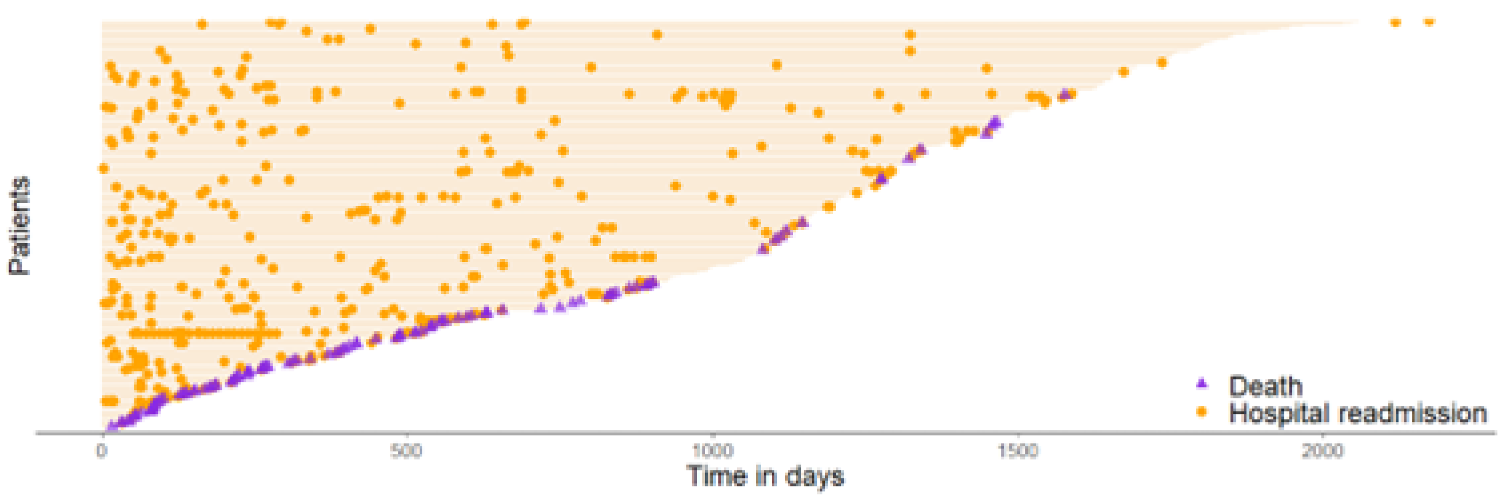
Predicting Hospital Readmission after Cancer Surgery with Survival Analysis and Machine Learning

AliH 2024 : Artificial Intelligence in Healthcare
4 | 6 SEPTEMBER

Juliette Murris*, Sandrine Katsahian†, Audrey Lavenu‡

* HeKA, Inserm, Inria, Université Paris Cité, Pierre Fabre R&D, †Centre d'Investigation Clinique 1418 Épidémiologie Clinique, Paris, France ‡ Institut de Recherche Mathématique de Rennes (IRMAR), Rennes, France

Background, data and objectives



- Readmission dataset from the frailtypack R package,
- Multiple rehospitalizations after surgery,
- 403 patients diagnosed with colorectal cancer,
- In average, there were 1.13 hospital readmissions per patients, with 199 patients with no admission and a total of 106 deaths.

Available options within a survival framework

- Time-to-first event (either readmission or death)
- Time-to-reccurence, with or without death

The advent of machine learning

- Usual machine learning algorithms have been extended to account for survival data
- But not to account for survival data and recurrent events, with or without a terminal event.

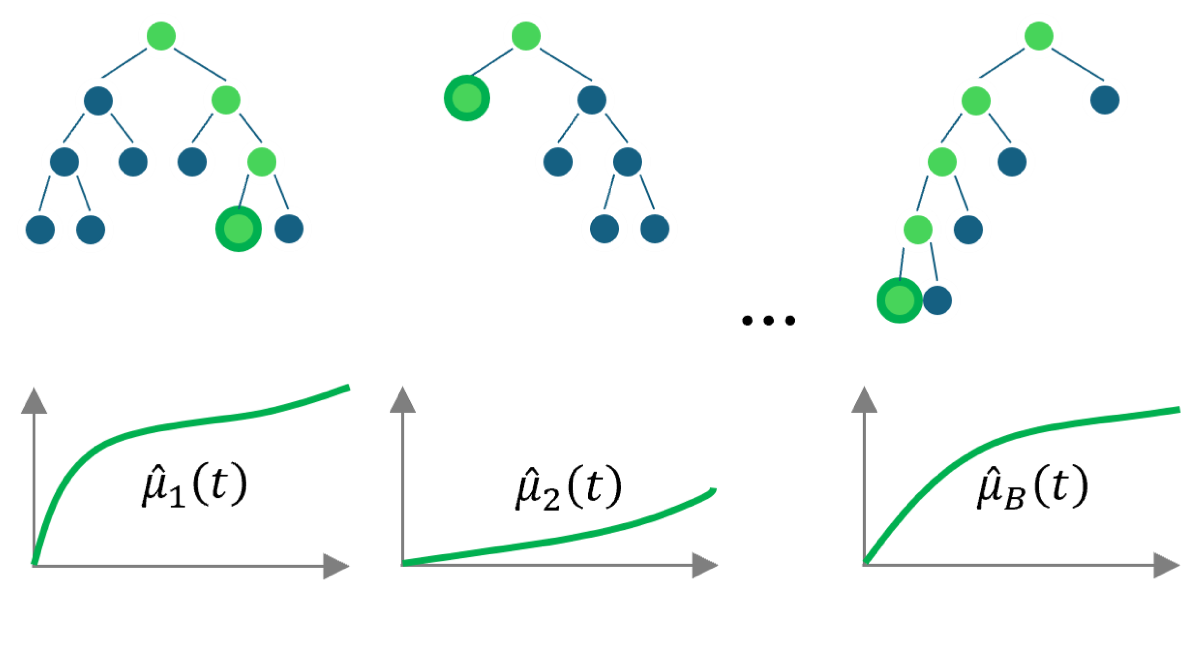
Objectives

- Introduce a **new approach** to model recurrent events using **ensemble methods**
- Application on hospital readmission after cancer surgery

METHODS

RecForest Algorithm

RecForest Algorithm	Without a terminal event	With a terminal event
(1) Draw B **bootstrap** samples from the learning data;		
(2) Grow a **survival tree** b extended to recurrent events;		
Splitting rule	Maximize the test statistic	
At each node, $mtry$ predictors are randomly selected with $mtry \in \mathbb{N}$	Pseudo score test from NP estimates	Wald test from Ghosh-Lin model
Terminal node estimator for tree b	$\hat{\mu}_b(t\|\mathbf{x}) = \hat{R}_b(t\|\mathbf{x}) = \int_0^t \frac{N_b(\mathrm{d}u\|\mathbf{x})}{Y_b(\mathrm{d}u\|\mathbf{x})}$	$\hat{\mu}_b(t\|\mathbf{x}) = \int_0^t \hat{S}_b(u\|\mathbf{x})\mathrm{d}\hat{R}_b(u\|\mathbf{x})$
Pruning strategy	A minimal number of events and/or a minimal number of individuals	
(3) **Estimate** $\hat{\mathbb{M}}$ is computed over the B trees.		



$$\hat{\mathbb{M}}(t|\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B}\hat{\mu}_b(t|\mathbf{x})$$

Results

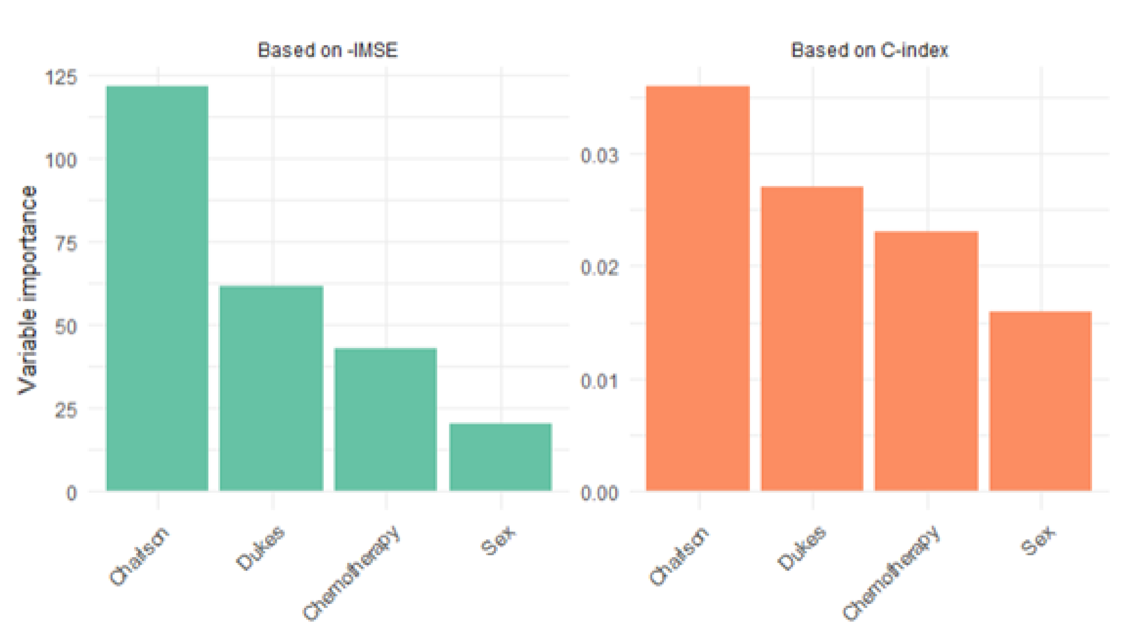
1. Performances, using adapted versions of C-index and MSE

Metric	Np	GL1	GL2	GL3	GL4	RecForest	GL*
C-index ↑	0.58 (0.05)	0.53 (0.08)	0.48 (0.08)	0.48 (0.07)	0.45 (0.05)	**0.80 (0.04)**	0.60 (0.06)
IMSE ↓	7 883.50 (6 229.47)	7 843.99 (6 106.36)	8 361.16 (6 292.29)	8 229.08 (6 478.35)	9 981.50 (6 064.23)	**706.02 (508.96)**	7 934.28 (6 606.23)

- The non-parametric estimator registers a C-index = 0.58.
- RecForest outperforms with C-index = 0.80.
- All GL models with one to four covariates for adjustment, maintain relatively consistent C-indices around 0.45 to 0.53.
- IMSE for RecForest indicate lower margin of errors.

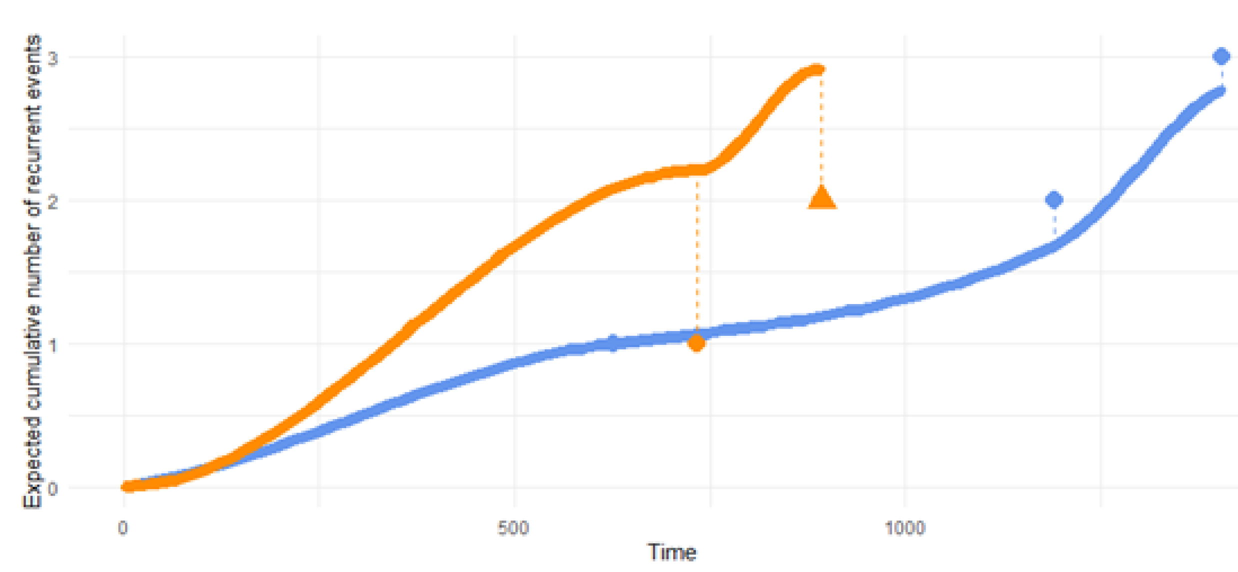
- Variable importance for RecForest was based on both the C-index and the opposite of the integrated MSE.
- Most important variable identified by RecForest was the Charlson comorbidity index.

2. Variable importance to measure impact on predictions



Factors are sex (M/F), chemotherapy treatment (Yes/No), Dukes tumoral stage (with levels A-B, C, and D), and comorbidity Charlson's index (with levels 0, 1-2, and ≥ 3).

3. Predictions for new data



- We build prediction curves for RecForest as the expected number of recurrent events.
- We focus on 2 patients :
  - one with the highest Charlson comorbidity score (in orange), the model predicted 3 readmissions as the patient dies after two observed readmissions.
  - and the other with the lowest Charlson comorbidity score (in blue), the patient in blue, the model predictions are in line with observed events.

DISCUSSION & CONCLUSION

- Our approach is **simple** and easily **accessible** in order to resolve high-dimensional problems involving recurrent events.
- Our algorithm benefits from random forests features (ability of handling missing data or multicollinearity, reducing overfitting with bagging principle).

RecForest is a **valuable contribution** for analysing **recurrent events** in **medical** research

BIBLIOGRAPHY

Andrews DF, Hertzberg AM (1985)
Bouaziz, O. (2023)
Breiman, L. (2001)
Cook, R. J., & Lawless, J. (2007)
Devaux, A, et. Al (2023)
Feurer, M., & Hutter, F. (2019)
Harrell Jr, F. E., Lee, K. L., & Mark, D. B. (1996)
Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009)
Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008)
Kaplan, E. L., & Meier, P. (1958)
Kim, S., Schaubel, D. E., & McCullough, K. P. (2018)
Kvamme, H., & Borgan, Ø. (2019)
Murris, J., Charles-Nelson, A., Lavenu, A., & Katsahian, S. (2022)
Nelson, W. B. (2003)
Therneau, T., Grambsch, P., & Fleming, T. (1990)