Mathematical Biosciences and Engineering

https://www.aimspress.com/journal/mbe

MBE, 21(7): 6603–6607.
DOI: 10.3934/mbe.2024288
Received: 18 June 2024
Revised: 04 July 2024
Accepted: 07 July 2024
Published: 10 July 2024

Editorial

Advances in computational methods for process and data mining in healthcare

Marco Pegoraro[1,*], Elisabetta Benevento[2], Davide Aloini[2], and Wil M.P. van der Aalst[1]

[1] Chair of Process and Data Science (PADS), RWTH Aachen University, Ahornstraße 55, 52074 Aachen, Germany

[2] Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Largo Lucio Lazzarino 1, 56122, Pisa, Italy

* **Correspondence:** Email: pegoraro@pads.rwth-aachen.de.

Keywords: Process Mining, Process Science, Data Mining, Healthcare, Clinical Data, Medical Data, Computational Methods

1. Introduction

In the course of the 20th century, nearly all aspects of operations in businesses, organizations, and other entities have been disrupted by the introduction of computers and information systems. Early support systems for processes, often implemented on paper or through other analog means, have been progressively superseded by digital computer systems. This has not always occurred in a linear, homogeneous fashion; at times, technology has progressed in bursts, often in correspondence to decade- or century-defining historical moments. For instance, the immediate post-World War II period has seen a dramatic acceleration of computing technology and related research fields.

In the early twenties, a similar phenomenon was induced by the public health crisis caused by the COVID-19 pandemic. In the effort to combat the spread of the epidemic, and in the immediate aftermath, medical practitioners and scientists have attempted to improve the effectiveness of treatments by incorporating recently-developed technologies. A prominent such technology is data science, of which the application has enabled the analysis of healthcare operations with a level of detail never seen before.

Process mining [1] is a branch of data science that enables data-driven analytics based on discrete event data. Through recordings of past process executions collected in event logs, process mining

allows for the automatic discovery of process models and for the measurements of potential deviations between normative models and real-life execution of processes. With the steep increase in demand of domain-specific, multi-perspective data science and analytics in industry, commercial tools based on process mining are enjoying a remarkable success.

While process mining techniques have been developed in the context of business processes, recent years have also seen a proliferation of process mining techniques in a number of new fields of application. One such field is healthcare. A wide array of process mining approaches for analyzing both clinical and administrative data from the healthcare domain have now been introduced [2–4], and the application of process science techniques to the healthcare and medical domains is now an established and rapidly-growing branch of data science [5–7].

In light of all the above, we invited the community of process and data mining researchers to submit novel scientific results on the topic of computational methods for the health sciences to this special issue of the *AIMS Mathematical Biosciences and Engineering* journal. The papers accepted for publication in the special issue introduce new frameworks for clinical and organizational healthcare event data analysis, or obtain novel significant insights on the context of a case study in the healthcare domain.

2. Milestones of process mining in healthcare

The *Process Mining Manifesto*, largely recognized as one of the most influential publications in the history of the discipline, mentions healthcare as one of the non-traditional applications of process mining [8, p. 180]. In it, the process mining community specifically highlights how data from medical record is very difficult to mine, for multiple reasons—processes are partially recorded on paper, data stream connector from/to medical information systems are lacking or downright absent, information is recorded with little structure (e.g., in textual notes or emails).

Throughout the 2010s, improvements in the field have been slowly driven by process mining practitioners and by new data recording technologies adopted by healthcare institutions. Notably, early attempts focused on performing process-oriented analysis on administrative hospital data, which were supported much more accurately. This early research acted as an outpost, allowing data scientists to build bridges with the medical community. Much of this early work is summarized in [9].

In the late 2010s, structured and organized efforts by process scientists led to the establishment of process mining in healthcare as a mature sub-discipline. In 2018, the first dedicated academic event took place, the *Process-Oriented Data Science for Healthcare* workshop (PODS4H) [10]. The initiative was and still is led by the homonymous international interest group, the PODS4H Alliance*.

In 2022, a special issue dedicated to process mining in healthcare significantly expanded the amount of literature available on the topic [11]. Such special issue also contains the paper *"Process mining for healthcare: Characteristics and challenges"* [12], the de-facto manifesto of the field. This article shapes the near future of the discipline, and summarizes the main challenges of process-oriented analysis in healthcare for the next years ahead. Later in the same year, the first summer school on process mining took place; the teaching of process mining applied to healthcare domain had a prominent role in the empirical section of the school. An article version of such lecture about process mining in healthcare has been published as chapter of the *Process Mining Handbook*.

*https://pods4h.com/

3. Articles included in the special issue

In this special issue, four articles have been accepted for publication:

- In *"Identification of drug side effects with a path-based method"*, Jiang et al. [14] builds a set of networks to compactly represent drugs, side-effects, and drug side-effects in a clinical setting. They then used a path-based exploration technique to estimate and assess the possible side-effects of a given farmaceutical. They show that their method surpasses state-of-the-art approached for network-based side-effect estimation.
- In *"A framework for multi-perspective process mining into a BPMN process model"*, Tiftik et al. [15] introduce a framework that merges process discovery with multi-perspective mining, obtaining a tool able to couple additional process perspectives (control-flow, data, organizational) to the output of a selection of discovery algorithms, to obtain a holistic, multi-perspective view of processes. The authors showcase their tool on a medical event log.
- In *"Reconstructing invisible deviating events: A conformance checking approach for recurring events"*, Grüger et al. [16] tackle the problem of reoccurring activities in a process—connected to 1-loops in a correspondent process model—and the issues that such activities cause in the context of conformance checking with temporal rules. The authors process the event log with the insertion of invisible deviating events, additional activities which, instead of being part of the original process, mark the occurrences of missing behavior.
- In *"Analysis of single-cell RNA-sequencing data identifies a hypoxic tumor subpopulation associated with poor prognosis in triple-negative breast cancer"*, Shi et al. [17] perform a study of RNA data obtained through single-cell sequencing techniques. Using single-sample Gene Set Enrichment Analysis (ssGSEA) and cell-cell communication analysis, the authors identify two genetic markers (*ARTN* and *L1CAM*) the expression of which, verified against a risk-score model, indicates a dismal prognosis in patients affected by triple-negative breast cancer (TNBC).

4. Conclusion

The papers contributing to this special issue highlight the high relevance of the application of data mining and process mining to clinical and medical settings. We hope for this promising branch of science to live up to its clear potential and to flourish, bringing on further scientific achievements.

We would like to express our sincere gratitude to all the authors that contributed to this issue, for their effort in writing and revising the papers; to all reviewers, for providing feedback and valuable suggestions; and lastly, to the editorial team of the *AIMS Mathematical Biosciences and Engineering* journal, for their assistance and support in making this special issue possible.

Acknowledgments

This special issue is an initiative of the *Process-Oriented Data Science for Healthcare* Alliance, which is affiliated with the *IEEE Task Force on Process Mining*[†].

We thank the Alexander von Humboldt (AvH) Stiftung for supporting our research.

[†] https://www.tf-pm.org/

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. W.M.P. van der Aalst, Process mining: Data science in action, Second Edition, Springer International Publishing, (2016). https://doi.org/10.1007/978-3-662-49851-4
2. E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: A literature review, *J. Biomed. Inform.*, **61** (2016), 224–236. https://doi.org/10.1016/j.jbi.2016.04.007
3. M. Ghasemi, D. Amyot, Process mining in healthcare: A systematised literature review, *Int. J. Electron. Healthcare*, **9** (2016), 60–88. https://doi.org/10.1504/IJEH.2016.078745
4. T.G. Erdogan, A.K. Tarhan, Systematic mapping of process mining studies in healthcare, *IEEE Access*, **6** (2018), 24543–24567. https://doi.org/10.1109/ACCESS.2018.2831244
5. C. Fernandez-Llatas, Interactive process mining in healthcare, Springer International Publishing, (2020). https://doi.org/10.1007/978-3-030-53993-1
6. A. Guzzo, A. Rullo, E. Vocaturo, Process mining applications in the healthcare domain: A comprehensive review, *WIREs Data Mining Knowledge Discovery*, **12** (2022), e1442. https://doi.org/10.1002/widm.1442
7. E. De Roock, N. Martin, Process mining in healthcare - An updated perspective on the state of the art, *J. Biomed. Inform.*, **127** (2022), 103995. https://doi.org/10.1016/j.jbi.2022.103995
8. W. M. P. van der Aalst, A. Adriansyah, A. K. A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, et al., Process mining manifesto, in *Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I* (eds. F. Daniel, K. Barkaoui, S. Dustdar), **9** (2012), 169–194. https://doi.org/10.1007/978-3-642-28108-2_19
9. R. S. Mans, W. M. P. van der Aalst, R. J. B. Vanwersch, Process mining in healthcare: Evaluating and exploiting operational healthcare processes, Springer International Publishing, (2015). https://doi.org/10.1007/978-3-319-16071-9
10. F. Daniel, Q. Z. Sheng, H. Motahari, Business process management workshops - BPM 2018 international workshops, Sydney, NSW, Australia, September 9-14, 2018, revised papers, *Lecture Notes in Business Information Processing*, **342** (2019). https://doi.org/10.1007/978-3-030-11641-5
11. J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. Johnson, M. Sepúlveda, Innovative informatics methods for process mining in health care, *J. Biomed. Inform.*, **134** (2022), 104203. https://doi.org/10.1016/j.jbi.2022.104203

12. J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, et al., Process mining for healthcare: Characteristics and challenges, *J. Biomed. Inform.*, **127** (2022), 103994. https://doi.org/10.1016/j.jbi.2022.103994
13. N. Martin, N. Wittig, J. Munoz-Gama, Using Process Mining in Healthcare, in *Process Mining Handbook* (eds. W. M. P. van der Aalst, J. Carmona), Lecture Notes in Business Information Processing, **448** (2022), 416–444, Springer. https://doi.org/10.1007/978-3-031-08848-3_14
14. M. Jiang, B. Zhou, L. Chen, Identification of drug side effects with a path-based method, *Math. Biosci. Eng.*, **19** (2022), 5754–5771. https://doi.org/10.3934/mbe.2022269
15. M. N. Tiftik, T. G. Erdogan, A. K. Tarhan, A framework for multi-perspective process mining into a BPMN process model, *Math. Biosci. Eng.*, **19** (2022), 11800–11820. https://doi.org/10.3934/mbe.2022550
16. J. Grüger, M. Kuhn, R. Bergmann, Reconstructing invisible deviating events: A conformance checking approach for recurring events, *Math. Biosci. Eng.*, **19** (2022), 11782–11799. https://doi.org/10.3934/mbe.2022549
17. Y. Shi, X. Huang, Z. Du, J. Tan, Analysis of single-cell RNA-sequencing data identifies a hypoxic tumor subpopulation associated with poor prognosis in triple-negative breast cancer, *Math. Biosci. Eng.*, **19** (2022), 5793–5812. https://doi.org/10.3934/mbe.2022271

AIMS Press