Time-to-Event Interpretable Machine Learning for Multiple Sclerosis Worsening Prediction: Results from iDPP@CLEF 2023

Angela Lombardi , Maria Luigia Natalia De Bonis , Giuseppe Fasano, Alessia Sportelli, Tommaso Colafiglio, Domenico Lofù, Paolo Sorino, Fedelucio Narducci, Eugenio Di Sciascio, Tommaso Di Noia

CLEF 2023 Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 18th to 21st, 2023.

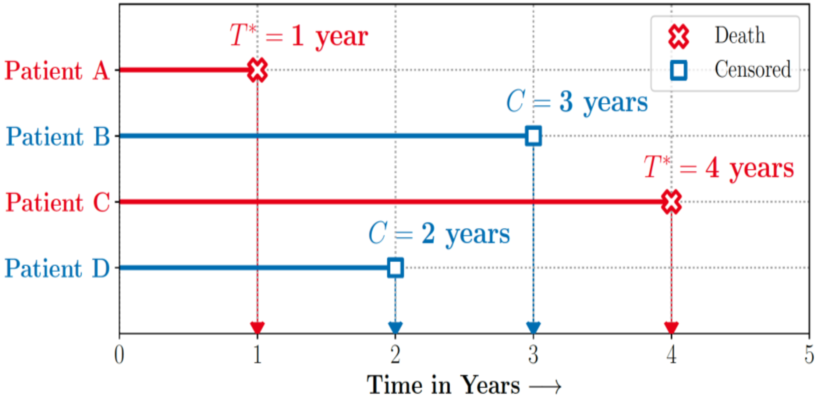
Outline

- Motivations
- Materials
- Algorithms
- Research questions
- Results
- Conclusion

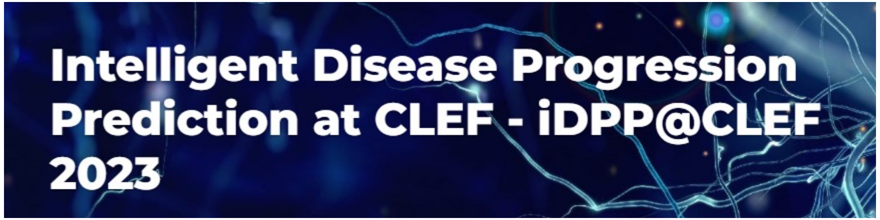
Motivations

Time-to-event Machine Learning Models

- Time-to-event machine learning models have shown greater performance than traditional parametric approaches for disease progression prediction tasks.
- Such models are capable of **handling high-dimensional data**, **nonlinear relationships**, and **interactions between predictors**.
- Machine learning survival models **can handle censoring**, i.e., the event of interest has not yet occurred at the end of the study.
- They can also incorporate **time-varying predictors**, allowing for dynamic predictions and accounting for changes in the predictors over time.



Goals



Brainteaser

Demographic and clinical characteristics of about 1800 subjects

↓

Human-Interpretable ML framework

Explore the potential of **different time-to-event ML approaches** for MS disease progression prediction

Highlights different aspects related to the **interpretability of the results** such as possible bias and the impact of clinical predictors on the achieved performance

Materials

Original Dataset

Static features about each patient with information on:

- age, sex, and others related to the onset of the disease such as the presence of certain symptoms, age of onset and the medical center;

Dynamic features:

- information about the relative start date of relapses;
- tests on evoked potentials;
- information on the areas on which MRIs have been performed and the observed lesions;
- information about the MS course;
- the relative date when EDSS scores were measured, together with the EDSS scores evaluated by clinicians;

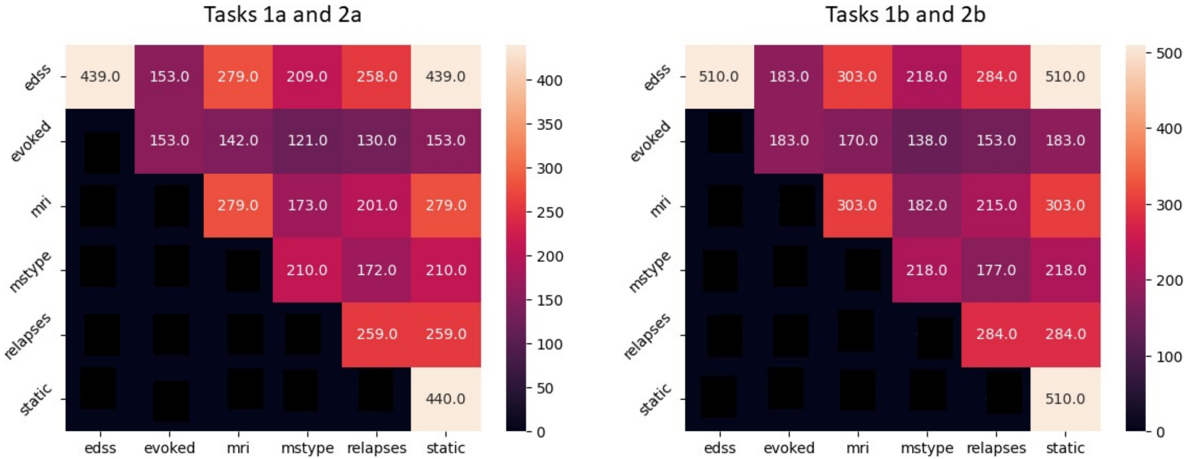
Outcomes:

- containing the patients' worsening occurrence, together with the time of occurrence.

Derived Dataset

Compare the number of subjects with static information and outcomes with respect to all the other dynamic information

Selecting for all tasks only the static features, relapses and MRI information



Derived Dataset

Features on lesions extracted by using MRI images were also excluded due to the high number of missing data

Derive clinical history of the EDSS scores by using a landmarks-style approach obtaining 16 temporal points for each patient

$EDSS\_t_0$	$EDSS\_t_1$		$\dots$	$EDSS\_t_{15}$	
$EDSS\_t_0$	$EDSS\_t_1$		$\dots$		0

Derived Dataset

Transform the information about the relapses for each patient i, into an additional relative frequency feature

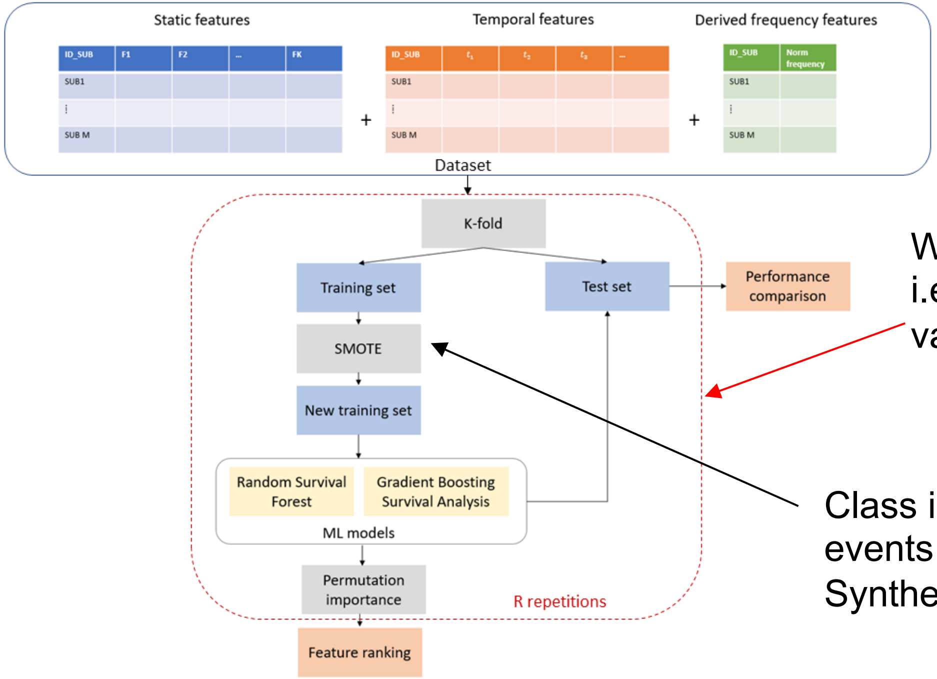
Impute missing values for each variable by using the median values.
Exclude variable "centre"

$$variation\_relapses_i = \frac{n\_relapses}{(t_{end} - t_1)},$$

$n\_relapses$: is the number of relapses between the first recorded time and the last recorded time

Algorithms

Proposed framework



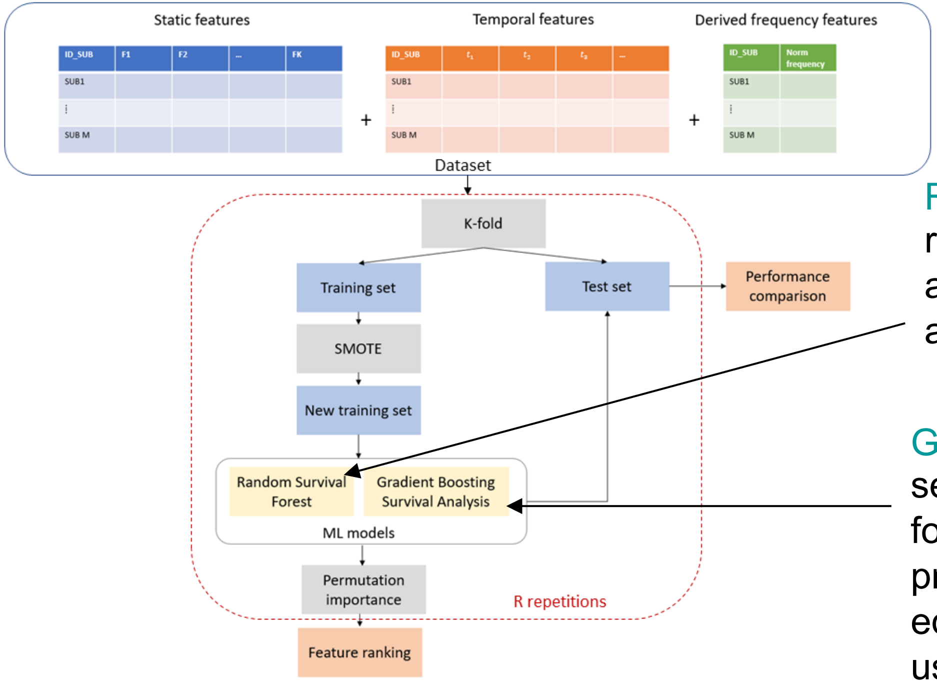
Validation scheme

We adopted a repeated k-fold cross-validation scheme, i.e., a resampling technique that combines k-fold cross-validation with repetition

Class imbalance

Class imbalance between censored and not-censored events oversampling the minority class by means of the Synthetic Minority Oversampling TEchnique (SMOTE)

Proposed framework

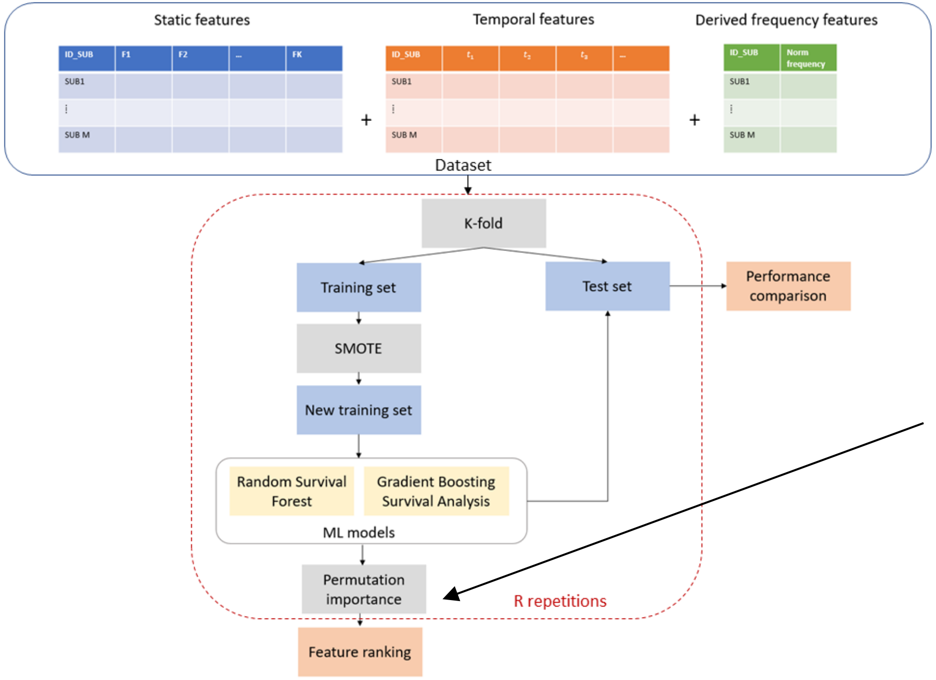


Time-to-Event Machine Learning models

Random Survival Forests (RSF) are an extension of random forests specifically designed for survival analysis. They combine the principles of survival analysis with the concept of decision trees.

Gradient Boosting (GB) Machines that create a sequence of weak learners, which are combined to form a strong learner. Boosting starts with initializing predictions for each sample. Initially, all samples have equal weights. Boosting builds a series of weak models, usually decision trees, in an iterative manner.

Proposed framework



Permutation feature importance

To identify which features have the most impact on the model's performance and provide insights into the underlying relationships between the features and the target variable.

Research questions

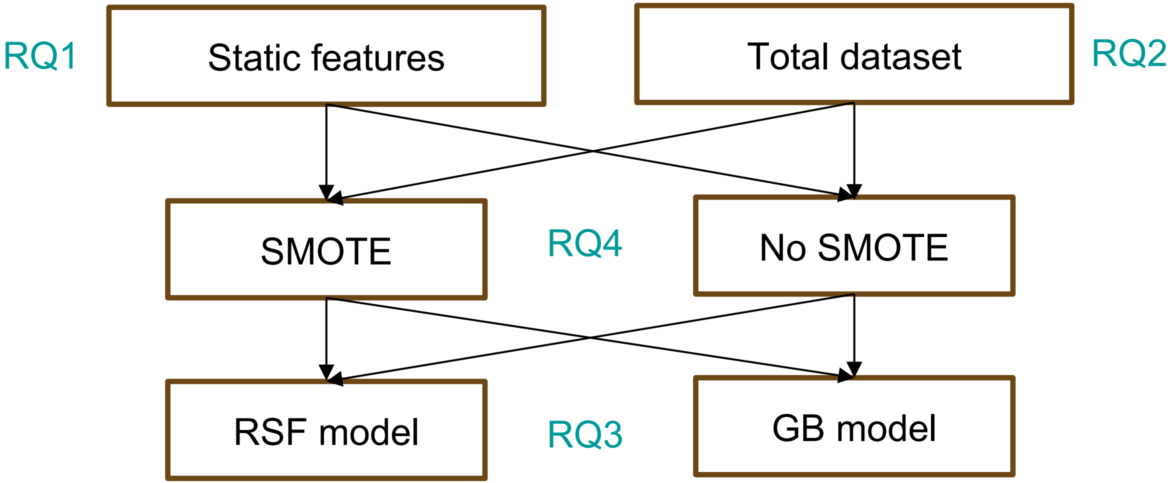
Experimental setup

We organized the setup of the experiments to address the following research questions:

- RQ1 What is the impact of the static features on the model performance for the two datasets A and B?
- RQ2 What is the additional contribution of the dynamic features?
- RQ3 Is there a significant ranking of all the features for both definitions of MS worsening?
- RQ4 Does the imbalance of the types of events ('censored' and non-censored) affect the performance of the models?

Experimental setup

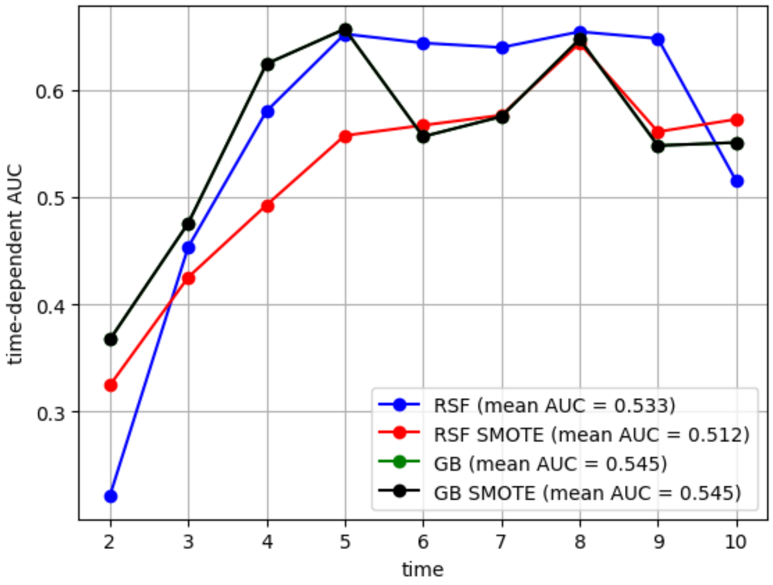
Trained 16 ML models (8 for each task):



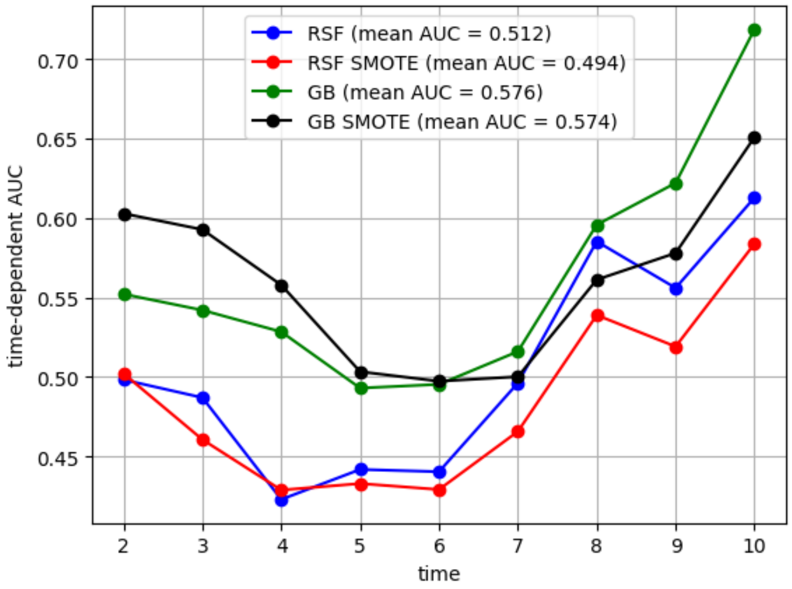
Results

RQ1: impact of the static features on the performance

Dataset A

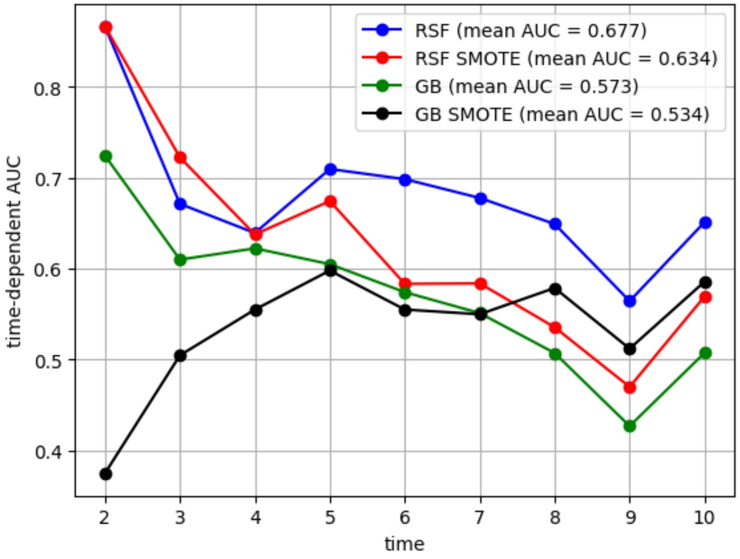


Dataset B

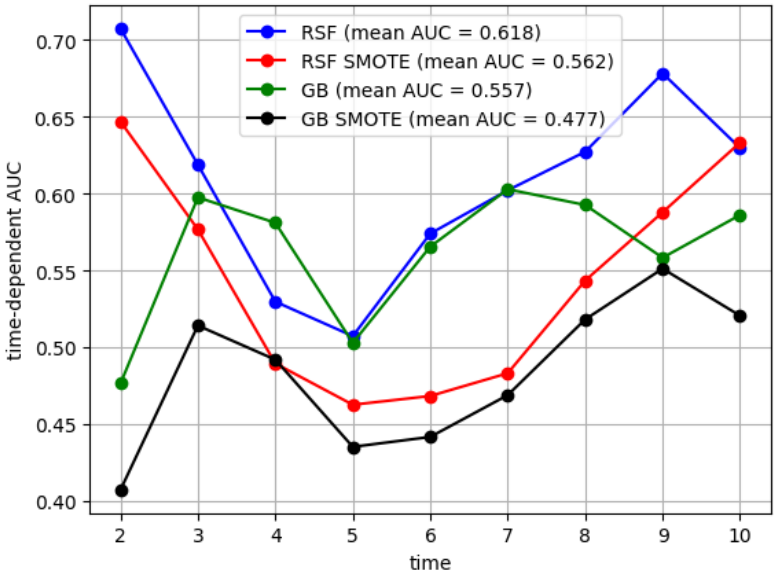


RQ2: the additional contribution of the dynamic features
RQ4 effect of class imbalance on the performance

Dataset A

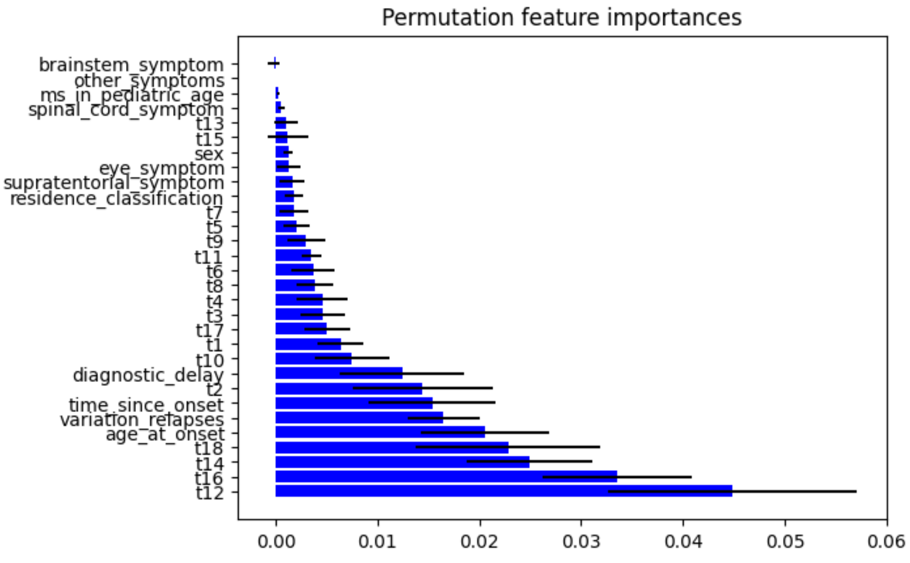


Dataset B

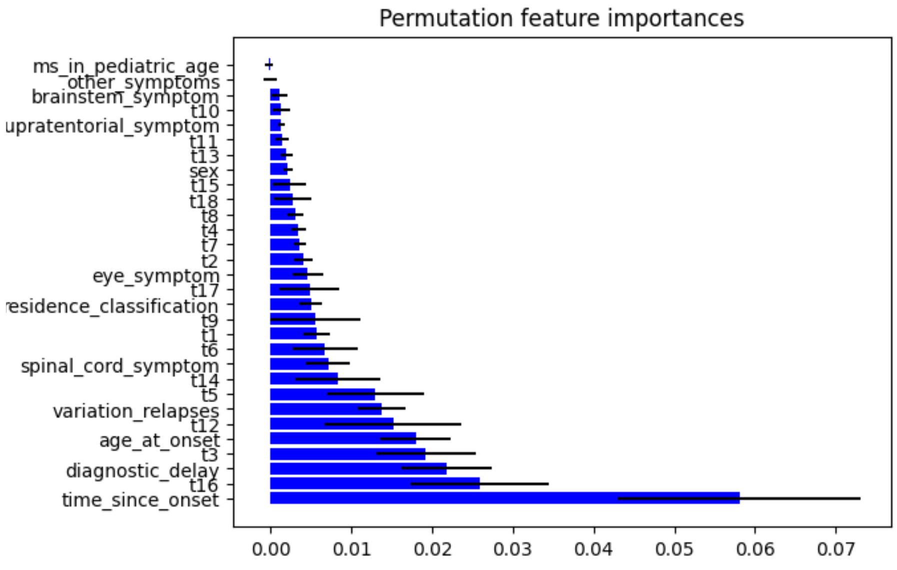


RQ3: ranking of the features for both tasks

Dataset A



Dataset B



Conclusion

Findings

- The proposed framework is modular and allows to investigate **the link between the provided static and dynamic features and the outcome to be predicted**.

- Our findings show that for a more complex and condition-dependent definition of worsening (tasks 1b and 2b) significantly lower results are obtained than those obtained with a simpler definition of worsening (tasks 1a and 2a).

Future work

- Only two ML methods with a **landmark approach** were considered: **different algorithms** that can automatically **model time series of different lengths** will be exhaustively explored in future work.

- We **excluded information about the patients' center**: future developments will involve the use of **site harmonization algorithms** prior to time-to-event analysis to remove potential bias related to this variable.

Thank you!

For further questions and information, don’t hesitate to contact us!

Angela Lombardi	angela.lombardi@poliba.it
Maria Luigia Natalia De Bonis	m.debonis4@studenti.poliba.it
Giuseppe Fasano	g.fasano8@studenti.poliba.it
Alessia Sportelli	a.sportelli3@studenti.poliba.it
Tommaso Colafiglio	tommaso.colafiglio@poliba.it
Domenico Lofù	domenico.lofu@poliba.it
Paolo Sorino	paolo.sorino@poliba.it
Fedelucio Narducci	fedelucio.narducci@poliba.it
Eugenio Di Sciascio	eugenio.disciascio@poliba.it
Tommaso Di Noia	tommaso.dinoia@poliba.it