**Introduction**

This report presents a comprehensive analysis of the handling of patients suspected of sepsis in one of the Dutch hospitals. Sepsis, a critical medical condition, is defined as the body's extreme response to an infection and is potentially life-threatening if not promptly diagnosed and treated. The significance of this study lies in its potential to improve patient care and outcomes in sepsis cases.

According to the research done on the subject [1], process mining techniques can provide vast and valuable insights into the trajectory of patients in a hospital setting. This innovative approach allows researchers to map out the complex journey of sepsis patients from admission to discharge. Moreover, it enables the verification of whether the daily clinical practice in the hospital aligns with established medical guidelines for sepsis management. This comparison between actual practice and guidelines is crucial for identifying areas of improvement in patient care protocols.

In the research that will be presented below, we performed several processes to gain insights into the process of treating sepsis patients. We started by understanding the database and cleaning it, from there we moved on to discovering the model and performing conformance checking against the database data, while choosing the optimal model and checking the processes that the patients go through in relation to it, and whether there is anything to improve in the existing processes in the organization. Finally, we tried to improve the model we received using additional techniques, for example: checking the time for a process that a certain patient goes through, mining decision points, and more.

**A literature survey on process mining techniques and their use in our project**

In this section, we present methods used during our project at different stages, such as discovering the model, performing conformance checking and improving the model. This literature survey examines four key process mining techniques: Alpha Miner, Inductive Miner, Conformance Checking, and Decision Point Mining.

The Alpha Miner [2] constructs a Petri net representation of a process by identifying direct succession, causality, and parallel relationships between activities in event logs. The algorithm examines the order of events in the log and infers the underlying process structure. While providing a basic understanding of process structure, it has limitations in handling complex behaviors such as loops and noise in real-world logs. We used Alpha Miner to create a basic model, from which we could advance to better discovery methods.

The Inductive Miner [3] addresses some of the Alpha Miner's limitations. It uses a divide-and-conquer approach to recursively split the event log and discover process fragments, which are then combined into a process tree. This technique is effective in handling large and complex event logs, making it suitable for analyzing intricate processes. We used inductive miner to more accurately discover our model, tested its soundness, and used the resulting model as we'll explain in the report.  
Conformance Checking [4] evaluates the alignment between process models and actual process executions. It compares the discovered process model with a predefined reference model or set of rules. This technique allows organizations to assess compliance with procedures, identify deviations, and measure the fitness of their process models. We used conformance checking for two purposes: to determine which model best fits our data, and mainly to examine an existing model's suitability against actual hospital behavior. This allowed us to map non-conformities and identify less suitable process parts, revealing opportunities for improvement.

Decision Point Mining [5] analyzes choices made at specific points in a process. This technique uncovers factors influencing decisions, providing insights into rules and patterns governing these choices. By examining case characteristics at decision points, it helps identify criteria used in process-related decisions. We tried to use this and other methods to try to improve our discovered model.

**Description of the database and preprocessing the database**

Our study examines sepsis cases in a Dutch emergency ward, utilizing a robust event log collected over 18 months from November 2013 to June 2015. The dataset encompasses 1,050 patient traces, comprising 15,214 events across 16 activity types and 39 data attributes per case. By integrating information from various hospital systems, including triage documents, blood tests, and financial records, the study enables a detailed analysis of patient journeys. Applying process mining techniques to this rich dataset allows researchers to assess guideline adherence, identify treatment protocol deviations, visualize care pathways, and analyze patterns of patient returns within 28 days. Ultimately, our research aims to enhance emergency care processes and improve patient outcomes by providing valuable insights into the complexities of sepsis treatment in an emergency setting .the possible steps in the process include: ​

1. Registration in the emergency room: This is the initial step where patients with symptoms of sepsis are registered upon arrival at the hospital's emergency ward.
2. Triage: A triage document is filled out for sepsis patients, which includes information such as the time of triage, symptoms present (SIRS criteria for sepsis), diagnostics ordered, and the time infusions of liquid and antibiotics were administered. ​
3. Medical activities: These activities include measurements of leukocytes, CRP (C-reactive protein), and lactic acid. ​These tests are likely performed to assess the severity of the sepsis condition and guide treatment decisions. ​
4. Admission to hospital wards: Patients may be admitted to different wards based on their condition. The document mentions admission to normal care wards and intensive care wards as possible trajectories. ​
5. Discharge: There are different variants of discharge from the hospital, which may depend on the patient's condition and response to treatment. ​
6. Returning patients: The document mentions investigating the trajectory of patients who return to the hospital within 28 days. ​ This suggests that some patients may require readmission or further treatment after their initial discharge.

we highlighted the most common process variants observed in the patient handling process for sepsis cases:

תמונה שמכילה טקסט, צילום מסך, גופן, לבן

התיאור נוצר באופן אוטומטי

As we can see, variants such as ER Registration, ER Triage, and ER Sepsis Triage were prevalent in the event log data, indicating the standard sequence of activities when managing sepsis patients. Interestingly, none of these common variants concluded with a discharge, suggesting a deviation from the expected endpoint in the patient trajectory. This discrepancy raised concerns about the completion of the patient handling process and the need for further investigation to address this issue.

As a result, we decided to filter our event log data using specific start and end activities. The valid start activity was ER Registration, marking the beginning of a patient's trajectory in the hospital. The valid end activities, including Release A, Release B, Release C, Release D, Release E, and Return ER, signified different points at which a patient's journey could conclude. By filtering the event log based on these start and end activities, a subset of 734 traces was isolated for further examination and process modeling. This filtering step was crucial in focusing the analysis on complete patient cases and understanding the flow of activities from registration to discharge in handling sepsis patients within the hospital environment.

In addition, we also analyzed the distribution of events that occur in varying periods of time (by month, by day, and by hour), and also the distribution of events by the total duration recorded in the database (Graphs for this section are in the Appendices section). We noticed that most events happen in the hospital in the morning. It can be concluded from this that people naturally come to the hospital more in the morning after they get up, even if something happens to them the night before.

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטי

**Process Discovery**

In the process analysis, we selected to focus on the three most frequent variants in the model.

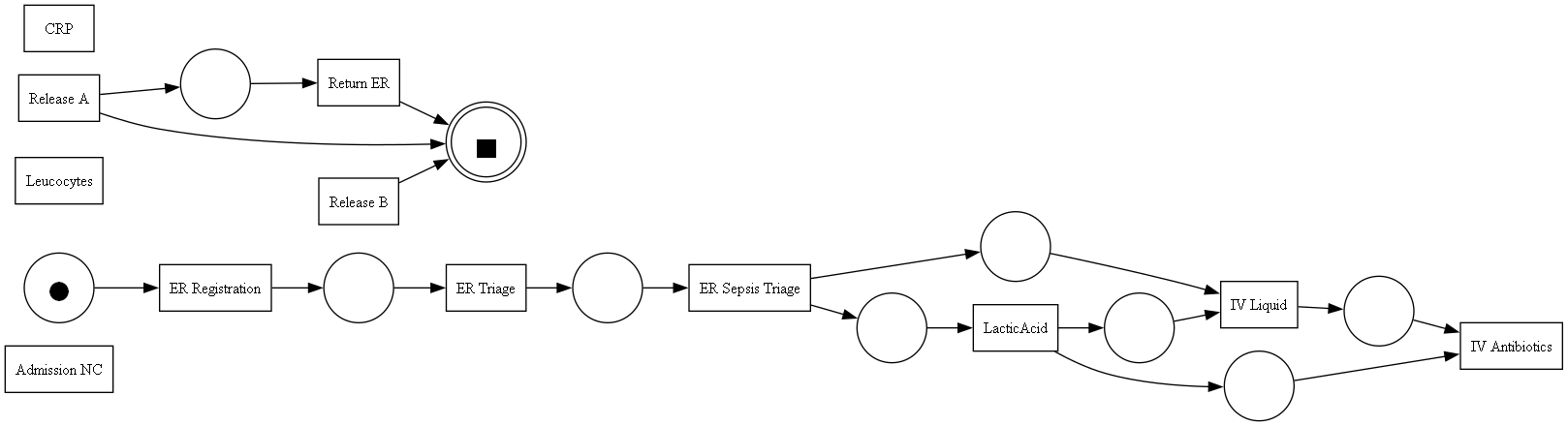
Our goal was to create a model that is easy to understand, work quickly, and give accurate results. To do this, we tried different ways to build the model. We experimented with using more or fewer steps in the process. After careful thought and testing, we found that using only the three most common steps was the best approach.

By limiting the model to these three variants, we maintained a high level of precision and simplicity while ensuring that the model accurately represents the key steps in the process flow of handling sepsis patients in the Emergency Room. This decision was made to optimize the model's effectiveness in capturing the essential activities and sequences without compromising on the overall quality and interpretability of the process representation.

We made a comparison between a different number of common variants and found out that building the model with more than three variants reduces precision and simplicity, so the final model uses only the top three variants with metrics presented in a table.

We wanted to make our model as strong and accurate as possible. So, we used inductive algorithm. This way was better than the alpha algorithm way because it could handle tricky parts of the process and make sure the model is correct. The alpha couldn't always do this, especially when there were steps that repeated or when the process was complicated. By using the inductive algorithm, we made sure our model was right and could handle all the different things that happen when caring for sepsis patients in the emergency room.

An example of petri net obtained using Alpha-



we can see that the result is wrong. Not all the parts are connected to each other as we would like to get.

להוסיף יוריסיטי

From now on, the Petri Net networks we will represent will be using an inductive algorithm only.

Using only the most common variant gave us this petri net-



As we expected- one variant gives very simple petri net, but it doesn't fit the whole data.

However, when we used 35 variants, we got this-

תמונה שמכילה תרשים, קו, לבן, שרטוט

התיאור נוצר באופן אוטומטי

Which is very complicated petri net, with much higher fitness (95.4%).

The 3 most common variants petri net-

תמונה שמכילה תרשים, קו, לבן, גופן

התיאור נוצר באופן אוטומטי

we think that this model is a suitable model. this model is simpler and faster, but it still gives us a good picture of how to care for sepsis patients in the emergency room.

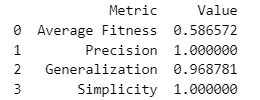
**Conformance Checking**

Conformance checking serves multiple purposes, as discussed in our course.

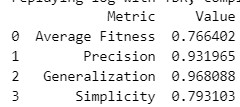
One primary goal is model discovery, which involves engineering a conformance checking algorithm to uncover a suitable process model. While there may not be a single "best" model, this approach allows for the identification of a model that accurately represents the observed behavior.

The second and more critical objective of conformance checking is to evaluate the conformance of an existing model. This evaluation is performed against a larger log of traces that were not used to generate the model.

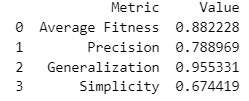
We evaluated four quality metrics for each model: Fitness, Precision, Generalization, Simplicity and we compared the results across different trace counts



1 most common trace:



3 most common trace:



5 most common trace:

Among the three models, the middle model with three variants seems like the most effective choice. It strikes a balance between achieving high performance across all metrics and maintaining a manageable level of complexity.

**Fitness**, which measures how well the model aligns with real-world data, reaches a 76.64% for our model. This indicates that the model accurately captures the essential steps involved in sepsis patient care.

**Precision**, reflecting the model's ability to correctly identify patterns, attains a 93.19%. This suggests that the model effectively distinguishes between valid and invalid process executions.

**Generalization**, assessing the model's ability to handle new, unseen data, achieves a 96.8%. This implies that the model can generalize beyond the training data and adapt to real-world scenarios.

**Simplicity**, evaluating the model's ease of understanding and interpretation, reaches a 79.31%. This indicates that the model is not overly complex and can be readily grasped by stakeholders.

In contrast, the model with one variant, while achieving high Fitness of 88.22% and 100% Precision, suffers from lower Generalization 95.53% and Simplicity 67.44%. This suggests that the model overfits the training data and may not generalize well to real-world scenarios.

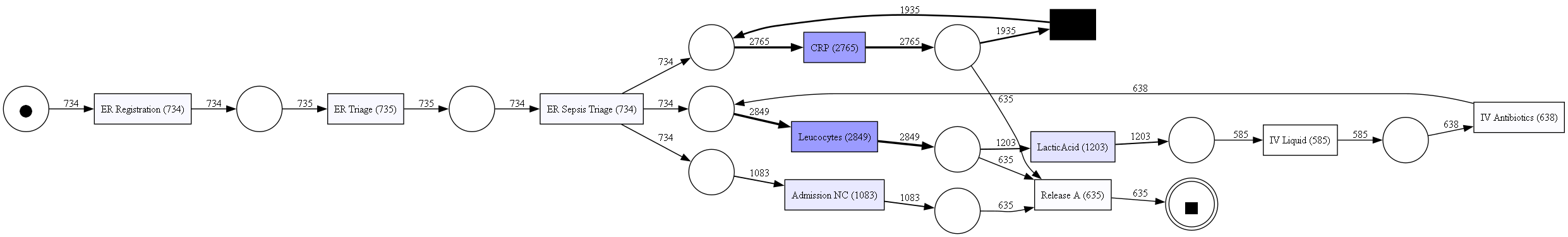
On the other hand, the model with five variants, while achieving high Generalization 98.32% and Simplicity 80.12% , is lower in Fitness 68.89% and Precision 78.89%. This indicates that the model may be too complex and may not accurately capture the essential steps in the process.

Therefore, the model with three variants strikes the optimal balance between performance and simplicity, making it the most suitable choice for representing the process flow.it does a good job of matching real data, is accurate, can handle new situations, and is easy to understand.

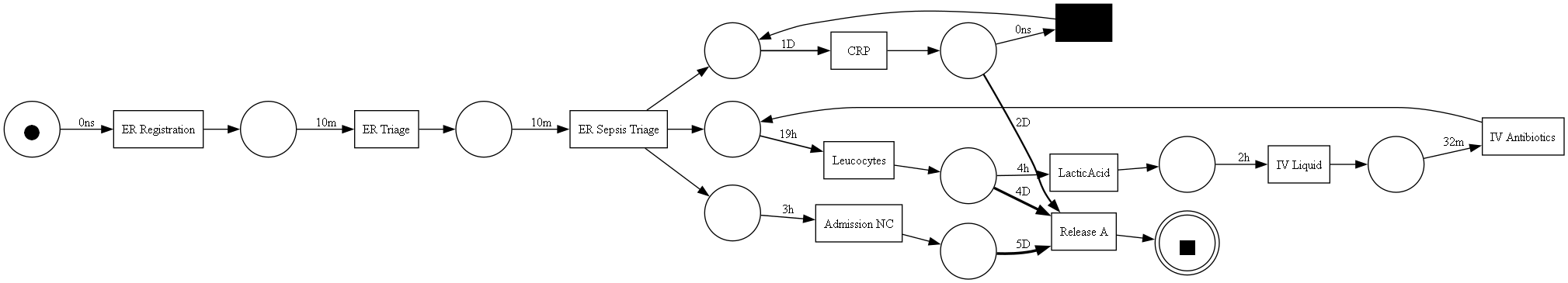
**6) Analysis of the model and insights into the processes.**

**7) Proposing improvements to the process based on methodologies learned in the course (for example, point mining)**

We created Frequency Information Plot to provide insights into the occurrence rates of different activities in the process model. This plot helps us understand the distribution of activities and their relative importance in the overall process.  
We wanted to check if there is an event that seems unusual - an event that repeats less often, but as we can see that all the events are relatively common and therefore we did not find any exceptions that can be removed

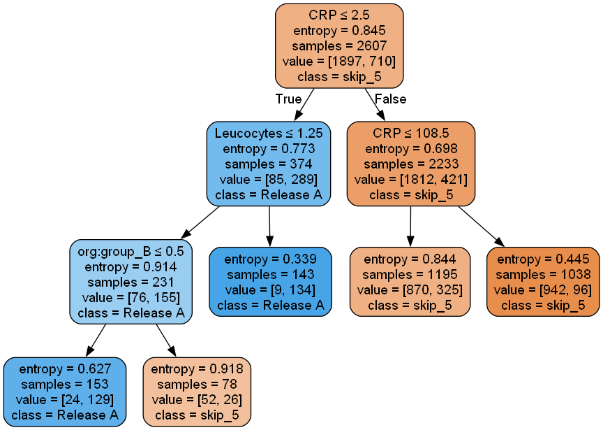


The Temporal Information Plot evaluate if the process models aligned with medical guidelines by visualizing the timing of critical actions. This analysis aimed to determine if prescribed treatments and tests were administered within recommended timeframes While initial observations indicated potential delays, further investigation revealed that repeated actions, rather than overall process failures, were the primary cause of these issues.



we created classifiers for the decision points in our model. There are two decision points of interest: one after CRP, and another one after Leucocytes.

תמונה שמכילה טקסט, גופן, צילום מסך, כתב יד

התיאור נוצר באופן אוטומטי

Making decisions based solely on these decision trees proves to be challenging. CRP and Leucocytes are important factors, but the trees show that there's a lot of uncertainty in the outcomes. This means that using only these two things might not be enough to make the right choice every time. We might need to consider other things or use more advanced methods to make better decisions.

**Summary**

**After references** 10) נספחים (לא נספרים במניין העמודים).

**References**

[1] F. Mannhardt and D. Blinde. “Analyzing the trajectories of patients with sepsis using process mining”. English. In: RADAR+EMISA 2017, Essen, Germany, June 12-13, 2017. CEUR Workshop Proceedings. RADAR + EMISA 2017 ; Conference date: 12- 06-2017 Through 13-06-2017. CEUR-WS.org, 2017, pp. 72–80.

[2] Van Der Aalst, W. (2016). Process mining: Data science in action. Springer, Berlin, Heidelberg. Pp. 167-177

[3] Van Der Aalst, W. (2016). Process mining: Data science in action. Springer, Berlin, Heidelberg. Pp. 222-235

[4] Van Der Aalst, W. (2016). Process mining: Data science in action. Springer, Berlin, Heidelberg. Pp. 243-274

[5] Van Der Aalst, W. (2016). Process mining: Data science in action. Springer, Berlin, Heidelberg. Pp. 294-296

**Appendices**

**A link to our code:**

<https://github.com/Noam-Diamant/Process-Mining/blob/main/sepsis%20project.ipynb> **Description of the database and preprocessing the database**

The most common variants of the database before the filtering we applied:

תמונה שמכילה טקסט, צילום מסך, גופן, לבן

התיאור נוצר באופן אוטומטיThe most common variants of the database after the filtering we applied:

תמונה שמכילה טקסט, צילום מסך, גופן, לבן

התיאור נוצר באופן אוטומטיDistribution of events by date:

תמונה שמכילה טקסט, תרשים, עלילה, קו

התיאור נוצר באופן אוטומטי

Distribution of events by month:

תמונה שמכילה טקסט, צילום מסך, גופן, עלילה

התיאור נוצר באופן אוטומטי

Distribution of events by day:

תמונה שמכילה טקסט, צילום מסך, גופן, תרשים

התיאור נוצר באופן אוטומטי

Distribution of events by hour:

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטי