aires, Argentina: AAAI Press, juill. 2015, p. 389-395.

[21]. X. Lin, Z. Yang, et Q. Zhang, « Pareto Set Learning for Neural Multi-Objective Combinatorial Optimization », présenté à International Conference on Learning Representations, oct. 2021.

[22]. D. M. Roijers, P. Vamplew, S. Whiteson, et R. Dazeley, « A Survey of Multi-Objective Sequential Decision-Making », J. Artif. Intell. Res., vol. 48, p. 67-113, oct. 2013, doi: 10.1613/jair.3987.

[23]. K. Van Moffaert, M. M. Drugan, et A. Nowé, « Scalarized multi-objective reinforcement learning:

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec1251

Novel design techniques », in 2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), avr. 2013, p. 191-199. doi: 10.1109/ADPRL.2013.6615007.

[24]. S. Natarajan et P. Tadepalli, « Dynamic preferences in multi-criteria reinforcement learning », in Proceedings of the 22nd international conference on Machine learning, in ICML '05. New York, NY, USA: Association for Computing Machinery, août 2005, p. 601-608. doi: 10.1145/1102351.1102427.

[25]. C. Liu, X. Xu, et D. Hu, « Multiobjective Reinforcement Learning: A Comprehensive Overview », IEEE Trans. Syst. Man Cybern. Syst., vol. 45, no 3, p. 385-398, mars 2015, doi: 10.1109/TSMC.2014.2358639.

[26]. P. Vamplew, R. Dazeley, A. Berry, R. Issabekov, et E. Dekker, « Empirical evaluation methods for multiobjective reinforcement learning algorithms », Mach Learn, vol. 84, no 1-2, p. 51-80, juill. 2011, doi: 10.1007/s10994-010-5232-5.