Universitätsmedizin GREIFSWALD

University Medical Center Groningen

PhysioNet CinC Challenge 2019 — Early Prediction of Sepsis from Clinical Data

Marcus Vollmer[1], Christian F. Luz[2], Philipp Sodmann[1], Bhanu Sinha[2], Sven-Olaf Kuhn[3]

[1] Institute of Bioinformatics, University Medicine Greifswald, Germany
[2] Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, The Netherlands
[3] Department of Anesthesiology and Intensive Care Medicine, University Medicine Greifswald, Germany

11 September 2019, CinC Singapore

Content

1. The challenge – the data

Challenge description

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

20th PhysioNet Computing in Cardiology Challenge – project page

Objective

"The goal [...] is the **early detection of sepsis using physiological data**. [...] we define sepsis according to the Sepsis-3 guidelines, i.e., a two-point change in the patient's Sequential Organ Failure Assessment (SOFA) score and clinical suspicion of infection [...]."

- Designing and implementation of a working, open-source algorithm
- Automatically identify a patient's risk of sepsis and make a positive or negative prediction of sepsis for every time interval based only on the clinical data
- Team with best predictions for the patients in the hidden test set wins

Challenge Data

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

Training dataset

40,336 patient ids | 1,552,210 dates | 2,932 sepsis patients

Vital signs	HR, O2Sat, Temp, SBP, MAP, DBP, Resp, EtCO2
Laboratory values	BaseExcess, HCO3, FiO2, pH, PaCO2, SaO2, AST, BUN, Alkalinephos, Calcium, Chloride, Creatinine, Bilirubin_direct, Glucose, Lactate, Magnesium, Phosphate, Potassium, Bilirubin_total, TroponinI, Hct, Hgb, PTT, WBC, Fibrinogen, Platelets
Demographics	Age, Gender, Unit1 (MICU), Unit2 (SICU), HospAdmTime, ICULOS
Outcome	SepsisLabel For sepsis patients, SepsisLabel is 1 if $t \geq t_{\text{sepsis}}-6$ and 0 if $t < t_{\text{sepsis}}-6$. For non-sepsis patients, SepsisLabel is 0.

Data Insight

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

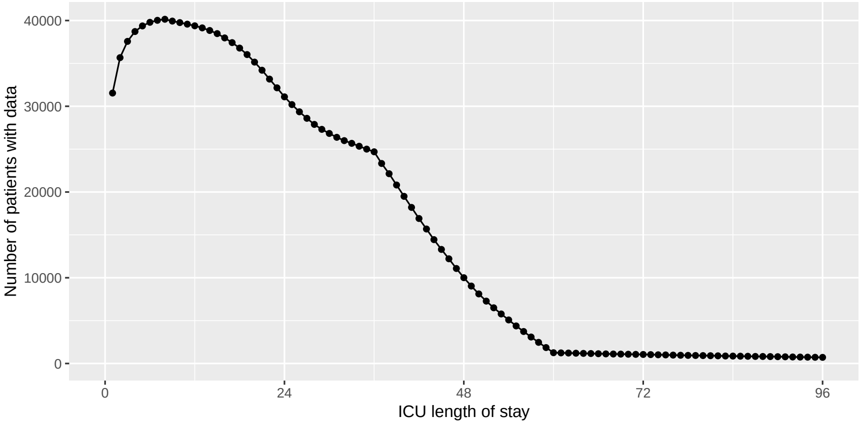
Training files are available as delimitered csv files:

HR	O2Sat	Temp	...	HospAdmTime	ICULOS	SepsisLabel
NaN	NaN	NaN	...	-50	1	0
86	98	NaN	...	-50	2	0
75	NaN	NaN	...	-50	3	1
99	100	35.5	...	-50	4	1

For data screening to manually revisit false predictions we implemented an interactive Sepsis Challenge Patient Explorer ▸ Sepsis Challenge Patient Explorer with R Shiny.

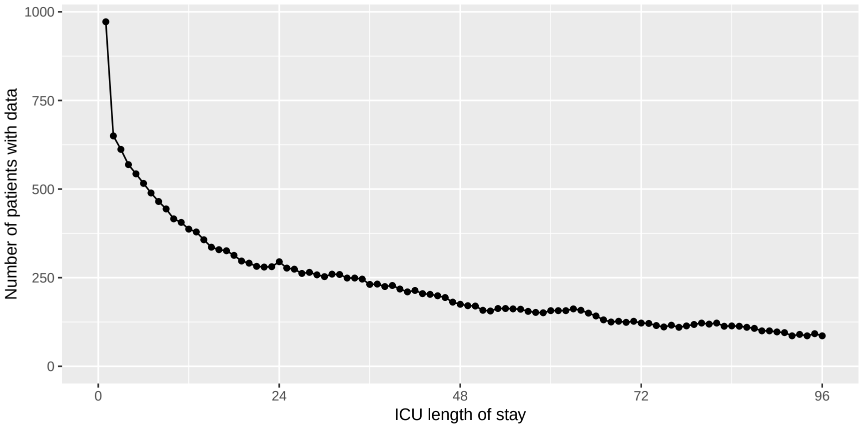
ICU Length of Stay

All patients



ICU Length of Stay

Sepsis patients

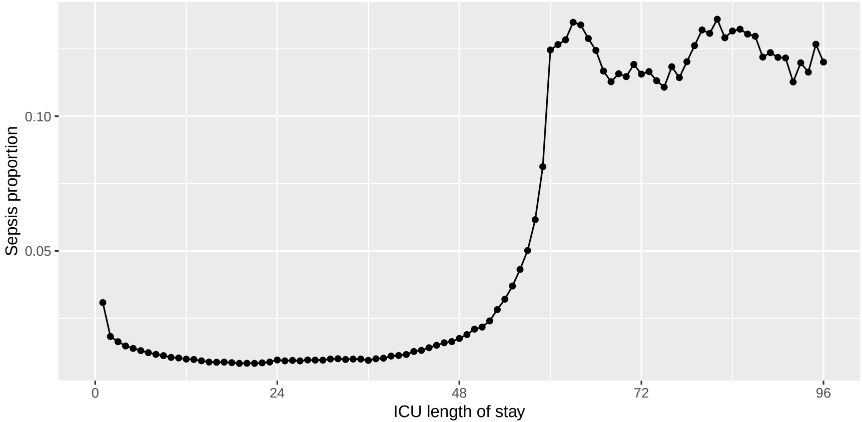


ICU Length of Stay

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

Sepsis ratio



Challenge Scoring

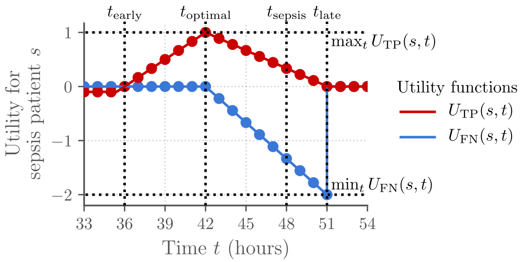
The early prediction of sepsis is potentially life-saving, and we challenge participants to **predict sepsis 6 hours before** the clinical prediction of sepsis. **The utility function rewards early predictions and penalizes late predictions as well as false alarms**. Late prediction of sepsis is potentially life-threatening. Predicting sepsis in non-sepsis patients (or predicting sepsis very early in sepsis patients) consumes limited hospital resources.

Challenge Scoring

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

The early prediction of sepsis is potentially life-saving, and we challenge participants to **predict sepsis 6 hours before** the clinical prediction of sepsis. **The utility function rewards early predictions and penalizes late predictions as well as false alarms**. Late prediction of sepsis is potentially life-threatening. Predicting sepsis in non-sepsis patients (or predicting sepsis very early in sepsis patients) consumes limited hospital resources.

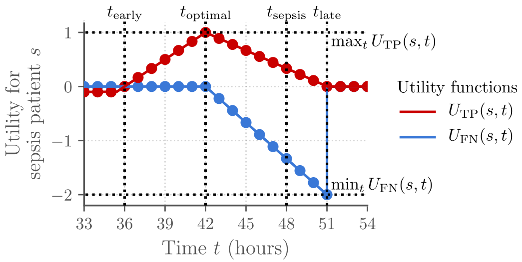


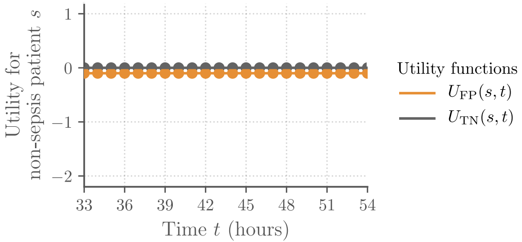
Challenge Scoring

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

The early prediction of sepsis is potentially life-saving, and we challenge participants to **predict sepsis 6 hours before** the clinical prediction of sepsis. **The utility function rewards early predictions and penalizes late predictions as well as false alarms**. Late prediction of sepsis is potentially life-threatening. Predicting sepsis in non-sepsis patients (or predicting sepsis very early in sepsis patients) consumes limited hospital resources.





Challenge Scoring

- Each prediction (each line in the data file) will be scored by the utility function $U(s, t)$ (patient s, time interval t):

$$U(s,t) = \begin{cases} UTP(s,t), & \text{positive prediction at time } t \text{ for sepsis patient } s \\ UFN(s,t), & \text{positive prediction at time } t \text{ for non-sepsis patient } s \\ UFP(s,t), & \text{negative prediction at time } t \text{ for sepsis patient } s \\ UTN(s,t), & \text{negative prediction at time } t \text{ for non-sepsis patient } s \end{cases} \tag{1}$$

- Score for a classifier:

$$U_{\text{total}} = \sum_{s \in S} \sum_{t \in T(s)} U(s,t) \tag{2}$$

- Normalized classifier score:

$$U_{\text{normalized}} = \frac{U_{\text{total}} - U_{\text{no predictions}}}{U_{\text{optimal}} - U_{\text{no predictions}}} \tag{3}$$

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

2. From data cleaning to feature generation

Data Cleaning

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

- Vitals and lab values were screened for physiological plausibility
- 2263 values (mainly within blood pressure variables, respiration rate and oxygen levels) removed

```
clean_vars <- function(x) {
  x %>%
    mutate(hr = if_else(hr > 180, NA_real_, hr),
           o2sat = if_else(o2sat < 50, NA_real_, o2sat),
           temp = if_else(temp < 32 | temp > 43, NA_real_, temp),
           map = if_else(map > 200, NA_real_, map),
           dbp = if_else(dbp > 150, NA_real_, dbp),
           resp = if_else(resp > 50, NA_real_, resp),
           base_excess = if_else(base_excess < -30 | base_excess > 30, NA_real_, base_excess),
           hco3 = if_else(hco3 < 10 | hco3 > 50, NA_real_, hco3),
           fi_o2 = if_else(fi_o2 > 100, NA_real_, fi_o2),
           sa_o2 = if_else(sa_o2 < 50, NA_real_, sa_o2),
           chloride = if_else(chloride < 70 | chloride > 145, NA_real_, chloride),
           potassium = if_else(potassium > 10, NA_real_, potassium),
           hct = if_else(hct < 15 | hct > 70, NA_real_, hct),
           platelets = if_else(platelets > 1000, NA_real_, platelets),
           ventilated = if_else(!is.na(et_co2), TRUE, FALSE))
}
```

Rhythms of Data Availability

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

- Identification of data availability rhythm from **remaining 12,036,860 values**

vars	n0	n1	n2	n3	q0	q1	q2
hr	1398740	87915	8904	6524	0.90	0.96	0.96
o2sat	1349202	94193	13086	9352	0.87	0.93	0.94
temp	525111	56587	49411	160226	0.34	0.37	0.41
sbp	1325945	97290	11865	8200	0.85	0.92	0.92
map	1358496	99527	11949	6842	0.88	0.94	0.95
dbp	1065282	68661	9376	7266	0.69	0.73	0.74
resp	1313516	100718	15486	10159	0.85	0.91	0.92
et_co2	57636	3407	562	367	0.04	0.04	0.04

Rolling windows

- Implementation of rolling windows of 6, 12, 24 and 48 hours for frequently repeated features: heart rate, oxygen saturation, temperature, systolic/diastolic/mean atrial blood pressure, respiration rate and serum glucose
- Compute quantiles, quantile ranges, and differences and quotients to the actual value: Quantiles $(0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95)$ represents the course of a disease without outliers

ICU LOS	1	2	3	4	5	6	7	8	9	10
map_raw		75.3	86.0		91.3		77.0	76.3	88.3	87.3
map_roll.t6.p50		75.3	80.7	80.7	86.0	86.0	81.5	81.5	82.7	87.3
map_roll.t6.p75		75.3	83.3	83.3	88.7	88.7	87.3	87.3	89.1	88.3

Imputation

- **Last observation carried forward** method copies last available lab or vital values to the next dates if new data is missing
- Close to the medical perspective of decision making
  (lab values from blood samples are usually measured unsteadily)

ICU LOS	1	2	3	4	5	6	7	8	9	10
map_raw		75.3	86.0		91.3		77.0	76.3	88.3	87.3
map_roll.t6.p50		75.3	80.7	80.7	86.0	86.0	81.5	81.5	82.7	87.3
map_roll.t6.p75		75.3	83.3	83.3	88.7	88.7	87.3	87.3	89.1	88.3
map_LOCF		75.3	86.0	86.0	91.3	91.3	77.0	76.3	88.3	87.3

Making Missingness Explicit

- Introducing binary variables to indicate whether the values were carry-forwarded
- introducing numerical variables showing the up-to-dateness such that machine learning models are able to learn the relevance of out-dated variables
(0 means newly measured, 6 means measured 6 hours ago)

ICU LOS	1	2	3	4	5	6	7	8	9	10
map_raw		75.3	86.0		91.3		77.0	76.3	88.3	87.3
map_LOCF		75.3	86.0	86.0	91.3	91.3	77.0	76.3	88.3	87.3
map_miss	T	F	F	T	F	T	F	F	F	F
map_miss_val	NA	0	0	1	0	1	0	0	0	0

Clinical Scores

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

- **ShockIndex** (hr/sbp)
- **qSOFA** (sbp and resp)
- **SOFA** and partial SOFA scores (respiration, renal function, platelets, liver function, sofa_renal, sofa_plate, mean arterial pressure), SOFA from worst 24h partial scores
- **SIRS** scores, SIRS criteria, worst 24h SIRS score, SIRS criteria with hard temperature thresholds,
- **NEWS** and partial NEWS scores
  (respiration, oxygen saturation, systolic blood pressure, pulse rate, temperature)
- **cNEWS** uses linear regression
  (gender, age, NEWS, log(resp), temp, log(sbp), log(dpb), log(hr), o2sat, o2support)

Clinical Scores

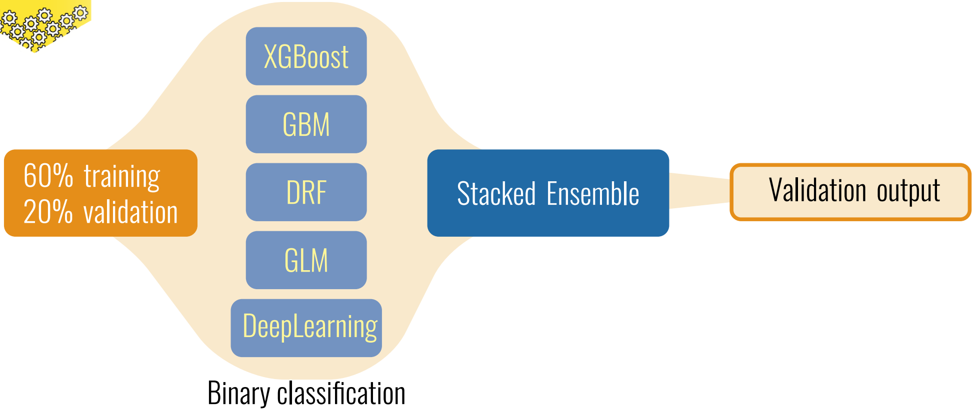
- **Rolling versions using robust measures**
  qSOFA_t6 uses 25% and 75% quantiles of last 6h
  shockIndex_t6 uses 25% and 75% quantiles of last 6h
  SIRS_t24 and partial scores uses 25%, 75% quantiles for temperature and 90% quantiles of last 24h for heart rate and respiratory rate
  NEWS_t6 uses 50% quantiles of respiratory rate, heart rate and systolic bp of last 6h

- **Size of the tibble is now 1,552,210 x 427 $\approx$ 578 MB**

3. Training of time-specific metalearners

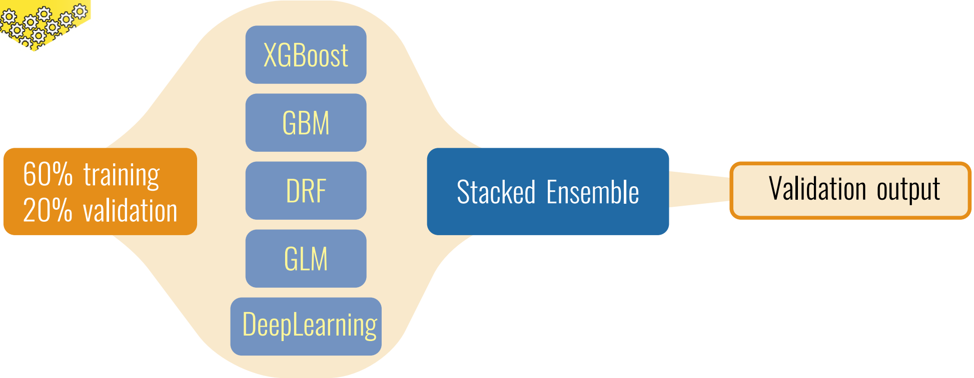
Ensemble Learning with H2O.ai

The H2O software provides a stacked ensemble implementation for Python, R, and Spark



Ensemble Learning with H2O.ai

The optimal H2O parameter setting was identified (stopping metric, feature set, balancing)



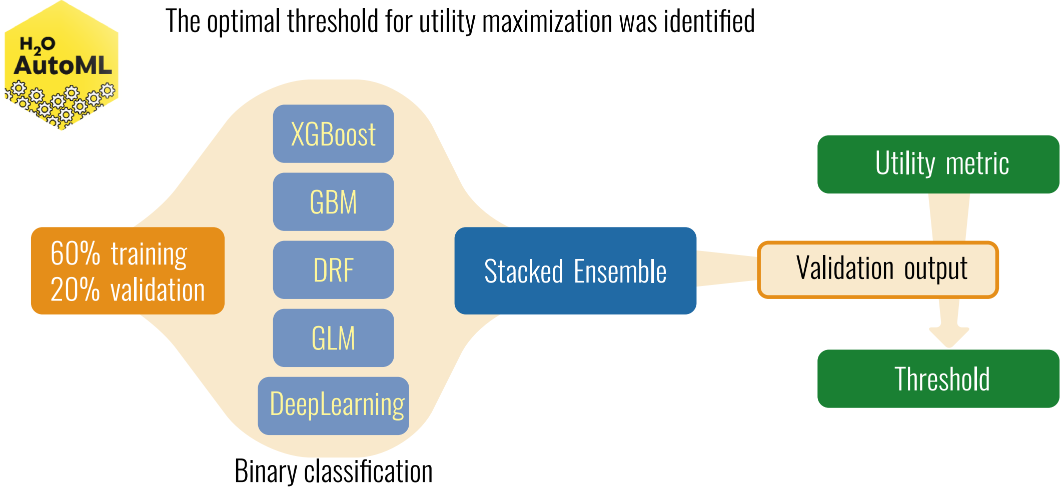
Binary classification, 5-fold cross-validation, logloss stopping metric, class-sampling (.1,2)

Ensemble Learning with H2O.ai

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

The optimal threshold for utility maximization was identified



Ensemble Learning with H2O.ai

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

60% training
20% validation

XGBoost

GBM

DRF

GLM

DeepLearning

Binary classification

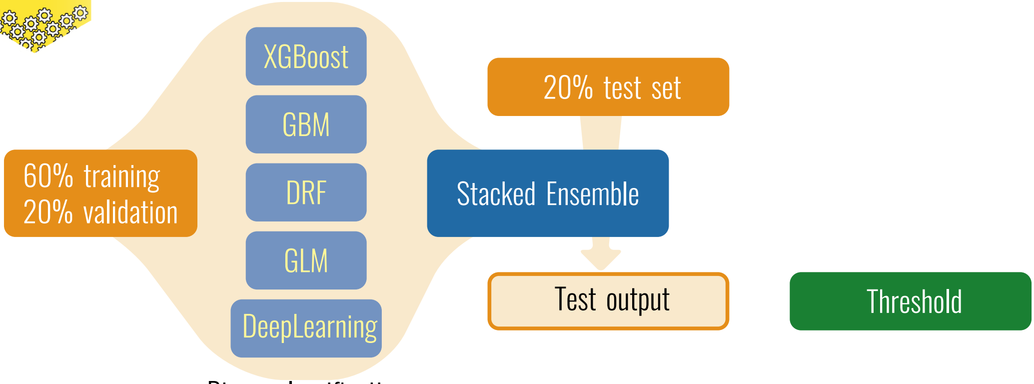
Stacked Ensemble

Threshold

Ensemble Learning with H2O.ai

University Medical Center Groningen

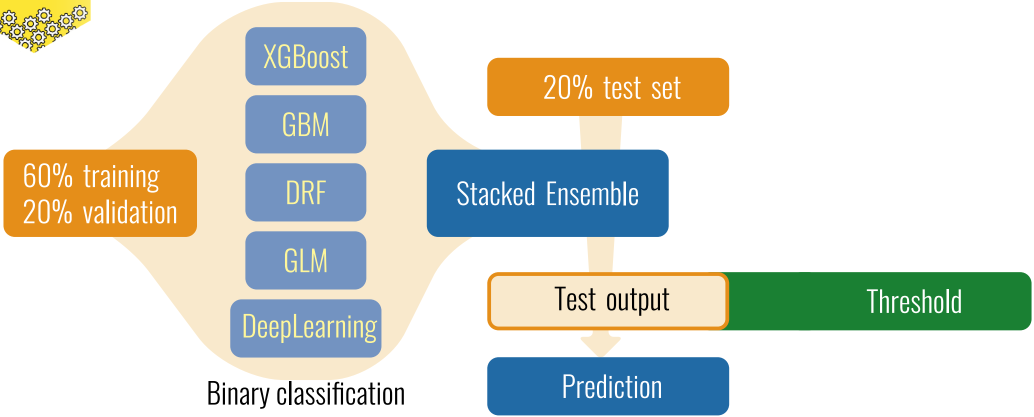
Universitätsmedizin GREIFSWALD



Binary classification

Ensemble Learning with H2O.ai

H2O AutoML



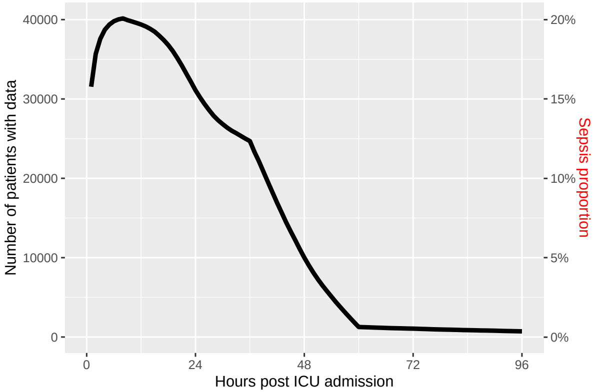
Training Data?

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

Contra using all the data:

- Long-stay patients over-represented
- Redundant data
- Mixing home-acquired, hospital-acquired and ICU-acquired sepsis
- Spectrum of pathogens and source of infection changes
- Long-stay patients are usually subject to a higher exposure to develop sepsis



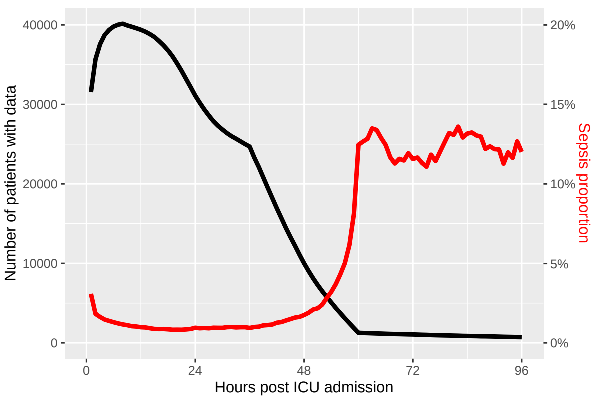
Training Data?

University Medical Center Groningen

Universitätsmedizin GREIFSWALD

Contra using all the data:

- Long-stay patients over-represented
- Redundant data
- Mixing home-acquired, hospital-acquired and ICU-acquired sepsis
- Spectrum of pathogens and source of infection changes
- Long-stay patients are usually subject to a higher exposure to develop sepsis



LOS-specific Metalearner

SE: ICU LOS <6

Trained on admission values (<20 NA)

SE: ICU LOS >48

Trained on data($ICULOS > 48$)

SE: ICU LOS = 6

Trained on data($1 \leq ICULOS \leq 12$)

SE: ICU LOS = 7

Trained on data($1 \leq ICULOS \leq 13$)

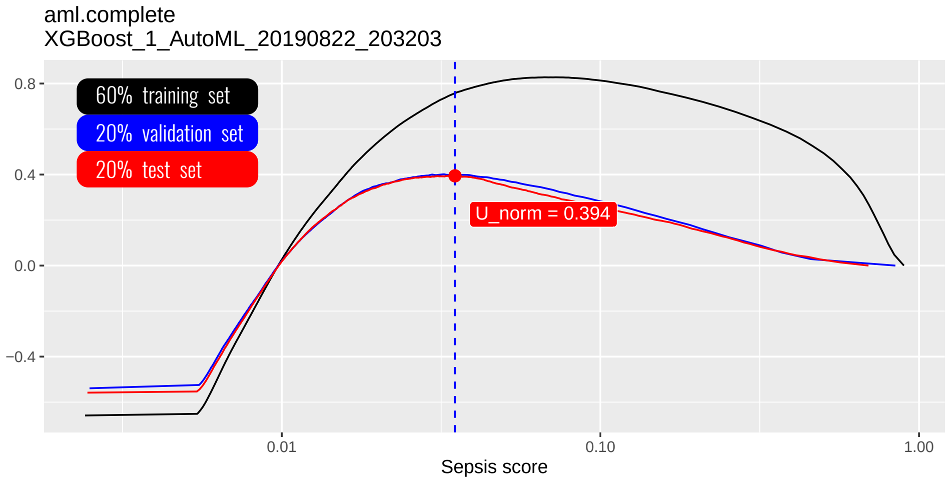
$\vdots$ $\vdots$

SE: ICU LOS = 48

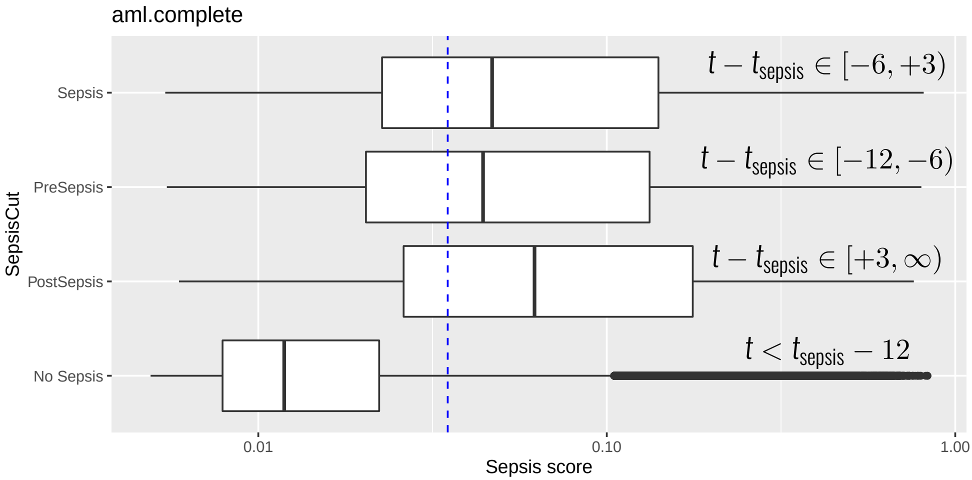
Trained on data($42 \leq ICULOS \leq 54$)

4. Results

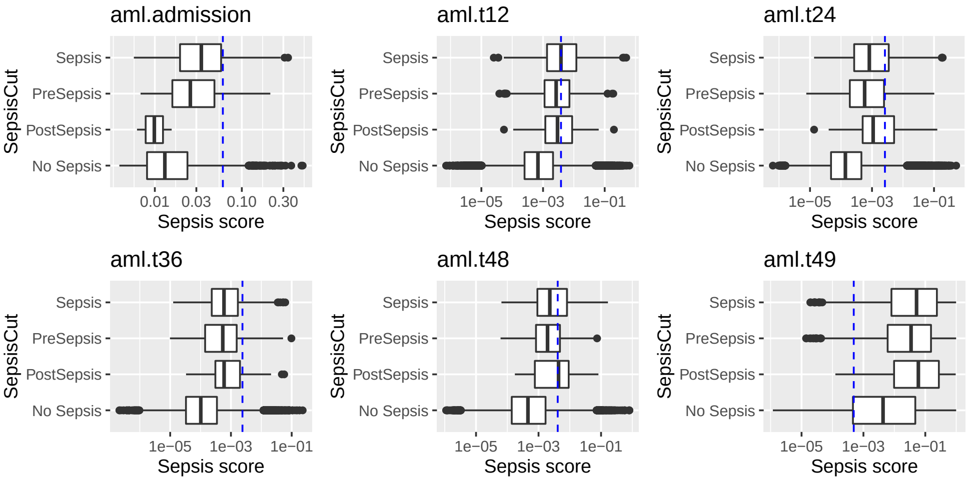
How Stable is the Thresholding?



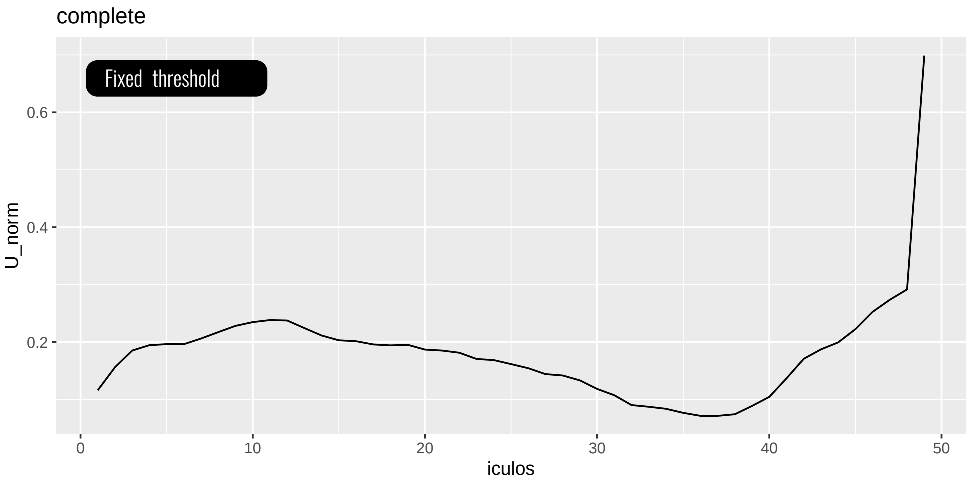
Is the Classification Good?



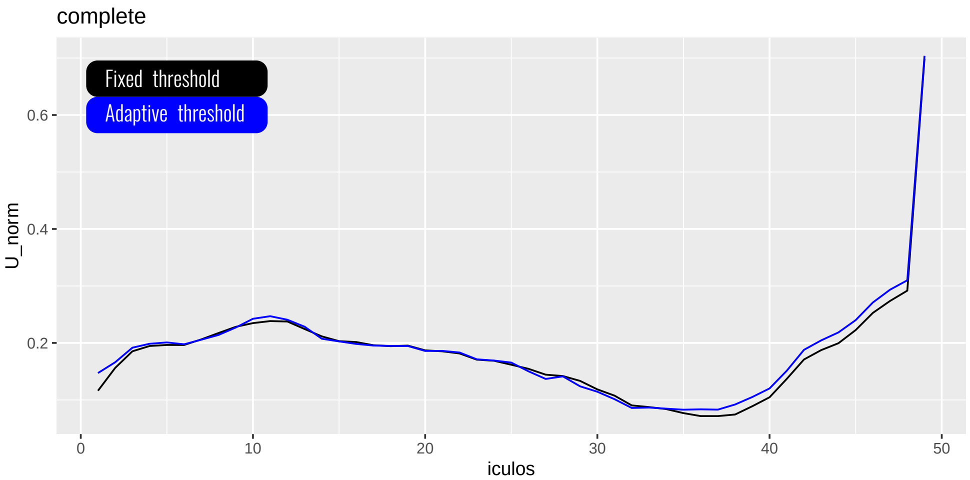
Is the Classification Good?



LOS-specific Utility Scores



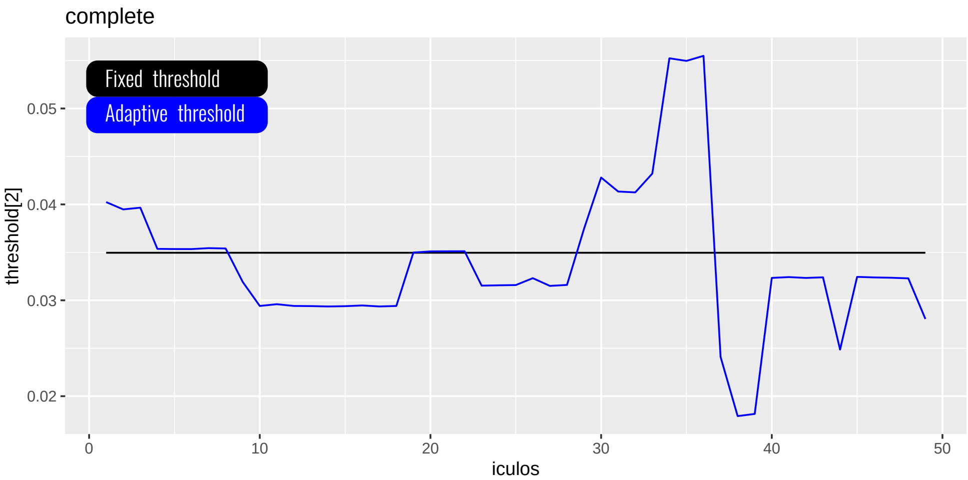
LOS-specific Utility Scores



LOS-specific Utility Scores

University Medical Center Groningen

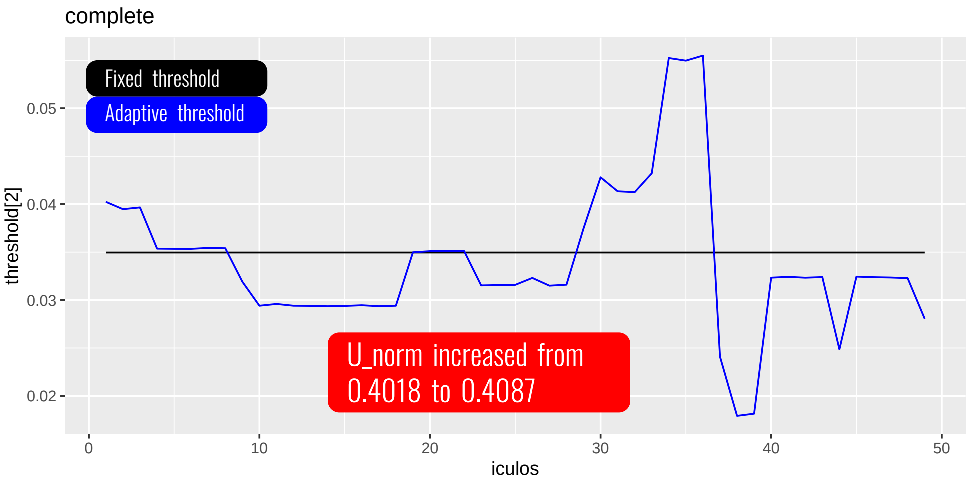
Universitätsmedizin GREIFSWALD



LOS-specific Utility Scores

University Medical Center Groningen

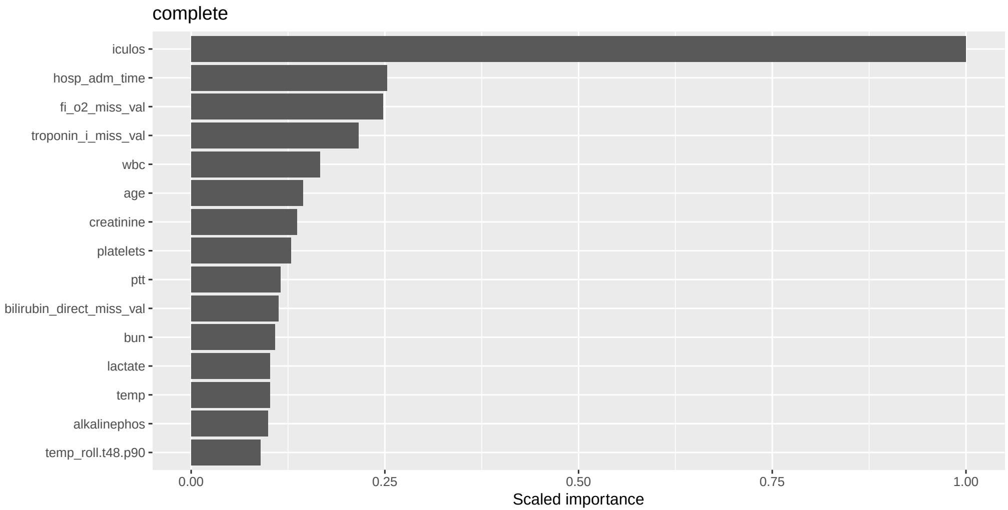
Universitätsmedizin GREIFSWALD



Which Variables are Important?

University Medical Center Groningen

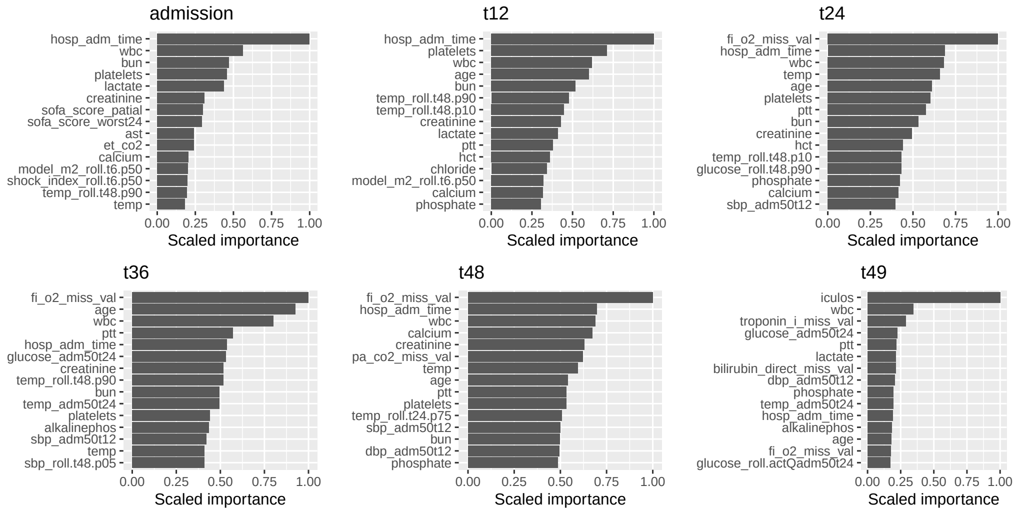
Universitätsmedizin GREIFSWALD



Which Variables are Important?

University Medical Center Groningen

Universitätsmedizin GREIFSWALD



Happy to discuss at Poster 52

4th Floor Foyer
12:00–14:00

Universitätsmedizin GREIFSWALD

University Medical Center Groningen

SEPSIS DIALOG

university of groningen