Process Mining for Healthcare: The Full Checkup

Artem Polyvyanyy[1]

[1]School of Computing and Information Systems,
The University of Melbourne, Parkville, VIC 3010, Australia.
artem.polyvyanyy@unimelb.edu.au

Sunday 8th December, 2024

Abstract

This chapter promotes the adoption of process mining methods, techniques, and tools for analyzing, monitoring, and forecasting healthcare processes and their outcomes. By taking the process-oriented view of the data stored and managed by information systems of healthcare organizations, process mining can support self-regulation and evidence-based practices of health professionals, teams, and institutions. The chapter reviews healthcare processes and data amenable to process mining analysis, discusses the prominent use cases and challenges of process mining in healthcare, and provides recommendations for the evolution of the process mining discipline to address the needs of the healthcare domain.

Keywords: Healthcare, process mining, data science, process science

1 Introduction

Healthcare aims to improve the health of individuals and communities and to reduce the burden of diseases and disabilities (Buchbinder, Shanks, & Kite, 2019). To achieve this goal, healthcare institutions engage in *self-regulation* and *evidence-based* practices. Self-regulation involves monitoring, evaluating, and reflecting on one's behaviors and processes to systematically identify and address biases and limitations, thus improving the outcomes (Knapp, Gottlieb, & Handelsman, 2017). Evidence-based practices support clinical decisions based on available evidence rather than intuition or beliefs (Li, Jeffs, Barwick, & Stevens, 2018). Such practices support continuous learning, allowing healthcare professionals to adapt rapidly to changing health demands and new technologies and treatment methods.

Effective implementation of self-regulation and evidence-based practices in healthcare relies on analyzing relevant data from historical observations of the activities performed by stakeholders and events triggered by health equipment and technology. A *health information system* is a data storage and tool for analysis of the data of a healthcare organization, for example, a hospital or a network of healthcare institutions (Winter et al., 2010). The data from such a system can be analyzed using conventional statistics, causal inference, machine learning and data mining techniques (Consoli, Recupero, & Petković, 2019), and more recently process mining (Mans, van der Aalst, & Vanwersch, 2015; Rojas, Munoz-Gama, Sepúlveda, & Capurro, 2016).

The research discipline of *process mining* studies methods, techniques, and tools that use data recorded by information systems that support operational processes of organizations aiming to understand and analyze past and current and to forecast and improve future real-world processes and their outcomes (van der Aalst, 2016). Insights produced by process mining approaches have delivered value in many domains, including retail, manufacturing, telecom, and healthcare (Reinkemeyer, 2020). Specific to healthcare, by taking the *process-oriented* view of the historical healthcare data managed by health information systems, process mining enables several dedicated use cases for monitoring, evaluating, and improving healthcare processes, thus supporting self-regulation and evidence-based practices (Rovani, Maggi, de Leoni, & van der Aalst, 2015).

Based on existing literature, this chapter provides a comprehensive review of healthcare processes and data types amenable to process mining analysis and use cases and exemplar applications of process mining analysis over healthcare data. In addition, it compiles the challenges and limitations of state-of-the-art process mining techniques in the healthcare context and provides recommendations for the process mining discipline aiming to address them.

The next section presents the process mining discipline. Section 3 discusses processes supported and data managed by health information systems that can be the focus of process mining studies. Then, Section 4 presents process mining use cases in healthcare and demonstrates exemplary process mining analytics. Section 5 discusses the challenges and limitations of state-of-the-art process mining for healthcare and gives recommendations for addressing them. Finally, Section 6 concludes the chapter.

2 Process Mining

Process mining combines studies of inferences from data in data mining and machine learning with process modeling and analysis. Process mining tackles, among other problems, the discovery, monitoring, and improvement of real-world processes. The context of three main process mining types, discovery, conformance, and enhancement, as proposed by van der Aalst (2016), is shown in Fig. 1. Process mining aims to understand how a real-world process has performed historically, is performing now, and will perform in the future, and to recommend how the future process can be improved.

A process mining study starts with an event log, a special format of data recorded by software systems, e.g., enterprise resource planning or customer relation management systems, that support organizations' business processes. An *event log* is a collection of traces, where each *trace* is a sequence of events that relate to the same *case*, e.g., an identifier of a patient admitted to a hospital, ordered by *timestamps* at which software systems recorded them. An *event* usually represents a transaction or a message and, in addition to the case and timestamp attributes, has the *activity* attribute that captures the nature of the process step, e.g., a transaction storing customer details or a message requesting additional information from a business partner. The three standard event at-

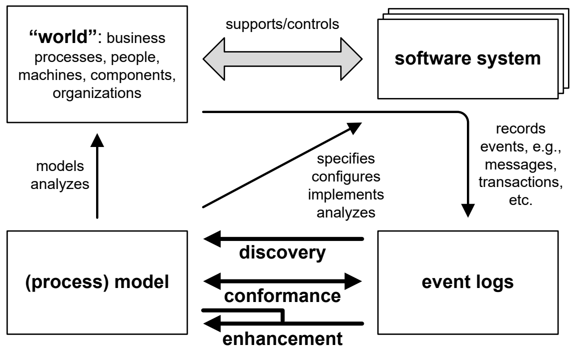


Figure 1: The context of process mining (van der Aalst, 2016).

Table 1: Ten events of a trace that capture a pathway of a sepsis patient

No.	Timestamp	Activity	Group	Age	CRP	L.Ac.	Leuk.
1	7 Oct 2014 07:24:16	ER Registration	A	70			
2	7 Oct 2014 07:28:59	ER Triage	C				
3	7 Oct 2014 07:29:18	ER Sepsis Triage	A				
4	7 Oct 2014 07:30:33	IV Liquid	A				
5	7 Oct 2014 07:30:36	IV Antibiotics	A				
6	7 Oct 2014 07:49:00	CRP	B		205.0		
7	7 Oct 2014 07:49:00	LacticAcid	B			3.4	
8	7 Oct 2014 07:49:00	Leukocytes	B				11.1
9	7 Oct 2014 09:56:09	Admission NC	Q				
10	9 Oct 2014 13:00:00	Release B	E				

tributes are often extended by further payload attributes, like resources used or involved in the execution of the activity that triggered the event.

Table 1 lists ten events constituting a single trace that describes a pathway of a sepsis patient through a Dutch hospital (Mannhardt & Blinde, 2017). Columns in the table encode event attributes, while rows specify the events. As all the events belong to the same case, the case attributed is omitted in the table. In addition to the standard attributes, events have payload attributes of an organization group in which the event originated, values of the patient's age and diagnostic measurements of C-reactive protein (CRP), and the levels of lactic acid and leukocytes in the blood.

Given an event log of multiple historical traces and possibly a process model, e.g., records of historical pathways of all sepsis patients in a hospital from a certain period and a normative description of which steps such patients should follow, the three main types of process mining address these problems:

- Given an event log, the *discovery* problem consists of constructing a process model that describes all the traces the process that generated the log can produce.
- Given an event log and a process model, the *conformance* problem studies how faithfully the model describes the process that generated the log.
- Given an event log and a process model, the *enhancement* problem studies ways to improve the accuracy of the model with respect to the process that generated the log.

A *process model* is a conceptual model that describes sequences of activities the system can generate. In process mining, several notations are used for capturing process models, but most commonly, Petri nets, business process modeling and notation, and event-driven process chains. Commercial process mining tools predominantly use directly-follows graphs (DFGs) for representing process models. Process models that follow the declarative modeling paradigm are also used in process mining and the healthcare applications of process mining (Rovani et al., 2015).

3 Healthcare Processes and Data

This section reviews common types of processes present in healthcare settings, refer to Section 3.1, presents different data types these processes generate and software systems that manage these data types, see Section 3.2, and, in Section 3.3, discusses considerations relevant to preparing healthcare data for process mining analysis.

3.1 Processes

Healthcare processes are broadly classified into *medical treatment processes* (MTP) that comprise treatment steps of patients and *organizational processes* (OP) that coordinate health practitioners and resources (Lenz & Reichert, 2007). Guzzo, Rullo, and Vocaturo (2021) specialize these core process types into seven subtypes listed below.

P1 Clinical pathways (MTP) capture interactions of patients with healthcare providers, for instance, to communicate symptoms and perform medical exams or surgeries.

P2 Patient pathways (MTP&OP) constitute steps patients follow in healthcare facilities, such as registration, diagnosis, and hospitalization.

P3 Patient behavior (MTP) is composed of activities performed over a patient's body, for example, temperature and blood pressure measurements.

P4 Personnel interactions (OP) comprise messages and information exchanges between hospital staff and departments.

P5 Medical processes (MTP) specify activities of medical procedures, for instance, therapies and surgeries.

P6 Human movements (OP) are sequences of patients' or medical staff's locations in healthcare facilities.

P7 Personnel tasks (OP) are composed of activities performed by medical staff, including interactions with information systems.

Patients, doctors, and practitioners can be further involved in the processes listed below (Lenz & Reichert, 2007; Mans et al., 2015), which we also attribute to the medical treatment and organizational process types.

P8 Administrative processes (OP) handle activities related to admission, discharge, scheduling, and billing of patients.

P9 Clinical trials (MTP&OP) are studies that involve human participants that aim to advance medical knowledge and improve patient care.

P10 Decision-making processes (MTP) govern the principles for implementing complex medical decisions and ethical dilemmas.

P11 Education processes (MTP&OP) aim to ensure medical staff meet existing qualifications and educate patients to engage in their healthcare management.

P12 Facility management processes (OP) address the management of infrastructure, facility, and equipment maintenance.

P13 Financial processes (OP) regulate budget planning and financial management.

P14 Information management processes (OP) manage data entry, data privacy, and data security within healthcare organizations.

P15 Medication management processes (MTP&OP) govern activities of prescribing, administering, and managing medications and pharmaceuticals.

P16 Patient flow processes (MTP&OP) aim to improve the flow of patients through healthcare departments and settings.

P17 Regulatory compliance processes (OP) ensure the maintenance of legal and regulatory compliance of various activities within healthcare facilities.

P18 Supply processes (OP) support the procurement of equipment and pharmaceuticals for implementing effective and efficient health care.

All the discussed process types can be the focus of process mining studies to improve the efficiency and effectiveness of organizational processes and clinical outcomes of medical treatment processes to promote job satisfaction of healthcare practitioners and the health of the population.

3.2 Data

The data generated by the healthcare processes is managed by dedicated software systems and is often stored in database management systems. The most common classes of software systems used in healthcare environments, including those discussed by Guzzo et al. (2021), are summarized below.

D1 Hospital management system (HMS) supports clinical (P1) and patient (P2) pathways, personnel interactions (P4) and tasks (P7), and administrative (P8) and financial (P13) processes within a healthcare institution. An HMS can include functionality for managing medications (P15) and inventory (P18), provides clinical decision support (P10), and supports reporting and adherence to regulatory compliance (P17). HMSs are actively involved in the management of medical processes (P5), including scheduling of medical procedures and allocation of resources to and documentation of these procedures. Finally, HMSs may include modules for supporting education processes for both medical staff and patients (P11). They may include features for scheduling training and management of medical licenses and qualifications.

D2 Electronic medical record (EMR) is a digital version of a medical history and information of a patient and, thus, stores data managed by clinical (P1) and patient (P2) pathways, medication management processes (P15), clinical decision management processes (P10), and clinical research and trials (P9). EMRs are also involved in the management of medical processes (P5). For example, they are used to prioritize medical procedures for patients and store their results.

D3 Electronic health record (EHR) system extends the functionality of EMR by managing information on healthcare practitioners involved in the diagnosis and treatment of patients (P4, P7, and P10) and medical billing and insurance management (P13), and staff involved in the patient's care. It can also share information with other healthcare facilities.

D4 Real Time Location System (RTLS) identifies and tracks the real-time location of people, resources, and further assets to monitor the location and safety of patients (P2, P3, and P6), patient flow (P16), staff movements (P6), interactions (P4), tasks (P7), and equipment utilization (P12).

D5 Picture Archiving and Communication System (PACS) manages access to medical images, for instance, X-rays, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans, and, thus, supports decision-making processes (P10), personnel tasks (P7), clinical trials (P9), and information management (P14).

D6 Body-sensor data, for example, data generated by electrocardiogram, temperature, and respiratory rate sensors, that describes physiological parameters and health metrics of patients supports patient pathways (P2) and behavior processes (P3), clinical trials (P9), and facilitates decision making (P10).

The data managed by all the discussed systems can be used for process mining analysis of the corresponding processes.

3.3 Considerations

Several standards have been proposed to systematize how event logs are stored and exchanged. The existing commercial and open-source tools that implement process mining techniques often rely on these standards to ensure data compatibility.

The first version of the IEEE eXtensible Event Stream (XES) standard was initiated in 2010 and officially published on November 11, 2016 (XES Standard, 2016). This standard includes an XML schema that captures the structure of an XES event log or an event stream and an XML schema describing the structure of a log or stream extension.

The process mining community regularly publishes industrial event logs in the XES format, including logs from the healthcare domain. A Dutch Academic Hospital event log contains 150 291 events distributed over 1 143 traces, each capturing one case of diagnosing and treating a patient of a gynecology department.[1] Another publicly available event log contains 15 214 events from 1 050 cases, each representing a pathway of a sepsis patient through a hospital (Mannhardt & Blinde, 2017).[2] The latter event log is also presented and discussed in Section 2.

Recently, the process mining community introduced an XES successor, the Object-Centric Event Log (OCEL) standard (Ghahfarokhi, Park, Berti, & van der Aalst, 2021). In object-centric event data, the notion of a case of an event is replaced with a group of

[1] https://doi.org/10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54
[2] https://doi.org/10.4121/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460

objects the event is associated with (van der Aalst, 2023). This shift in process scoping appears particularly interesting for the healthcare domain, as it supports flexible analysis of event data based on different equipment and healthcare team configurations (Senderovich et al., 2015). While most existing process mining tools work with XES logs, it is expected that, with time, more tools will adopt OCEL as the underlying standard for exchanging event data.

4 Use Cases and Exemplars

Process mining techniques support many use cases and applications in healthcare (Mans et al., 2015). Most frequently, process mining is used to perform process discovery and conformance analysis. Further applications include concept drift detection, predictive analysis, and simulation (Guzzo et al., 2021). Next, we discuss and exemplify the use of process discovery, refer to Section 4.1, and conformance analysis, see Section 4.2, based on the example event log of sepsis patients pathways.

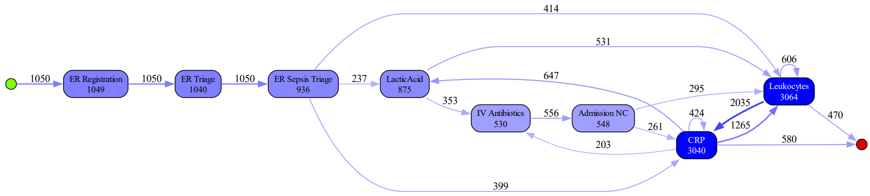
4.1 Discovery

Figures 2a and 2b show two directly-follows graphs (DFGs) discovered from the event log of 1 050 pathways of sepsis patients introduced in Section 2 using the Directly Follows visual Miner process discovery algorithm (Leemans, Poppe, & Wynn, 2019). A DFG is a special type of *process model*. A directed path in a DFG that starts in its source node, the only node without incoming arcs, and ends in the sink node, the only node without outgoing arcs, describes a possible sequence of activities that can be executed within a single process case. A DFG discovered from a given event log aims to describe the process cases that can be supported by the process, including the historical process executions recorded in the log the DFG was constructed from and the executions that can be observed in the future executions of the process.

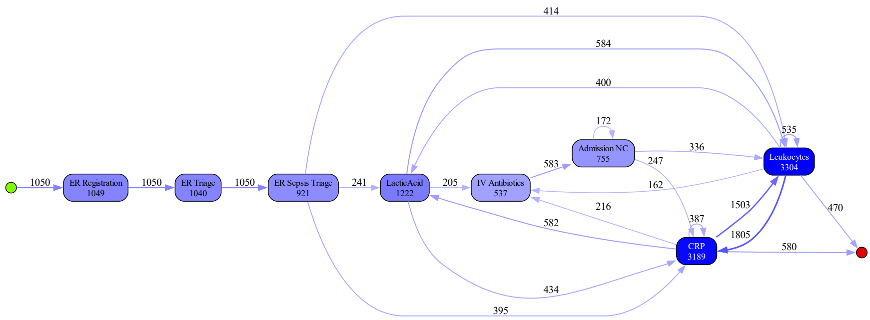
The graph in Fig. 2a was constructed to ensure it describes precisely at least 10% of the traces in the input event log. In turn, the DFG in Fig. 2b was constructed to guarantee to represent precisely at least 25% of the process executions recorded in the event log. The numbers on nodes of DFGs represent the historical frequencies of executing the corresponding activities, while the numbers on arcs describe the historical frequencies of executing the connected activities one after the other. In the figures, the source nodes and sink nodes use green and red backgrounds, respectively, while all other nodes and arcs are blue. The thickness of the arcs and the darkness of the nodes emphasize the corresponding frequencies.

The more of the distinct executions from the event log the discovered model represents precisely, the more complex it gets. It is often the task of the process analyst to identify a range of settings of the process discovery algorithm that result in the construction of simple models that faithfully describe the process that generated the log.

Process models discovered from historical executions of different types of processes discussed in Section 3.1 can be used to understand how they are performed in the real world based on the evidence recorded in the event logs. For example, based on the models in Fig. 2, one can conclude that the CRP and white blood cell (leukocytes)



(a)



(b)

Figure 2: Two directly-follows graphs discovered from historical pathways of sepsis patients at a hospital.

levels are monitored regularly at various stages of handling sepsis patients and always precede the discharge of a patient (the end of the process). Domain experts can use the process knowledge captured in the models to identify bottlenecks in the process and possibilities for improving it. Process discovery was reported as a key component in many case studies in healthcare, including the reduction of surgery times and risks of individual surgical manipulations (Blum, Padoy, Feußner, & Navab, 2008), supporting operations of active transhumeral prostheses (Su et al., 2023), and monitoring care processes in an oncology department (Caron et al., 2014). For a comprehensive review of process discovery applications in healthcare, refer to the work by Guzzo et al. (2021).

4.2 Conformance

Conformance techniques can be used to study deviations of historical process executions and the designed executions captured in models. Mannhardt and Blinde (2017) apply conformance techniques to demonstrate that the historical pathways of sepsis patients often deviate from the patient flow as specified by the process stakeholders.

To perform such analysis, one can compute *optimal alignments* (Adriansyah, van Dongen, & van der Aalst, 2011) between the evidence, historical process executions recorded in an event log, and the expectations, process executions supported by a designed model of the corresponding process. Figure 3 shows optimal alignments between two historical pathways of sepsis patients and the DFG from Fig. 2b. The alignment in Fig. 3a places the trace from Table 1 into the context of an execution described in the DFG. Each chevron arrow represents a move in the alignment, either an event

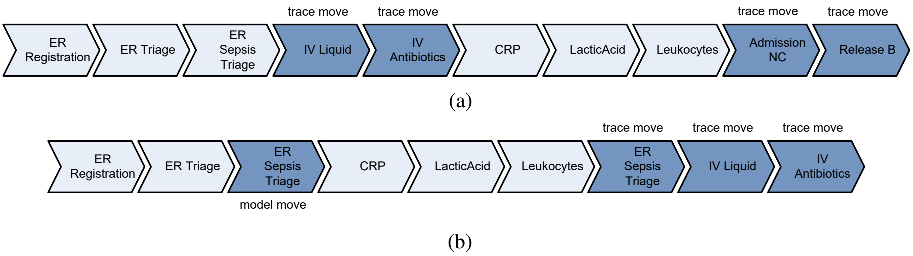


Figure 3: Optimal alignments between two historical pathways of sepsis patients and the DFG in Fig. 2b.

from the trace in Table 1 or an activity from the DFG. A chevron arrow with a light background represents a synchronous move, an event from the trace matched by an activity in the model. A chevron arrow with a dark background represents an asynchronous move, that is, an event that is not matched by an activity or an activity that is not matched by an event. The alignment in Fig. 3a has four asynchronous moves. All these are trace moves, denoted by the corresponding labels above the chevron arrows. A trace move signifies a trace event not matched by a model activity. The alignment in Fig. 3b has four asynchronous moves, three trace moves and one model move, denoted by the corresponding label below the chevron arrow. A model move signifies that a model activity is not matched by a trace event.

The optimal alignments in Fig. 3 are computed to contain minimal deviations between the trace and the model. In general, several optimal alignments between a trace and a model can exist with the same number, or cost, of asynchronous moves. Optimal alignments can be used to construct analytics that help to understand the differences between processes as they are executed in the real world and designed to be executed. These insights can be used, for instance, to enforce future real-world process executions to follow the guidelines prescribed in the model or to redesign the process model to incorporate best practices from historical real-world process executions (Fahland & van der Aalst, 2015; Polyvyanyy, van der Aalst, ter Hofstede, & Wynn, 2017).

Conformance analysis is used in healthcare, for instance, to assess deviations of care pathways from clinical guidelines (Caron et al., 2014; Kelleher, Jagadeesh Chandra Bose, Waterhouse, Carter, & Burd, 2014; Mannhardt & Blinde, 2017), conformance of medical treatment processes (Kirchner, Herzberg, Rogge-Solti, & Weske, 2013), and analysis of patients with similar treatment procedures (Dewandono, Fauzan, Sarno, & Sidiq, 2013). Again, for a comprehensive review of conformance applications in healthcare, refer to the work by Guzzo et al. (2021).

5 Discussion

This section discusses the challenges and limitations of process mining, see Section 5.1 and Section 5.2, respectively, and provides recommendations for advancing process mining in the healthcare domain, refer to Section 5.3.

5.1 Challenges

The general challenges of process mining, for example, dealing with complex event logs and creating representative benchmarks, were discussed by van der Aalst et al. (2012), whereas challenges specific to the healthcare domain were introduced by Martin et al. (2020) and Munoz-Gama et al. (2022). Next, we introduce challenges specific to process mining for healthcare that reuse some but also go significantly beyond the challenges discussed in the mentioned works. They were identified in the works on the use of process mining and the broader process analytics studies in the context of the healthcare domain. Addressing these challenges requires a multidisciplinary approach, including experts in process mining, healthcare, legal, and ethics.

C1 Event data quality. Healthcare data is often incomplete or inconsistent due to errors in handwritten or electronic health records, leading to challenges in obtaining accurate analysis results.

C2 Event data privacy and security. Healthcare data is highly sensitive and is subject to strict privacy and security regulations in many countries, posing significant challenges in sharing and analyzing healthcare data.

C3 Event data scarcity. Healthcare data can be imbalanced as specific outcomes of rare health conditions or diseases may manifest in very few positive or negative outcomes, leading to additional challenges for distinguishing such conditions from noise and drawing statistically significant conclusions based on them.

C4 Clinical data heterogeneity. Clinical data is highly heterogeneous due to patient demographics and genetic differences, making it challenging to identify common trends and underlying principles underpinning the data.

C5 Event data heterogeneity. Healthcare data comes in multiple types, including categorical, numerical, ordinal, geospatial, time series, and image data types, each with dedicated characteristics making it difficult to implement its homogeneous analysis.

C6 Longitudinal event data. Collected by monitoring the health of patients over extended time, healthcare data is often longitudinal, requiring special approaches for handling irregular measurement intervals, missing values, and attrition bias.

C7 Event data integration. Healthcare data is stored in different formats managed by multiple software systems, refer to Section 3.2 for details, and integrating these fragmented data sources poses significant practical challenges.

C8 Event data management compliance. Due to its high sensitivity, handling healthcare data is often governed by complex legal frameworks requiring non-trivial compliance with data handling protocols.

C9 Ethical considerations. Healthcare data is subject to ethical considerations to avoid unfair treatment of patient cohorts due to biases of analysis algorithms.

C10 Explainability considerations. To ensure the wide adoption of healthcare data analysis techniques, their results must be explainable and, thus, trustworthy to healthcare practitioners and patients. However, it is well known that the higher the quality of automated data analysis results, the harder it is to ensure the underlying principles can be communicated to humans.

C11 Domain knowledge. Healthcare data interpretation often requires specialized domain knowledge by dedicated healthcare practitioners, and it is challenging to access them due to their scarcity and busy work schedules.

C12 Patient-centered analysis. Healthcare data often needs to be analyzed to obtain insights about the health and needs of a concrete patient, making it challenging to generalize the analysis principles across multiple patients.

C13 Multi-perspective analysis. Healthcare data must often be analyzed from multiple perspectives to understand the important data patterns and obtain comprehensive insights, including process control-flow, organizational view, and data decisions.

C14 Analysis scope. Given the plethora of healthcare data sources and types available in modern healthcare facility settings, it is challenging to define the scope of data that will maximize the effectiveness of the envisioned analysis.

C15 Analysis timeliness. Given the importance of in-action reflective practices in healthcare, it is often essential to provide timely data analysis insights, which imposes requirements on the runtime efficiency of the analysis techniques.

C16 Analysis urgency. Healthcare data may require urgent analysis, for instance, in situations of sudden health conditions requiring urgent attention from doctors or specialists, such as in the case of emergency room treatment (Rojas et al., 2019). Providing accurate insights for informed decision-making under time pressure is challenging.

C17 Innovation drifts. The healthcare environment is knowledge-intensive, with new treatment methods and innovative approaches to diagnostics appearing regularly, leading to concept drifts in healthcare processes and, consequently, in the corresponding healthcare data footprints. It is thus challenging to identify and account for such drifts during healthcare data analysis (Bose, van der Aalst, Zliobaite, & Pechenizkiy, 2014; Yeshchenko, Ciccio, Mendling, & Polyvyanyy, 2022).

C18 DIY (Do-It-Yourself) analysis. To cover a wide range of use cases, including urgent and ad-hoc situations when the analysis results can be of interest, healthcare practitioners should be able to perform the analysis with little assistance. This requires tailoring the analysis techniques to this group of end users and a widespread process mining literacy of healthcare staff, which is not trivial to achieve.

5.2 Limitations

At the time of writing this chapter, several important limitations still prevent process mining techniques from being widely used in healthcare. These are summarized below.

L1 Methodologies. Several methodologies have been developed to support standard process mining analysis use cases. However, the choice of process mining methodologies that target analysis of healthcare environments is somewhat limited (Fernández-Llatas, Lizondo, Sanchez, Benedí, & Traver, 2015; Rebuge & Ferreira, 2012; Vathy-Fogarassy, Vassányi, & Kósa, 2022), and no widely accepted tailored methodologies that address the particularities of event data analysis in healthcare are known todate. The best practices for performing process mining studies to support various use cases in healthcare are still to be understood and systematized.

L2 Skills. Today, healthcare organizations lack staff with expertise in process mining. Due to the absence of dedicated process mining methodologies for healthcare, broad expertise beyond process mining are often required to perform various data analysis studies before valuable results can be obtained. Consequently, in the healthcare context, it is currently not cost-effective to specialize only in process mining, which limits the availability of experts in this field.

L3 Practical adoption. The absence of dedicated methodologies and the skill gap prevent a comprehensive implementation and use of process mining techniques in healthcare. Despite being not rare, process mining projects in healthcare are not performed systematically.

L4 Categorical data. The majority of existing process mining techniques work with categorical data, for example, records of performed activities and resources involved in the execution of the activities, and do not analyze the processes as the evolutions of characteristics of patients measured as, for instance, numerical or ordinal data. For example, consider the process of the evolution of the CRP level of a patient reported in Table 1 over time. Many insights in clinical settings relate to the series of numerical measurements obtained during the diagnosis and monitoring of patients.

We envision these limitations will be addressed and overcome in the coming years.

5.3 Recommendations

Based on the analysis of processes and data types, see Section 3.1 and Section 3.2, this section provides recommendations for the development of process mining discipline to address the pertinent challenges of process mining in the healthcare setting reviewed in Section 5.1. The recommendations are summarized schematically in Fig. 4.

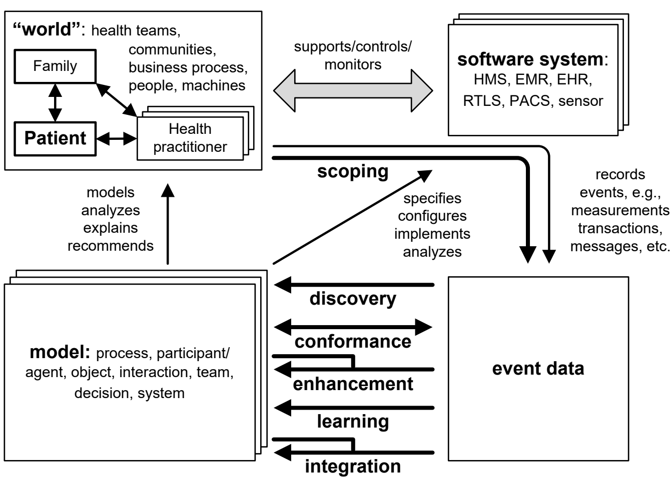


Figure 4: The recommended context of process mining for healthcare.

The variety of the data types in healthcare (C4, C5, and C7) calls for techniques to process categorical, discrete and continuous numerical, ordinal, time series, geospatial, and sensor data, and expand the analysis from structured to semi-structured and unstructured data captured in text and images. In Fig. 4, we refer to such data as *event data* and envision that event data is provided as a collection of triples, each linking an event, its attribute, and the value of this attribute. For example, the triplets (1,Timestamp,7 Oct 2014 07:24:16), (1,Activity,ER Registration), (1,Group,A), and (1,Age,70) characterize event number 1 from Table 1.

The event data in healthcare is collected and produced by specialized software systems, which include systems like HMS (D1), EMR (D2), EHR (D3), RTLS (D4), PACS (D5), and body sensors (D6), which can collectively be referred to as health information systems. These systems monitor, support, and help to control special scopes of interest (C14) that relate to patients and their families, health practitioners and health teams, people and (autonomous) machines, and broader communities, with the focus on the health of each particular individual (C12).

To address the complexity of the healthcare environments and the need for their multi-perspective analysis (C13), the types of models addressed by process mining techniques must be expanded significantly. In addition to conventional process models studied by process mining techniques, models of participants or agents (Tour, Polyvyanyy, & Kalenkova, 2021) involved in healthcare episodes, objects involved in healthcare manipulations (Berti & van der Aalst, 2023), interactions between patients and doctors (Tour, Polyvyanyy, Kalenkova, & Senderovich, 2023), and clinical decisions can be studied to provide a nuanced analysis. For example, the analysis of individual healthcare professionals and their interactions studied in agent system mining (Tour et al., 2021) can support understanding the dynamics and effectiveness of temporal team formations, common in healthcare, providing input for both in-action and on-action reflections and subsequent evidenced-based learning.

Due to the multiple perspectives (C13) and aspects (C14) related to data analysis, to avoid data overload, relevant event data must be identified. We refer to this problem as scoping, refer to Fig. 4. An effective scoping of event data can improve the performance of process mining (Han, Kamber, & Pei, 2011; Tour et al., 2021), including efficiency (C15 and C16), interpretability (C10), and handling of heterogeneous and imbalanced data (C3, C4, C5, and C6).

As multiple models of different types are discovered and analyzed, there is a need for their integration (C7), also based on the evidence present in event data. For example, agent and interaction models can be integrated into agent system models (Tour et al., 2021), providing a comprehensive description of the healthcare environment, including agents like patients, family members, doctors, nurses, healthcare equipment, and information flow between them via interactions. Consequently, we see *integration* as an important process mining problem.

Like machine learning, process mining can explore ways to use event data to learn patterns and discover models for predicting future events and supporting decision-making. We embrace these envisioned capabilities under the class of learning techniques and extract a dedicated problem of *learning* from event data, solutions to which should aim to construct such prediction and decision models.

The six problems in the context of process mining for healthcare are listed below:

- Given software systems, event data they record, and the aims of a process mining project, the *scoping* problem studies how to identify data and its sources that are relevant to the purpose of the project.
- Given event data, the *discovery* problem consists of constructing models that describe the phenomena, for example, processes and agents, that generated the data.
- Given event data and models, the *conformance* problem studies how faithfully the models describe the phenomena that generated the data.
- Given event data and models, the *enhancement* problem studies ways to improve the accuracy of the models with respect to the phenomena that generated the data.
- Given event data that describes examples of inputs and outputs of executed historical instances of a task, the *learning* problem consists of constructing models for predicting and making decisions when solving the future instances of the task.
- Given event data and models, the *integration* problem studies how to integrate the models to obtain an accurate multi-perspective description of the phenomenon that generated the data.

6 Conclusions

This chapter reviews the use of process mining techniques in healthcare settings. Specifically, it discusses healthcare processes that can be the focus of process mining studies aiming to understand the historical and current procedures and operations in healthcare organizations and improve how they are carried out in the future. It also discusses data readily stored and managed by health information systems that can inform process mining analysis. Finally, the chapter reviews the existing challenges and limitations associated with the use of state-of-the-art process mining techniques in the healthcare context and recommends studying healthcare processes as interactions of the stakeholders, including patients, their family members and communities, and healthcare practitioners, addressing perspectives that go beyond the behavior of process participants, using various data sources and non-standard data types.

References

Adriansyah, A., van Dongen, B. F., & van der Aalst, W. M. P. (2011). Conformance checking using cost-based fitness analysis. In *IEEE international enterprise distributed object computing conference EDOC* (pp. 55–64). IEEE Computer Society. doi: 10.1109/EDOC.2011.12

Berti, A., & van der Aalst, W. M. P. (2023). OC-PM: analyzing object-centric event logs and process models. *International Journal on Software Tools for Technology Transfer*, *25*(1), 1–17. doi: 10.1007/s10009-022-00668-w

Blum, T., Padoy, N., Feußner, H., & Navab, N. (2008). Workflow mining for visualization and analysis of surgeries. *International Journal of Computer Assisted Radiology and Surgery*, *3*(5), 379–386. doi: 10.1007/s11548-008-0239-0

Bose, R. P. J. C., van der Aalst, W. M. P., Zliobaite, I., & Pechenizkiy, M. (2014). Dealing with concept drifts in process mining. *IEEE Trans. Neural Networks Learn. Syst.*, *25*(1), 154–171. doi: 10.1109/TNNLS.2013.2278313

Buchbinder, S. B., Shanks, N. H., & Kite, B. (2019). *Introduction to health care management*. Jones & Bartlett Learning, LLC.

Caron, F., Vanthienen, J., Vanhaecht, K., Van, E. L., Weerdt, J. D., & Baesens, B. (2014). Monitoring care processes in the gynecologic oncology department. *Computers in Biology and Medicine*, *44*, 88–96. doi: 10.1016/j.compbiomed.2013.10.015

Consoli, S., Recupero, D. R., & Petković, M. (Eds.). (2019). *Data science for healthcare*. Springer International Publishing.

Dewandono, R. D., Fauzan, R., Sarno, R., & Sidiq, M. (2013). Ontology and process mining for diabetic medical treatment sequencing. In *Icts* (pp. 171–178).

Fahland, D., & van der Aalst, W. M. P. (2015). Model repair — aligning process models to reality. *Information Systems*, *47*, 220–243. doi: 10.1016/j.is.2013.12.007

Fernández-Llatas, C., Lizondo, A., Sanchez, E. M., Benedí, J., & Traver, V. (2015). Process mining methodology for health process tracking using real-time indoor location systems. *Sensors*, *15*(12), 29821–29840. doi: 10.3390/s151229769

Ghahfarokhi, A. F., Park, G., Berti, A., & van der Aalst, W. M. P. (2021). OCEL: A standard for object-centric event logs. In *ADBIS (short papers)* (Vol. 1450, pp. 169–175). Springer.

Guzzo, A., Rullo, A., & Vocaturo, E. (2021). Process mining applications in the healthcare domain: A comprehensive review. *WIREs Data Mining and Knowledge Discovery*, *12*(2). doi: 10.1002/widm.1442

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques, 3rd edition* (3rd ed. ed.). Morgan Kaufmann. Retrieved from `http://hanj.cs.illinois.edu/bk3/`

Kelleher, D. C., Jagadeesh Chandra Bose, R. P., Waterhouse, L. J., Carter, E. A., & Burd, R. S. (2014). Effect of a checklist on advanced trauma life support workflow deviations during trauma resuscitations without pre-arrival notification. *Journal of the American College of Surgeons*, *218*(3), 459–466. doi: 10.1016/j.jamcollsurg.2013.11.021

Kirchner, K., Herzberg, N., Rogge-Solti, A., & Weske, M. (2013). Embedding conformance checking in a process intelligence system in hospital environments. In (pp. 126–139). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-36438-9_9

Knapp, S., Gottlieb, M. C., & Handelsman, M. M. (2017). Enhancing professionalism through self-reflection. *Professional Psychology: Research and Practice*, *48*(3), 167–174.

Leemans, S. J. J., Poppe, E., & Wynn, M. T. (2019). Directly follows-based process mining: Exploration & a case study. In *International conference on process mining (ICPM)* (pp. 25–32). IEEE. doi: 10.1109/ICPM.2019.00015

Lenz, R., & Reichert, M. (2007). IT support for healthcare processes — premises, challenges, perspectives. *Data Knowl. Eng.*, *61*(1), 39–58. doi: 10.1016/j.datak.2006.04.007

Li, S.-A., Jeffs, L., Barwick, M., & Stevens, B. (2018). Organizational contextual features that influence the implementation of evidence-based practices across healthcare settings: A systematic integrative review. *Systematic Reviews*, *7*(1).

Mannhardt, F., & Blinde, D. (2017). Analyzing the trajectories of patients with sepsis using process mining. In *18th international working conference on business process modeling, development and support (bpmds)* (Vol. 1859, pp. 72–80). CEUR-WS.org.

Mans, R., van der Aalst, W. M. P., & Vanwersch, R. J. B. (2015). *Process mining in healthcare—evaluating and exploiting operational healthcare processes*. Springer. doi: 10.1007/978-3-319-16071-9

Martin, N., Weerdt, J. D., Fernández-Llatas, C., Gal, A., Gatta, R., Ibáñez, G., … Van Acker, B. B. (2020). Recommendations for enhancing the usability and understandability of process mining in healthcare. *Artificial Intelligence in Medicine*, *109*.

Munoz-Gama, J., Martin, N., Fernández-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., … Zerbato, F. (2022). Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, *127*.

Polyvyanyy, A., van der Aalst, W. M. P., ter Hofstede, A. H. M., & Wynn, M. T. (2017). Impact-driven process model repair. *ACM Transactions on Software Engineering and Methodology*, *25*(4), 1–60. doi: 10.1145/2980764

Rebuge, Á., & Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, *37*(2), 99–116. doi: 10.1016/j.is.2011.01.003

Reinkemeyer, L. (Ed.). (2020). *Process mining in action*. Springer International Publishing.

Rojas, E., Cifuentes, A., Burattin, A., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2019). Performance analysis of emergency room episodes through process mining. *International Journal of Environmental Research and Public Health*, *16*(7), 1274. doi: 10.3390/ijerph16071274

Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *J. Biomed. Informatics*, *61*, 224–236. doi: 10.1016/j.jbi.2016.04.007

Rovani, M., Maggi, F. M., de Leoni, M., & van der Aalst, W. M. P. (2015). Declarative process mining in healthcare. *Expert Syst. Appl.*, *42*(23), 9236–9251. doi: 10 .1016/j.eswa.2015.07.040

Senderovich, A., Weidlich, M., Gal, A., Mandelbaum, A., Kadish, S., & Bunnell, C. A. (2015). Discovery and validation of queueing networks in scheduled processes. In *Caise* (pp. 417–433). Springer. doi: 10.1007/978-3-319-19069-3_{2}{6}

Su, Z., Yu, T., Lipovetzky, N., Mohammadi, A., Oetomo, D., Polyvyanyy, A., … van Beest, N. (2023). Data-driven goal recognition in transhumeral prostheses using process mining techniques. In *2023 5th international conference on process mining (ICPM)*. IEEE. doi: 10.1109/icpm60904.2023.10271945

Tour, A., Polyvyanyy, A., & Kalenkova, A. A. (2021). Agent system mining: Vision, benefits, and challenges. *IEEE Access*, *9*, 99480–99494. doi: 10.1109/ACCESS .2021.3095464

Tour, A., Polyvyanyy, A., Kalenkova, A. A., & Senderovich, A. (2023). Agent miner: An algorithm for discovering agent systems from event data. In *Business pro-*

cess management - 21st international conference, BPM 2023, utrecht, the netherlands, september 11-15, 2023, proceedings (Vol. 14159, pp. 284–302). Springer Nature Switzerland. doi: 10.1007/978-3-031-41620-0_{1}{7}

van der Aalst, W. M. P. (2016). *Process mining - data science in action*. Springer.

van der Aalst, W. M. P. (2023). Object-centric process mining: Unraveling the fabric of real processes. *Mathematics*, *11*(12), 2691. doi: 10.3390/math11122691

van der Aalst, W. M. P., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., … Wynn, M. (2012). Process mining manifesto. In *Business process management workshops* (pp. 169–194). Springer Berlin Heidelberg.

Vathy-Fogarassy, Á., Vassányi, I., & Kósa, I. (2022). Multi-level process mining methodology for exploring disease-specific care processes. *Journal of Biomedical Informatics*, *125*, 103979. doi: 10.1016/j.jbi.2021.103979

Winter, A., Haux, R., Ammenwerth, E., Brigl, B., Hellrung, N., & Jahn, F. (2010). Introduction. In *Health information systems* (pp. 1–2). Springer London. doi: 10.1007/978-1-84996-441-8_1

XES Standard. (2016). *IEEE standard for eXtensible event stream (XES) for achieving interoperability in event logs and event streams.* doi: 10.1109/IEEESTD.2016.7740858

Yeshchenko, A., Ciccio, C. D., Mendling, J., & Polyvyanyy, A. (2022). Visual drift detection for event sequence data of business processes. *IEEE Trans. Vis. Comput. Graph.*, *28*(8), 3050–3068. doi: 10.1109/TVCG.2021.3050071