Improve Hospital Management Through Process Mining, Optimization, and Simulation: the CH4I-PM Project

Published version:

DOI:10.1007/s13218-024-00882-5

(Article begins on next page)

Noname manuscript No.
(will be inserted by the editor)

Improve hospital management through process mining, optimization, and simulation: the CH4I-PM project

Massimiliano Ronzani · Emilio Sulis

Received: date / Accepted: date

Abstract The growing digitalization of society opens up the exploitation of new IT techniques in the healthcare sector. This report presents an application of AI techniques such as prediction, optimization, and automated knowledge extraction with process mining from hospital information system data. In addition, a simulation effort with Building Information Modeling and Agent-Based Modeling techniques has been performed. The present report describes practical cases and the lesson learned from planning, management, and coordination activities of the project as a whole.

Keywords Process mining · Simulation · Healthcare · Project report

1 Introduction

The application of Artificial Intelligence (AI) techniques has already demonstrated remarkable success in health sector applications [18], including the organization of medical services with decision support systems. Improving healthcare processes can be achieved through the exploitation of data recorded in hospital information systems (HIS). The relatively new discipline of Process Mining (PM) [1] is a data-driven methodology that provides tools and algorithms for analyzing and visualizing event data, aiming to optimize and enhance business and healthcare processes.

Massimiliano Ronzani
Fondazione Bruno Kessler (FBK)
Via Sommarive 18, 38123, Povo, TN, Italy
E-mail: mronzani@fbk.eu

Emilio Sulis
Computer Science Department of the University of Torino
Corso Svizzera 185, 10149, Torino, Italy
E-mail: emilio.sulis@unito.it

This report describes the contribution of a recent application of PM and AI techniques in a foundation-funded research project (CH4I-PM) in northern Italy[1]. The CH4I-PM project has activated the collaboration of various research bodies and public institutions, with the aim of proposing suggestions for hospital managerial decision-making. Section 2 describes the background and the practical cases (i.e., real-world applications of the proposed research methods), while methods and main scientific results are described in Section 3. Finally, Section 4 discusses lessons learned from the project.

2 Background and Practical Cases

2.1 A project for the healthcare organisations

The general project had a planned duration of 2 years and was funded with 1 million euros from a tender in 2020 in Turin, one of the main Italian cities. Having gone through the pandemic emergency period, an additional 6 months were used, from March 2021 to August 2023. The project focused on using data and AI techniques to improve the organisation of industries for human healthcare.

The partnership proposed by the CH4I-PM project includes different public and private entities with the following hierarchical structure. The financing body interfaced by mean of a *node manager* with a representative of the *Lead group* of the project, i.e. the group which includes the University of Turin (UniTo), an evaluation body (CNR) for the monitoring and final evaluation of the proposed research, and a Competence Centre for networking and results dissemination (CIM). The

[1] Project website: `https://ch4i.di.unito.it/` - Timing of the project and images are in the folder: `https://t.ly/_LiJ8`.

Lead group had a guidance and support role. UniTo also offered scientific expertise by actively collaborating on some of the sub-projects involving hospital research. On the operative side, a *Project leader* coordinates the scientific research activities interacting predominantly with territorial entities, e.g. hospitals. The representative of the *Lead group* belonged to UniTo, while the *Project leader* was the private research centre Fondazione Bruno Kessler (FBK) having a well recognised experience in process-oriented data science. Finally, UniTo strongly collaborated directly with FBK, focusing research efforts on the main hospital complex in Turin "City of Health and Science" (CHS), and the medium-sized hospital "Cottolengo".

2.2 Methods

We summarize all research areas and the relevant methodologies used in the project, based on the expertise of the researchers involved. The order of presentation corresponds to the importance within the project, as determined by the allocation of funds.

Process mining. Process mining is a versatile discipline with growing adoption in healthcare [12]. It leverages data recorded in a HIS to extract event logs that capture the key events that occurred. These logs enable the application of process discovery techniques to analyze real-world processes based on data. Variant analysis can examine variations in processes, which is particularly significant in healthcare for exploring patient pathways and the actual procedures followed [11]. In the CH4I-PM project, the discovered processes have been represented using languages like Directly-Follows Graphs (DFG) or standard modeling languages such as Business Process Model and Notation (BPMN). The combination of machine learning and process mining algorithms supports the development of *predictive process monitoring (PPM)* techniques. These techniques are valuable for forecasting outcomes, identifying the next tasks, estimating time remaining to completion, and other related aspects of processes. This blend of approaches offers a comprehensive and adaptable toolkit for understanding and improving healthcare processes [7].

Explainable AI. Explainable AI (XAI) denotes a collection of techniques used to extract meaningful explanations about the choices made by a machine learning or deep learning model. Applying XAI to PPM technique enables the identification of primary factors contributing to specific predictions made by the PPM model. Recent research shows that XAI can enhance model accuracy by mitigating the influence of features that often lead to prediction errors [13].

Operation research. This discipline exploits mathematical algorithms and analytical methods to make better decisions in complex situations. Several Operation Research (OR) applications are in the areas of optimization problems in Business and Management, Economics, Computer Science, Engineering. In this project, we adopted integer linear programming, combined with PM and PPM, to improve healthcare management [2].

Simulation. Within the framework of computational simulations, it is possible to create a digital twin of health services in order to monitor the progress of the process, capture its costs and possible indicators, aimed at their optimisation. In the project, we explored Building Information Modelling (BIM) to perform a realistic simulation based on the model of the healthcare building [5]. We also explore the adoption of agent-based business process simulation [16] to perform stochastic simulations used to study the interactions between patients, nurses, doctors, modelled as individual agents in a healthcare setting [15,17].

2.3 Practical cases

In the following, we provide a brief description of scientific ideas for each practical case, outlining the primary challenge to be addressed and offering an overview of the methods used to tackle it.

PC1 - Decision support to the home hospitalisation service. The home hospitalisation service (HHS) is a service provided by the CHS. It is a valid alternative to the standard hospitalisation for many acute and chronic exacerbated diseases [9]. The challenge of this practical case is two-fold: the identification of the main reasons which conditions the doctor to select HHS for a patient, and the identification of bad HHS decisions, that consists in patients that have to return to the hospital after a short time. Wrong HHS assignment can, in fact, implicate extra costs for the hospital that needs to hospitalize again the patient, as well as a great inconvenience for the patient. The main challenges in this case is predicting the outcome of home hospitalization for the patient, as well as identifying the key factors that influence doctors' decisions. To address this problem, we utilize PPM and XAI techniques.

PC2 - Optimisation of operating rooms usage. This case focuses on the usage of operating rooms on the Diagnostic Imaging and Interventional Radiology (IR) department of CHS. IR procedures are minimally invasive procedures which allow to achieve a therapeutic goal with the minimum trauma for the patient. Each procedure can be compose by different diagnostic and therapeutic treatment. The main challenge of this case is the optimization of operating rooms. The goal of the optimization is twofold: reducing costs for the hospital and shortening the waiting list of patients in need of IR services. The optimization must also consider potential unforeseen events and delays inherent in daily routines. We address these challenges by combining OR and PPM.

PC3 - Automatic generation of rostering plans. The aim of this case is the automation of the assignment of work shifts to the medical staff at Cottolengo hospital. Many explicit and implicit factors have to be considered when generating a rostering plan, such as operative and contractual constraints (e.g. maximum number of consecutive working days and night), as well as personal needs and the habits of staff which are usually unspoken. In addition, considering the high complexity of the problem, a manually assembled rostering plan may not always be a good plan. For this reason another challenge of the present case is the generation of optimised rostering plans which satisfy at best implicit and explicit constraints. We address this challenge by extracting common patterns of personal preferences from historical data using PM, and applying OR techniques to optimize the rostering plan accordingly.

PC4 - Therapeutic decision support. A fourth practical case focuses on the care pathways of patients with breast cancer within the Breast Unit of the Cottolengo hospital. The aim is to support physicians by extracting relevant patterns from the patient's healthcare pathways which are useful to identify the most suitable diagnostic and therapeutic strategy. The challenge consists in managing the extreme complexity and variety of patient medical histories, which often span many years and are highly individualized. The goal is to gain insights from this data to assess the feasibility of developing a decision-support tool. We employ PM to analyze the care pathway of patients.

PC5 - Simulation for scenario analysis. The last practical case focuses on supporting the analysis of patient movements and activities in healthcare facilities. The method consists of simulation to verify current conditions, identifying bottlenecks and crowded situations that could affect hospital activities efficiency and compliance with COVID-19 pandemic regulations. Based on the results of the simulations, improvements in the management of spaces, activities, and people flows can be proposed, e.g. by modifying the building layout or implementing new user flow management strategies. The hospital management was interested in performing a spatial analysis of the functioning of the service, given the possibility of moving the service to a new building.

3 Scientific results

The outcome of the research in the projects brought different scientifically relevant results, which were published in appropriate venues.

PC1 - Decision support to the home hospitalisation service. Two PPM tasks were evaluated, the first on the outcome of home hospitalisations, the second on the prediction of doctors' choice of hospitalisation (home or ordinary) [4,3]. In the first task, we investigated three dimensions of analysis: 1) the exploitation of the unstructured information contained in the handwritten textual first aid diagnoses. We considered four options: not using the information, using the unstructured textual diagnoses without any transformations, map textual diagnoses into standard ICD-9-CM codes using two NLP techniques based on lexicographic and semantic analysis. 2) The level of generality of the diagnoses: using the hierarchical structure of ICD-9-CM taxonomy, whereby full codes correspond to specific diagnoses, while truncating one or two digits make it possible to return to more general diagnoses (e.g. 491.20 Chronic obstructive bronchitis, without exacerbation, 491.2 Chronic obstructive bronchitis, 491 Chronic bronchitis). 3) The confidence level on the ICD-9-CM code associated to the textual diagnosis by the chosen NLP algorithms: raising the confidence level reduce the number of diagnoses that the algorithm can transform into ICD-9-CM codes but increase the quality of the associations. The results [14] show how on average better performances are obtained by: 1) Considering the transformation of textual first aid diagnoses into ICD-9-CM codes using lexicographic NLP algorithm; 2) considering more general diagnoses (truncated ICD-9-CM codes); and 3) considering only high confidence ICD-9-CM code associations given by the NLP algorithm. The results indicate that incorporating information extracted from textual notes with high confidence significantly improve the quality of the prediction.

The second predictive model aims at reproducing the decision of the doctors about the home hospitalisation of a patient. The PPM model has been trained on a dataset composed by patients which have either been sent to home hospitalisation or to ordinary hospitalisation. The relevant attributes used by the predictive model are personal attributes of the patient, the five main diagnoses assigned to they and the five main procedures to which they has undergone. Both diagnoses and procedures are encoded by the ICD-9-CM taxonomy. The model obtained scores a good accuracy of about 75%. Afterwards, XAI techniques have been adopted to extract from the model the principal causes which condition the model prediction, and which have been assessed by the doctors. This procedure allows the doctors to uncover some unacknowledged causes strongly related to the home hospitalisation decision.

PC2 - Optimisation of operating rooms usage. The initial results on a pipeline combining PM and OR for the optimisation of IR scheduling seem a promising line of research. In particular, process discovery techniques obtained meaningful results about main causes for delays and lagging cases, i.e. the IR procedures requiring more time. In Figure 1 the traces of patient with delay are depicted. This information is integrated into the OR integer linear programming model to determine patients' priority, resulting in a more robust scheduling. A preliminary analysis clearly shows that PM enables the identification of useful knowledge to be exploited in the design of the optimisation model [6].

Another effort in this direction focuses on the integration of PPM models and OR optimisation techniques. We have built an optimisation pipeline in which the optimized schedule generated by the OR algorithm is inputted into a PPM model. This model predicts the delay resulting from the schedule's implementation. If the predicted delay exceeds a certain threshold the schedule is optimised again by the OR algorithm in a more robust way. The cycle stops when the delay predicted by the PPM model is below the given threshold.

PC3 - Automatic generation of rostering plans. A three-step methodological framework has been applied to our real-world scenario. Firstly, a rostering optimisation produces an initial scheduling. A multi-criteria mixed integer linear programming model balances the monthly working hours of the healthcare personnel in accordance with a list of operative constraints. In a second step, pattern extraction techniques identified most frequent habits or patterns included in the rostering plans, which represent implicit constraints or desiderata. Finally, pattern adaptation exploiting a mechanism called "rectangular-adaptation" produces the best plan with refined staff work-shifts satisfying the implicit constraints found during the previous step [8]. The results

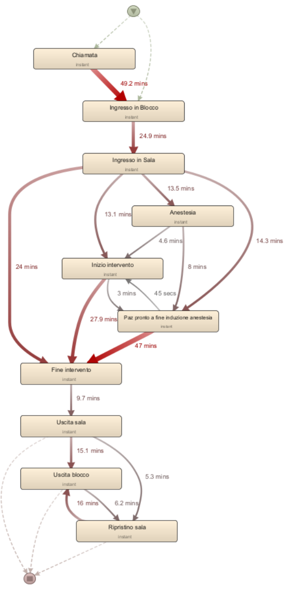

Fig. 1 Directly-follows graph for the activities of patients with delays in the IR department.

show that the proposed framework produces rostering plans that, on average, satisfy 20% desiderata patterns compared to traditional planning.

PC4 - Therapeutic decision support. Clinical pathways of patients in the breast cancer unit are reconstructed from datasets containing radiological, histological, surgical, and oncological reports. Process discovery is then applied to identify the most common diagnostic and therapeutic pathways, while anomalous pathways are identified and discussed with domain experts. In addition, PPM techniques have been applied on historical process data to infer possible evolution of running processes. These techniques provide decision support to physicians regarding diagnostic or therapeutic treatments to be prescribed to patients. More specifically, two PPM models are trained to predict the necessity of a histological exam and the need for surgical therapy for the patient.

A critical issue that is widely present in the predictive field, especially in the medical scenario, is data imbalance. For example, a negative outcome of a histological examination is much more likely than a positive one. These types of data imbalances produce less accurate prediction models. We applied re-sampling techniques that, by removing or generating data, improve the training of models resulting in more robust prediction algorithms. Nevertheless, the PPM results are far to be satisfactory, with an f1 score of 0.55. An effort on this front requires a larger dataset as future work.

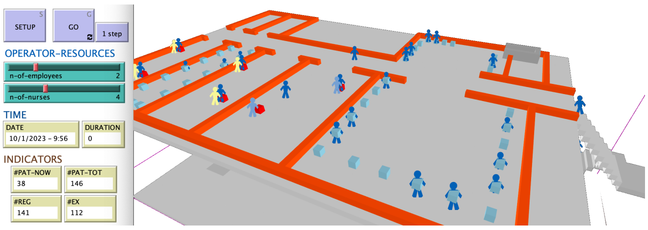

Fig. 2 Agent-based simulation of accesses to the blood sampling area

PC5. Simulation for scenario analysis. The integration of BIM and crowd simulation (simulation of large groups of people moving together) made it possible to virtually verify building usage patterns, user flows and layouts during the design or utilisation phase. The agent-based analysis allows parameters related to the number of operators or patient influx to be modified. In this way, different scenarios can be tested and suggested to healthcare managers. The two methodologies have been applied with different tools, i.e. Pedestrian Dynamics (InControl) and NetLogo open-source simulation software, as represented in Figure 2. Doctors tested different configurations by changing patient arrivals, operation durations and route lengths as part of the scenario analysis. The selected tools were compared through two simulations with the same parameters, highlighting the merits and shortcomings of each [10].

4 Lessons learned

Planning. The initial high-level planning and subsequent identification of practical cases during the project proved to be functional. Rather than starting with a single established practical case, it was preferred to proceed with multiple smaller ones in parallel. The initial presentations of the skills by the FBK and UniTo researchers to the doctors served to better identify needs and interests. A lesson learned is that a clear and full involvement of doctors as domain experts is necessary from the very beginning of the project. Conversely, a generic support is not enough: a clear involvement and strong interest in the specific research proposed is needed. Another aspect that worked was the fact that meetings were held from the beginning with the few partners relevant to individual practical cases, so as to optimize everyone's time.

Structuring. The structuring of the project included a typical division into work packages, with deliverables (some in double version, e.g. preliminary at month 6 and definitive at the end of the project) and milestones to be achieved, well distributed over the two years. A star structure, with FBK in the center supported by UniTo, and several cases that were relatively small and autonomous from each other, facilitated the management and avoided stagnation (e.g., when waiting for data in one case, one could proceed with the others.)

The practical cases were organized in a matrix structure, where each study could have multiple leaders (either a Lead group or Project leader). These leaders shared specialized skills across various cases (such as process mining experts) and maintained strong coordination among team members to promote communication and information sharing.

Development. An aspect that was highly appreciated by the hospital managers and staff involved in the return and evaluation of the scientific outcome was the ease of reading the results. This is evident both in the analyses with activity diagrams (DFG, BPMN) for the process discovery from the log data, and in the simulations with BIM and ABM, for the immediacy and simplicity of interpreting the results. An aspect that worked well in the project was the willingness of the partners to alter the initially scheduled project times as well as to work even at unusual times. For instance, project partners adapted to the needs of the doctors' working times (e.g., project meetings during lunch time or in the late evening), and to extend the project activities by six months. As European legislation is very restrictive on privacy concerns, an important aspect that has worked in data management is the possibility of using a High Performance Computing research infrastructure (HPC4AI) aiming at designing, developing, and operating experimental innovative applications and services in a privacy-compliant manner. This infrastructure stored the models and enabled the execution of AI algorithms and simulations within the project.

Coordinating. A good practice has been the role of the *Lead group*. On the one side, this group and the responsible partner (UniTo) maintained relations with the financing institution, via the *node manager* as a facilitator (not a controlling role). On the other side, it remains the unique referent of the partners working on the project topics. In such a way, operative partners have a single referent coordinating the activities. This structure in our view largely facilitate the development of individual activities. The evaluation partner, CNR, worked throughout the project with the coordination group to identify, in agreement with the *Project leader*, the best possibilities for proposing an evaluation of the different practical cases. This co-participation in the direction of the activities of the *Lead group* from the very first project action, also preparing the monitoring plans to be carried out in itinere, was instrumental in facilitating the preparation of the evaluation plans, at the same time preventing this partner from being seen as a burden, appearing to the research partners only towards the end of the project, as is often the case with those who have to deal with evaluation.

5 Conclusions

The aim of the project was to support hospital decision-making. This report described the main scientific and project activities, highlighting the lessons learnt. There was no shortage of critical issues typical of this type of project: slow and difficult arrival of the necessary data, anonymisation of the data sometimes not perfect and lengthy, need to communicate between different disciplines and combine medical and computer terminology, availability of medical staff depending on their workload, need to meet project deadlines. We focused this paper on describing the methods adopted in the project, allowing to obtain valuable knowledge and suggestions on eligible patients, on staff rotation, on operating rooms management, as well as on optimizing patient access to medical care. In continuity with the project, some research activities proceeded with the same hospitals, also based on new funded projects that built on the above mentioned results.

Acknowledgements The 'Circular Health for Industry' project has been funded by Compagnia di San Paolo. The research work was partially conducted as part of Circular Health European Digital Innovation Hub (CHEDIH) - Grant Agreement n. 101083745.

References

1. van der Aalst, W.M.P.: Process Mining - Data Science in Action, Second Edition. Springer (2016). DOI 10.1007/978-3-662-49851-4
2. Ali, I., Kannan, D.: Mapping research on healthcare operations and supply chain management: a topic modelling-based literature review. Ann. Oper. Res. **315**(1), 29–55 (2022). DOI 10.1007/S10479-022-04596-5
3. Aringhieri, R., Boella, G., Brunetti, E., Caro, L.D., Francescomarino, C.D., Dragoni, M., Ferrod, R., Ghidini, C., Marinello, R., Ronzani, M., Sulis, E.: Leveraging structured data in predictive process monitoring: the case of the ICD-9-CM in the scenario of the home hospitalization service. In: P. Lucas, F. Stella (eds.) Proc. of the Workshop on Towards Smarter Health Care, AIxIA, Nov. 29th, 2021, vol. 3060, pp. 48–60. CEUR-WS.org (2021). URL `https://ceur-ws.org/Vol-3060/paper-6.pdf`
4. Aringhieri, R., Boella, G., Brunetti, E., Caro, L.D., Francescomarino, C.D., Dragoni, M., Ferrod, R., Ghidini, C., Marinello, R., Ronzani, M., Sulis, E.: Towards the application of process mining for supporting the home hospitalization service. In: A. Marrella, D.T. Dupré (eds.) Proc. of the 1st Italian Forum on BPM, Rome, Italy, Sept. 10th, 2021, vol. 2952, pp. 33–38. CEUR-WS.org (2021). URL `https://ceur-ws.org/Vol-2952/paper_295a.pdf`
5. Choi, J., Leite, F., de Oliveira, D.P.: Bim-based benchmarking system for healthcare projects: Feasibility study and functional requirements. Autom. Constr. **96**, 262–279 (2018)
6. Cunzolo, M.D., Guastalla, A., Aringhieri, R., Sulis, E., Amantea, I.A., Ronzani, M., Francescomarino, C.D., Ghidini, C., Fonio, P., Grosso, M.: Combining process mining and optimization: A scheduling application in healthcare. In: Cabanillas C., et al. (ed.) BPM 2022 International Workshops, Münster, Germany, September 11-16, 2022, *LNBIP*, vol. 460, pp. 197–209. Springer (2022). DOI 10.1007/978-3-031-25383-6_15
7. Di Francescomarino, C., Ghidini, C.: Predictive process monitoring. In: W.M.P. van der Aalst, J. Carmona (eds.) Process Mining Handbook, vol. 448, pp. 320–346. Springer (2022). DOI 10.1007/978-3-031-08848-3_10
8. Guastalla, A., Sulis, E., Aringhieri, R., Branchi, S., Francescomarino, C.D., Ghidini, C.: Workshift scheduling using optimization and process mining techniques: An application in healthcare. In: WSC, Singapore, December 11-14, 2022, pp. 1116–1127. IEEE (2022). DOI 10.1109/WSC57314.2022.10015405
9. Isaia, G., Bertone, P., Isaia, G.C., Ricauda, N.: Home care for patients with chronic obstructive pulmonary disease. Arch Phys Med Rehabil **100**, 664–5 (2010)
10. Locatelli, M., Pellegrini, L., Accardo, D., Sulis, E., Tagliabue, L.C., di Giuda, G.M., et al.: People flow management in a healthcare facility through crowd simulation and agent-based modeling methods. In: J. Phys. Conference series, pp. 1–6 (2023)
11. Mans, R., van der Aalst, W.M.P., Vanwersch, R.J.B.: Process Mining in Healthcare - Evaluating and Exploiting Operational Healthcare Processes. Springer Briefs in BPM. Springer (2015). DOI 10.1007/978-3-319-16071-9
12. Munoz-Gama, J., et al: Process mining for healthcare: Characteristics and challenges. J. Biomed. Informatics **127**, 103994 (2022). DOI 10.1016/J.JBI.2022.103994
13. Rizzi, W., Di Francescomarino, C., Maggi, F.M.: Explainability in predictive process monitoring: When understanding helps improving. In: Dirk Fahland et al. (ed.) BPM Forum, pp. 141–158. Springer, Cham (2020). DOI https://doi.org/10.1007/978-3-030-58638-6_9
14. Ronzani, M., Ferrod, R., Francescomarino, C.D., Sulis, E., Aringhieri, R., Boella, G., Brunetti, E., Caro, L.D., Dragoni, M., Ghidini, C., Marinello, R.: Unstructured data in predictive process monitoring: Lexicographic and semantic mapping to ICD-9-CM codes for the home hospitalization service. In: Bandini, S. et al (ed.) AIxIA 2021, Dec. 1-3, 2021, *LNCS*, vol. 13196, pp. 700–715. Springer (2021). DOI 10.1007/978-3-031-08421-8_48
15. Sulis, E., Mariani, S., Montagna, S.: A survey on agents applications in healthcare: Opportunities, challenges and trends. Comput. Methods Programs Biomed. **236**, 107525 (2023). DOI 10.1016/J.CMPB.2023.107525
16. Sulis, E., Taveter, K.: Agent-Based Business Process Simulation - A Primer with Applications and Examples. Springer (2022). DOI 10.1007/978-3-030-98816-6
17. Sulis, E., Terna, P.: An agent-based decision support for a vaccination campaign. J. Medical Syst. **45**(11), 97 (2021). DOI 10.1007/S10916-021-01772-1
18. Yu, K.H., Beam, A.L., Kohane, I.S.: Artificial intelligence in healthcare. Nature biomedical engineering **2**(10), 719–731 (2018). DOI 10.1038/s41551-018-0305-z