Simulation, optimization, and process mining: practicals applications in healthcare

Roberto Aringhieri[1], Matteo Di Cunzolo[1], Matteo Dutto[1], Laura Genga[2], Alberto Guastalla[1], Mirko Locatelli[3], Laura Pellegrini[3], Giulia Ruffini[1], Emilio Sulis[1,*], Lavinia Chiara Tagliabue[1] and Adriana Boccuzzi[4]

[1]*Computer Science Department - University of Turin, Via Pessinetto 12, 10152, Torino, Italy*
[2]*Eindhoven University of Technology, Groene Loper 3, 5612, The Netherlands*
[3]*Department of Management - University of Turin, Corso Unione Sovietica 218 bis, 10134, Torino, Italy*
[4]*A.O.U. San Luigi Hospital, Regione Gonzole 10, 10043, Orbassano, Italy*

Abstract
The paper describes the contribution of three methodologies in process research projects through practical applications in healthcare. On the simulation side, an integration of Building Information Modelling in the medical field has been proposed, besides discrete-event and agent-based modeling. Processes can be obtained from hospital information systems in order to derive log files and studied with process mining tools. Regarding optimization of healthcare processes, analyses have focused on the integration of process mining and optimization.

Keywords
Simulation, Process Mining, Optimization, Building Information Modelling

1. Introduction

Healthcare organisations are increasingly aware of the importance of focusing on medical and decision-making processes, including through the application of Artificial Intelligence (AI) techniques. This contribution includes research approaches in healthcare by applying different AI techniques: simulation, process mining, optimization. At the core there is the use of data stored in an Hospital Information System (HIS). Analytical and decision-making tools can improve healthcare management in daily work, as well as in the major emergencies, such as the recent COVID-19 pandemic. The following sections focus on different techniques applied on healthcare processes.

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
roberto.aringhieri@unito.it (R. Aringhieri); matteo.dicunzolo@edu.unito.it (M. Di Cunzolo); matteo.dutto174@edu.unito.it (M. Dutto); l.genga@tue.nl (L. Genga); alberto.guastalla@unito.it (A. Guastalla); mirko.locatelli@unito.it (M. Locatelli); laura.pellegrini@unito.it (L. Pellegrini); giulia.ruffini@unito.it (G. Ruffini); emilio.sulis@unito.it (E. Sulis); laviniachiara.tagliabue@unito.it (L. C. Tagliabue); adriana.boccuzzi@unito.it (A. Boccuzzi)
0000-0002-5170-2630 (R. Aringhieri); 0009-0000-7182-3897 (M. Dutto); 0000-0001-8746-8826 (L. Genga); 0000-0003-1852-2221 (A. Guastalla); 0000-0002-3059-4204 (M. Locatelli); 0000-0002-8729-4907 (L. Pellegrini); 0009-0003-4179-496X (G. Ruffini); 0000-0003-1746-3733 (E. Sulis); 0000-0003-0100-3169 (L. C. Tagliabue); 0000-0002-5789-1997 (A. Boccuzzi)

CEUR Workshop Proceedings (CEUR-WS.org)

2. Simulation and Building Information Modeling in Healthcare

Crowd modeling and simulation are mainly applied in the Architecture and Construction field to model and simulate crowd and pedestrian dynamics to optimize design solutions and to support safety management strategies in large spaces and buildings [1, 2]. On the other hand, the Building Information Modeling (BIM) methodology is employed as a proficient means to manage complex buildings, as it serves as a relational database repository that integrates several data typologies, e.g., dimensional data, space functions, and spaces and building elements characteristics [3]. Furthermore, BIM can serve as a foundation for conducting simulations, which then enable the analysis of different design scenarios, potential hazards, and other factors that may impact the performance and safety of a building.

We applied crowd simulations on an existing healthcare facility retrieving data from a BIM model aiming to optimize patients, members of the staff, and external people flows. The goal is to model and simulate the occupant movements to identify the less invasive interventions on the building layout and space organization to support the optimal use of the spaces.

2.1. Background and Methodology

Case study. BIM and crowd simulation techniques are applied to the Cottolengo Hospital for scenario analysis, verifying the patients' paths, travel times, and crowding levels of the hospital blood drawing center. The simulation includes all the activities performed by the patients from the arrival to the exit from the hospital. Two scenarios are simulated: the standard conditions of the blood drawing center with a total of 140 patients and the scenario with a peak attendance of 220 patients.

Building Information Modelling and crowd simulation. The first step of the methodology is the definition of the BIM model of the blood drawing center with Autodesk Revit. The BIM model represents a unique database of the building for the following data:

- space geometrical data;
- space functions;
- maximum number of patients allowed in each space (in standard conditions and according to the COVID-19 restrictions);
- location of stairs, elevators, entrances, and exits.

The BIM model is used as a basis for the simulation model creation regarding the space characteristics and data. The simulation is performed with Incontrol Simulation Pedestrian Dynamics. The modeled spaces are:

- on the ground floor: entrance, reception, and corridor connecting with the elevator and stairs;
- on the first floor: two waiting areas (W1 and W2), a registration area, and two blood draw rooms. Once defined the space geometry and related static data from the BIM model, the simulation is set defining the following parameters:
- agent profiles: three agent profiles (i.e., able-bodied patients, disabled or pregnant patients, and in-hospital patients), each with specific walking speed and size (e.g., to consider the wheelchair for disabled patients);
- activities: entrance, admission, moving from the ground floor to the first floor, waiting for the registration, registration, waiting for the blood test, blood test, and exit.
- activity routes (i.e., the sequences of activities): each agent profile has its own activity route, considering that able-bodied patients are expected to use the stairs, while disabled or pregnant patients are more likely to use the elevator to reach the first floor. In-hospital patients, unlike the other two agent profiles, come from other hospital departments, are not required to check in and register, and have priority for the blood test. A total of three activity routes are defined.
- agent generators (i.e., the number of users created in the simulation and the time they enter the simulation): each scenario, i.e., the standard conditions and peak attendance scenario, requires three generators, one for each agent profile, according to the average number of patients that are typically registered in the blood drawing center in the standard and peak conditions respectively.

In addition, the simulation is set up for all agents to respect the interpersonal distance of 1 meter according to COVID-19 pandemic requirements. The outputs of each of the two simulated scenarios are the following:

- density maps of each floor plotting at each point in the space the maximum level of people per square meter recorded during the whole simulation;
- charts of the travel times between two activities;
- charts of the number of users in the whole simulation space or in specific areas during the whole simulation;
- 2D and 3D videos of the simulation.

2.2. Analysis of blood drawing center travel times and crowding levels

The simulations of the two scenarios enabled the identification of overcrowded areas and critical activities causing crowding phenomena. Figure 1 shows the density maps of the ground and first floor in the peak attendance scenario. The waiting area W2 is the most overcrowded area exceeding the limit of 2 people per square meter as imposed by COVID-19 pandemic requirements. In addition, Figure 1 includes a chart showing the average and maximum number of people during the whole simulation in waiting area W2. According to the chart, the maximum capacity of the room (i.e., 13 patients) is exceeded by almost double for three out of four hours of the simulation. The activity identified as the cause of the overcrowding phenomenon is blood testing.

To minimize the overcrowding phenomenon different hypotheses are made prioritizing the ones with the lowest impact on the building layout and proposing simple changes in the functions of the rooms. In particular:

- the patient flows can be better managed via a booking system enabling the staggered entry of patients, thereby preventing overcrowding and reducing the initial flow.
- a room, located between the two current blood draw rooms and currently not utilized, could be repurposed to accommodate one or two additional blood testing areas, thereby expediting the blood test process, or could at least be used as an additional waiting area.

The proposals only involve changes in the intended use of the rooms without modifying the layout of the spaces. This allows for the implementation of the proposed hypotheses in a short time, without interrupting or disturbing hospital activities, and with minimal costs. Summarizing, the proposed methodology, through the integration of BIM and crowd simulation, enables the verification of the building usage patterns, user flows, and layout during the design or in-use phase in a virtual way. Consequently, it is possible to test and verify different scenarios and hypotheses without inconveniencing users and interrupting building activities.

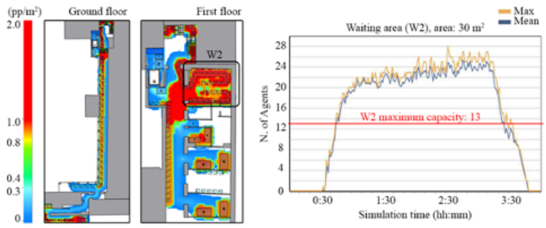

Figure 1: Ground and first-floor density maps (on the left) and agent count during the simulation in the waiting area W2 (on the right)

3. Simulation and process mining

Simulations can play an important role in healthcare process management [4]. Data stored in an HIS can also be used to address a discrete-event simulation [5]. The output of the simulation can be exploited to generate an event-log concerning the activities performed by an instance (i.e., a patient) in the care pathway.

Agent-based modeling. Agent-based simulation focuses on *agents* to model the healthcare activities [6], e.g. to reconstructs the path taken by patients from the ground floor to the second floor of the Hospital and the main variables distinguishing operators and patients, as well as the times related to patient arrivals, working hours, number of seats in the two waiting rooms and in the corridor. Figure 2 describes an ABM developed with NetLogo 3D, which also facilitates the representation of the process for stakeholders.

4. Healthcare information systems and process mining

Recent techniques exploit data stored in an HIS to explore a Process Mining (PM) perspective [7]. To address process-oriented analysis in healthcare [8], we started from discover the real ED processes of a medium-size city hospital. From the organizational perspective, it is relevant that procedures and resources are well organised and properly distributed within the ED.

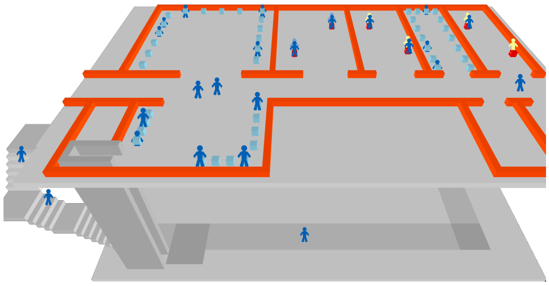

Figure 2: ABM simulation to generate a synthetic event-log

4.1. Background and Methodology

Case study. The HIS of the Orbassano San Luigi Hospital[1] includes dates of the main activities performed for patients from the arrival in the ED to the discharge. A dataset consists of 3,479 patients that have accessed the Emergency Department during a single month (September 2022) with an average of 116 patients per day. Each patient-case have a specific path, related to the different services the patient needs.

Healthcare information system extraction. First, we obtained an event-log from the HIS. The main information in the log are: ID of patient, the name of the activity, the timestamp.

We are provided with a database characterized by

- **key**, with the cases ID of single admission in ED;
- **event**, in which we have different activities, such as triage, take charge, laboratory performance, consultations, other medical services, OBI activity and discharge:
- **medical services**, with the various health services to which the patient is subjected;
- **begin date**, a timestamp column of the begin of each activity;
- **end date/report**, a timestamp column of the end of almost all activities;
- **discharge outcome**, in which we find some possible outcome, such as ”at home”, ”admitted”, ”deceased”, ”discharge against medical advice”.

In order to see in the event column the precise medical services, we have substituted each row with ”other services” with the same row of the medical services column.

Event log construction. In a pre-processing step, we found some possible errors from the dataset, e.g. few cases do not contains the hour of timestamp. We opted to remove these cases from the log. In addition, filtering

[1]https://www.sanluigi.piemonte.it/web/it

on the activity "discharge" we found that a case has been inserted by mistake, so we decided to remove also this case. In order to have a more consistent event log, we decided to eliminate all cases with total performance (time spent in ED) less than 45 minutes. It is unlikely that an ED patient takes only 5 or even 15 minutes in the ED structure. Analyzing the event log, we notice some cases with registration errors (e.g. in a case, the laboratory activity has been taken after the discharge activity). For that reason, we decided to filtered our log on endpoints considering only cases with discharge as endpoint.
The final event-log includes 3,042 traces, i.e. patients. The total number of activities is 116, with a minimum of 3 and a maximum of 32 tasks for each trace - patient, depending on special needs. The number of different variants is 1,142. This shows the complexity of an Emergency Department and the difficulty to provide a standard path to follow. Indeed, each patient is different from one another and requires different services.

Discovery analysis. The three main activities are quite evident in the log: triage, take charge and discharge, then different paths are more complicated due to the patient needs. We used Disco for event-log analysis, as well as process discovery with *fuzzy miner* algorithm[2].

4.2. ED process diagrams

Process discovery allows to focus on real healthcare processes emerged in the event-log. When considering all the activities and paths in the event-log, we obtain process representations very difficult to understand (*spaghetti-like* diagrams). Figure 3 shows the performance of cases, focus the attention on readable process diagram, with 25% of activities and paths.

Describing the process, we notice that all the patients begin with *Triage* and end with *Discharge*. After *Triage*, each patient waits until being taken in charge and then he follows a specific path. The most frequent activities, in this particular case, are related to X-Ray of different part of the body (thorax, foot, hand, ankle, wrist...). Additionally, some patients go directly from *Triage* to *Electrocardiogram* and others pass through the *Short-Stay Observational Unit*. Among the main activities, we find *Laboratory*, which indicates special exams, like blood or other body fluid tests. This activity is quite slow, indeed between *Laboratory* and subsequent activities the patient could wait 45 - 60 minutes in median.

Next steps. To improve the analysis, we could consider some additional features. In the like the priority label. This makes it possible to establish a call order for patients based on urgency. Another interesting task is to categorise each patient according to type of discharge, indeed some patients are discharge at home, others die in ED, some others are hospitalised.

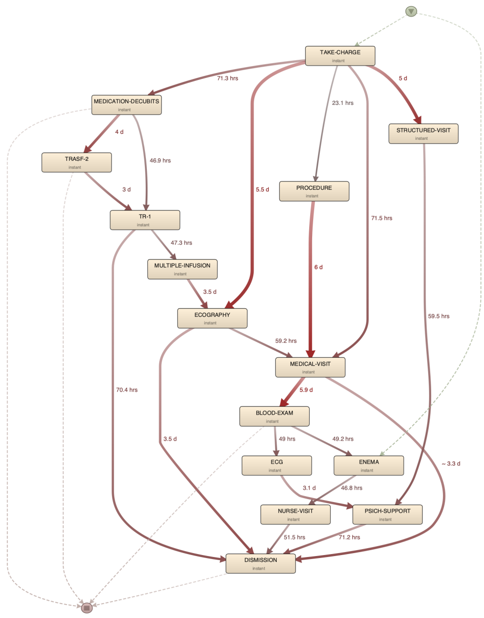

Figure 3: Performance analysis in process discovery

5. Optimization in healthcare

Process mining can be used to strengthen optimization algorithms. Furthermore, online optimization algorithms with look-ahead are an effective and efficient methodology to manage processes in real time, which have been proved to be effective in the management of health processes (operating room planning, emergency medical services, emergency departments, radiotherapy scheduling, ...) ensuring the process management improvement in terms of efficiency and, by consequence, in the quality of the health service provided [9]. In this section, we briefly report some examples of such a combination of methodologies to deal with complex healthcare problems.

Scheduling fair workshifts. This case study has been illustrated in details in [10] and carried on in collaboration with the Cottolengo Hospital in Turin, Italy. Workshift scheduling is particularly relevant for healthcare organizations due to the complexity of managing medical care. Recent research pointed out the importance of workload balancing [11] in the scheduling surgical procedures.

[2]https://fluxicon.com/disco/

We presented a novel support system that automatically generating rostering plans by combining optimization and process mining methodologies. In our approach we exploit the idea for which the patterns included in the realised rostering plans could represent the personal needs and the unspoken habits of the personnel. Based on this remark, we propose a three-step methodological framework – rostering optimization, pattern extraction, pattern adaptation – that it was applied to a real-world scenario.

The first step consists in a multi-criteria mixed integer linear programming model, which models the problem of determining the monthly rostering balancing the monthly working hours of the healthcare personnel in accordance with a list of operative and contractual constraints. The second step consists in a sequence pattern mining algorithm to mine frequent contiguous sequential patterns from data, that is to identify constraints that are possibly not explicitly reported as domain knowledge. The third and last step consists in the adaptation of the rostering plan provided by the multi-criteria model in such a way to increase the number of the patterns mined in the second step while maintaining the feasibility and the optimality of the initial rostering solution.

We reported our decision support method for healthcare management based on real event logs. The computational results proved the capability of our approach to highly improve the quality of the initial rostering plan.

Scheduling interventional radiology procedures. This case study has been illustrated in details in [12] and carried on in collaboration with the Hospital Department of Diagnostic Imaging and Interventional Radiology of the City of Health and Science (CHS) of Turin, Italy.

Optimizing the scheduling of surgery procedure is quite a challenging task, as different aspects, some of which the medical personnel is not completely aware of, may have a strong impact on the scheduling and need to be taken into account. We addressed such a problem by proposing a pipeline combining process mining and optimization techniques We proposed a proof-of-concept of the proposed pipeline applied to the scheduling of the interventional radiology procedures. Leveraging a real-life dataset we built a healthcare event log, and analyzed it in order to discover the main causes for delays and lagging cases. The discovered information – such as the IR procedures requiring more time – is then used to generate an optimized scheduling able to take into account all these aspects.

Figure 4 illustrates an example of the solution computed by the proposed pipeline. In our case study, three different type procedures can be scheduled, that is a clean procedure (on time or delayed is denoted with the colour navy and dark green, respectively), a dirty procedure (on time or delayed is denoted with the colour light green and yellow, respectively), and a Covid procedure (on time or delayed is denoted with the colour orange and red, respectively). Figure 4 proves that the proposed approach determines a more robust scheduling as soon as the number of light green procedures increases.

Figure 4: Illustrative example of the proposed pipeline

Managing the emergency department patient flow. This case study has been illustrated in details in [13, 14] and carried on in collaboration with the the hospital Sant'Antonio Abate di Cantù, Italy. An Emergency Department (ED) operates 24 hours a day, providing initial treatment for a broad spectrum of illnesses and injuries with different urgency. Such treatments require the execution of different activities, such as visits, exams, therapies and intensive observations. Therefore human and medical resources need to be coordinated in order to efficiently manage the patient flow, which varies over time for volume and characteristics. Overcrowding affects EDs through an excessive number of patients in the ED, long patient waiting times and patients leaving without being visited, and also imposing to treat patients in hallways and to divert ambulances. From a medical point of view, when the crowding level raises, the rate of medical errors increases and there are delays in treatments, that is a risk to patient safety.

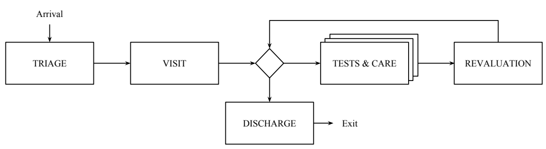

Figure 5: A generic care pathway for a patient within the ED

Because of the wide variety of different patient paths within the ED process (Fig. 5) and the missing of data or tools to mine them, strong assumptions and simplifications are usually made, neglecting fundamental aspects, such as the interdependence between activities and accordingly the access to resources. The access to the usually limited ED resources is a challenge issue: as a matter of fact, the resources needed by each patient are known only after the visit (Fig. 5) while are unknown for all the triaged patients, which could be the majority in a overcrowded situation.

The proposed approach in [14] is based on an online optimization approach with look-ahead embedded in a simulation model: exploiting the prediction based on ad hoc process mining model [13], the proposed online algorithm is capable to pursue different policies to manage the access of the patients to the critical resources. The quantitative analysis – based on a real case study – proves the feasibility of the proposed approach showing also a consistent crowding reduction on average, during both the whole day and the peak time. The most effective policies are those that tend to promote patients (i) needing specialized visits or exams that are not competence of the ED staff, or (ii) waiting for their hospitalization. In both cases the simulation reports a reduction of the waiting times of more than 40% with respect to the actual case study under consideration.

6. Conclusions

Research has approached healthcare processes from different perspectives. The integration of different approaches and disciplines has seen the growing interest of stakeholders (medical staff, hospital executives) in applications of AI to healthcare organizations.

Acknowledgments

This research has been partially carried out within the "Circular Health for Industry" project, funded by "Compagnia di San Paolo" under the call "Intelligenza Artificiale, uomo e società", as well as the "RIsPOSTE" project. We thank the management of "San Luigi" Hospital in Orbassano, CHS, and Cottolengo Hospital.

References

[1] H. Zhang, H. Liu, X. Qin, B. Liu, Modified two-layer social force model for emergency earthquake evacuation, Physica A: Statistical Mechanics and its Applications 492 (2018) 1107–1119.

[2] H. Mayer, W. Klein, C. Frey, S. Daum, P. Kielar, A. Borrmann, Pedestrian simulation based on bim data, 2014.

[3] H. T. T. Le, V. Likhitruangsilp, N. Yabuki, A bim-integrated relational database management system for evaluating building life-cycle costs, Engineering Journal 24 (2020) 75–86.

[4] E. Sulis, A. D. Leva, Public health management facing disaster response: a business process simulation perspective, in: B. Johansson, S. Jain (Eds.), WSC, IEEE, 2018, pp. 2792–2802. doi:10.1109/WSC.2018.8632534.

[5] I. A. Amantea, M. Arnone, A. D. Leva, E. Sulis, D. Bianca, E. Brunetti, R. Marinello, Modeling and simulation of the hospital-at-home service admission process, in: M. S. Obaidat, T. I. Ören, H. Szczerbicka (Eds.), Proceedings of SIMULTECH, SciTePress, 2019, pp. 293–300. doi:10.5220/0007928602930300.

[6] E. Sulis, I. A. Amantea, G. Fornero, Risk-aware business process modeling: A comparison of discrete event and agent-based approaches, in: WSC, IEEE, 2019, pp. 3152–3159. doi:10.1109/WSC40007.2019.9004822.

[7] W. M. P. van der Aalst, Process Mining - Data Science in Action, Second Edition, Springer, 2016.

[8] J. Munoz-Gama, et al., Process mining for healthcare: Characteristics and challenges, J. Biomed. Informatics 127 (2022) 103994.

[9] R. Aringhieri, Online optimization in health care delivery: Overview and possible applications, Operations Research Proceedings 2019 (2020) 357–363.

[10] A. Guastalla, E. Sulis, R. Aringhieri, S. Branchi, C. D. Francescomarino, C. Ghidini, Workshift scheduling using optimization and process mining techniques: An application in healthcare, in: WSC, IEEE, 2022, pp. 1116–1127. doi:10.1109/WSC57314.2022.10015405.

[11] R. Aringhieri, D. Duma, P. Landa, S. Mancini, Combining workload balance and patient priority maximisation in operating room planning through hierarchical multi-objective optimisation, European Journal of Operational Research 298 (2022) 627–643.

[12] M. Di Cunzolo, A. Guastalla, R. Aringhieri, E. Sulis, I. A. Amantea, M. Ronzani, C. Di Francescomarino, C. Ghidini, P. Fonio, M. Grosso, Combining process mining and optimization: A scheduling application in healthcare, in: C. Cabanillas, N. F. Garmann-Johnsen, A. Koschmider (Eds.), BPM Workshops, Springer International Publishing, Cham, 2023, pp. 197–209.

[13] D. Duma, R. Aringhieri, An ad hoc process mining approach to discover patient paths of an emergency department, Flexible Services and Manufacturing Journal 32 (2020) 6–34.

[14] D. Duma, R. Aringhieri, Real-time resource allocation in the emergency department: a case study, Omega 117 (2023) 1–16.