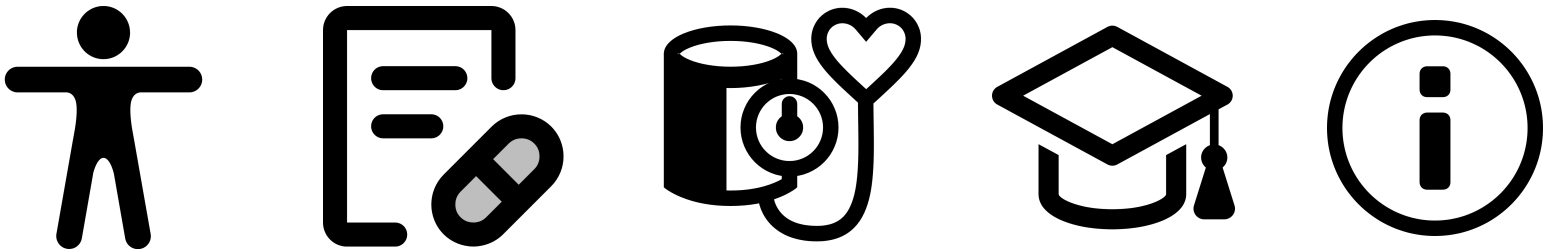
Predicting survivability of waiting heart transplant patients

Intro to UNOS Data Set



Preparing the Data Set

- Choosing relevant features (available at t=0)
- Unwanted data - 18+, duplicate entries, uninteresting patient outcomes
- Outlier removal
- Imputation of missing data

Preparing the Data Set

- Normalization
- Encoding - One-hot
- Undersampling
- Other feature engineering efforts

Metric - F1 Macro Score

$$\text{precision} = \frac{\text{true positives}}{\text{positives}}$$

$$\text{recall} = \frac{\text{true positives}}{\text{true posititives} + \text{false negatives}}$$

$$\texttt{f1_score} = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$

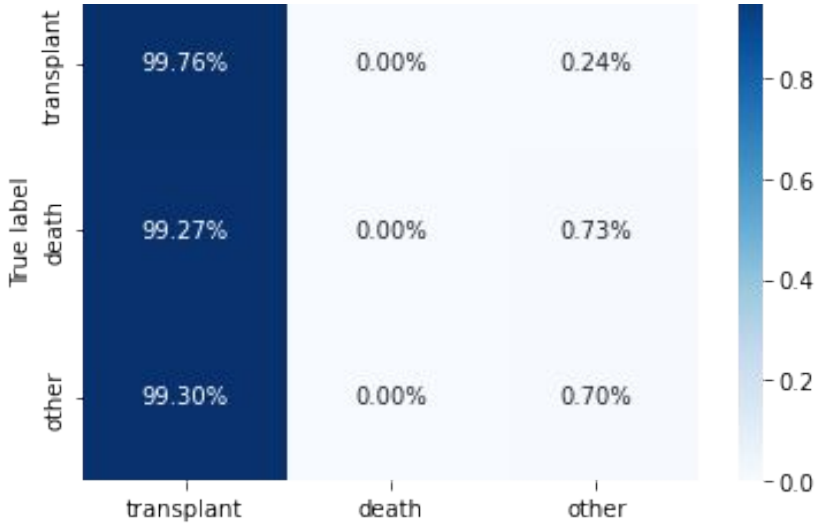
$$\texttt{f1_macro} = \texttt{avg}(\texttt{f1_score})$$

First (very poor) model

Logistic Regression

	precision	recall	f1-score	support
Transplanted	0.66	1.00	0.79	37905
Died	0.00	0.00	0.00	7690
Other	0.36	0.01	0.02	12181

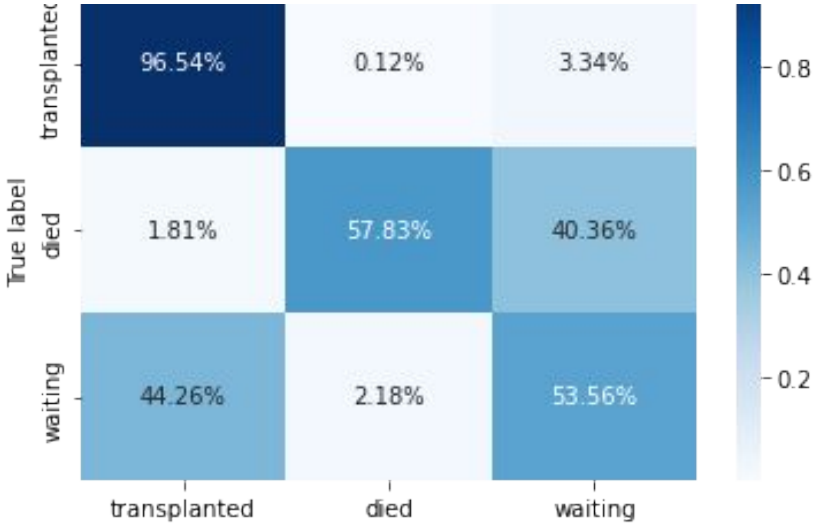
F1 macro: 0,27



Waitlist Model Single Layer Neural Network

	precision	recall	f1-score	support
Transplanted	0.89	0.97	0.92	13219
Died	0.79	0.58	0.67	607
Still Waiting	0.74	0.54	0.62	3667

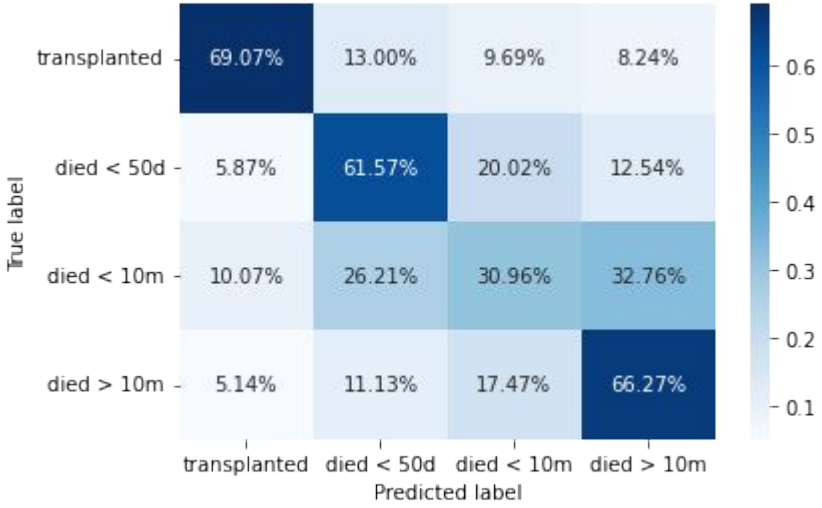
F1 macro: 0,74



Death Interval Model Logistic Regression

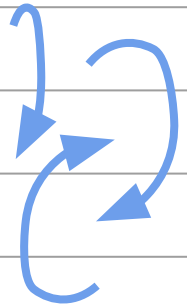
	precision	recall	f1-score	support
Transplanted	0.78	0.69	0.73	1177
Died < 50d	0.55	0.62	0.58	1124
Died < 10m	0.38	0.31	0.34	1053
Died > 10m	0.57	0.66	0.61	1168

F1 macro: 0,57

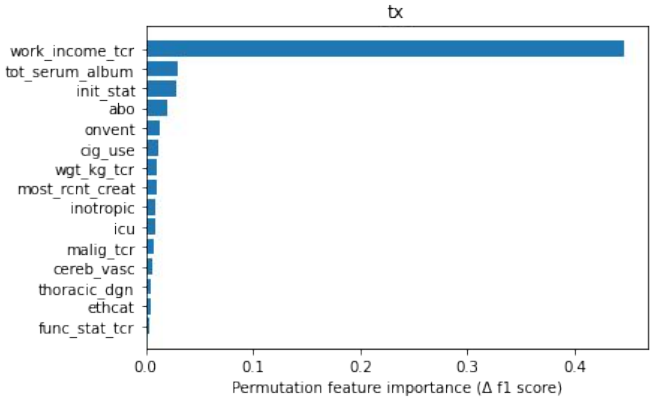


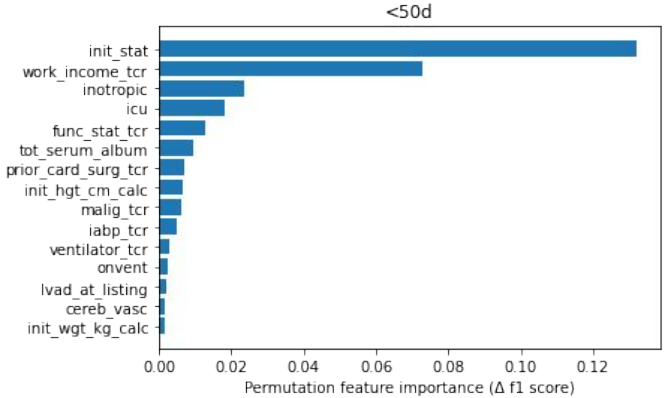
Permutation Feature Importance

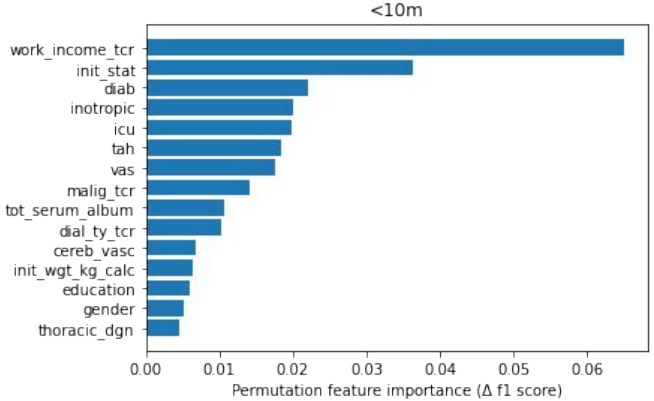
Weight (kg)	In Intensive Care?	…	# Bananas peeled (lifetime)
61	No	…	280
…	…	…	…
125	Yes	…	48
57	No	…	43

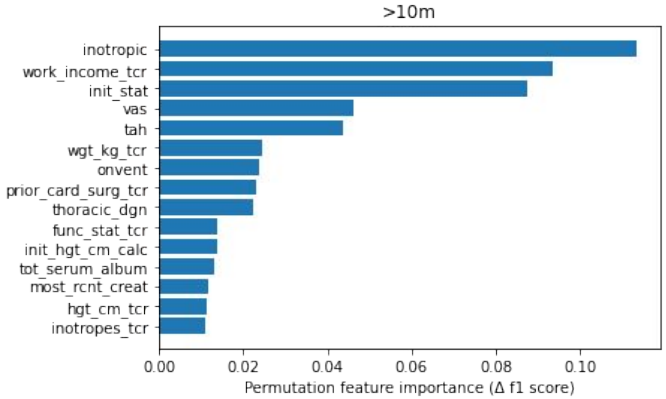


Permutation Feature Importance









Thank you