Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 615–620
DOI: 10.15439/2024F8855
ISSN 2300-5963 ACSIS, Vol. 39

Efficient Maritime Healthcare Resource Allocation Using Reinforcement Learning

Tehreem Hasan
Quaid-i-Azam University
Islamabad Pakistan
Email: tehreemhasan@ele.qau.edu.pk

Farwa Batool
Quaid-i-Azam University
Islamabad Pakistan
Email: farwabatool@ele.qau.edu.pk

Mario Fiorino
Politecnico di Torino, Italy
Email: mario.fiorino@polito.it
ORCID: 0009-0007-9393-7095

Giancarlo Tretola
Department of Computer Engineering
Università Giustino Fortunato
Benevento Italy
Email: g.tretola@unifortunato.eu

Musarat Abbas
Quaid-i-Azam University
Islamabad Pakistan
Email: mabbas@qau.edu.pk

***Abstract*—The allocation of healthcare resources on ships is crucial for safety and well-being due to limited access to external aid. Proficient medical staff on board provide a mobile healthcare facility, offering a range of services from first aid to complex procedures. This paper presents a system model utilizing Reinforcement Learning (RL) to optimize doctor-patient assignments and resource allocation in maritime settings. The RL approach focuses on dynamic, sequential decision-making, employing Q-learning to adapt to changing conditions and maximize cumulative rewards. Our experimental setup involves a simulated healthcare environment with variable patient conditions and doctor availability, operating within a 24-hour cycle. The Q-learning algorithm iteratively learns optimal strategies to enhance resource utilization and patient outcomes, prioritizing emergency cases while balancing the availability of medical staff. The results highlight the potential of RL in improving healthcare delivery on ships, demonstrating the system's effectiveness in dynamic, time-constrained scenarios and contributing to overall maritime safety and operational resilience.**

I. Introduction

THE allocation of healthcare resources is an important and critical task for provision of quality health services [1]. This task in the restricted and often isolated setting of ship is not simply a matter of convenience; rather, it is an essential requirement that directly influences the safety and well-being of all individuals on board[2]. Unlike on land, where medical facilities are usually easily accessible, ships operate in environments where the availability of external aid can be significantly limited or delayed. Consequently, the distribution of healthcare resources becomes not just significant but paramount for mitigating risks and ensuring the uninterrupted continuation of maritime activities.

In maritime settings, the presence of proficient medical staff on board is comparable to having a mobile healthcare facility[3]. Physicians, nurses, and paramedics have crucial roles, accountable not only for immediate treatment during crises but also for preserving overall health and wellness throughout journeys. Their expertise, coupled with a variety of medical services ranging from basic first aid to complex procedures, establishes the foundation of a ship's healthcare framework[4].

Maritime healthcare encounters challenges beyond the provision of services. Efficient resource allocation requires ongoing monitoring of the deployment of medical personnel[5]. This includes ensuring sufficient staff numbers strategically located to promptly respond to emergencies anywhere on the ship. It also involves forecasting changes in demand based on the duration of the voyage, the nature of the cargo, and the demographics of the crew and passengers[6].

Furthermore, the distribution of healthcare resources surpasses mere logistics; it entails incentivizing effective decision-making. Immediate incentives and delayed penalties act as stimulants for proactive resource management, cultivating a culture of safety and accountability[7]. By acknowledging and reinforcing positive actions, ship operators ensure effective resource utilization, enhancing the overall resilience of the healthcare system[8].

Fundamentally, the allocation of healthcare resources in maritime settings demands careful planning, constant monitoring, and proactive decision-making[9]. It demonstrates the flexibility and resourcefulness of maritime experts navigating intricate connections among personnel, services, data, and measures to safeguard health and well-being.

Moreover, onboard medical facilities serve broader objectives of safety, security, and operational effectiveness[10]. They function as crucial support systems during crises, mitigating the impacts of adverse events. Nevertheless, the distinctive maritime environment presents challenges such as limited space, harsh weather conditions, and isolation, which magnify medical risks. Therefore, resource allocation must address these challenges, ensuring the preparedness of personnel to deliver efficient care[11].

Additionally, the distribution of healthcare resources on ships intertwines with risk management and compliance with regulations[12]. Maritime authorities impose stringent criteria on medical care provision and facility upkeep. The failure to

meet requirements may result in severe outcomes, compelling operators to comply with regulatory standards and industry best practices, thereby safeguarding the health and safety of individuals aboard.

The paper organized as follows: **Section 2** provides the Related Works. In **Section 3**, the problem definition and background of the healthcare system on ships are described. The system model and experimental setup, including the Q-learning algorithm, are presented in **Section 4**. Computational results based on simulated scenarios are provided in **Section 5**. Finally, **Section 6** concludes the paper.

II. Related Works

This section presents some state of the art on the use of AI-empowered solutions for healthcare problems.

A machine learning method upper confidence bound is utilized in [13] to assist patients during their medication process at home. Authors considered the cognitive and physical impairments of the patients in the training of the machine learning model. A similar work is also done in [14] but with the help of Thompson sampling method. However, these solutions are applicable to certain scenarios during medication at home.

Dynamic Treatment Regime (DTR) is has an importance in healthcare as well as for medical research. DTR are considered as sequence of alternative treatment paths and any of these treatments can be adapted depending on the patient's conditions [15]. Therefore, the authors in [16] apply a cooperative imitation learning approach to utilize information from both negative and positive trajectories to learn the optimal DTR. The given framework minimizes the chance of choosing any treatment that results in a negative outcome during the medical examination. However, the proposed work is not suitable to employ for the medication emergency on ships.

The works in [17] and [18] use AI techniques for risk management in nuclear medication department. The later will is the extension of former one and discuss the risk cases during examination at such departments. Although, the proposed systems are useful to avoid possible risk at nuclear medication departments but are not useful for healthcare solutions at ships.

Moreover, there are some AI based solutions for the continuous and remote monitoring of unpredictable health issues. Such a failure mode and effect analysis is given in [19], [20] and [21] for a specific mobile health monitoring system. Both of these systems were designed to provide remote healthcare solutions but these are for certain cases and environments and cannot be generalised for other cases.

The proposed work examines managing healthcare resources on ships for safety. It tackles challenges with planning, monitoring, and decision-making. Using reinforcement learning, the system optimizes doctor-patient assignments in real-time. Patient urgency and doctor availability impact the allocation process[10]. By employing Q-learning, the system learns optimal strategies for maximizing rewards in urgent situations. Simulations show improved resource use and patient care. It highlights the importance of efficient resource allocation and decision-making in maritime healthcare for enhancing safety and well-being on ships.

III. Background

Our ship's healthcare system utilizes reinforcement learning (RL) to optimize doctor-patient assignments and resource allocation, a branch of machine learning focusing on decision-making through environment interaction[22]. RL is beneficial for dynamic, uncertain healthcare settings requiring sequential actions to achieve long-term goals [23]. At the core of RL is the agent concept, learning decision-making through environment feedback [24]. The agent in our scenario assigns doctors to patients within the ship's healthcare infrastructure, influenced by factors like patient urgency and treatment outcomes[25]. A key RL component is the reward signal, offering feedback on action desirability based on factors like patient conditions and treatment efficiency[26]. The RL agent maintains a policy for actions in each environment state, aiming to learn an optimal policy for maximizing cumulative rewards over time using algorithms like Q-learning, popular for discrete state and action spaces.

Q-learning iteratively updates action value estimates (Q-values) based on observed rewards and state transitions, enabling the agent to improve decision-making and reach an optimal policy [27]. In our ship's healthcare scenario, Q-learning assists in adapting to changing conditions and making informed decisions about doctor-patient assignments. By learning from experiences and exploring strategies, the system can identify effective healthcare delivery patterns and policies[28]. Reinforcement learning provides a framework for optimizing decision-making in dynamic healthcare environments[29], enhancing efficiency, patient outcomes, and resource utilization.

IV. System model

The objective of the proposed is to tackle the complex challenge of efficiently allocating physicians to patients within a time-critical framework during a medical emergency on ships[30]. The system functions dynamically throughout a 24-hour cycle, where the availability of medical staff and the influx of patients exhibit significant variability[31]. At any specific moment, the system has the maximum capacity of 10 patients and a team of 5 doctors.

Upon arrival at the medical facility, patients present a range of medical conditions, classified into emergency and general categories as also demonstrated in the Figure 1. The urgency level for treatment varies between these categories, with emergency situations like abrupt illnesses or injuries necessitating immediate action, while general cases encompass issues such as seasickness, infections, dehydration, and fever. Each patient category is linked to specific rewards, reflecting the importance of timely treatment and the resources allocated to address their needs[32].

To replicate the patterns of patient arrivals and doctor availability, we use simulated data through a sequence of scenarios. Each scenario shows a situation where patients

come to the facility in need of medical care. The scenario begins by setting the current time in the 24-hour cycle and determining the size of the patient queue, which fluctuates based on temporal elements. During daylight hours, when patient influx is typically higher, the queue tends to be more extensive compared to quieter periods.

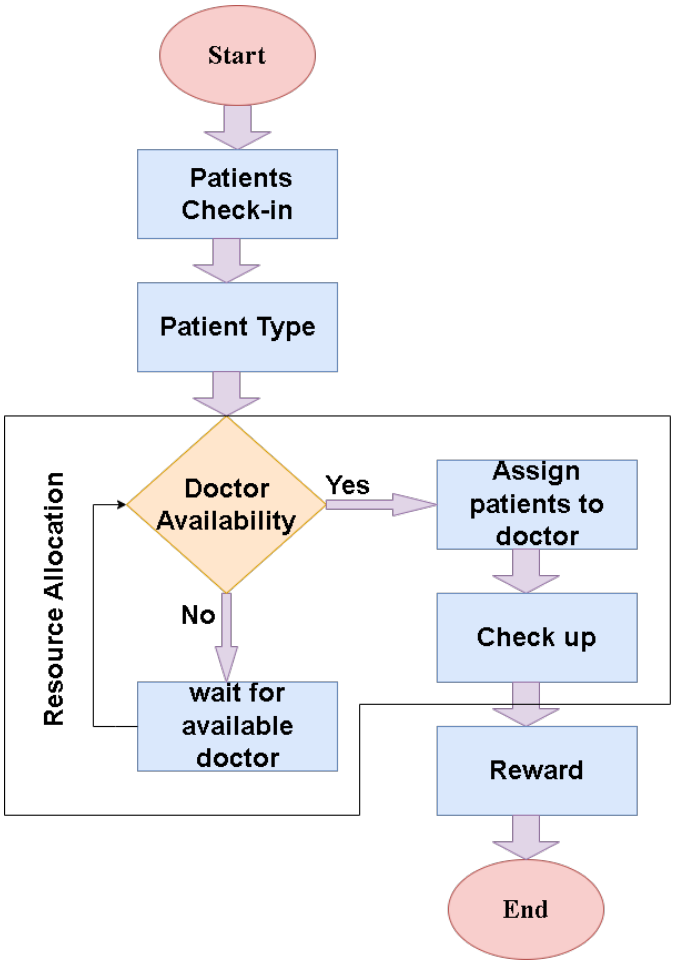

Fig. 1. System Model

The assignment of physicians to patients is influenced by various factors, including the urgency of patient conditions and the availability of medical staff[33]. Emergency situations are prioritized to ensure that patients with critical conditions receive immediate medical attention[34]. Doctor availability fluctuates throughout the day, with a higher probability of doctors being available during standard working hours. Hence, the allocation process seeks to strike a balance between the urgency of patient needs and the availability of medical personnel, aiming to enhance the number of patients treated while optimizing resource utilization[35].

To support decision-making processes within the system, we employ a Q-learning algorithm, which is a RL technique that progressively acquires optimal strategies through trial and error [36]. The state space comprises patient indices, representing the order of patients in the queue, while the action space includes potential doctor assignments. The Q-learning algorithm adjusts Q values based on the rewards gained from treating patients, with the aim of acquiring an optimal policy that maximizes the accumulation of overall rewards.

The system's performance is assessed using various metrics, such as the total rewards accumulated across multiple scenarios and the average reward per scenario. By examining the acquired Q-values and doctor-patient allocations, valuable insights can be derived on effective approaches to enhance healthcare delivery in time-constrained scenarios. Ultimately, the system model acts as a foundation for investigating and enhancing strategies to improve patient care and resource distribution in healthcare environments.

Experimental Setup The experimental setup involves the utilization of Q-learning, a reinforcement learning technique, to optimize the allocation of doctors to patients on board[37]. The primary goal is to enhance the overall rewards obtained by efficiently managing the treatment of different patient categories within specified constraints. This experimental arrangement encompasses the establishment of the environment, initialization of parameters, preprocessing of the dataset, and execution of the Q-learning algorithm to acquire the optimal policy for doctor assignments.

Environment: The environment comprises patients and doctors with specific conditions and availability, respectively, where the maximum number of patients and doctors is limited to 10 and 5, respectively, operating within a 24-hour window. Patients are categorized into emergency conditions (e.g., sudden illnesses, injuries) and general conditions (e.g., seasickness, infections, dehydration, fever), each associated with a predefined reward indicating the priority of treating that condition, with emergency conditions offering higher rewards.

Q-learning Parameters

To train the Q-learning model, we define several parameters:

- **Alpha (α):** The learning rate, set to 0.1, determining the significance of new information over old information.
- **Gamma (γ):** The discount factor, set to 0.9, reflects the importance of future rewards.
- **Epsilon (ε):** The epsilon-greedy parameter, set to 0.1, balances exploration (choosing random actions) and exploitation (choosing the best-known actions).

Q-learning Algorithm The Q-learning algorithm is employed to iteratively learn the optimal doctor-patient assignment policy. The key steps in the algorithm include:

1) **State Initialization:** For each episode (a single simulation run), a random initial state representing a patient index is selected.
2) **Action Selection:** The epsilon-greedy policy is used to choose an action (doctor assignment) based on the current state. With a probability of ε, a random action is selected; otherwise, the action with the highest Q-value is chosen.
3) **Reward Observation:** The reward for the chosen action is determined based on the patient type.

4) **Next State Calculation:** The next state is determined by checking doctor availability. If the selected doctor is available, they become busy, and the state progresses to the next patient. If no doctor is available, the state remains unchanged.
5) **Q-value Update:** The Q-value is updated using the Bellman equation, incorporating the observed reward and the maximum expected future reward.

Update rule for Q-learning

The basic update rule for Q-learning is as follows:

Q[state, action] = Q[state, action] + lr * (reward + gamma * np.max(Q[new_state, :]) — Q[state, action])

- **alpha** is the learning rate.
- **reward** is the reward received for taking the action in the current state.
- **gamma** is the discount factor.
- **np.max(Q[next_state, :])** computes the maximum Q-value for the next state over all possible actions.
- **Q[state, action]** is the current Q-value

Episode Termination: The episode ends when all patients have been assigned or a maximum iteration limit is reached.

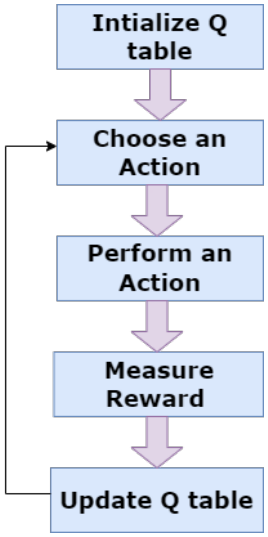

Fig. 2. Q Learning Algorithm

This experiment demonstrates the application of Q-learning in a simulated healthcare environment on a ship. By learning from multiple episodes of patient-doctor interactions, the algorithm aims to maximize the total reward, ensuring efficient and effective medical care. The results showcase the potential of reinforcement learning in optimizing resource allocation and decision-making in real-world scenarios.

V. Results

The graph shows the cumulative average reward per episode during the Q-learning process. The average reward per episode increases steadily as the agent learns and improves its policy, indicating that the Q-learning algorithm is effectively optimizing the agent's behavior.

Algorithm 1 Q-learning Algorithm

```
Initialize Q(s, a) arbitrarily
Set learning rate α, discount factor γ, and exploration rate ϵ
for each episode do
    Initialize state s
    while state s is not terminal do
        if a random number < ϵ then
            Choose random action a
        else
            Choose action a = arg max_a' Q(s, a')
        end if
        Take action a, observe reward r and next state s'
        Update Q(s, a):
        Q(s, a) ← Q(s, a) + α[r + γ max_a' Q(s', a') − Q(s, a)]
        s ← s'
    end while
end for
```

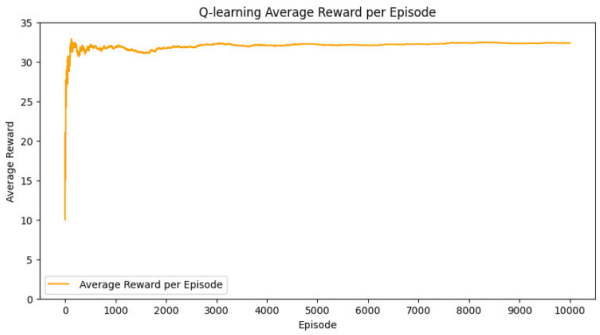

Fig. 3. Q Learning performance

Epsilon-Greedy Training vs. Greedy Evaluation

The graph compares the cumulative average rewards per episode for Q-learning with epsilon-greedy (training) and greedy (evaluation) policies over episodes. The orange line, representing epsilon-greedy, shows a steady increase in average rewards, indicating effective learning and exploration. The blue dashed line for greedy policy shows consistent but lower average rewards, highlighting the impact of exploration on learning performance.

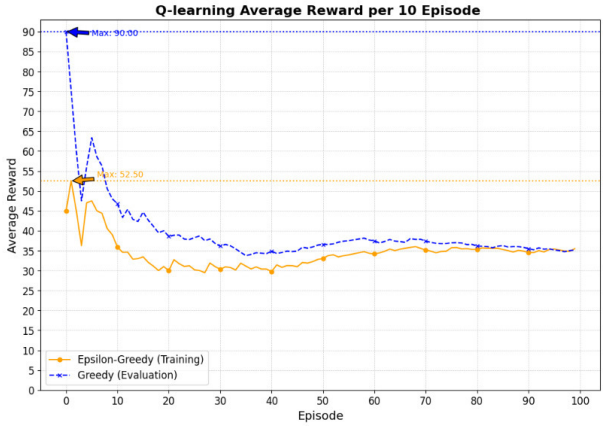

Fig. 4. Average Reward vs 100 episodes

The graph presents the cumulative rewards achieved by the Q-learning algorithm with an epsilon-greedy policy (epsilon = 0.1) over 100 episodes. It demonstrates the algorithm's convergence in optimizing patient assignment to available doctors, balancing exploration and exploitation to effectively utilize medical resources.

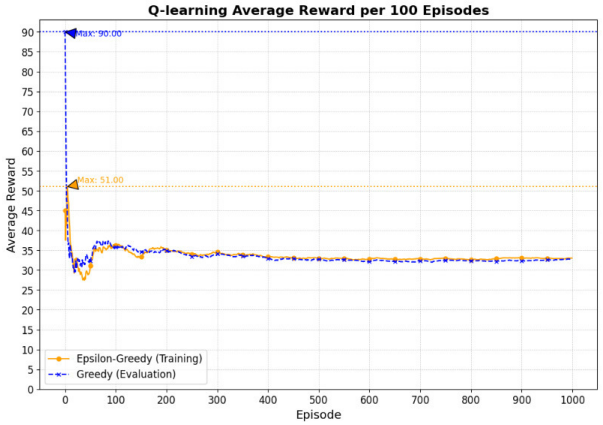

Fig. 5. Average Reward vs 1000 episodes.

The plot visualizes Q-learning performance with epsilon-greedy policy across 100 episodes, revealing cumulative rewards per episode. It demonstrates stable learning convergence in medical resource allocation tasks, reflecting effective policy optimization and resource management.

VI. Challenges of Proposed Model

The current model's limitations include its reliance on Q-learning, which may not handle large state and action spaces efficiently, potentially leading to slow learning and suboptimal performance in complex, real-world scenarios[38]. Furthermore, the model assumes static conditions for medical emergencies and staff availability, which may not accurately reflect the dynamic nature of healthcare needs on ships.

Deep Q-learning techniques could enhance the model by leveraging deep neural networks to approximate the Q-value function, enabling it to manage more complex and high-dimensional state spaces. This approach could improve the system's ability to generalize from past experiences and make more informed decisions in varied and unpredictable environments.

However, practical deployment of this model in the shipping industry faces several challenges. Firstly, ensuring real-time data collection and processing for accurate decision making could be difficult due to potential connectivity issues and limited computational resources on ships. Secondly, integrating the system with existing healthcare frameworks and protocols requires careful coordination and regulatory compliance. Additionally, the variability in medical emergencies and staff expertise may introduce further complexity, necessitating continuous training and adaptation of the model to maintain optimal performance. Lastly, gaining trust and acceptance from maritime healthcare professionals and stakeholders is crucial for successful implementation, requiring demonstrable reliability and effectiveness of the proposed system in real-world conditions.

VII. Conclusion and Future Work

In conclusion, our study demonstrates the potential of Q-learning within reinforcement learning to optimize healthcare resource allocation on ships. By dynamically assigning doctors based on the urgency of patient conditions and their availability, we can significantly enhance patient care and overall resource utilization. The experimental results and simulations validate the effectiveness of this approach, showcasing improved decision-making capabilities in healthcare management.

The application of Q-learning in maritime healthcare environments addresses the unique challenges posed by limited medical resources, fluctuating patient inflow, and the critical nature of onboard medical emergencies. This methodology provides a robust framework for optimizing resource distribution, ensuring that medical personnel can respond effectively to both routine and urgent healthcare needs.

Future developments in this field could explore the integration of more advanced reinforcement learning techniques, such as deep Q-learning or actor-critic methods, to further enhance the system's performance. Additionally, incorporating real-time data from onboard health monitoring systems could improve the accuracy and responsiveness of the resource allocation process. Expanding this research to include other critical aspects of maritime operations, such as disaster response and long-term health monitoring, could further enhance the safety, security, and operational effectiveness of ships, ultimately ensuring the well-being of all individuals on board.

Acknowledgment

This project has been partially funded by the "Programma Nazionale Ricerca, Innovazione e Competitività per la transizione verde e digitale 2021/2027 destinate all'intervento del FCS "Scoperta imprenditoriale" - Azione 1.1.4 "Ricerca collaborativa" - with the project SIAMO (Servizi Innovativi per

l'Assistenza Medica a bOrdo) project number F/360124/01-02/X75.

REFERENCES

[1] M. Ciampi, A. Coronato, M. Naeem, and S. Silvestri, "An intelligent environment for preventing medication errors in home treatment," *Expert Systems with Applications*, vol. 193, p. 116434, 2022.

[2] C. Hetherington, R. Flin, and K. Mearns, "Safety in shipping: The human element," *Journal of Safety Research*, vol. 37, no. 4, pp. 401–411, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022437506000818

[3] S. Nazim, V. K. Shukla, F. Beena, and S. Dubey, "Smart intelligent approaches for healthcare management," in *Computational Intelligence in Urban Infrastructure*. CRC Press, 2024, pp. 189–211.

[4] D. Martínez-Méndez and M. Bravo-Acosta, "The challenges faced after a major trauma at an expedition ship at a remote area. report of one case," *Revista Medica de Chile*, vol. 151, no. 2, pp. 255–258, 2023.

[5] K. Zong and C. Luo, "Reinforcement learning based framework for covid-19 resource allocation," *Computers & Industrial Engineering*, vol. 167, p. 107960, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835222000304

[6] M. M. S. L. T. R. Benjamin Rolf, Ilya Jackson and D. Ivanov, "A review on reinforcement learning algorithms and applications in supply chain management," *International Journal of Production Research*, vol. 61, no. 20, pp. 7151–7179, 2023.

[7] P. Fiorucci, F. Gaetani, R. Minciardi, R. Sacile, and E. Trasforini, "Real time optimal resource allocation in natural hazard management," 2004.

[8] C. Yu, J. Liu, S. Nemati, and G. Yin, "Reinforcement learning in healthcare: A survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–36, 2021.

[9] E. Aktaş, F. Ülengin, and Ş. Ö. Şahin, "A decision support system to improve the efficiency of resource allocation in healthcare management," *Socio-Economic Planning Sciences*, vol. 41, no. 2, pp. 130–146, 2007.

[10] O. Elfahim, E. M. B. Laoula, M. Youssfi, O. Barakat, and M. Mestari, "Deep reinforcement learning approach for emergency response management," in *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 2022, pp. 1–7.

[11] J. Zhang, M. Zhang, F. Ren, and J. Liu, "An innovation approach for optimal resource allocation in emergency management," *IEEE Transactions on Computers*, 2016.

[12] T. Ø. Kongsvik, K. Størkersen, and S. Antonsen, "The relationship between regulation, safety management systems and safety culture in the maritime industry," *Safety, reliability and risk analysis: Beyond the horizon*, pp. 467–473, 2014.

[13] M. Naeem, A. Coronato, and G. Paragliola, "Adaptive treatment assisting system for patients using machine learning," in *2019 sixth international conference on social networks analysis, management and security (SNAMS)*. IEEE, 2019, pp. 460–465.

[14] A. Coronato and M. Naeem, "A reinforcement learning based intelligent system for the healthcare treatment assistance of patients with disabilities," in *International Symposium on Pervasive Systems, Algorithms and Networks*. Springer, 2019, pp. 15–28.

[15] A. Coronato, M. Naeem, G. De Pietro, and G. Paragliola, "Reinforcement learning for intelligent healthcare applications: A survey," *Artificial Intelligence in Medicine*, vol. 109, p. 101964, 2020.

[16] S. I. H. Shah, A. Coronato, M. Naeem, and G. De Pietro, "Learning and assessing optimal dynamic treatment regimes through cooperative imitation learning," *IEEE Access*, vol. 10, pp. 78 148–78 158, 2022.

[17] G. Paragliola, A. Coronato, M. Naeem, and G. De Pietro, "A reinforcement learning-based approach for the risk management of e-health environments: A case study," in *2018 14th international conference on signal-image technology & internet-based systems (SITIS)*. IEEE, 2018, pp. 711–716.

[18] S. I. H. Shah, M. Naeem, G. Paragliola, A. Coronato, and M. Pechenizkiy, "An ai-empowered infrastructure for risk prevention during medical examination," *Expert Systems with Applications*, vol. 225, p. 120048, 2023.

[19] M. Cinque, A. Coronato, and A. Testa, "A failure modes and effects analysis of mobile health monitoring systems," in *Innovations and advances in computer, information, systems sciences, and engineering*. Springer, 2012, pp. 569–582.

[20] M. Bakhouya, R. Campbell, A. Coronato, G. d. Pietro, and A. Ranganathan, "Introduction to special section on formal methods in pervasive computing," pp. 1–9, 2012.

[21] M. Cinque, A. Coronato, and A. Testa, "Dependable services for mobile health monitoring systems," *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 4, no. 1, pp. 1–15, 2012.

[22] I. Sanz *et al.*, "Resource allocation in home care services using reinforcement learning," in *Artificial Intelligence Research and Development: Proceedings of the 25th International Conference of the Catalan Association for Artificial Intelligence*, vol. 375. IOS Press, 2023, p. 173.

[23] M. Fiorino, M. Naeem, M. Ciampi, and A. Coronato, "Defining a metric-driven approach for learning hazardous situations," *Technologies*, vol. 12, no. 7, p. 103, 2024.

[24] M. Naeem, S. T. H. Rizvi, and A. Coronato, "A gentle introduction to reinforcement learning and its application in different fields," *IEEE access*, vol. 8, pp. 209 320–209 344, 2020.

[25] F. Masroor, A. Gopalakrishnan, and N. Goveas, "Machine learning-driven patient scheduling in healthcare: A fairness-centric approach for optimized resource allocation," in *2024 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2024, pp. 01–06.

[26] S. Bharti, D. S. Kurian, and V. M. Pillai, "Reinforcement learning for inventory management," in *Innovative Product Design and Intelligent Manufacturing Systems: Select Proceedings of ICIPDIMS 2019*. Springer, 2020, pp. 877–885.

[27] G. Paragliola and M. Naeem, "Risk management for nuclear medical department using reinforcement learning algorithms," *Journal of Reliable Intelligent Environments*, vol. 5, pp. 105–113, 2019.

[28] T. Li, Z. Wang, W. Lu, Q. Zhang, and D. Li, "Electronic health records based reinforcement learning for treatment optimizing," *Information Systems*, vol. 104, p. 101878, 2022.

[29] K. Gai and M. Qiu, "Optimal resource allocation using reinforcement learning for iot content-centric services," *Applied Soft Computing*, vol. 70, pp. 12–21, 2018.

[30] A. Alelaiwi, "Resource allocation management in patient-to-physician communications based on deep reinforcement learning in smart healthcare services," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2020, pp. 1–5.

[31] C. Shyalika, T. Silva, and A. Karunananda, "Reinforcement learning in dynamic task scheduling: A review," *SN Computer Science*, vol. 1, no. 6, p. 306, 2020.

[32] S. Liu, K. C. See, K. Y. Ngiam, L. A. Celi, X. Sun, and M. Feng, "Reinforcement learning for clinical decision support in critical care: comprehensive review," *Journal of medical Internet research*, vol. 22, no. 7, p. e18477, 2020.

[33] E. Cabrera, M. Taboada, M. L. Iglesias, F. Epelde, and E. Luque, "Simulation optimization for healthcare emergency departments," *Procedia computer science*, vol. 9, pp. 1464–1473, 2012.

[34] R. Fujimori, K. Liu, S. Soeno, H. Naraba, K. Ogura, K. Hara, T. Sonoo, T. Ogura, K. Nakamura, T. Goto *et al.*, "Acceptance, barriers, and facilitators to implementing artificial intelligence–based decision support systems in emergency departments: quantitative and qualitative evaluation," *JMIR formative research*, vol. 6, no. 6, p. e36501, 2022.

[35] N. Sahota, R. Lloyd, A. Ramakrishna, J. A. Mackay, J. C. Prorok, L. Weise-Kelly, T. Navarro, N. L. Wilczynski, R. Brian Haynes, and C. S. R. Team, "Computerized clinical decision support systems for acute care management: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes," *Implementation Science*, vol. 6, pp. 1–14, 2011.

[36] M. Jamal, Z. Ullah, M. Naeem, M. Abbas, and A. Coronato, "A hybrid multi-agent reinforcement learning approach for spectrum sharing in vehicular networks," *Future Internet*, vol. 16, no. 5, p. 152, 2024.

[37] E. B. Laber, K. A. Linn, and L. A. Stefanski, "Interactive model building for q-learning," *Biometrika*, vol. 101, no. 4, pp. 831–847, 2014.

[38] E. Riachi, M. Mamdani, M. Fralick, and F. Rudzicz, "Challenges for reinforcement learning in healthcare," *arXiv preprint arXiv:2103.05612*, 2021.