Use Case Description

The underlying use case is from a Dutch hospital with about 50,000 patients per year and 700 beds. The process focuses on sepsis cases. Sepsis can be life-threatening and occurs when the body reacts to an infection and injures its organs or tissue.

To track all performed events, the hospital uses an Enterprise Resource Planning system. The process contains logistical activities, i.e., the patient's pathway through the hospital, and medical, i.e., which blood values were measured and which treatments were performed.

First, patients are registered at the Emergency Room (ER). Next, they go through ER triage and ER sepsis triage. Their blood values in regards to Lactic Acids, Leukocytes and CRP are measured recurringly throughout the process. Further, patients can receive IV Liquid, IV Antibiotics or both during the treatment process. Based on their blood values, patients are admitted to Normal Care (NC) or Intensive Care (IC). Patients are released from the hospital with one of five release types. Finally, it is tracked if patients return to the ER after the release.

The event log representing the use case contains of 1,050 process instances with 846 different variants over a time span of 1.5 years. Each process instance consists of four to 185 events with a mean of 14 events per process instance. However, in order to limit computational time, we filtered out process instances longer than 20 events. The events can be differentiated into 16 activities with the following purposes:

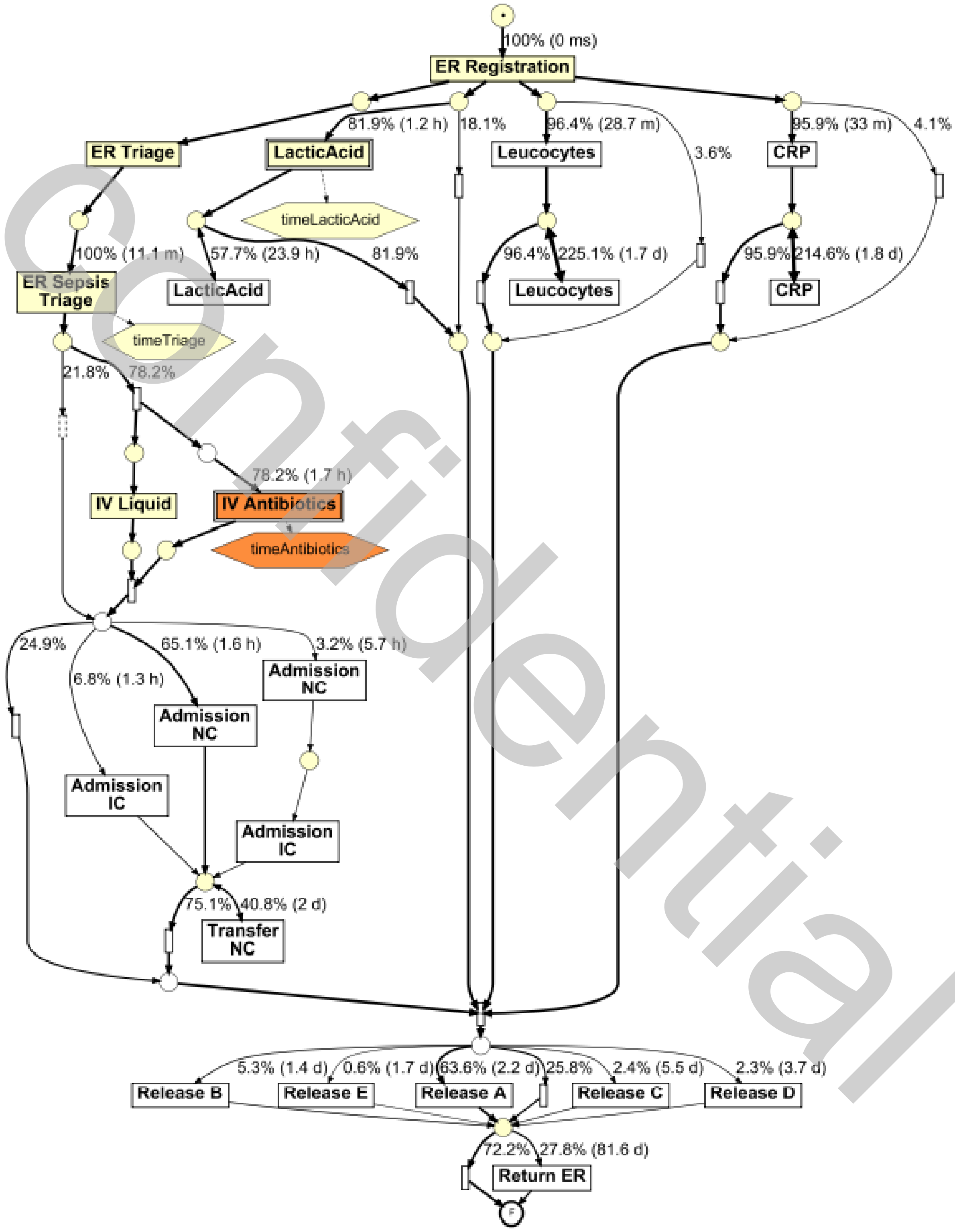
- 5 activities regarding the release type
- 3 activities regarding the measurement of CRP, Leukocytes and Lactic Acid
- 3 activities regarding the registration and triage
- 2 activities stating if the patient was admitted to IC or NC
- 2 activities regarding the treatment with IV
- one activity stating whether the patient had to return to ER

Besides the control flow information, the event log contains 29 context attributes. These are, e.g., the organizational unit responsible for an activity, the patient's age, the measured values of the CRP, Leukocytes and lactic acid, or the diagnosis. The remaining are binary attributes stating, e.g., whether the patient received an infusion.

Note: Some data was anonymized due to data security issues. E.g., Release Type A means a patient was released from the hospital as healthy. Further release types can be death, transferral to other hospitals, etc.

1. Process Model

The process model looks as follows (with rectangles representing activities in the process); one walk through the process model refers to one patient.



Source: Mannhardt, Felix & Blinde, Daan. (2017). Analyzing the Trajectories of Patients with Sepsis using Process Mining.

2. Prediction

In our paper, we aim at predicting patient pathway. Therefore, we create a method based on Artificial Intelligence. This method uses the data gathered in the process as input and creates a prediction. This data is not only the conducted activities but also further information like age, blood values, etc. In the underlying use case, we differentiate between two prediction tasks. *(Note: There are further tasks that are out of scope for this paper.)* The first one is the prediction of Release Type A, i.e., will the patient be released from the hospital as healthy. The second one predicts if the patients will be admitted to Intensive Care during their stay at the hospital.

Prediction task: Patient will be released as Type A

Q1: *Is the prediction of the release type relevant in general? If so, in what way would it support medical staff in their decisions/ tasks?*
Q2: *Would you trust the prediction without further explanation on how it was created?*

Prediction task: Patient will be admitted to Intensive Care

Q3: *Is the prediction of the release type relevant in general? If so, in what way would it support medical staff in their decisions/ tasks?*
Q4: *Would you trust the prediction without further explanation on how it was created?*

General:

Q5: *Can you think of other prediction tasks which may be of use for practitioners?*

3. Prediction with Explanation

Research has shown that practitioners do not trust such predictions because the reasons remain unclear to the user. Therefore, we have created a method that not only predicts the patients' pathways but also provides an explanation of the prediction to the user.

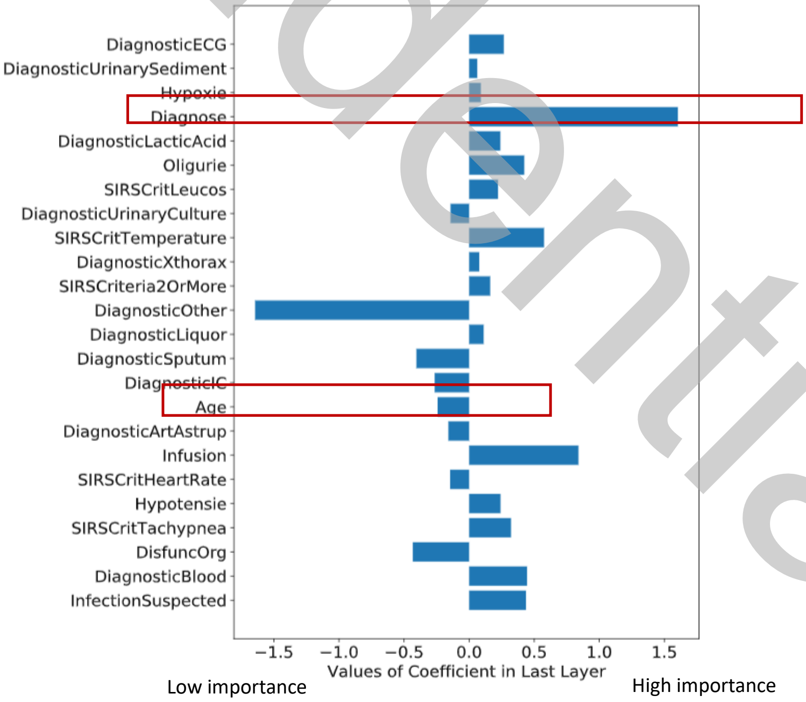
3.1 Static patient information

First, we focus on static data, i.e. data that remains the same throughout the process, such as the patient's age.

For both prediction tasks, we have created a plot stating which information in the data set is relevant for the prediction.

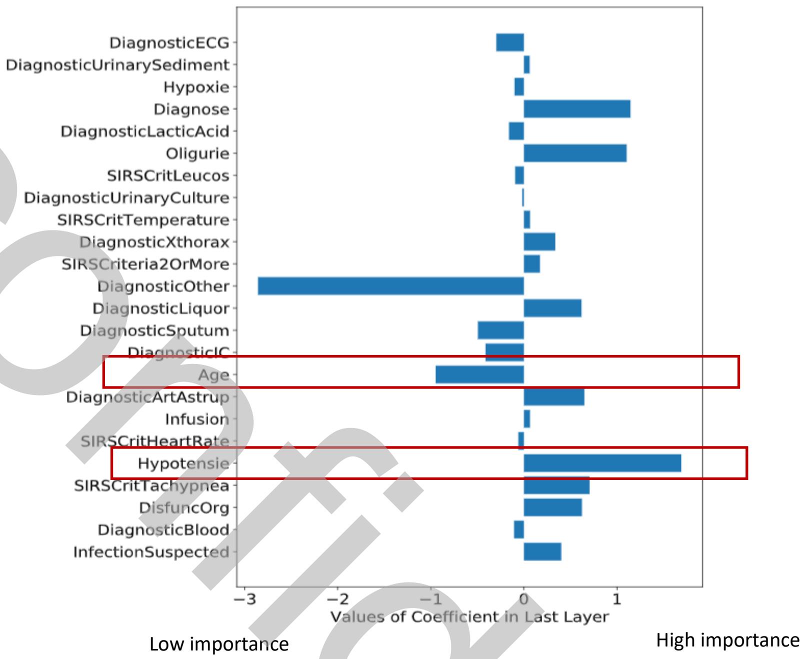
Prediction task: Patient will be released as Type A

For example, as you can see in the following plot, the diagnose is the most important information. It is, for example, more important than the patient's age in order to predict if the patient was released as healthy.



Prediction task: Patient will be admitted to Intensive Care

In the second case, for example, an occurring hypotensie is the most important information for the prediction. Compared to the figure above, the diagnose is not as important. The patient's age is once again less important.



Overall, we can see that the influencing factors on the prediction vary across prediction tasks.

> ***Q6:*** *In general, are explanations (e.g., in the form provided above) helpful to you and provide additional insights?*
> ***Q7:*** *Would explanations (e.g., in the form provided above) influence your trust/ acceptance of predictions?*

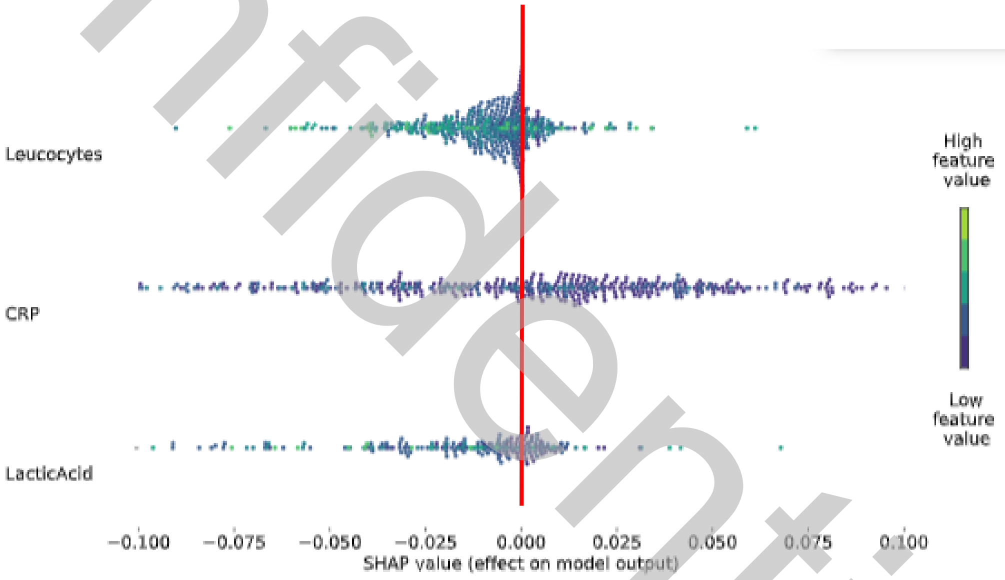
3.2 Non-Static patient information

Non-static information is information that changes throughout the process. For example, if the blood value is measured multiple times throughout the process, it brings back different values with every measurement.

For both prediction tasks, we have created a plot stating which information in the data set is relevant for the prediction. Here one data point represents one patient.

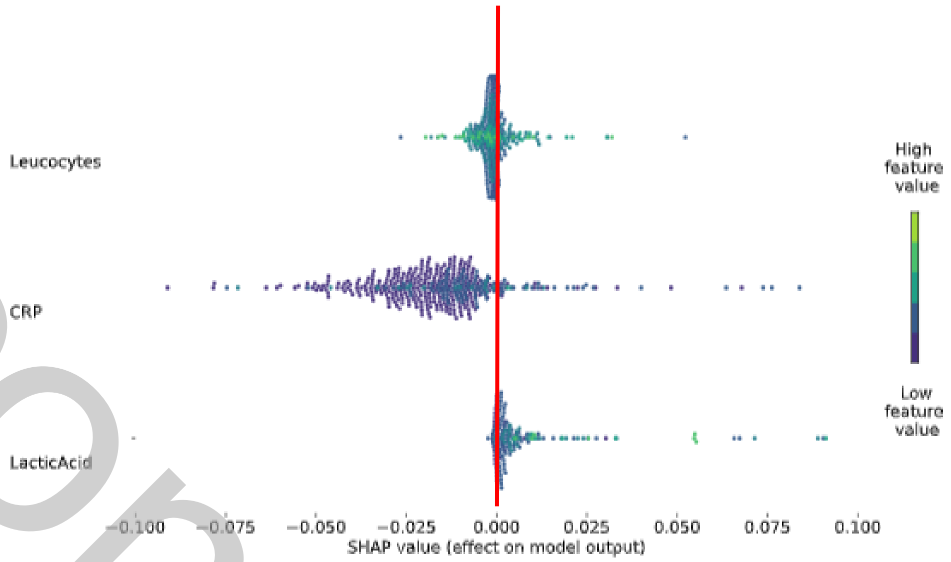
Prediction task: Patient will be released as Type A

Regarding the Leucocytes, more bright green points are negative. Therefore, a higher Leucocyte level would motivate the model not to predict Relaease type A. For CRP, the values are more spread, with more values being positive. Therefore, we can see that the CRP values have higher importance for the prediction. Lactic acid has low importance as the data points are close to zero.



Prediction task: Patient will be admitted to Intensive Care

The Leucocytes value has low importance for the prediction in regards to a visit at the intensive care. The CRP values are mostly negative and also have low importance. The results for Lactic acid values are mainly positive and therefore are important for the prediction. Further, as these positive values are bright green, the high values of lactic acid motivate the model to predict that a patient was sent to IC. (High levels of lactic acid can cause a life-threatening condition, called lactic acidosis.)



Overall, we can see that the influencing factors on the prediction vary across prediction tasks.

> ***Q8:*** *In general, are explanations (e.g., in the form provided above) helpful to you and provide additional insights?*
> ***Q9:*** *Would explanations (e.g., in the form provided above) influence your trust/ acceptance of predictions?*