Diagnostic Expert Advisor: Progress

- Last experiments were conducted with Horowitz only. A new model was trained per patient.
- Now all available data is utilized and a global model is trained.

Previously	Now
Only Horowitz Index	All available timeseries data
No static data	Static data included (Age, Sex, Weight, …)
No Surprise Loss	Surprise Loss included
New model per patient	Consistent model trained on all patients

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Diagnostic Expert Advisor: Setup and new Results

- Cohort:
  - Roughly 1.000 ARDS patients
  - Small cohort to limit training times
- Prediction:
  - 70%/10%/20% split for training, testing and validation
  - Prediction Horizon up to 12 hours
- Temporal Fusion Transformer (TFT):
  - Multi-Horizon, Uncertainty Modeling, Flexible history length, …
  - Easy to use data format

Temporal Fusion Transformers
for Interpretable Multi-horizon Time Series Forecasting

Bryan Lim[a,1,*], Sercan Ö. Arık[b], Nicolas Loeff[b], Tomas Pfister[b]

[a]University of Oxford, UK
[b]Google Cloud AI, USA

Abstract

Multi-horizon forecasting often contains a complex mix of inputs – including static (i.e. time-invariant) covariates, known future inputs, and other exogenous time series that are only observed in the past – without any prior information on how they interact with the target. Several deep learning methods have been proposed, but they are typically 'black-box' models which do not shed light on how they use the full range of inputs present in practical scenarios. In this paper, we introduce the Temporal Fusion Transformer (TFT) – a novel attention-based architecture which combines high-performance multi-horizon forecasting with interpretable insights into temporal dynamics. To learn temporal relationships at different scales, TFT uses recurrent layers for local processing and interpretable self-attention layers for long-term dependencies. TFT utilizes specialized components to select relevant features and a series of gating layers to suppress unnecessary components, enabling high performance in a wide range of scenarios. On a variety of real-world datasets, we demonstrate significant performance improvements over existing benchmarks, and showcase three practical interpretability use cases of TFT.

Keywords: Deep learning, Interpretability, Time series, Multi-horizon forecasting, Attention mechanisms, Explainable AI.

1. Introduction

Multi-horizon forecasting, i.e. the prediction of variables-of-interest at multiple future time steps, is a crucial problem within time series machine learning. In contrast to one-step-ahead predictions, multi-horizon forecasts provide users with access to estimates across the entire path, allowing them to optimize their actions at multiple steps in future (e.g. retailers optimizing the inventory for

Preprint submitted to Elsevier *September 29, 2020*

GEFÖRDERT VOM
MEDIZIN INFORMATIK INITIATIVE
Bundesministerium für Bildung und Forschung

Smart Medical Information Technology for Healthcare

Diagnostic Expert Advisor: Setup and new Results

Old Results

Model Name	RMSE	MAE
Linear Regression	78	71
ARIMA	55	45
Facebok Prophet	301	266
Neural Prophet	70	63
Autoregressive Deep Learning	76	66
TFT Hyperparameters tuned, transfer learning no static data, no SL	68	40
TFT (preliminary) transfer learning static data, SL	54	25

New Results

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Diagnostic Expert Advisor: Setup and new Results

- Observation:
  - MAE improved with ML Model using all available data
  - RMSE barely changed

Model	RMSE	MAE
ARIMA	55	45
TFT	54	25

- Interpretation:
  - Models focuses on fitting to stationary parts of Horowitz
  - We should focus on the phase transitions!
  - The improvement is misleading (we mostly improve stationary), but we can trust that the method can learn effectively!
  - *Now we need to teach it to learn the right thing!*

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Diagnostic Expert Advisor: Next Steps

- Split prediction task into two phase:
  - Phase I: Predict volatile regions
  - Phase II: Predict the direction of change
- Leaving out the stationary parts for the training of Phase II and changing the prediction task for Phase I
- The two phases can both be handled by the existing, known to be working, model architecture, altering the target functions and input data

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Diagnostic Expert Advisor: Hardware Requirements

- Approximations based on current experiences training/runnig the DEA
- Training/Development mostly done with ~1.000 patients, as using all available 18.000+ patients increases training times to days even on the Cluster (10+ 24-core machines, Tesla V100 GPUs)
- Without optimization (e.g. SL approximation) running the model for a hundred beds can quickly require around 80 CPU hours to complete

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Diagnostic Expert Advisor: Hardware Requirements Summary

- Optimizations (e.g. ablation studies) can bring down required resources for inference (in-hospital prediction)
- Surprise Loss is computationally expensive and needs to regularly be updated
- Depending on how frequent the predictions should be updated this step alone would require multiple dedicated state-of-the-art workstations
- These estimation don't yet factor in any kind of local training, which would increase the resource requirements even more.

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Calibration data

Smart Medical Information Technology for Healthcare

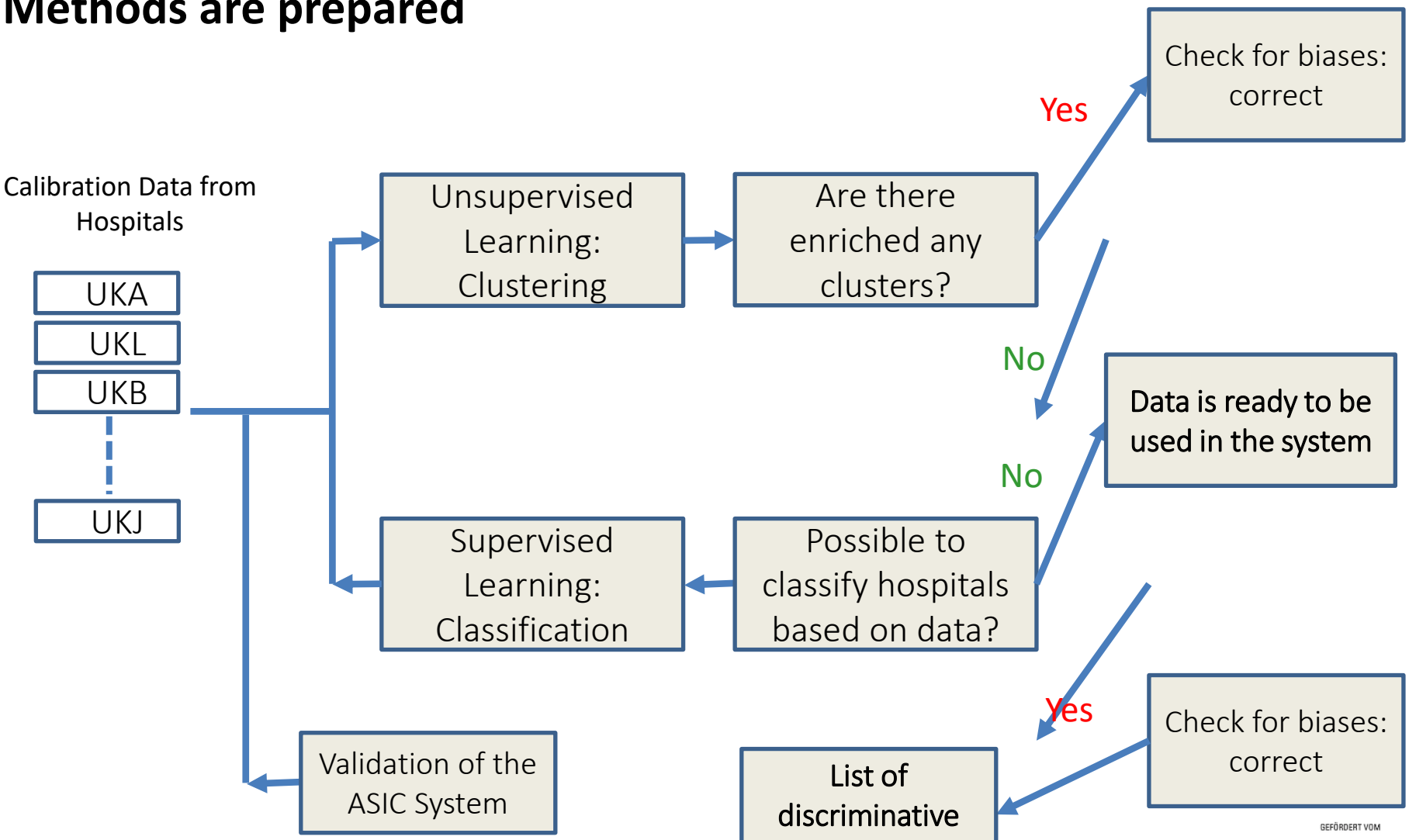
- Data available from 4 sites (5th site - only 150 patients)
- Sample sizes:
  - uk_00 - 13068 patients
  - uk_01 - 1024 patients
  - uk_02 - 2977 patients
  - uk_03 - 1367 patients
- Overall 18000 MV patients satisfying inclusion criteria for the ASIC use case
- Still a lot of opened questions for some of sites --> reminders have been sent on 26.04.2021

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Calibration data: Strategy

Smart Medical Information Technology for Healthcare

- Methods are prepared

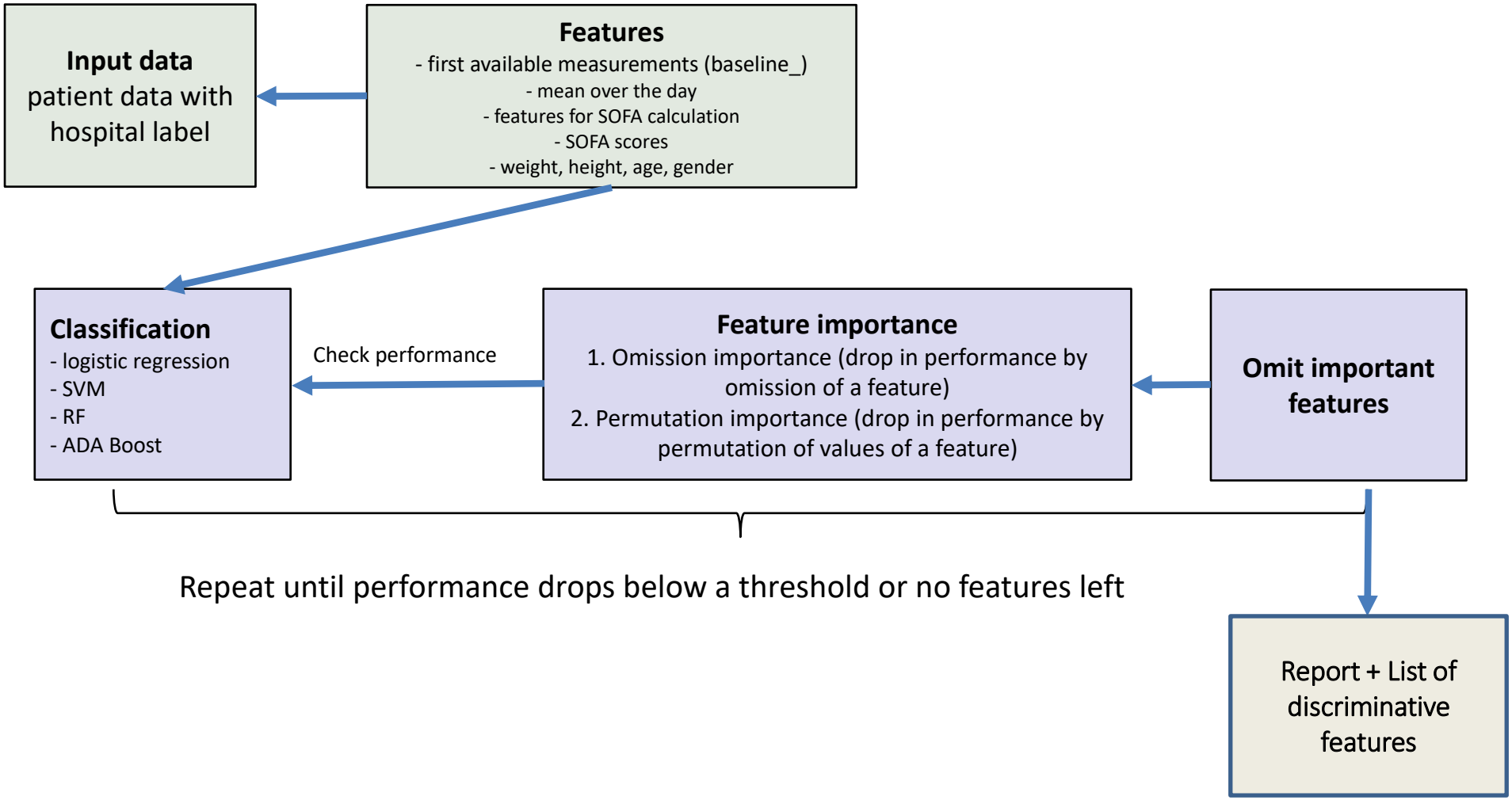


GEFÖRDERT VOM
MEDIZIN INFORMATIK INITIATIVE
Bundesministerium für Bildung und Forschung

Calibration data: Supervised learning

Smart Medical Information Technology for Healthcare

- Methods are prepared



GEFÖRDERT VOM
MEDIZIN INFORMATIK INITIATIVE
Bundesministerium für Bildung und Forschung

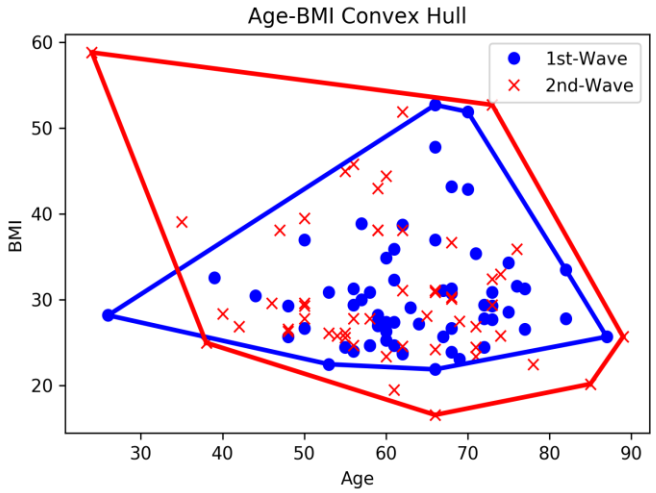
Calibration data: Convex hull analysis

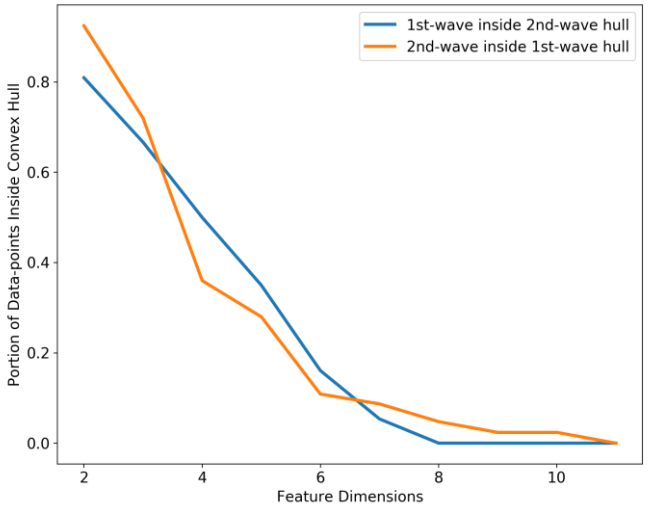
Smart Medical Information Technology for Healthcare

- the smallest convex set enclosing all data points
- data-driven method can make predictions inside the convex hull of the given training data
- transfer learning is possible if **data of new hospital is inside of the convex hull of research data repository**

Convex hull of COVID-19 Data
- 1st and 2nd wave MV COVID-19 patients consisting of 63 and 54 patients
- for low-dimensional feature spaces, e.g. 2D Age and BMI, convex hulls cover each other

Convex Hull in Higher Dimensions
- for higher dimensional feature spaces, the convex hulls of 1st and 2nd wave data get separated
- results are shown for the most important features for mortality classification
- larger data sets are needed for a classification task using data-driven models





MEDIZIN INFORMATIK INITIATIVE

Bundesministerium für Bildung und Forschung

Calibration data: next steps

Smart Medical Information Technology for Healthcare

- Final curation after feedback from the sites
- Parameters for calibration found
  - test with other features/on different cohorts (MIMIC/ICCA, COVID, etc..)
- Decide how to treat respective parameters
- Convex hull analysis of available calibration data - general tool for population comparison and transfer learning

MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Virtual Patient: Fitting a VP to the real patient

Parameters to optimize for

1. Define an initial state of the VP with **measurable global parameters** of the real patient:

	CO	FatmO2	Hb	IE	PaCO2	PaO2	PEEP	Temperature	VentRate	Vtidal	HCO3a	Bea	scPressure	SaO2	pHa
1	10000.0	0.500000	143.755008	0.343615	3.946272	10.652268	8.582611	36.312500	17.041667	505.296	20.4	-2.45	24.675008	93.075	7.4325

2. Optimization parameters: (anatshunt, RQ, VO2, m_1, m_2, m_3, alpha, c_1, c_2)

Non-measurable global parameters:

- Anatomical Shunt
- Respiratory quotient
- Metabolic rate of O2 in the body

Non-measurable parameters reflecting distributions of compartmental parameters:

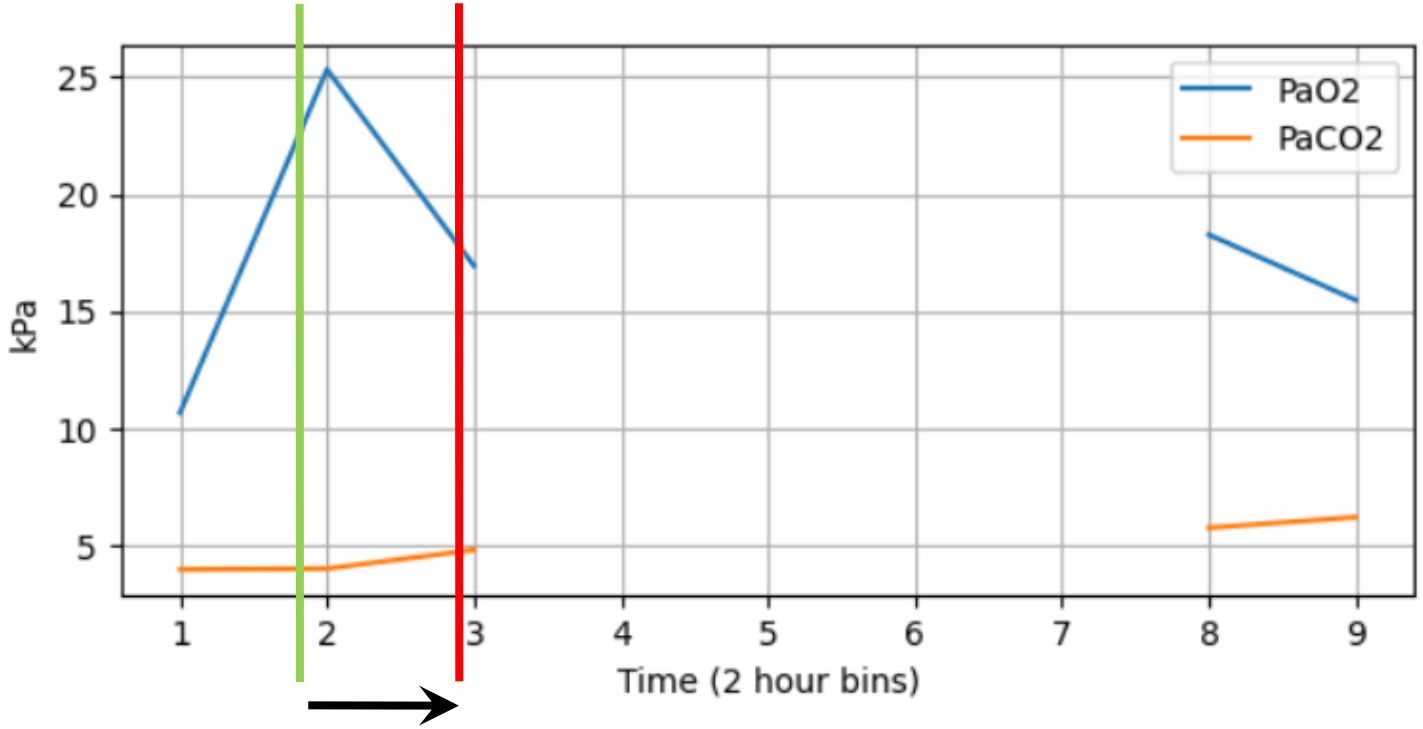
- # of closed compartments (controlled through Pext - alpha closed, 1-alpha opened compartments)
- m_1, m_2, m_3
- where $R_{ards} = m_1*N + c_1$, $VR_{ards} = m_2*N + c_2$, $STalv_{ards} = m_3*STalv_{healthy}$

3. Objective function:

$$\min_{params}((PaO2_{real} - PaO2_{sim})^2 + (PaCO2_{real} - PaCO2_{sim})^2)^{0.5}$$

Virtual Patient: Fitting a VP to the real patient

Objective function: $\min_{params}((PaO2_{real} - PaO2_{sim})^2 + (PaCO2_{real} - PaCO2_{sim})^2)^{0.5}$

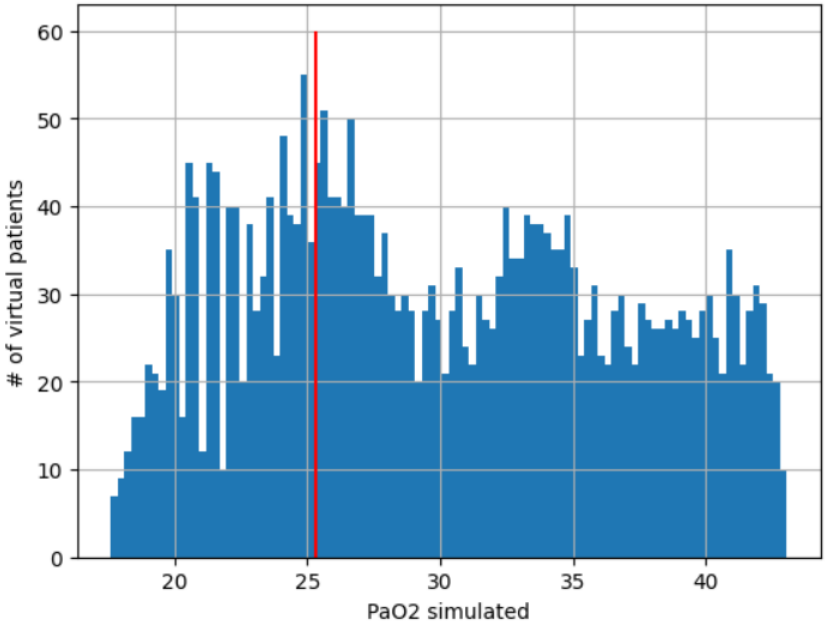


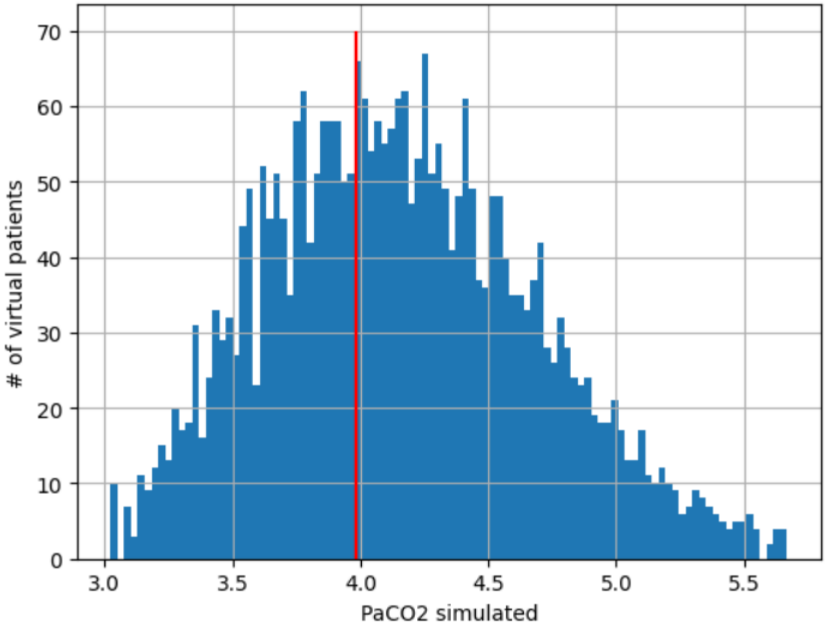
Simulation for the next 2 hours

- Potentially for every time step we can find a VP configuration, which suits
- Multiple time points increase confidence for a particular VP configuration
- Final configuration defines the VP

Virtual Patient: Fitting a VP to the real patient

- Currently defined a cohort of 3000 VP based on one real patient and varying optimization parameters
- Investigated whether simulations cover values of a real patient
  - A lot of VPs, which showed similar behaviour --> not unique solution
- Next step: implement objective function and perform optimization algorithm





Virtual Patient: Fitting a VP to the real patient

UNIKLINIK RWTH AACHEN

Issues:

- Ill-posed problem (solution is not unique)
- Multiple time points - multiple configurations possible
  - Dynamic configuration - parameters of VP do change with time
  - Sort optimization parameters based on the time scale for changes (i.e. Shunt changes slowly - should be fixed, N closed compartments can change)

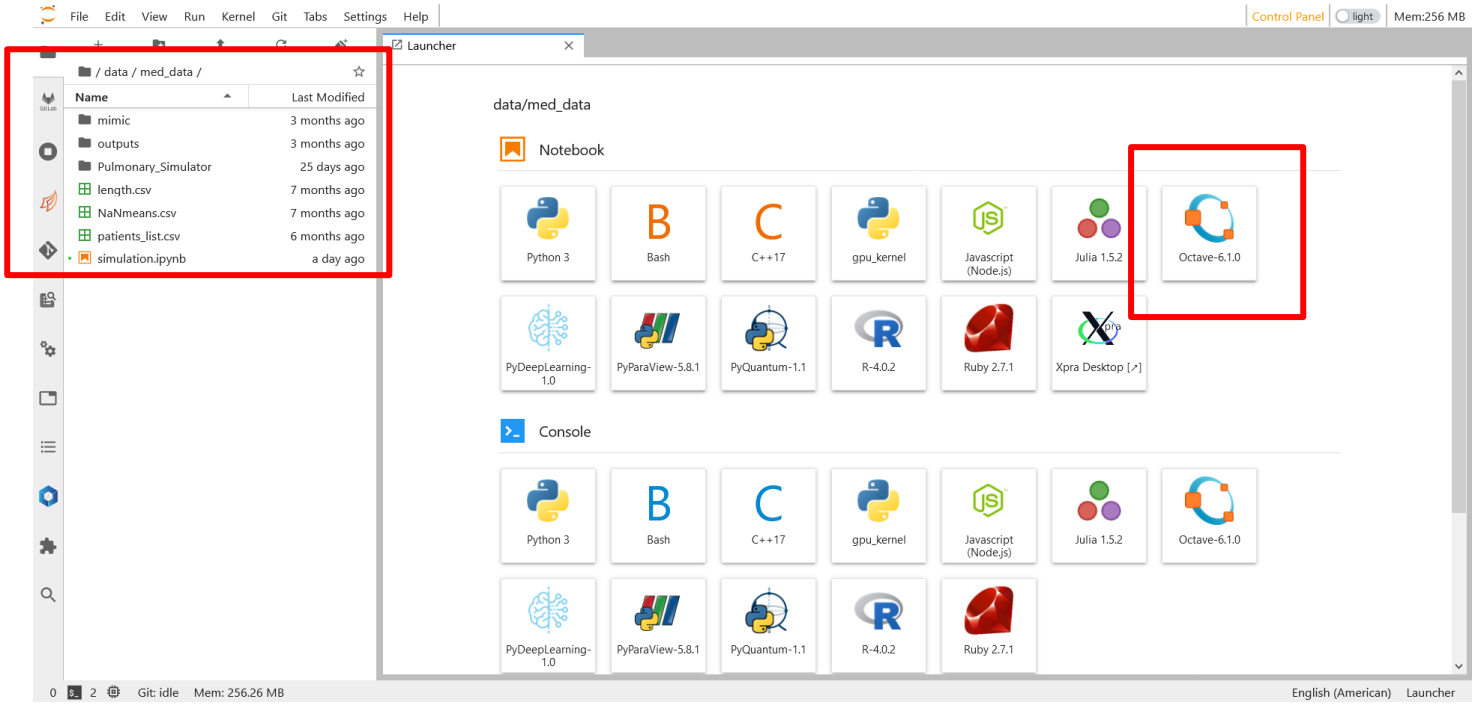
Work in progress:

- Objective function, which parameters to choose
- Define final set of parameters for optimization
- Choose and implement an optimization algorithm

Virtual Patient: HPC Implementation

Steps:

- Prepare a dedicated environment (storage, processors) for source data and for outputs.
- Set up platform to run scripts.
- Implement multirun matlab script on Cluster.
- Prepare outputs for model training.



Virtual Patient: HPC Implementation

Roadblocks:

- Since no Matlab implementation exists on the cluster we try to adapt the available Octave implementation.

- Had to import several packages that weren't originally available.

- Made changes to the original script to adapt it to these packages.

Result:

```
[3]: run('Pulmonary_Simulator/scripts/multi_run/multiRun_main.m')

warning: genvarname is obsolete; use matlab.lang.makeValidName or matlab.lang.makeUniqueStrings instead
T = dataframe with 24 rows and 28 columns
Src: ../input_table/patients_ID.csv
_1 MATLAB_NO                    ID anatShunt chPressure     CO ColVol FatmO2     Hb     IE  Ncomp PaCO2  PaO2   PEEP     RQ run_time scPres
sure Temperature VDphys VentRate    VO2 Vtidal                    comp_ID             P_ext            Rcomp            ST_alv            Stick_t
ime_rand            TOPcomp            VRcomp
Nr    double                   char    double     double double double double double double double  char  char double double   double     do
uble      double double   double double double                        char             char             char              char
char              char              char
 1        1                       A  0.010000          1  11100      6 0.8000    105 0.2759    100     -      -      5 0.6000      120
20      37.200     50   12.253 294.25    420                          A                A                A                 A
A                 A                 A
 2        2                       B  0.080000          1   7200      6 0.9000    108 0.2500    100     -      -      5 0.6500      120
20      37.200     60   15.800 300.00    420                          B                B                B                 B
B                 B                 B
 3        3                       H  0.020000          1   5000      6 0.2100    160 0.3300    100     -      -      5 0.8000      120
20      37.200     60   12.000 250.00    400                          H                H                H                 H
H                 H                 H
 4        4                       A  0.010000          1  11100      6 0.8000    105 0.2759    100   4.1 13.95      5 0.6000      120
20      37.200     50   12.253 294.25    420                          A                A                A                 A
A                 A                 A
 5        5                       B  0.080000          1   7200      6 0.9000    108 0.2500    100   4.1 13.95      5 0.6500      120
20      37.200     60   15.800 300.00    420                          B                B                B                 B
B                 B                 B
 6        6                       H  0.020000          1   5000      6 0.2100    160 0.3300    100   4.1 13.95      5 0.8000      120
20      37.200     60   12.000 250.00    400                          H                H                H                 H
H                 H                 H
 7        7               A_permuted  0.010000          1  11100      6 0.8000    105 0.2759    100   4.1 13.95      5 0.6000      120
20      37.200     50   12.253 294.25    420                 A_permuted       A_permuted       A_permuted        A_permuted                A_pe
rmuted         A_permuted        A_permuted
 8        8               B_permuted  0.080000          1   7200      6 0.9000    108 0.2500    100   4.1 13.95      5 0.6500      120
20      37.200     60   15.800 300.00    420                 B_permuted       B_permuted       B_permuted        B_permuted                B_pe
```

Virtual Patient: HPC Implementation

UNIKLINIK RWTH AACHEN

Next Steps:

- Performance analysis on the cluster:
  - Single run duration in JupyterLab.
  - Multi-run duration.
  - Adapting scripts to the cluster.

- Tweaking and testing parallelisation methods.

UNIKLINIK RWTH AACHEN

Thank you for your attention!