Hugo De Oliveira[1,2], Martin Prodel[2], Ludovic Lamarsalle[2], Andi Orlowski[3]
1. Mines Saint-Etienne, France; 2. HEVA, France; 3. Imperial College Health Partners, England
Contact: hdeoliveira@hevaweb.com

CONFERENCE OF THE EUROPEAN WORKING GROUP
ON OPERATIONS RESEARCH APPLIED TO HEALTHCARE SERVICES
28 JULY - 02 AUGUST, 2019 - KARLSRUHE, GERMANY

Process mining for predictive analytics: a case study on NHS data to improve care for sepsis patients

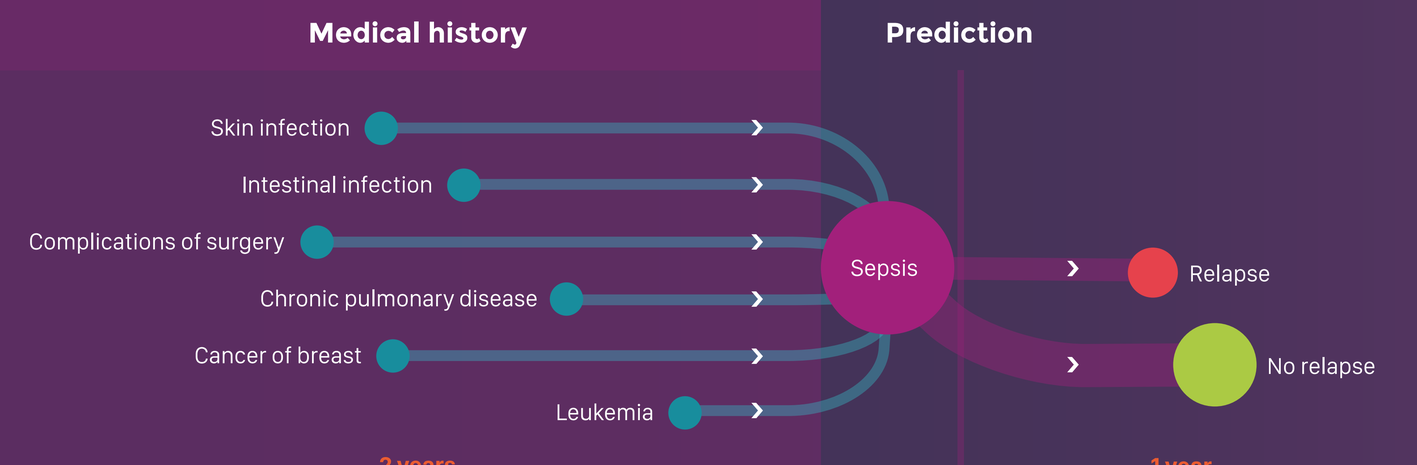


Fig 1. Illustration of Process Mining results for patients' pathways analysis

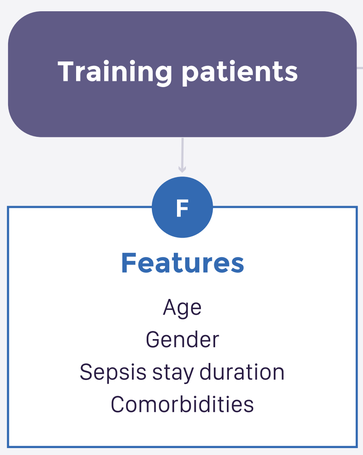
Context

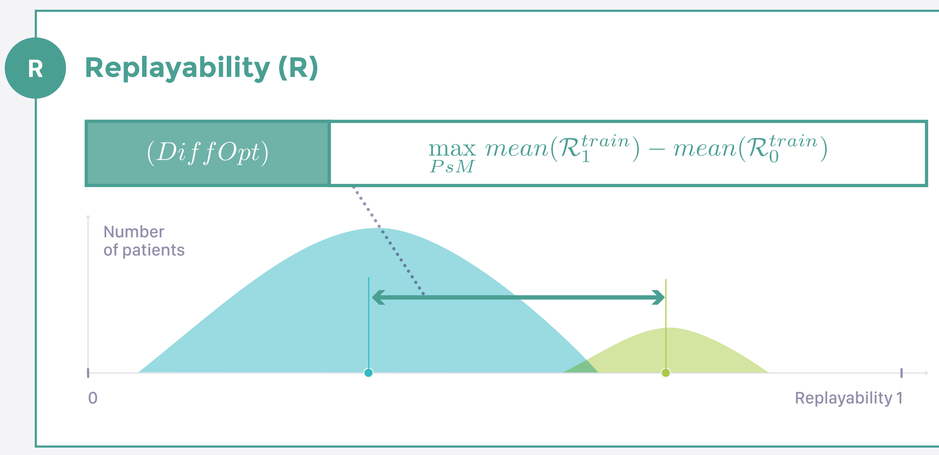
A definition for sepsis is 'life-threatening organ dysfunction caused by a dysregulated host response to infection' [1]. Understanding patients' pathways for this pathology is relevant to improving care.

Claims data from the NHS have been used to map sepsis patients' pathways. Patients with a sepsis episode in 2016 have been selected for inclusion. Medical events during the two years before and one year after sepsis have been analysed.

As the risk of sepsis relapse is challenging to identify, the presented study focuses on sepsis relapse prediction using patients' characteristics. For patients with an exhaustive medical history, inclusion information is not sufficient to predict sepsis relapse. An innovative methodology is proposed to use medical history structured in event logs, to improve prediction performances while maintaining the ability to explain predictions.

Pattern extraction from medical history using process model optimization





Medical history of training patients is structured in an event log $L = (L_0^{train}, L_1^{train})$ of class 0 (no relapse within 1 year) and 1 (relapse within 1 year).

The replayability $R(PsM, \sigma) \in [0, 1]$ is a measure, computed using an algorithmic procedure, which quantifies the ability of a process model PsM to represent a trace[2].

The idea behind the proposed method is the construction of a process model which well represents patients from L_1^{train} (relapse) while less representing traces from L_0^{train} (no relapse). This optimization function $(DiffOpt)$ used during graph construction incorporates this idea. Thus, discriminative patterns from pathways of patients having relapse are extracted.

Experiment

For all patients, features are used to train a decision tree for relapse prediction at sepsis episode release. For patients with exhaustive medical history (5 or more medical events within the 2 years before sepsis episode), the proposed methodology of pattern extraction is used to enrich feature data with the replayability score **(fig.2)**.

For each configuration, 80% of patients have been randomly selected as a 'train' set, the remaining 20% forming a 'test' set where area under the roc curve (AUC**) is computed as a performance measure. Decision tree (DT*) is used as a predictive algorithm, with maximum depth fixed as 4 for all models.

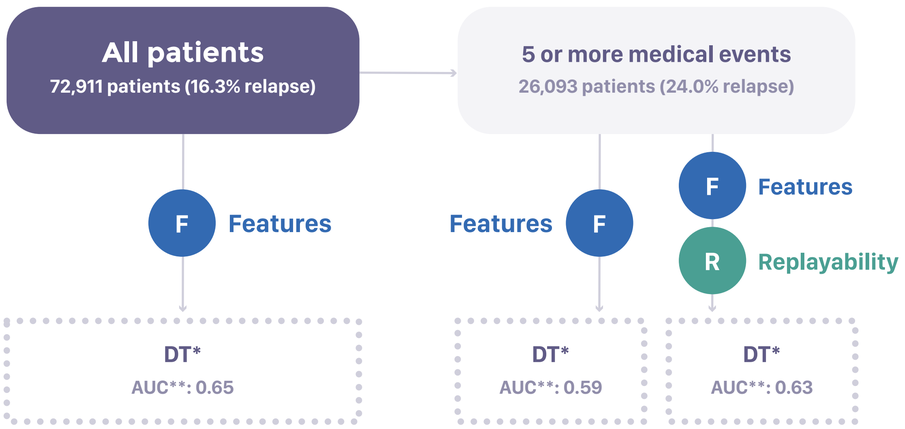


Fig 2. Schematic representation of the experiment

Results

Results show that for patients with 5 or more medical events, the addition of replayability within features increases performances in term of AUC** (from 0.59 to 0.63).

Moreover, replayability takes an important place in decision tree construction as the first split of the tree is performed using replayability (R > 0.003) **(fig 3.)**.

The result of optimization is a process model which shows extracted patterns of sepsis relapse within the medical history of patients. Thus, the visualization of such a process model provides insights to aid understanding of the risks of sepsis relapse. As a result, Leukemias, NHL*** or Neoplasms appear as particular events in medical history which are highly correlated with sepsis relapse **(fig 4.)**.

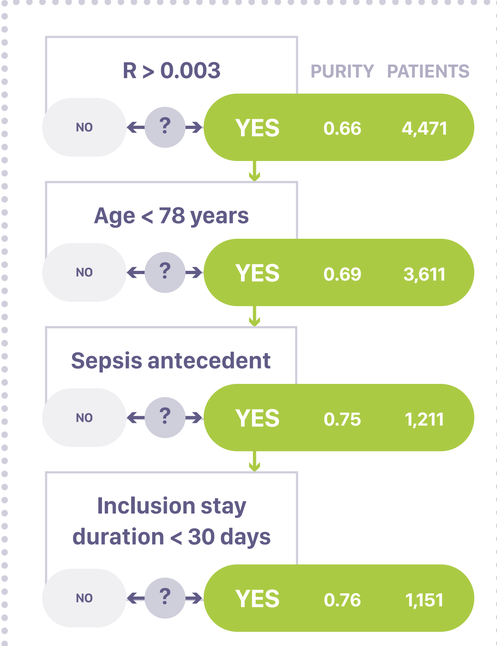


Fig 3. First 4 conditions of decision tree in favor of sepsis relapse, obtained for patients with 5 or more medical events (features + replayability)

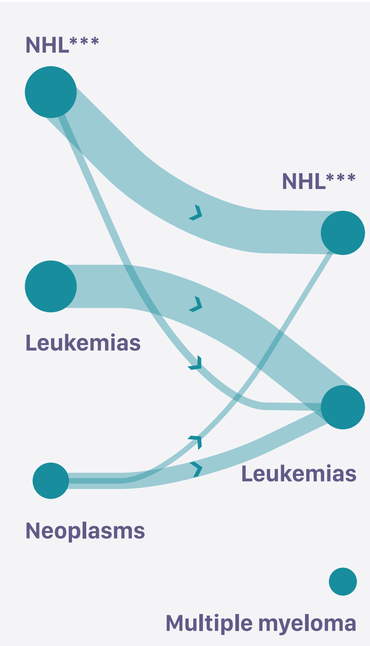


Fig 4. Resulting process model: pattern of sepsis relapse risk extracted from patients' medical history

Conclusion

Across this work, a methodology of pattern extraction from event log data (medical history) using process mining is presented. This method has been applied on a study case using NHS data to improve sepsis relapse prediction. Moreover, the obtained process model highlights some particular medical events within patients' history which will impact sepsis relapse risk.
The presented study is a proof of concept. Performances are not sufficient to be used in routine, but as the use of replayability increases performances, the methodology is encouraging. Future work will be focused on working with more precise data in order to improve performances and develop a tool to be used in practice, to identify at an early stage patients with a risk of sepsis relapse.

HEVA
IMPERIAL COLLEGE HEALTH PARTNERS

References
* DT: Decision Tree
**AUC: Area Under the Curve
*** NHL: Non-Hodgkin lymphoma

[1] Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 2016;315:801–10.
[2] M. Prodel, V. Augusto, B. Jouaneton, L. Lamarsalle, and X. Xie. Optimal process mining for large and complex event logs. IEEE Transactions on Automation Science and Engineering, 2018.