Predicting the risk of & time to impairment for ALS patients

Intelligent Disease Progression Prediction Lab @ CLEF 2022

Aidan MANNION, Thierry CHEVALIER, Didier SCHWAB, Lorraine GOEURIOT

Overview

- Tasks:
  1. rank subjects suffering from ALS based on the risk of the occurrence of adverse events (T1)
  2. predict the time window of occurrence for the most-likely event (T2)
- Our approach:
  - Survival analysis modelling to estimate risk of different events
  - Regression models to estimate time to event
  - Combination of risk scores & time predictions to calculate time-window predictions
  - Both prediction aspects based on ensemble learning via gradient boosted trees

Modelling & Problem Formulation

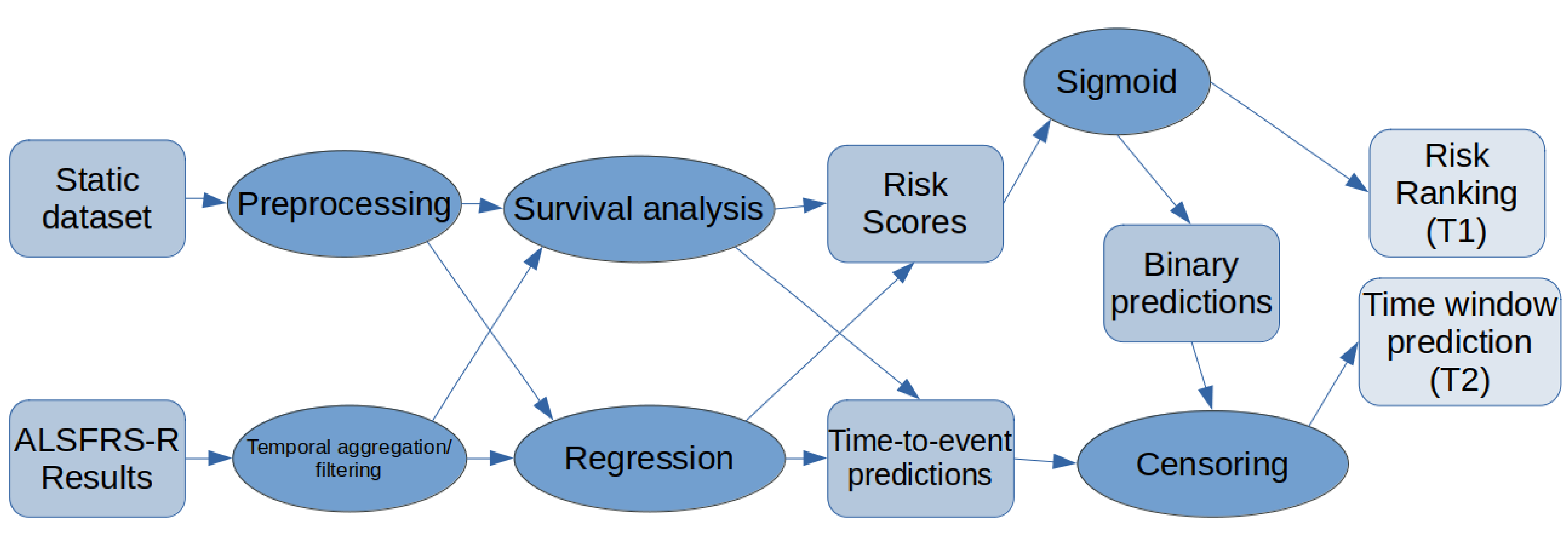


Figure: A general overview of the prediction architecture

Training Data

+ Two modalities;
  + Static dataset: patient demographic data & details of medical history recorded at the outset of treatment; (age at onset, weight, etc.); mainly composed of numerical quantities & binary indicators – 94 variables total
  + Temporal visit dataset: series of ALSFRS-R scores (12 integers 1-4) or spirometry data (1 float)

Exploratory Data Analysis: Static Dataset

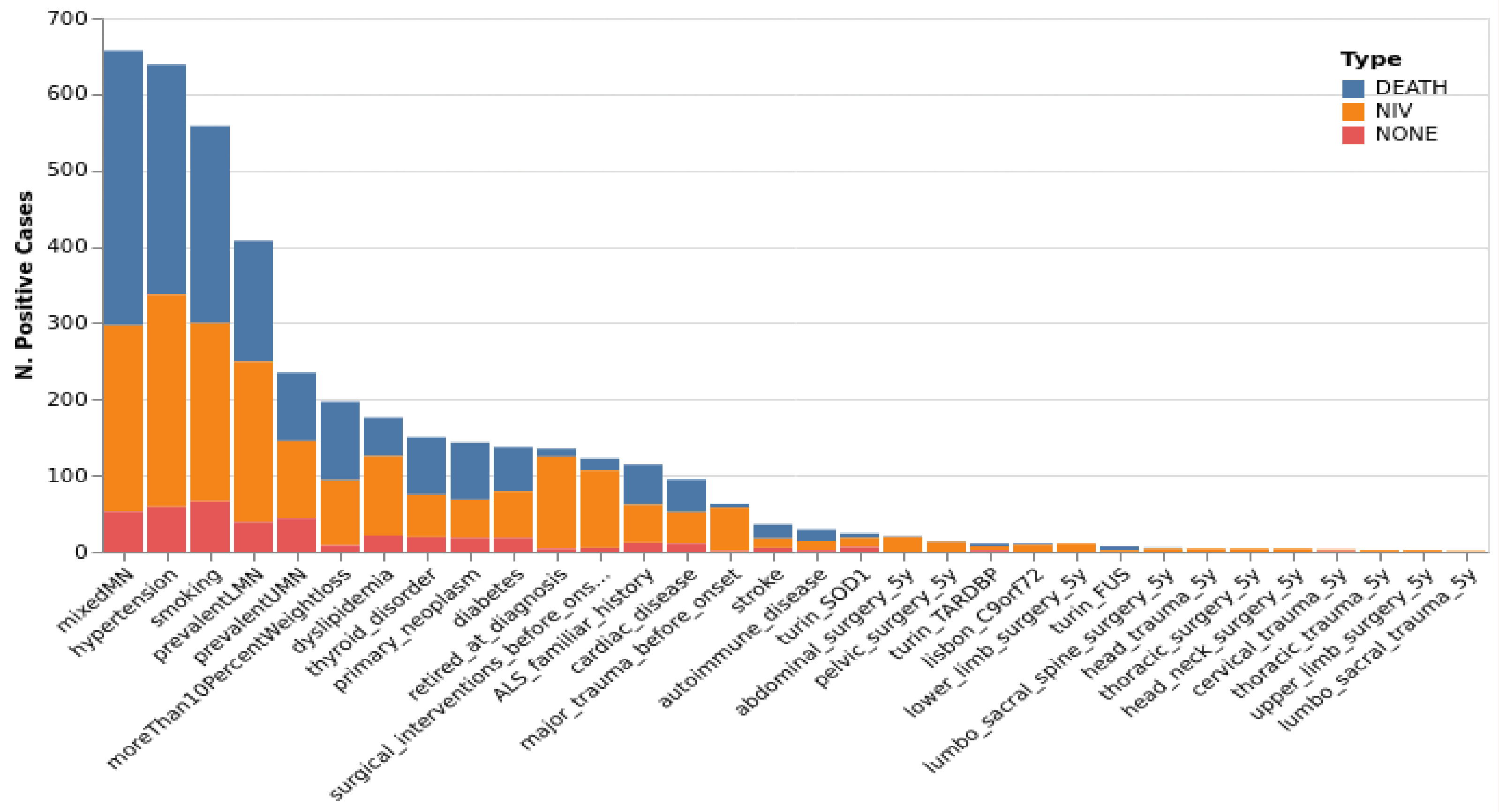


Figure: number of positive cases of each **binary** indicator variable in the static dataset A *(n=1454).*

Feature Selection: Static Dataset

- Based on variance, correlation with target outcome variables, & feature importances calculated via classifiers trained on entire dataset, 27 variables were selected for inclusion in the training data
- Weight loss following diagnosis, major-trauma history & retirement status seem to be strong predictors, based on initial analysis

Feature Engineering: Static Dataset

+ Imputation of missing values for weight (before & after ALS onset) and height – using MIDAS encoder-decoder technique (Lall & Robinson 2021)
+ Aggregation of these 3 fields to get the patient's change in BMI to use as a feature

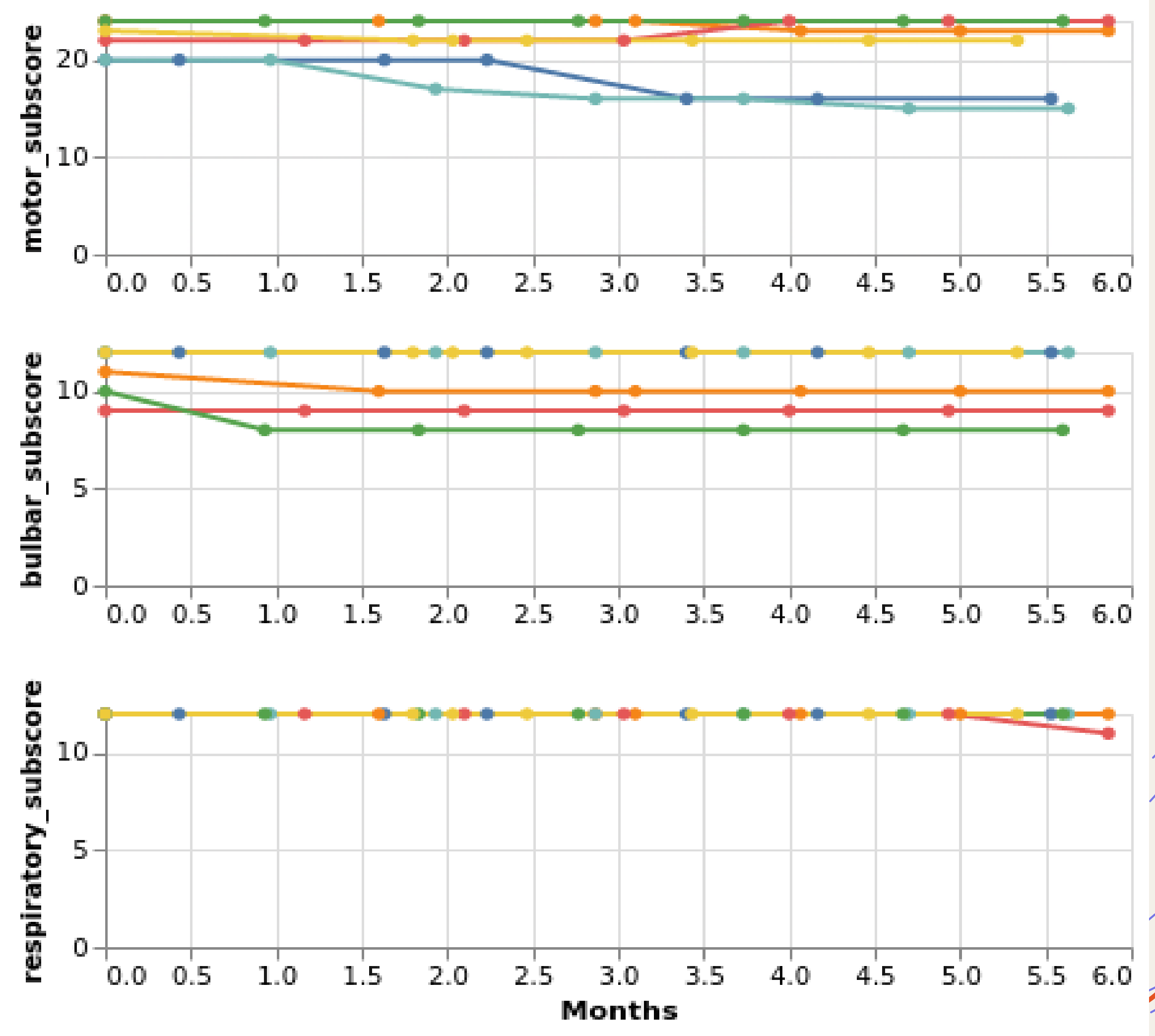
Exploratory Data Analysis: Visit Dataset

+ Very little variation in measurements over time => prediction at M6 shouldn't show much improvement over prediction at M0
+ Spirometry data was found to have a negligible effect on predictive performance and was excluded

Exploratory Data Analysis: Visit Dataset

Right: The evolution over time of the ALSFRS-R scores for the "top 6" patients, i.e. the only ones with 7 visits recorded in the dataset.

- Changes are rare & small
- Most of the rest of the patients (with <7 visits) follow a similar pattern
- Difficult to infer disease progression.



Feature Engineering: Visit Dataset

+ We experimented with four different levels of granularity for the ALSFRS-R responses;
  1. Each question individually: d=12 (gave best predictive performance)
  2. Bulbar, fine motor, gross motor, and respiratory subscores: d=4
  3. Bulbar, motor, and respiratory subscores: d=3
  4. Total score: d=1
+ M6 risk calculation: compared four methods of temporal aggregation;
  1. Average
  2. Max.
  3. Sum
  4. Temporally-weighted average (used for submission)

Learning Algorithm

+ Survival analysis with gradient-boosted decision trees as base (weak) learners; gradient boosting algorithm fit to the Cox proportional hazard loss function:

$$H(t|x) = \exp(f_\theta(x))H_0(t)$$

Experiment Setup

- Hyperparameter search for learning rate, # estimators & maximal depth of regression trees
- 5-fold cross-validation
- Evaluation metrics (parameter selection criteria);
  - Task 1: C-index (maximised)
  - Task 2: mean absolute error (minimised)

Ranking Patients by Risk of Impairment

1. Generate time-independent risk scores (log-hazard ratio) for each outcome using decision-tree ensembles trained on all data (competing events handled independently by different models)
2. Associate each patient with their highest relevant risk score and map the scores to pseudo-probabilities via a sigmoid function
3. Choose classification threshold *c* using the ROC curve from training – label subjects with probability <*c* with NONE in place of the adverse event
4. Sort the full dataset under consideration according to the risk scores.

ROC Curve

Selecting the classification threshold based on this curve for each relevant target variable allowed us to optimise the trade-off between the rates of true & false predictions, thus improving the placement of the cut-off point for "NONE" in the ranking

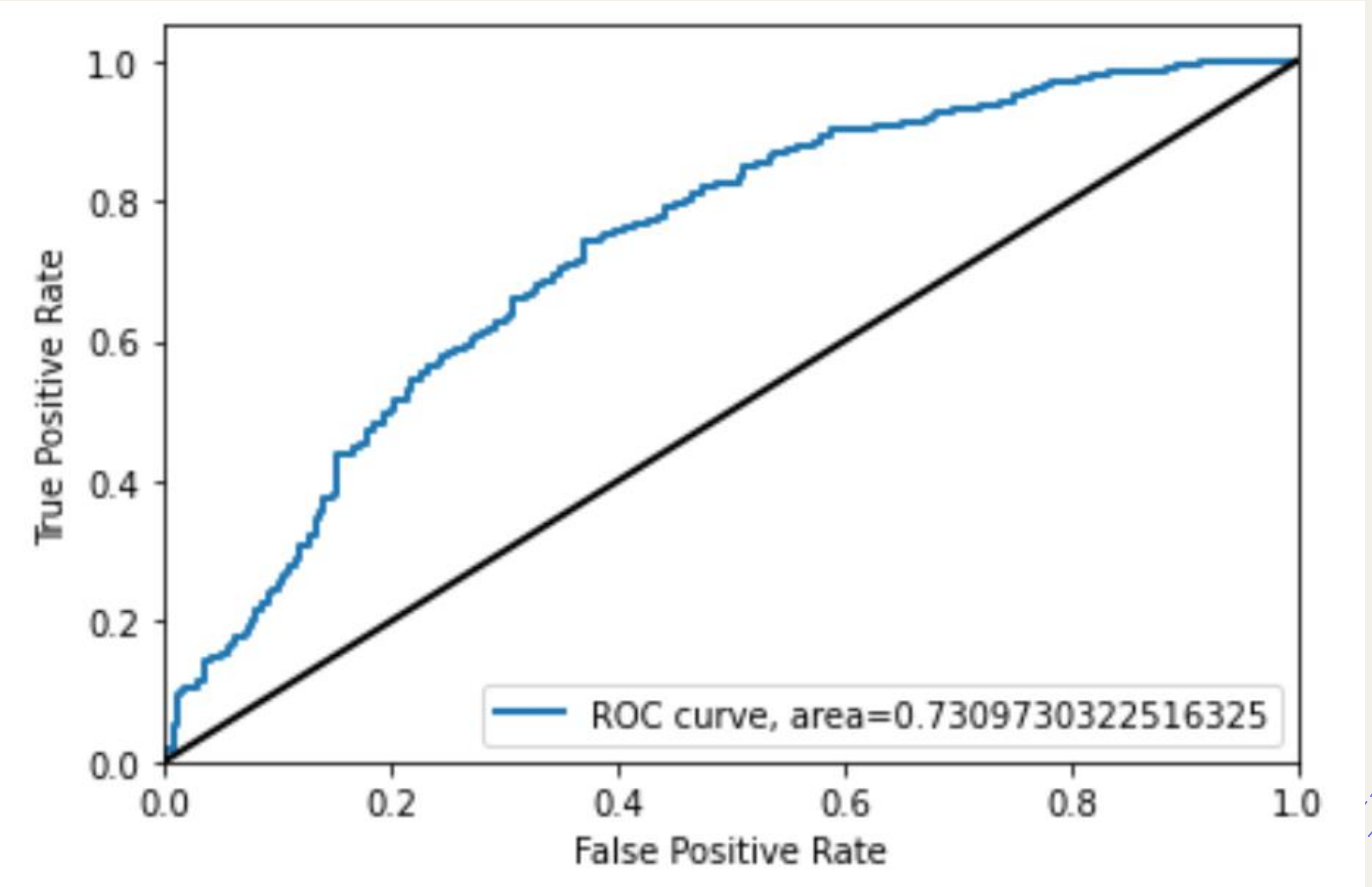


Figure: Receiver Operating Characteristic curve for the survival model trained on dataset C at month 6

Predicting Time of Impairment

- Gradient-boosted regression: decision-tree ensemble applied to regularised MSE loss function
- Standard regression model trained directly on the "Time" variable => does not account for censoring
- Occurrence predictions from survival analysis used to "censor" the regression output

Results

+ General theme: somewhat encouraging AUROC scores for Task 1 & specificity scores for Task 2, poor recall in Task 2 (particularly for time windows 24-30 & 30-36 months)
+ Best results for Task 2b @ month 0
+ Performance @M6 usually slightly better than @M0
+ Hyperparameters: "wide & shallow" approach gave better results (large tree ensembles with low maximum depth)

Task 1: Area under ROC

Task	Train	Dev	Test	Best
T1a_M0	0.848	0.645	0.760	0.842
T1a_M6	0.856	0.658	0.802	0.867
T1b_M0	0.856	0.639	0.795	0.870
T1b_M6	0.855	0.641	0.811	0.877
T1c_M0	0.910	0.666	0.767	0.866
T1c_M6	0.920	0.682	0.793	0.871

Task 2: Specificity

Task	Train	Dev	Test	Best
T2a_M0	0.812	0.798	0.854	0.864
T2a_M6	0.782	0.663	0.850	0.876
T2b_M0	0.812	0.628	0.865	0.865*
T2b_M6	0.763	0.647	0.865	0.872
T2c_M0	0.817	0.684	0.851	0.863
T2c_M6	0.820	0.652	0.864	0.866

Conclusions

- Gradient-boosting-based survival analysis seems promising as a method for ranking ALS patients according to risk
  - Performance not yet satisfactory for a real clinical setting
  - Proportional hazards assumption: potentially an oversimplification?
- Time-to-event prediction is a more challenging problem
  - More detailed visit data/longer observation period potentially required
  - More sophisticated temporal prediction should be used