Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 00 (2023) 000–000

www.elsevier.com/locate/procedia

CIRP BioM 2024

Reinforcement Learning Based Resource Management for CAR T-Cell Therapies

Szabolcs Szentpéteri[a], Krisztián B. Kis[a], Péter Egri[a], Carmen Sanges[b], Sophia Danhof[b], Katrin Mestermann[b], Michael Hudecek[b], Sergio Navarro Velázquez[c], Manel Juan[c], Balázs Cs. Csáji[a,d,*]

[a]*Institute for Computer Science and Control (SZTAKI), Hungarian Research Network (HUN-REN), Kende str 13-17, Budapest, 1111, Hungary*
[b]*Universitätsklinikum Würzburg, Versbacher Straße 5, Würzburg, 97078, Germany*
[c]*Hospital Clínic Barcelona, Villarroel 170 - Escala 3, Planta 1, Barcelona, 08036, Spain*
[d]*Institute of Mathematics, Eötvös Loránd University (ELTE), Budapest, 1117, Hungary*

* Corresponding author. *E-mail address:* csaji@sztaki.hu.

Abstract

This paper focuses on optimizing resource management strategies in chimeric antigen receptor (CAR) T-cell therapies using reinforcement learning (RL). CAR T-cell therapy is an innovative and promising treatment within the field of immunotherapy, which is based on the isolation of patients' own T-cells, genetically modifying these cells to express a CAR for tumor recognition, cultivating and expanding these T cells and infusing them back to the patient. These therapies require several reusable but scarce resources, including special equipment, such as the bioreactor that is used to expand CAR T cells, and medical resources, such as the hospital staff, e.g., doctors and nurses. Considering the stochastic nature of medical procedures and the production of CAR T cells, the scheduling of the therapies and the efficient allocation of the required resources pose a significant challenge inside the hospital environment. Here, we propose a derivative-free policy gradient algorithm that utilizes a simulation model of the therapy, built using real-world data, to obtain efficient control policies for the resource management problem. The proposed method is designed to minimize the expected number of deviations from the therapy protocol as well as the expected overall completion time of the therapies for every patient. The effectiveness of the proposed resource management approach is demonstrated via simulation experiments.

© 2024 The Authors. Published by ELSEVIER B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the CIRP BioM 2024

Keywords: patient scheduling; reinforcement learning; therapy simulation; CAR T-cell therapy;

1. Introduction

Chimeric antigen receptor (CAR) T-cell therapy is a revolutionary novel treatment in hematology [1] and oncology [2]. In this type of cell therapy, immune cells (T lymphocytes) are engineered to express a synthetic chimeric antigen receptor (CAR) and are redirected to seek and destroy cancerous cells in the patient's body. These receptors are artificial molecules that do not exist in nature, they are a mixture of the variable domain of heavy and light chains of an antibody, which represents the external domain of the CAR, and the zeta chain of CD3 molecule, which represents the internal domain of the CAR [1]. CAR T-cell therapy is indeed a transformative new treatment in hematology with clinical proof-of-concept in patients with acute leukemia [3] lymphoma (lymph node cancer) [4] and multiple myeloma (bone marrow cancer) [5]. A conceptual appeal of this treatment is that the patient's own immune cells (T cells) are genetically engineered – providing a personalized treatment – and that a single infusion of CAR T cells can be sufficient to eliminate the cancerous cells and to establish an immunological memory that protects the patient from tumor relapse (i.e., "living drug paradigm") [6].

An illustration of the CAR T-cell therapy is presented in Fig. 1. As the figure shows, in the first step, blood is collected from

2212-8271 © 2024 The Authors. Published by ELSEVIER B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the CIRP BioM 2024

the patient and subsequently, T cells are isolated. Then, these T cells are genetically modified to express a non-canonical molecule on their surface, called CARs, which can specifically recognize their respective antigen on the surface of cancer cells. In the next step, millions of these CAR T cells are grown by expanding the cells in a bioreactor. Finally, these CAR T cells are infused back to the patient to eliminate the cancerous cells.

CAR T-cell therapy in acute leukemia and lymphoma is now an approved treatment that is highly demanded by patients and caregivers [7]. The success of CAR T-cell therapy as a role model for gene-engineered cellular immunotherapy in hematology and oncology has spurred the development of this treatment for novel applications in infectious diseases [8], chronic and autoimmune diseases [9].

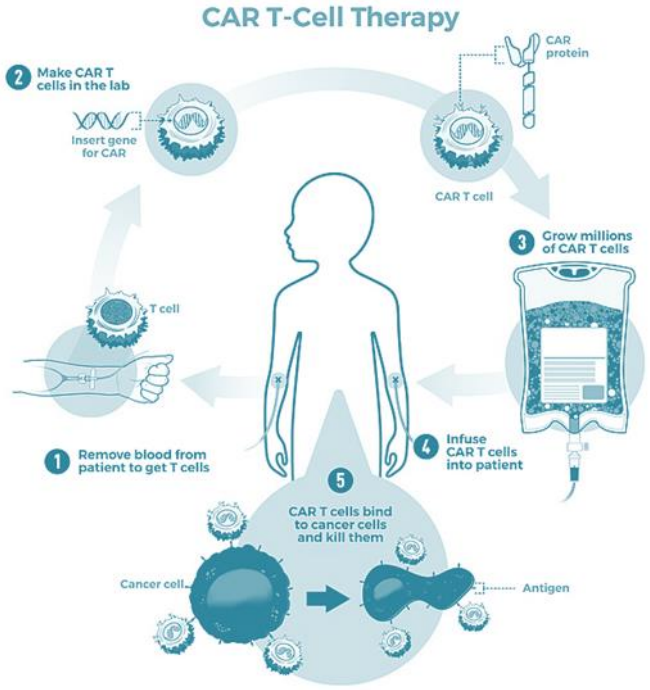

Fig. 1. CAR T-cell therapy (source: National Cancer Institute)

Conventional CAR T-cell therapy is complicated by complex logistics and supply from centralized production facilities, inflexible manufacturing and clinical use schemes that disregard patient and cell product characteristics, thereby limiting patient access and therapeutic outcome. These also make the therapy very expensive. The aim of the EU funded AIDPATH project is to establish personalized treatments directly at the clinical sites, and to accomplish end-to-end automation of hospital-based CAR T-cell manufacture. One of the challenges in developing such a system is to optimize scheduling and resource planning to reduce costs, to increase hospital resource utilization and to augment patient access.

From the perspective of resource management and patient scheduling, the hospital environment is a complex system with high uncertainty. Currently, this task is typically done manually, which means that the decisions regarding resource allocation and optimization are made based on human judgment and experience rather than automated processes. In this paper a *machine learning* (ML) approach is suggested, in order to solve scheduling problems in highly complex and stochastic environments. A branch of ML algorithms that can be effectively applied in these scenarios is *reinforcement learning* (RL). RL deals with the problem of learning a control policy (decision strategy) for sequential decision making by interacting with an uncertain and dynamic environment.

Standard resource allocation problems, where various RL algorithms were investigated is the *Job Shop Scheduling* (JSP) problem and the *Resource Constrained Scheduling Problem* (RCSP). The most popular value-iteration type model-free RL algorithm is Q-learning. Several variants of this algorithm were studied to solve resource allocation problems, such as JSP. One of the earliest solutions to the problem can be found in [10]. Kernel methods and clustering were applied to improve fitted Q-learning based resource allocation in [11]. A deep learning based approach, called *Deep-Q-Networks* (DQN), were studied for the JSP in [12][13]. Policy gradient type methods for solving JSP were also investigated: in [14] the authors proposed a multi-agent policy gradient method and in [15] a policy-iteration type actor-critic algorithm was presented.

A more healthcare oriented application of RL based resource allocation was presented in [16], where the hospital resource management problem was treated as a business process and a Q-learning based solution was used. This method was then applied to address the problem of optimizing resource allocation in a radiology CT-scan examination process. A simulation-based approximate dynamic programming (ADP) approach was suggested in [17], which considered both stochastic service times and uncertain future arrival of clients. In that work experimental investigations were concluded using data from the radiology department of a hospital. A deep RL based solution to patient scheduling in emergency departments was investigated in [18], where the scheduling problem was formulated as a Markov decision process and a DQN was designed to determine an optimal scheduling policy. In [19], an advantage actor-critic algorithm was applied to schedule appointments in various challenging hospital environments. Experimental comparisons with heuristics are also shown.

Here, we introduce an SPSA (Simultaneous Perturbation Stochastic Approximation) [21] based policy gradient type RL method for the resource management of CAR T-cell therapies. Policy gradient methods are preferred in continuous state spaces over value function based methods, and building on an efficiently parametrized policy makes the need of black-box type function approximators, such as deep neural networks, superfluous. Furthermore, since we do not necessarily have access to the derivatives of the control policy, REINFORCE type methods are infeasible, therefore, we work with direct estimates of the gradient provided by the SPSA method in a dimension-independent way.

The *main contributions* of the paper are as follows:

1. A flexible *simulation model* of the CAR T-cell therapy is presented. It allows the parallel execution of several simulation instances, to speed up the learning process.
2. A closed-loop *RL-based resource management and scheduling strategy* is proposed whose parameters are optimized by an SPSA-based *policy gradient* method.
3. The effectiveness of the resource management system is demonstrated via *numerical experiments*.

2. The AIDPATH project

In this section we describe the overall infrastructure of the AIDPATH project and the automated CAR T-cell platform.

The concept of AIDPATH, which is an EU funded H2020 project, consists of a manufacturing infrastructure installed at the clinical environment for the automated, data-driven CAR T-cell production, considering all physical assets, for example patients, materials, devices and specialized staff, furthermore the existing IT and logistic systems in the hospital.

On the physical level, the automated CAR T-cell production consists of heterogeneous machines and devices, which are connected via a standardized interface in the COPE integration framework, see Fig. 2. In this framework, all devices are controlled through a service-oriented software, which allows centralized management and supervision of the manufacturing processes. The control software is supported by a digital twin (AI1) that defines the initial process parameters and an ML-based process control algorithm (AI2) that adaptively adjusts these parameters during the operation.

The manufacturing and the resource planning systems optimize the logistic processes with AI-based solutions assuring an unobstructed, efficient production process. The production scheduling (AI3) module is closely linked to the control software and aims at generating production schedules for the manufacturing platform based on durations and characteristics of the biological processes. The resource management (AI4) module focuses on the personalized CAR T-cell therapies themselves, optimizing the usage of equipment and staff and aligns them with the production schedule.

The AIDPATH infrastructure also consist of a decision support system (AI5), which provides a model for personalizing the CAR T-cell product and treatment, based on a historical dataset, which includes information about the previous patients, their reaction to the infusion and their recovery during the follow-up period.

The LogiqCare Platform provides a secure and reliable cloud-based data management system in the AIDPATH project. LogiqCare stores and processes historical data for the training of ML algorithms, as well as establishes a data pipeline for the AI-based applications. It also lays the foundation of the data exchange between different hospitals in the future.

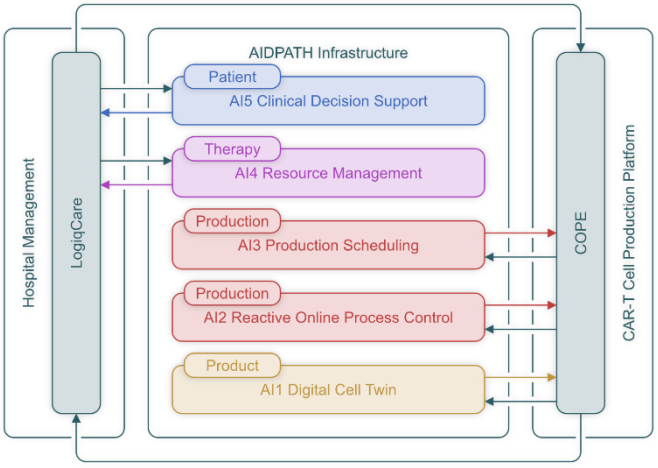

Fig. 2. AIDPATH platform

3. Problem setting

3.1. The CAR T-cell therapy protocol

In this section, we outline a treatment protocol of CAR T-cell therapies. In addition to the six FDA-approved products, a variety of novel CAR T-cell products are being developed and clinical protocols are still under development. Hence, treatment schedules can vary between centers, products and protocols. Here, we overview the current version of the CARAMBA protocol which is used in the Würzburg University Clinic.

The therapy consists of several medical procedures, Table 1 presents the main steps of the therapy along with the required resources and optimal time frame for each step. The equipment and staff required for each step can be divided into two groups: resources that are only used for the CAR T-cell therapies, and resources that are used by other departments in the hospital, as well. Shared resources include the ECG, MRT, PET imaging and the Apheresis units. In Table 1 the “specialist” and the “study nurse” abbreviate a physician and a nurse specialized in immunotherapies, working full-time on these therapies.

Table 1 Therapy overview

Therapy step	Equipment and staff	Time frame
Screening/Eligibility assessment	Specialist Doctor in Training Study Nurse ECG MRT	Day -31 to day -29
Eligibility check for Leukapheresis	Specialist Study Nurse	Day -29 to day -26
Leukapheresis	Apheresis	Day -26
CAR T Production	Production	Day -26 to Day -12
Quality Control	-	Day -12 to Day -5
Baseline assessment	Specialist Doctor in Training Study Nurse MRT Routine Nurse PET Imaging	Day -8 to Day -5
Lymphodepleting Chemotherapy assessment	Specialist	Day -5
Lymphodepleting Chemotherapy preparation	Study Nurse	Day -5
Lymphodepleting chemotherapy	Specialist Study Nurse	Day -5 to Day -2
Break	Specialist Routine Nurse	Day -2 to Day 0
Pre-infusion assessment	Specialist Study Nurse ECG	Day 0
CAR T product infusion	Specialist Study Nurse Study Nurse	Day 0
Post-infusion assessment	Specialist Study Nurse ECG Routine Nurse	Day 0
Follow-up	Specialist Routine Nurse	Day 1 to Day 28

3.2. Resource management for CAR T-cell therapies

A resource allocation problem typically includes scarce, reusable resources and non-preemptive, time-dependent, interconnected tasks (e.g., with precedence constraints) and seeks to find an allocation of these resources to tasks, such that a given objective is optimized subject to the constraints. Such problems typically have the following components:

- a set of available *resource types* with their *amounts*;
- a set of *tasks* to be completed, each using a specified number of resources from each resource type;
- a set describing the *interconnections* of the tasks;
- and an *objective function* (for example, a set of costs or returns for each task and resource).

In the context of resource management for personalized CAR T-cell therapies, tasks are the procedures that a patient has to undergo and the resources are the key staff and medical equipment, detailed in Table 1. The interconnection of the tasks is given by the order of the procedures. Finally, the aims of the resource management process are as follows:

1. To minimize the *number of protocol deviations* during the CAR T-cell therapies because of unavailable shared and non-shared resources at the hospital.
2. To reduce the *completion time of the treatment* of the last scheduled patient (closely related to maximizing the efficient utilization of the available resources).

A “protocol deviation” is a change, divergence, or departure from the study design or procedures defined in the protocol, e.g., a treatment is not done in its prescribed time window.

4. Reinforcement learning based resource management

In this section we present our reinforcement learning (RL) based resource management solution, which is a simulation-based optimization method. Our solution consists of two main parts: (1) a *simulation model* of the therapy, which works with stochastic task durations and even allows executing several simulations simultaneously, and (2) an SPSA-based policy gradient type *RL algorithm* that optimizes the scheduling of the patients via a priority matrix and therapy start times, exploiting the flexible simulation model. An overview of the two components integrated in the infrastructure of the AIDPATH project is shown in Fig. 3, where P1-P3 denotes the patients.

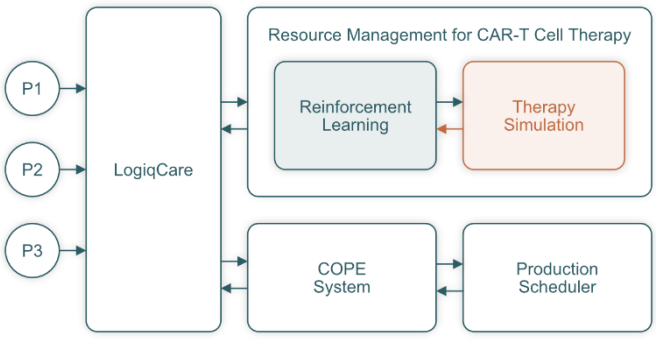

Fig. 3. Integration of the resource management module in AIDPATH

4.1. Simulation model of the therapy

The simulation is modelling the CAR T-cell therapy based on the clinical protocol, detailed in Table 1, and generates a schedule based on *patient-procedure priorities* and *patient start dates*. The simulation model has other uses, as well, such as generating test data and visualizing the therapy process. The model also includes a data interface, which serves as an API (application programming interface) for the machine learning algorithm, furthermore, it can be configured through the data interface, so that different medical procedures, resource quantities and uncertainty settings could be analyzed. Configurable reservations for the resources and eight-hour work shifts are also included in the simulation model. With the configurable reservations, e.g., the unavailability of the shared resources used by other departments can be simulated.

In our resource management concept, the order in which the therapy procedures are executed is controlled by priorities. There are two parameters influencing the priorities: the patients and the procedures. This enables to express cases, when some patients are considered to be more urgent due to their health status, or when some procedures are more important to be executed. Although, this priority-based representation provides a flexible framework in a technical sense, any decision regarding the actual urgency of individual patients or procedures are made outside the resource management module, by the physicians supervising the treatment.

Unlike standard priority-based approaches which only consider patients waiting for the procedures, and choosing the patient-procedure pair with the highest priority, our simulation model also operates with *partial prediction*, i.e., not only waiting patients are considered, but also patients with an ongoing procedure and patients waiting for the allotted day of the next step. In this way, even if a procedure could be started for a patient, it will not be started if that would cause delay for a patient-procedure pair with a higher priority in the future. This prediction only considers the next procedures and assumes the average (i.e., expected value of) procedure times, since the realized times are stochastic. The execution of the simulation is further influenced by the therapy start days. This means that not every therapy starts at the beginning of the simulation, but it can be specified, on which day a patient’s therapy is started.

The simulation applies the AnyLogic toolkit which provides three different paradigms that can be used independently or combined: discrete event, agent based and system dynamics. The CAR T-cell therapy simulation primarily applies the discrete event paradigm, where the therapy is described by the process modelling library. The procedures of the therapy are represented by services that require one or more resources from the defined resource pools. Since the priority-based control is quite complex with partial prediction included, the standard priority queues are not used. The agents are waiting in wait blocks and the custom priority-based control is implemented uniquely. The resources and the patients are represented as agents, but there is neither any decision inside the agents nor communication between them. A simulation server is also applied which can run several simulations in parallel. This can be utilized to accelerate the proposed learning process.

4.2. SPSA-based policy gradient for scheduling

In order to solve the therapy scheduling problem, we have developed a policy gradient-based solution. *Policy gradient* methods [20] are a branch of RL algorithms that directly optimize a parametrized control policy, by using gradient descent type recursive optimization techniques. Unlike standard RL solutions, in our proposed method, the input of the environment (hospital simulation) is not the actions (selecting a patient for a given resource), but the parametrized policy, and the simulation does the action selection based on this control policy. The output of the simulation is the therapy schedule that is generated for the given policy. Consequently, it is an offline RL solution, since the algorithm does not update the policy after every state transition of the simulation, but only after the simulations are finished, and the total cost is revealed. An illustration of the RL agent - hospital simulation interaction is illustrated in Fig. 4. In every optimization iteration the agent updates the control policy parameters θ_k based on the calculated costs C_k for the given schedule S_k generated by the simulation.

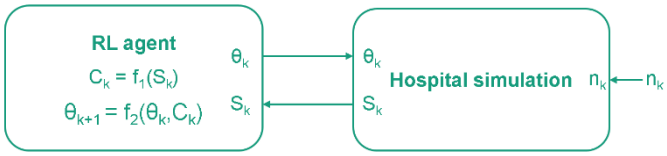

Fig. 4. RL agent-simulation interaction

In our offline RL solution, we use an *SPSA-based policy optimization* method. This algorithm estimates the gradient by computing finite differences from the observed (cumulative) costs. An advantage of the SPSA method is that it only needs two observations to get a gradient estimate, independently of the number of parameters (i.e., the dimension of the problem).

We further enhance the estimate by using mini-batches. The pseudocode of our proposed method is shown in Fig. 5.

At the beginning of the optimization process an initial policy is generated. Then, in each iteration, the actual control policy is perturbed and pairs of alternative policies are generated. With these perturbed resource control policies, the simulation environment is evaluated. From the schedules S_k, returned by the simulations (i.e., in a mini-batch), a cost is computed, which is determined by the optimization objectives detailed in Section 3.2. Using these costs, gradient estimates of the objective function in θ_k, are calculated. Finally, the policy parameters are updated using mini-batch (stochastic) gradient descent.

The policy parameter vector θ_k consists of two parts, a start day vector $\theta_{k,s}$ and a priority vector $\theta_{k,p}$, therefore $\theta_k := [\theta_{k,s}, \theta_{k,p}]$. The start day vector is an *n*-sized real vector, where *n* is the number of patients, and it determines the time when a patient's therapy should start relative to the current date in days. The priority vector is a vector of size $n \cdot m$, where *m* is the number of procedures, and it defines a priority value for every patient-procedure pair. The simulation environment expects a positive integer value given in minutes as the start day input. We apply a transformation to the start day vector $\theta_{k,s}$, where the absolute value of every element in the vector is weighted by a proper constant (conversion from days to minutes) and rounded up to fit the input requirement of the simulation.

Algorithm SPSA Policy Gradient Algorithm for Resource Management

1: Initialize the parameter vector of the resource management policy, θ_1
2: **for** $k = 1$ to $k_{\max}$ **do**
3: Generate multiple perturbed parameter vector pairs centered at θ_k: $(\theta^+_{k,1}, \theta^-_{k,1}), \ldots, (\theta^+_{k,b}, \theta^-_{k,b})$, where b is the size of the mini-batch
4: Simulate the hospital environment with each resource control policy defined by these perturbed parameter vectors to obtain cost-pairs, $(C^+_{k,1}, C^-_{k,1}), \ldots, (C^+_{k,b}, C^-_{k,b})$, where $C^{\pm}_{k,m}$ is the cost we got for $\theta^{\pm}_{k,m}$
5: For each cost-pair, estimate the gradient vector of the objective function in θ_k using the SPSA method, to obtain $g_{k,1}, \ldots, g_{k,b}$
6: Average these gradient estimates to get an improved estimate $\bar{g}_k$
7: Update the parameters of the policy via gradient descent using $\bar{g}_k$
8: **return** $\arg\min_{\theta_k \in \theta_1 \ldots \theta_{k_{\max}}} C(\theta_k)$, the policy with the smallest cost

Fig. 5. Pseudocode of the SPSA-based policy gradient algorithm

5. Numerical experiments

In this section we present some demonstrative simulation experiments about the effectiveness of our proposed solution. We mainly investigated the performance of our scheduling (resource management) optimization algorithm in the case when both the start day and the priorities were optimized.

We studied a configuration with seven patients and three production- and apheresis units. All of the other resource quantities were set to their minimum values. As the main objective of resource management is to minimize the expected number of protocol deviations, we investigated this quantity in our experiments. The total number of protocol deviations for all of the patients during training is illustrated in Fig. 6, where we have averaged the results of ten independent optimization runs. It can be seen that the number of protocol deviations approached zero during the training phase, demonstrating the viability of the proposed approach. Note that a protocol deviation can also occur as a result of the uncertainty of the medical procedures, hence even an optimal schedule could sometimes produce protocol deviations, e.g., when the variance of the procedure durations is large. Hence, the results indicate that our algorithm finds a quasi-optimal schedule.

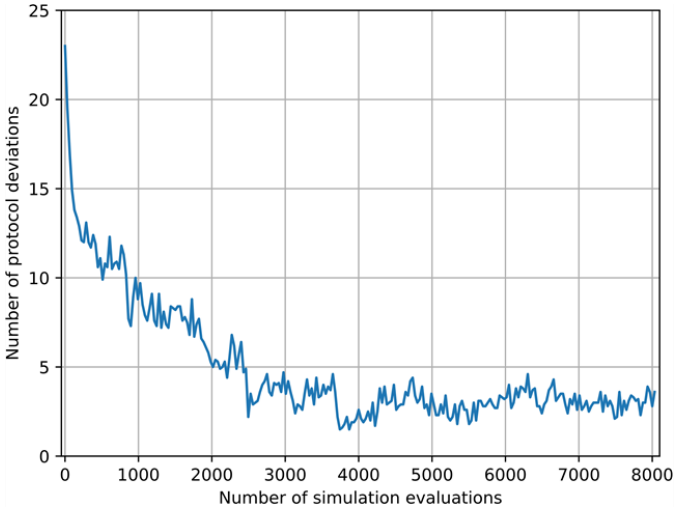

Fig. 6. Protocol deviations during training

We also studied the number of protocol deviations patient-wise for the initial and the optimized resource management

policies, which is shown in Fig. 7. The initial policy is a First Come-First Served (FCFS) policy (dispatching rule), hence the patients who were admitted earlier had higher priorities. In Fig. 7 it can be observed that patients 5, 6 and 7 have no protocol deviations, since they are prioritized to get all the resources. Since we configured the number of production units to three, which is the main bottleneck of the therapy, it follows that the expected number of patients who can be treated without protocol deviations is also three. All the other patients have a high number of protocol deviations, because waiting for the production causes a huge delay. Fig. 7 also illustrates that the resource control policy generated by our algorithm is quasi-optimal since only one patient has one protocol deviation after the optimization (the numbering of the patients is irrelevant).

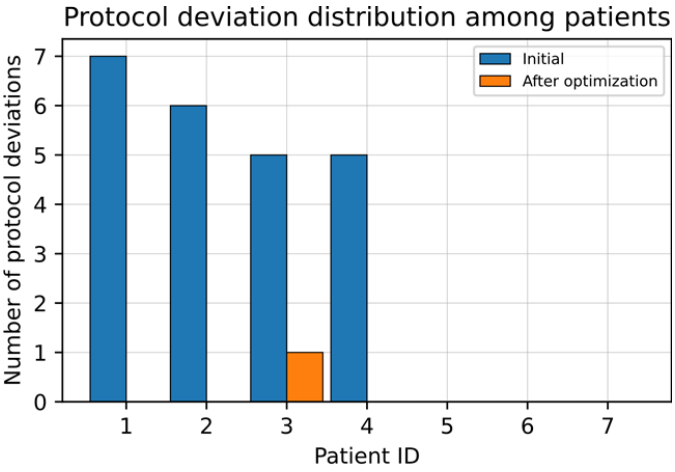

Fig. 7. Distribution of protocol deviations among patients

6. Conclusions

CAR T-cell therapies are transformative new treatments in the field of immunotherapy for which the efficient allocation of the required resources is challenging. In this paper, a novel reinforcement learning (RL) based resource management strategy was proposed, developed within the framework of the AIDPATH project. For this, first a flexible simulation model of the therapy was developed. Then, we suggested a method to optimize a priority-based resource control policy. This was achieved by a derivative-free policy gradient method which builds on the SPSA (Simultaneous Perturbation Stochastic Approximation) algorithm. Numerical experiments were also presented supporting the effectiveness of the approach. The results are indicative of the phenomenon that our method finds quasi-optimal solutions and significantly reduces the protocol deviations during the therapies. Expanding the approach to other kinds of therapies is a subject of future research.

Acknowledgements

This research was supported in part by the European Commission's Horizon 2020 program within the framework of the AIDPATH project, grant agreement number 101016909.

Sz. Szentpéteri, K. B. Kis, P. Egri and B. Cs. Csáji were also supported in part by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory program of Hungary.

References

[1] Gill, S., June, C. H., 2015. Going Viral: Chimeric Antigen Receptor T-Cell Therapy for Hematological Malignancies. Immunological Reviews, 263. Pages 68-89.

[2] Barrett, D. M., Singh, N., Porter, D. L., Grupp, S. A., June, C. H, 2014. Chimeric Antigen Receptor Therapy for Cancer. Annual Review of Medicine, 65. Pages 333–347.

[3] Ortíz-Maldonado, V., Rives, S., Castellà, et al, 2021. CART19-BE-01: A Multicenter Trial of ARI-0001 Cell Therapy in Patients with CD19+ Relapsed/Refractory Malignancies. Molecular Therapy : The Journal of the American Society of Gene Therapy, 29(2). Pages 636–644.

[4] Martínez-Cibrián, N., Ortiz-Maldonado, V., Español-Rego, M., et al, 2023. The Academic Point-of-Care Anti-CD19 Chimeric Antigen Receptor T-Cell Product Varnimcabtagene Autoleucel (ARI-0001 cells) Shows Efficacy and Safety in the Treatment of Relapsed/Refractory B-Cell Non-Hodgkin Lymphoma. British Journal of Haematology. Pages 1–9.

[5] Oliver-Caldés, A., González-Calle, V., Cabanas, V., et al., 2023. Fractionated Initial Infusion and Booster Dose of ARI0002h, a Humanised, BCMA-Directed CAR T-Cell Therapy, for Patients with Relapsed or Refractory Multiple Myeloma (CARTBCMA-HCB-01): A Single-Arm, Multicentre, Academic Pilot Study. The Lancet Oncology 24(8). Pages 913-924.

[6] McLellan, A.D., Ali Hosseini Rad, S.M., 2019. Chimeric Antigen Receptor T-Cell Persistence and Memory Cell Formation. Immunology & Cell Biology, 97. Pages 664-674.

[7] Sharma, P., Kasamon, Y. L., Lin, X., Xu, Z., Theoret, M. R., Purohit-Sheth, T, 2023. FDA Approval Summary: Axicabtagene Ciloleucel for Second-Line Treatment of Large B-Cell Lymphoma. Clinical Cancer Research, 29(21). Pages 4331-4337.

[8] Seif, M., Einsele, H., Löffler, J, 2019. CAR T Cells Beyond Cancer: Hope for Immunomodulatory Therapy of Infectious Diseases. Frontiers in Immunology, 10.

[9] Schett, G., Mackensen, A., Mougiakakos, D., 2023. CAR T-Cell Therapy in Autoimmune Diseases. Lancet 402. Pages 2034–2044.

[10] Aydin, M., Öztemel, E., 2000. Dynamic Job-Shop Scheduling Using Reinforcement Learning Agents. Robotics and Autonomous Systems 33, Issues 2–3, 30 November 2000, Pages 169-178.

[11] Csáji, B. Cs., Monostori, L., 2008. Adaptive Stochastic Resource Control: A Machine Learning Approach. Journal of Artificial Intelligence Research (JAIR), 32. Pages 453–486.

[12] Luo, S., 2020. Dynamic Scheduling for Flexible Job-Shop with New Job Insertions by Deep Reinforcement Learning. Applied Soft Computing, 91. June 2020, 106208.

[13] Han, B., Yang, J., 2020. Research on Adaptive Job-Shop Scheduling Problems Based on Dueling Double DQN. IEEE Access 8. October 9, 2020. Pages 186474 – 186495.

[14] Gabel, T., Riedmiller, M., 2011. Distributed Policy Search Reinforcement Learning for Job-Shop Scheduling Tasks. International Journal of Production Research, 50.

[15] Liu, C., Chang, C., Tseng, C., 2020. Actor-Critic Deep Reinforcement Learning for Solving Job Shop Scheduling Problems. IEEE Access, 8.

[16] Huang, Z., Aalst, W., Lu, X., Duan, H., 2011. Reinforcement Learning Based Resource Allocation in Business Process Management. Data & Knowledge Engineering, 70.

[17] Schuetz, H. J., Kolisch, R., 2012. Approximate Dynamic Programming for Capacity Allocation in the Service Industry. European Journal of Operational Research, 218.

[18] Lee, S., Lee, Y. H., 2020. Improving Emergency Department Efficiency by Patient Scheduling Using Deep Reinforcement Learning. Healthcare, 8.

[19] Gomes, T.T., 2017. Reinforcement Learning for Primary Care E-Appointment Scheduling. Technical Report.

[20] Sutton, R. S., Barto, A. G., 2018. Reinforcement Learning: An Introduction (2nd ed.). A Bradford Book, Cambridge, MA, USA.

[21] Spall, J. C., 2003. Introduction to Stochastic Search and Optimization (1st. ed.). John Wiley & Sons, Inc., USA