Process Mining in Sepsis Treatment Management: A Step-by-Step Approach to Discovery

Eugênio Rocha da Silva Júnior[1*], Uanderson Lima Santos[1], Marcos Batista Figueredo[1]
[1]UNEB, PPGMSB; Alagoinhas, Bahia, Brazil

This article explores the management of sepsis treatment processes, a prominent contributor to hospital morbidity and mortality on a global scale. The research delves into the application of process mining techniques during the discovery phase, utilizing anonymized data from Mannhardt and Blinde (2017). The impetus for this investigation stems from the imperative to pinpoint inefficiencies within sepsis treatment protocols, with the overarching goal of enhancing care quality while curbing expenses. A primary challenge lies in the need for a comprehensive overview of hospital processes for sepsis management. The study employs discovery algorithms such as Alpha Miner, Heuristic Mining, and DFG (Direct Flow Chart) to address this challenge. By leveraging these methodologies, the research endeavors to foster efficient allocation of costs and resources, elevate the standard of patient care, and foster operational efficacy within public health institutions.
Keywords: Sepsis, Process Mining, Hospital Efficiency.

Introduction

Sepsis, characterized by a dysregulated response of the body to infections, remains a prominent contributor to global hospital morbidity and mortality. The intricate nature of sepsis necessitates a meticulous and well-coordinated treatment approach, encompassing timely identification, prompt administration of appropriate therapies, and continuous patient monitoring. Nonetheless, hospitals encounter formidable challenges in effectively managing these care processes, directly impacting the quality of patient care.

This study is motivated by the quest for practical solutions to address the persistent challenge of sepsis within hospital settings. Sepsis poses significant challenges to healthcare professionals, resulting in considerable treatment expenses and, more critically, preventable loss of life. Given the complexity of sepsis treatment processes, a thorough analysis is imperative to pinpoint inefficiencies and bottlenecks that impede hospitals' ability to deliver high-quality care.

Received on 30 October 2023; revised 18 November 2023.
Address for correspondence: Eugênio Rocha da Silva Júnior. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: eugeniojr.dev@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):51-59

The central issue addressed in this study is the need for a comprehensive and detailed overview of hospital processes concerning sepsis treatment. This gap leads to inefficiencies, difficulty identifying bottlenecks, and constraints in implementing continuous improvements. Furthermore, more detailed analyses are needed to improve cost management, efficient resource allocation, and care coordination efforts.

The primary objective of this study is to conduct a comparative analysis of different process discovery algorithms within the field of process mining, utilizing hospital data pertaining to the treatment of sepsis. The anonymized daily records utilized in this study were sourced from Mannhardt (2016). Our primary emphasis was evaluating three distinct process discovery algorithms: Alpha Miner, Heuristic Mining, and DFG - Direct Flow Graph. It is essential to underscore that this investigation was confined to the discovery phase of process mining, excluding the compliance and improvement stages. Our principal anticipated outcomes encompass the identification of inefficiencies and bottlenecks within sepsis treatment processes, along with proposing enhancements and optimizations. Through the comprehension and standardization of daily records, we aim to provide valuable insights into the actual execution of processes, thereby facilitating more effective management of costs and resources, enhancing the quality of care, and fostering efficiency within public health facilities.

Related Works

In an environment where efficiency, precision, and quality are paramount, the healthcare sector continually seeks innovative methods to streamline its processes and enhance patient care delivery. Within this context, Process Mining has emerged as a promising approach capable of providing profound and informed insights into workflows and operations within healthcare systems [1]. While traditional process analysis often relies on theoretical models that may not fully capture the intricacies of clinical practice, Process Mining leverages real-world data collected through event logs, offering a robust and grounded perspective on actual process occurrences.

Analogous to a magnifying glass, Process Mining enables healthcare professionals to delve into the inherent details within event records, revealing patterns, temporal variations, activity interactions, and potential gaps that may impact care quality. By elucidating the authentic trajectory of clinical processes, Process Mining empowers professionals to pinpoint opportunities for improvement, optimize resource utilization, and ensure adherence to guidelines and regulations.

Ghasemi and Amyot [2] emphasized that the process mining framework can be delineated into three overarching categories. While additional steps may be integrated, these three categories constitute the foundational pillars of process mining method. Within these categories, we have the following:

Discovery: This stage unveils the underlying structure of processes, shedding light on their intricate dynamics and sequences.

Conformance: This phase is dedicated to scrutinizing the degree to which processes adhere to pre-established standards and specifications, ensuring compliance and consistency.

Enhancement: This aspect aims to perpetually refine processes and bolster operational efficiency through iterative optimizations and enhancements.

Process Mining

The origins of Process Mining can be traced back to the 1950s, marked by the generation of finite-state machines from sequence examples. Carl Adam Petri introduced the initial modeling language capable of effectively capturing competition dynamics, while Mark Gold delved into various notions of Machine Learning in 1967. Despite the burgeoning growth of Data Mining in the 1990s, minimal attention was directed toward processes. Process Mining gained notable traction and achieved a significant milestone in 2003 by publishing the first comprehensive survey in the field [3].

Over the years, Process Mining techniques have matured, bolstered by developing numerous specialized tools. Moreover, the discipline has undergone substantial expansion in scope. Initially centered on Process Discovery, Process Mining has gradually evolved to encompass many facets, including Compliance Checking, multi-perspective approaches, and Operational Support [3].

According Ghasemi and Amyot [2], Process Mining is an approach leveraging event logs to glean knowledge and insights into organizational processes. This method analyzes data collected during activities and workflows to discern real-world dynamics. The overarching aim of Process Mining is to unveil patterns, trends, variations, and potential inefficiencies inherent in processes, thereby enabling organizations to streamline their operations, enhance compliance with objectives, and pinpoint avenues for refinement.

Discovery

According to Rudnitckaia [4], the initial phase of process mining is termed the discovery stage. During this phase, a discovery technique utilizes event records as the foundation for constructing a process model without reliance on prior knowledge. An exemplary illustration of this method is the Alpha algorithm, which, when supplied with an event log, frequently generates a process model, typically a Petri net, to elucidate the behaviors captured within the log. As highlighted by Santos Garcia [1], this phase entails extracting process models from temporal records or event logs, culminating in a structured and visual depiction of organizational activities. A model is derived from the existing organizational processes as manifested in task execution. Subsequently, the extracted model undergoes a validation process through the conformance stage.

According to Rezende and colleagues [3], the Process Discovery stage is widely acknowledged as one of the most intricate phases within Process Mining. During this pivotal stage, a process model is meticulously crafted, leveraging event records as the foundation to encapsulate the behaviors documented within those events accurately. This form of Process Mining (PM) assumes a crucial role in discerning the authentic process model. It relies solely on an event log, from which a process model is synthesized to encapsulate the behaviors manifested within that log. [4] enumerates several prominent algorithms employed to accomplish this aim, including:

- Alpha Miner;
- Alpha+, Alpha++, Alpha#;
- Fuzzy Miner;
- Heuristic Miner;
- Multi-phase Miner;
- Genetic Process Mining;
- Region-based Process Mining (State-based regions and Language-based regions);
- Classical approaches do not deal with competitors.

Moreover, it is imperative to employ an appropriate notation to depict the process model in a manner that is comprehensible to end-users. Notations such as Workflow Networks, Petri Nets, Transition Systems, YAWL, BPMN, UML, Causal Networks (C-nets), and Event-Driven Process Chains (EPCs) are utilized. This selection of notation is essential not only to confine the search space for candidate models but also to favor specific types of models [4].

Rudnitckaia [4] also outlines that a process model generated during the Process Discovery phase can be evaluated based on the following quality criteria. Four primary dimensions of quality can be discerned and often compete with one another:

Adequacy (Fitness): The capability of the process model to accurately reproduce event records;

Simplicity: The selection of the most straightforward process model that still effectively explains the behavior observed in event records;

Precision: The capacity of the process model to disallow behavior not observed in event logs;

Generalization: The capacity of the process model to accommodate behavior not observed in event logs.

Materials and Methods

This endeavor culminated in an experimental investigation utilizing openly available data supplied by Mannhardt (2016) as the foundation to scrutinize the care trajectory for patients exhibiting suspected sepsis.

Moreover, our endeavor aims to delve into the three principal process mining techniques delineated in the literature: Alpha Miner, Inductive Miner, Heuristic Miner, and DFG - Direct Flow Graph. We endeavor to elucidate their distinctive characteristics and behaviors when applied to the dataset while showcasing their potential in facilitating process enhancement. To implement these techniques, we harness Python 3.7 and leverage the PM4Py library to execute the algorithms effectively.

The DataSet

The dataset comprises the trajectory of 1050 patients from an undisclosed Dutch hospital, all presenting with suspected sepsis. This dataset encompasses screening, laboratory, and financial records. However, our analysis was restricted to examining three primary pieces of information: the patient identifier, the activity undertaken, and the event date. These variables are denoted as case:concept: name, time: timestamp, and concept: name to ensure compatibility with the library's data requirements (Table 1).

Results and Discussion

Through the application of the three discovery models mentioned in the previous chapters, it was possible to identify points of attention in the care process and demonstrate the main flows and activities that occurred. Table 2 presents the initial activities of the process, the final ones, and the occurrence of each activity. This initial analysis allows for visualization of the patient's first and last interaction in the care unit. This analysis is essential to visualize whether the data follows the standard the unit recommends.

Table 2 delineates the start and end activities; however, more than this level of analysis is needed to illustrate the flow or pinpoint inefficiencies within the process.

Table 1. Simple event log.

Case:Concept:Name	Time: Timestamp	Concept: Name
A	2014-10-22 09:15:41+00:00	ER Registration
A	2014-10-22 09:27:00+00:00	Leucocytes
A	2014-10-22 09:27:00+00:00	CRP
A	2014-10-22 09:27:00+00:00	Lactic Acid
A	2014-10-22 09:33:37+00:00	ER Triage

Table 2. Start and end activities existing in the process.

Initials		Finals	
Activity	**Occurrence**	**Activity**	**Occurrence**
ER Registration	995	Release A	393
IV Liquid	14	Return ER	291
ER Triage	6	IV Antiniotics	87
CRP	10	Release B	55
ER Sepsis Triage	7	ER Sepsis Triage	49
Leucocytes	18	Leucocytes	44
		IV Liquid	12
		Release C	19
		CRP	41
		Lactic Acid	24
		Release D	14
		Admission NC	14
		Release E	5
		ER Triage	2

Table 3 showcases the most prevalent variants observed in the process. These variants represent the diverse pathways traversed by patients from their initial consultation to their final one. Our analysis revealed 85 distinct alternative flows, underscoring the significant variability inherent in the process. Such variations may stem from the intricacies of the care protocol or potential errors encountered during care.

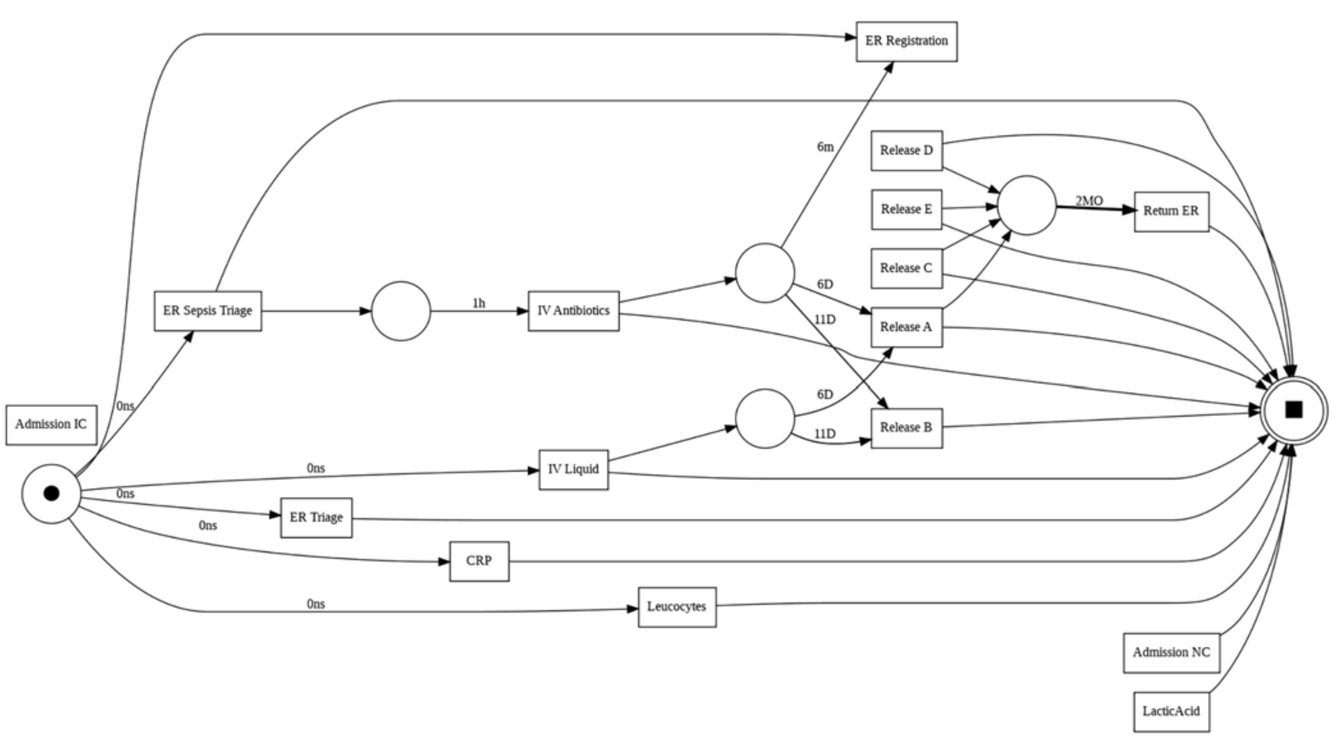
This analysis furnishes managers with insights into the activities warranting heightened attention and those of lesser frequency, thereby facilitating the optimization of resource allocation across human, technological, and financial domains. Furthermore, it aids in investigating potential aberrations within the process, given that activities occurring with greater frequency are more likely to recur. This sheds light on the proportion of patients referred for more intricate activities and the representativeness of such referrals.

In Figure 1, including transition durations between two activities enhances professionals' understanding by visually depicting the varied interactions among activities, along with any

Table 3. Variant table.

Variant/Sequence of Activities	Occurrence
'ER Registration' - 'ER Triage' - 'ER Sepsis Triage'	35
'ER Registration' - 'ER Triage' - 'ER Sepsis Triage' - 'Leucocytes' - 'CRP'	24
...	...
'ER Registration' - 'ER Triage' - 'Leucocytes' - 'CRP' - 'Lactic Acid' - ER Sepsis Triage' - 'IV Liquid' - 'IV Antibiotics'	3
'ER Registration' - 'ER Triage' - 'Lactic Acid - 'Leucocytes' - 'CRP' - 'ER Sepsis Triage' - 'IV Liquid' - 'IV Antibiotics'	3

Figure 1. Work process using the Alpha Miner algorithm.



anomalous occurrences. This visualization streamlines the analysis process, enabling professionals to discern patterns and anomalies more readily.

Figure 1 illustrates that throughout the process flow, specific transitions between activities exhibit longer durations compared to others. For instance, the transition from IV antibiotics to discharge A and discharge B spans 6 days and 11 days, respectively. Additionally, the transition from discharge to return activities averages 6 months, underscoring critical bottlenecks in the process. Furthermore, activities needing clear outcomes, such as the transition from Admission to ER Registration, necessitate further specialist analysis.

In Figure 2, the outcome of applying the Heuristic algorithm from Minas Gerais is depicted, revealing the identification of rework instances and the average time required for these reworks within the healthcare unit. This identification serves as a metric for gauging inefficiencies within the process, thereby enabling managers to initiate actions to enhance efficiency and address the root causes

Figure 2. Work process using the Heuristic Mining algorithm.

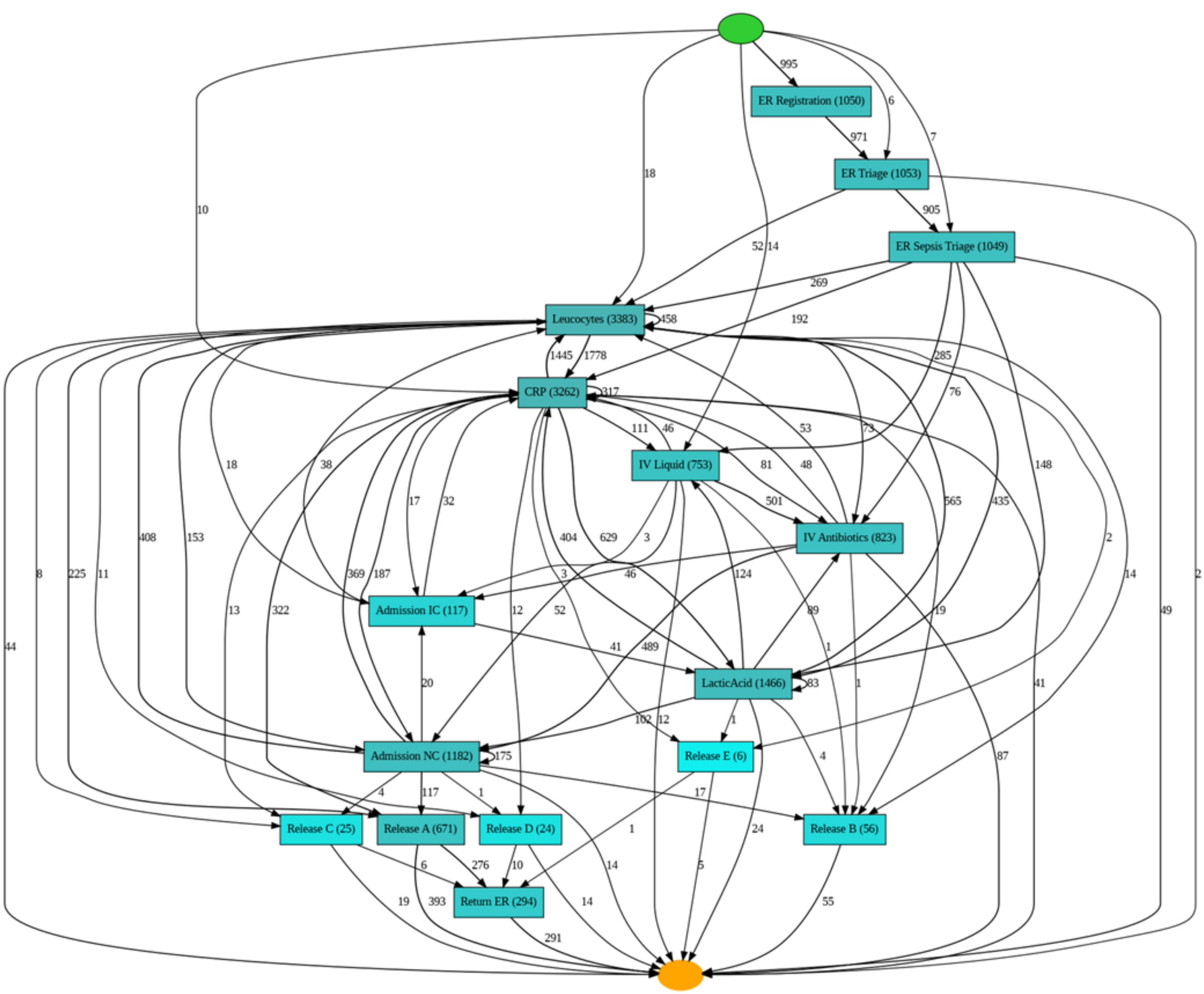
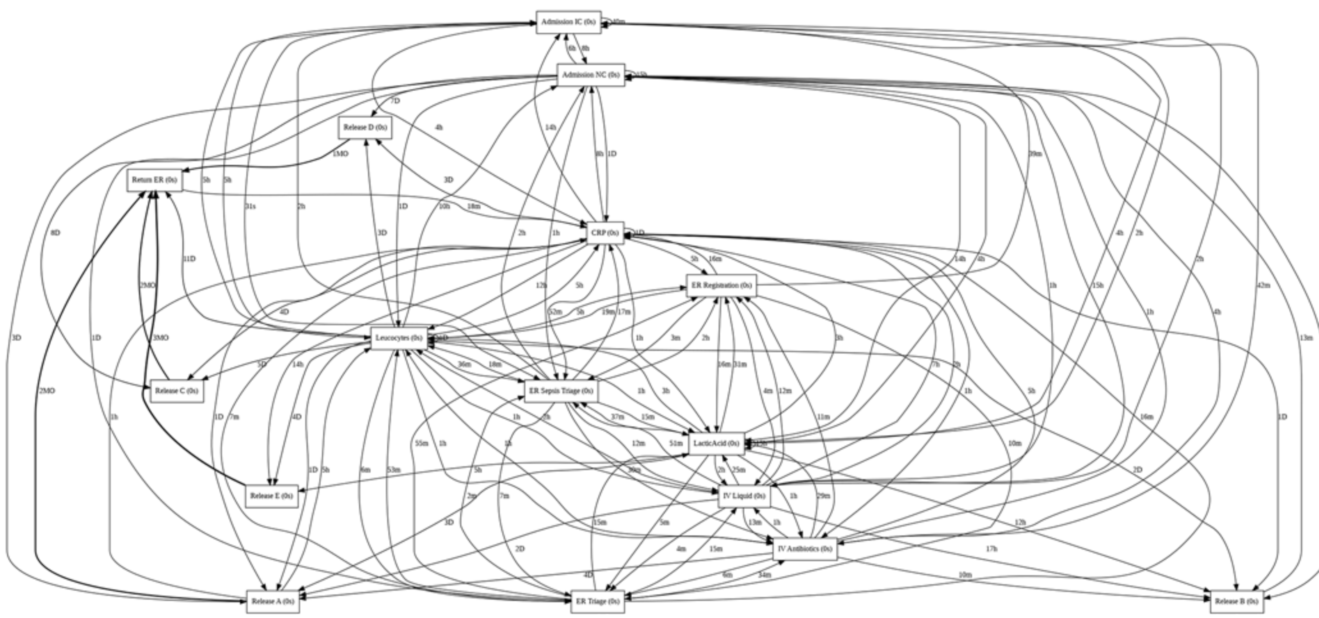


Table 4. Main reworks in the process.

Activity	Occurrences	Average Time per Occurrence
Leucocytes	458	ID
CRP	317	ID
Admission NC	175	15 h
Lactic Acid	83	15 h

Figure 3. Work process using the DFG algorithm - Direct flow graph.



of these rework occurrences (Table 4). As well in Figure 3, we have the same process using DFG algorithm.

In these studies, we also explored the inductive discovery technique, which facilitated identifying activities that may transpire concurrently or mutually exclusively. Figure 4 presents a simplified depiction of the most recurrent activities and their correlations, elucidating whether they unfold sequentially, in parallel, or encompass decision loops. This insight holds paramount significance at the strategic level, aiding in estimating personnel requirements across various sectors based on activity patterns.

Figure 5 adds information about the sequence of activities. However, it needs to clarify whether the gateways are exclusive and inclusive or the activities are sequential. This visualization mode could be more efficient for this analysis due to its limitation, compromising experts' analysis.

Conclusion

This study aimed to investigate three distinct process discovery techniques using data related to suspected sepsis. Applying these techniques elucidated attention points requiring careful consideration by healthcare professionals during patient care. These include instances of rework generated during treatment and prolonged transition durations. However, it is crucial to note that determining the root cause of such distortions,

Figure 4. Work process using the Heuristic Mining algorithm.

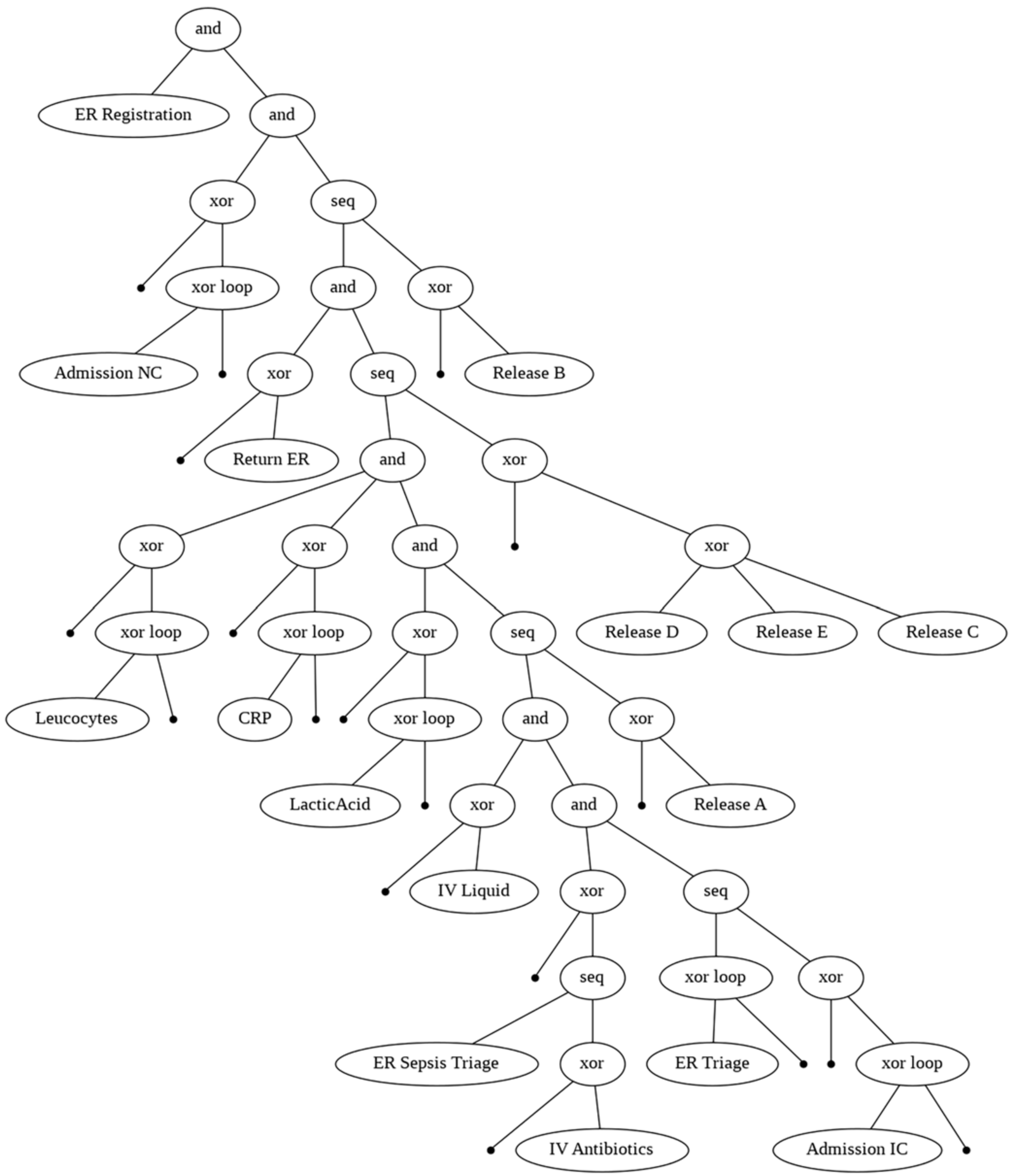
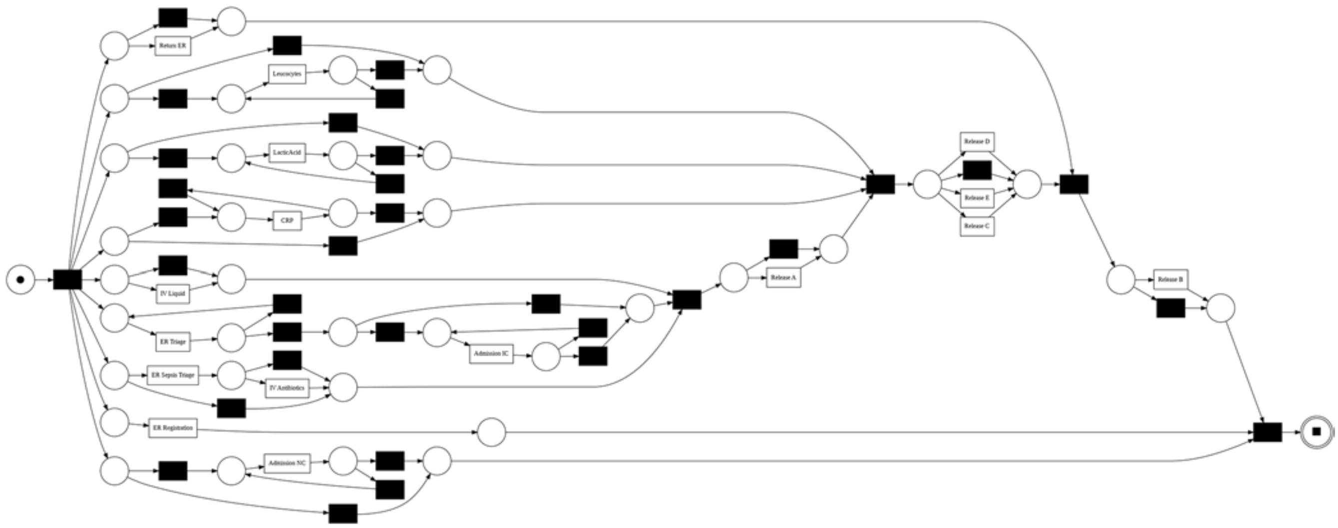


Figure 5. Work process using the Heuristic Mining algorithm.



whether they stem from errors or represent regular occurrences, necessitates collaboration with medical specialists. Moreover, these distortions may vary within each healthcare unit.

For future research endeavors, the focus will shift towards a neighboring healthcare unit, with the active involvement of healthcare professionals. This collaborative approach will not only encompass the utilization of discovery techniques but also aim to enhance the services provided to users by leveraging insights gleaned from the professionals' expertise.

References

1. dos Santos Garcia C, Meincheim A, Junior, ERF, Dallagassa MR, Sato DM.V, Carvalho DR, Santos EAP, Scalabrin EE. Process mining techniques and applications–a systematic mapping study. Expert Systems with Applications 2019;133:260–295.
2. Ghasemi M, Amyot D. From event logs to goals: a systematic literature review of goal-oriented process mining. Requirements Engineering 2020;25(1):67–93.
3. Rezende CA et al. Arcabouço de classificação e escolha de algoritmos de descoberta de processos. 2017.
4. Rudnitckaia J. Process mining. data science in action. University of Technology, Faculty of Information Technology, 2016:1–11.