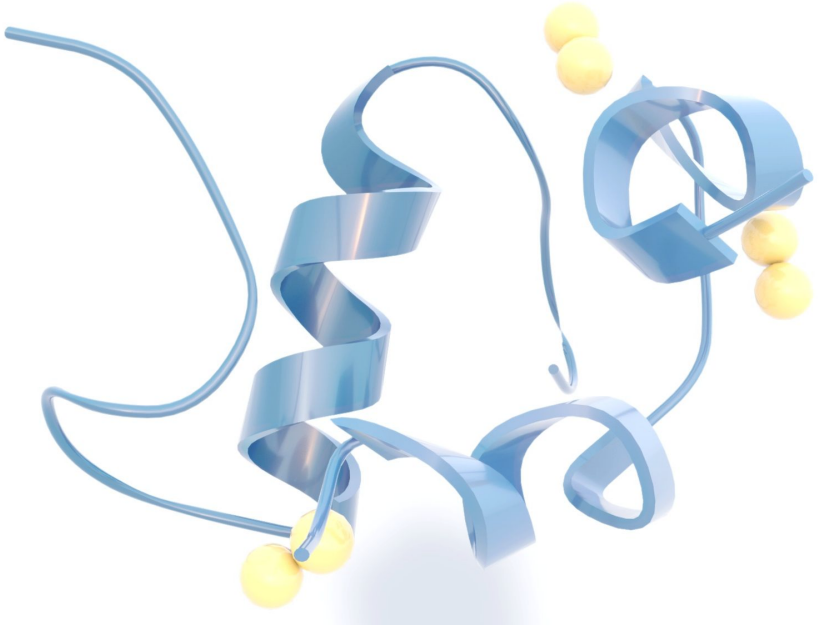
MUNI
FI

Glycemia Forecasting

Andrej Kubanda

Sugars, insulin, energy

- humans need sugar for energy
- insulin hormone regulates blood glucose levels (Glycemia)
  - hypo- and hyperglycemia



MUNI
FI

Diabetes

Type 1

- autoimmune condition
- insulin-producing cells in pancreas attacked
- lifetime insulin therapy required for survival

Type 2

- little production or resistance to insulin
- often caused by obesity
- treatment: lifestyle changes

BGLP Challenge: OhioT1DM dataset

- 8 weeks of data of 12 patients
- glycemia measured every 5 mins
- insulin doses (bolus & basal)
- self-reported meal times & estimates
- physiological data
  - exercise, sleep, stress, work

ID	Gender	Age	Pump Model	Sensor Band
540	male	20–40	630G	Empatica
544	male	40–60	530G	Empatica
552	male	20–40	630G	Empatica
567	female	20–40	630G	Empatica
584	male	40–60	530G	Empatica
596	male	60–80	530G	Empatica
559	female	40–60	530G	Basis
563	male	40–60	530G	Basis
570	male	40–60	530G	Basis
575	female	40–60	530G	Basis
588	female	40–60	530G	Basis
591	female	40–60	530G	Basis

MUNI FI

BGLP Challenge

Task: predict Glycemia 30 and 60 minutes into the future

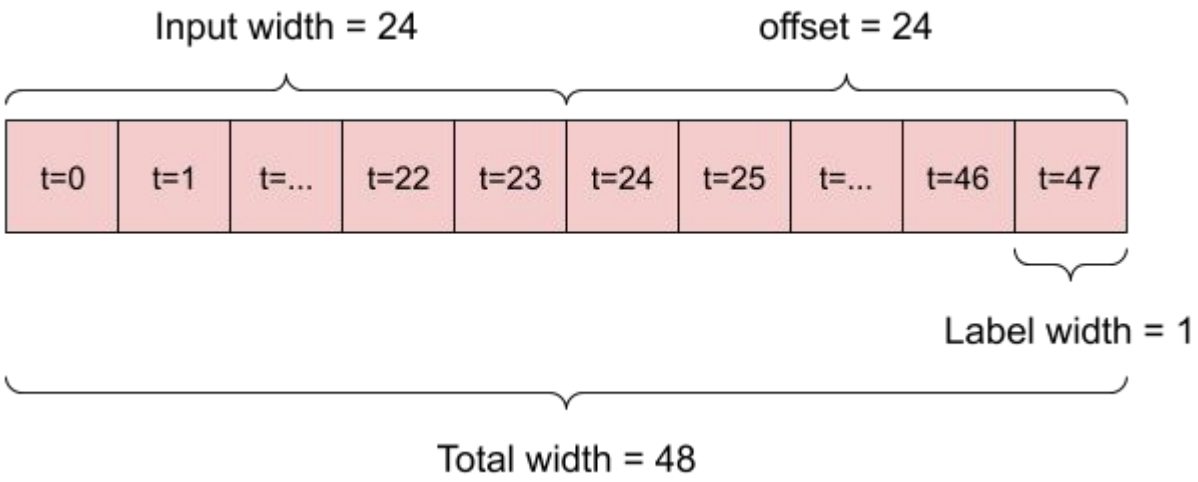
- 1st cohort available for training
- 2nd cohort split into train & test set
- per-patient evaluation on 2nd cohort test sets
  - RMSE & MAE

BGLP Challenge: Results

	30 min		60 min				
Paper ID	RMSE	MAE	RMSE	MAE	**Overall**	Online	Personalized
5	17.45	11.22	33.67	23.25	85.59	No	Yes
13	18.22	12.83	31.66	23.60	86.31	No	Yes
6	19.21	13.08	31.77	23.09	87.15	No	Yes
16	18.34	13.37	32.21	24.20	88.12	No	Yes
15	19.05	13.50	32.03	23.83	88.41	No	Yes
1	18.23	14.37	31.10	25.75	89.45	No	No
14	19.37	13.76	32.59	24.64	90.36	Yes	Yes
8	19.01	13.73	33.37	24.98	91.09	No	Yes
11	18.99	13.73	33.39	25.04	91.15	No	Yes
4	19.79	13.62	33.73	24.54	91.68	No	Yes
7	19.60	14.25	34.12	25.99	93.96	No	Yes
9	20.03	14.52	34.89	26.41	95.85	Yes	Yes
2	21.80	15.00	35.00	25.00	96.80	Yes	Yes

MUNI FI

Forecasting as Supervised Problem



MUNI FI

Convolutional RNN

Method

- multi-step forecast
- 2h history
- transfer learning

Features

- glycemia
- insulin basal
- insulin bolus
- meal carbs

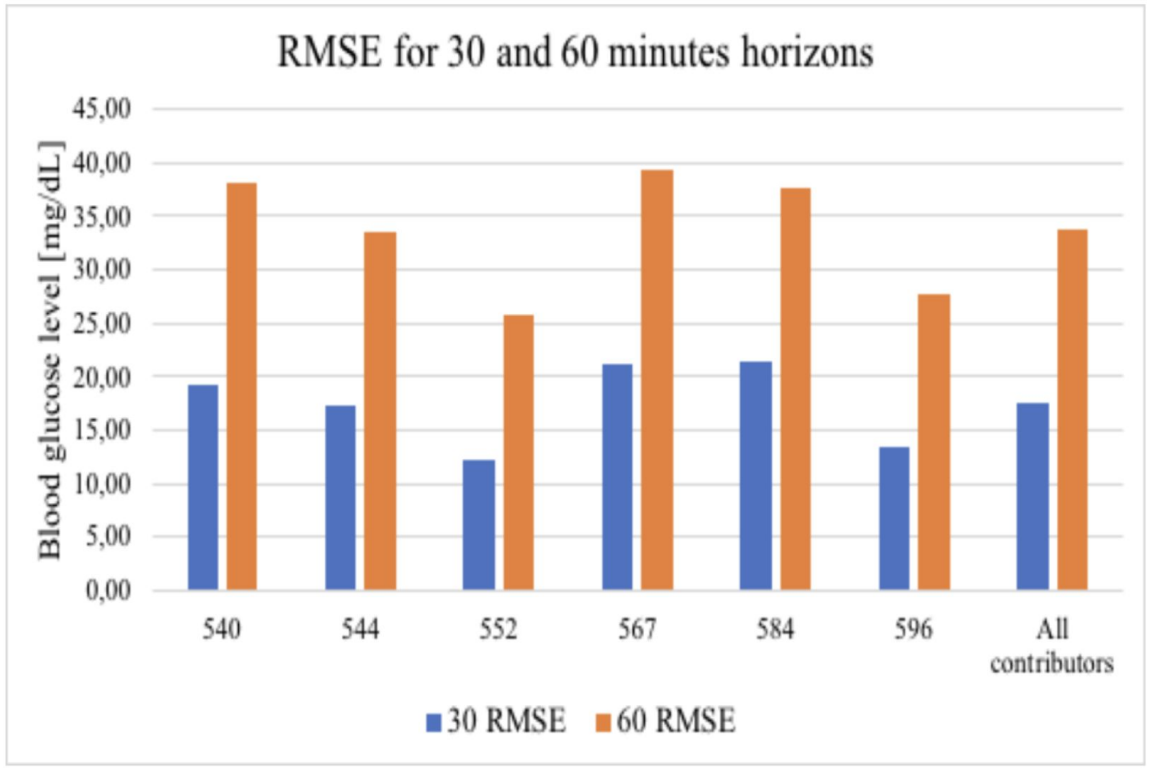
Preprocessing

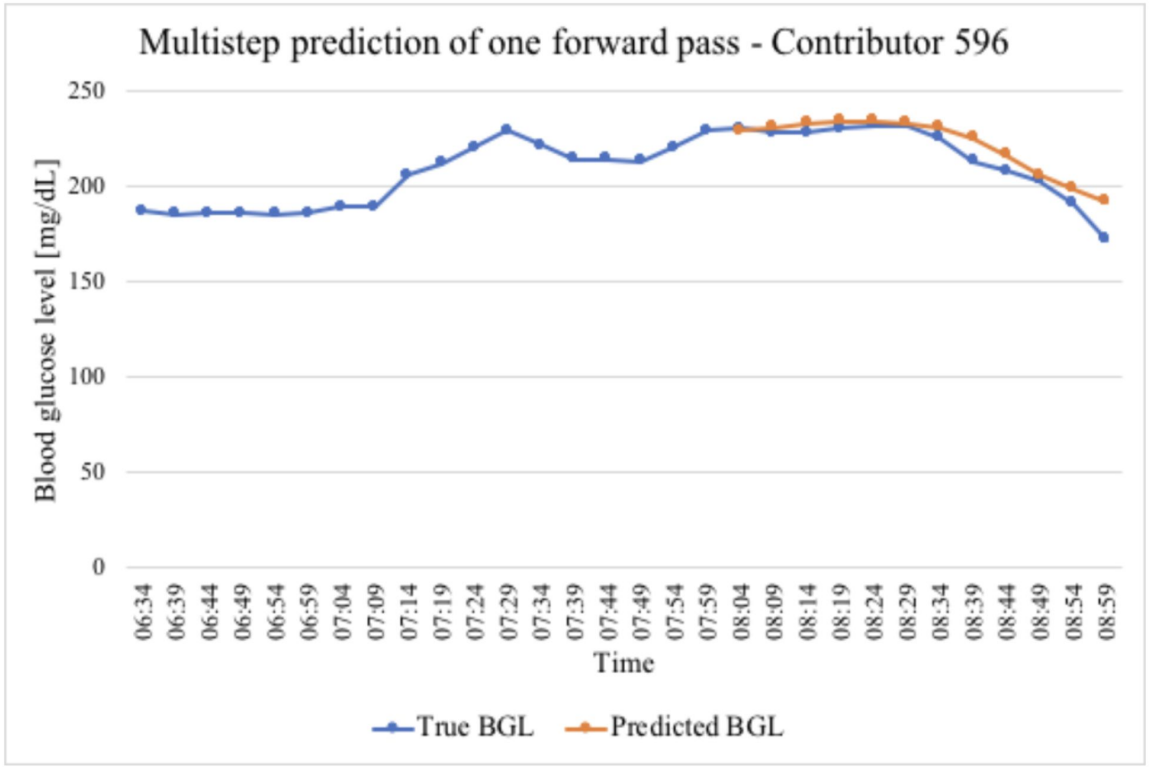
- ffill
- Gaussian filter smoothing
- glycemia differencing

Model Architecture

Layer description	Output dimension
Convolution 1D	(Batch size, 24, 8)
Max pooling 1D	(Batch size, 12, 8)
Convolution 1D	(Batch size, 12, 16)
Max pooling 1D	(Batch size, 6, 16)
Convolution 1D	(Batch size, 6, 32)
Max pooling 1D	(Batch size, 3, 32)
LSTM	(Batch size, 64)
Dense	(Batch size, 256)
Dense	(Batch size, 32)
Dense	(Batch size, 12)

Convolutional RNN: Evaluation





MUNI FI

LSTM Attention

Method

- non-personalized
- single-step forecast
- 30min history

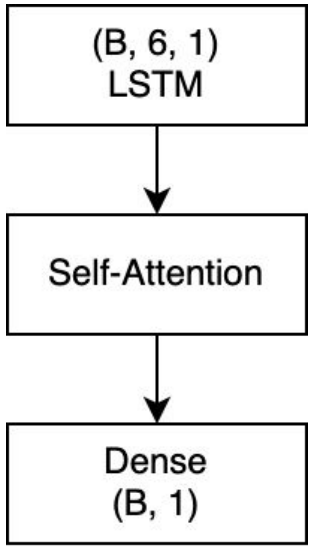
Preprocessing

- interpolation
- standardization
- discard or 0-replace missing values

Features

- glycemia

Model Architecture



MUNI FI

LSTM Attention: Evaluation

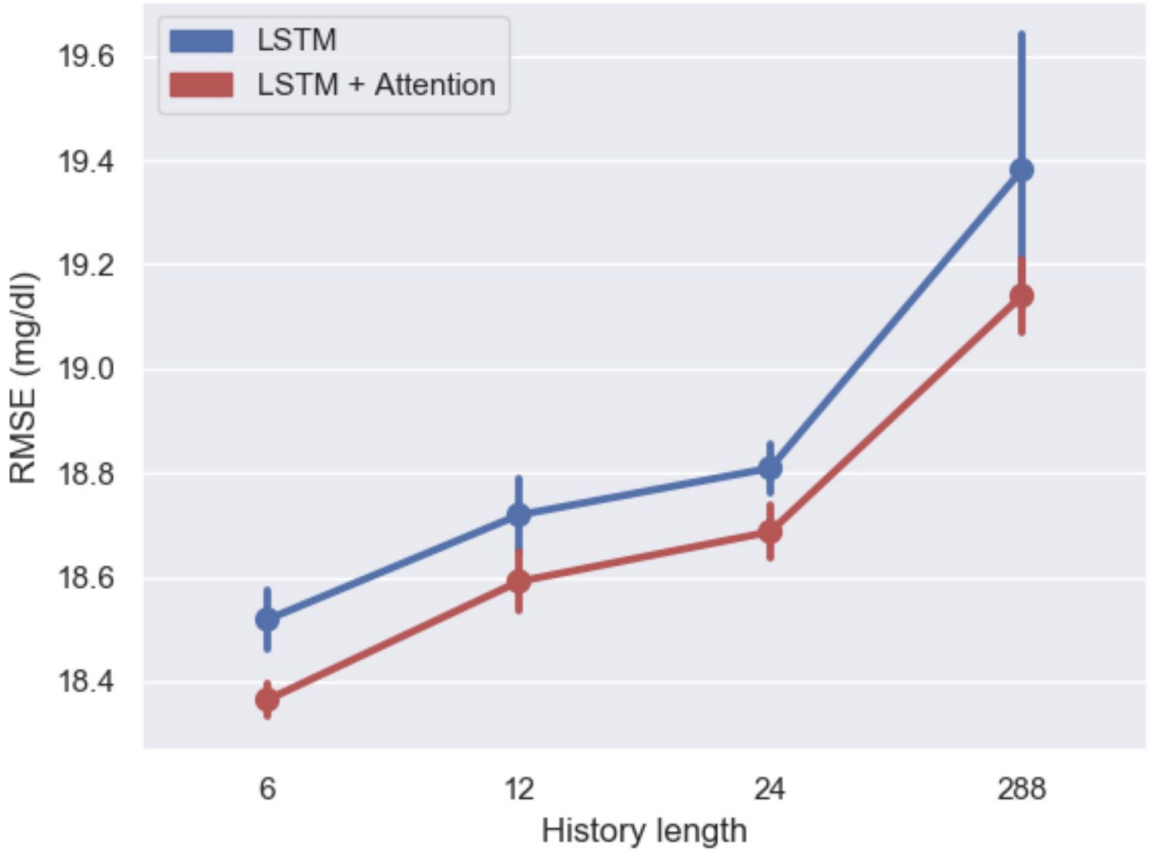
	RMSE		MAE	
Patient ID	**PH=30**	**PH=60**	**PH=30**	**PH=60**
540	21.03 (0.07)	37.37 (0.09)	16.64 (0.1)	30.8 (0.13)
544	16.14 (0.12)	28.4 (0.14)	12.85 (0.11)	23.57 (0.16)
552	15.82 (0.06)	27.6 (0.15)	12.43 (0.12)	22.78 (0.16)
567	20.29 (0.08)	34.28 (0.18)	15.9 (0.12)	28.95 (0.13)
584	20.39 (0.07)	32.97 (0.09)	15.99 (0.03)	27.04 (0.07)
596	15.7 (0.03)	25.99 (0.12)	12.4 (0.04)	21.33 (0.13)
AVG	18.23 (2.36)	31.1 (4.05)	14.37 (1.83)	25.75 (3.43)

Patient ID	**Exclude missing data**	**Include missing data**
540	21.45 (0.06)	**21.03 (0.07)**
544	16.79 (0.06)	**16.14 (0.12)**
552	16.27 (0.13)	**15.82 (0.06)**
567	21.19 (0.1)	**20.29 (0.08)**
584	21.16 (0.06)	**20.39 (0.07)**
596	16.08 (0.07)	**15.7 (0.03)**
AVG	18.82 (2.45)	**18.23 (2.36)**

MUNI FI

LSTM Attention: Cont.

- longer history leads to worse performance
- long patterns are more personalized



MUNI FI

Dicatil Project

- data collection & domain knowledge

1. LÉKAŘSKÁ FAKULTA
Univerzita Karlova

- data storage & administration

MagicWARE

- data analysis & predictor

MUNI FI

MUNI FI

Dicatil Project: Task

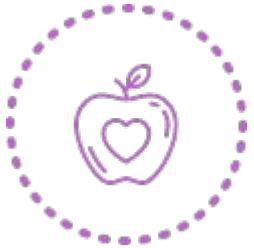
- 30-90 min glycemia forecasts
  - long Glycemia forecasts are useless

- forecasts using planned meals, steps & activities
- forecasts of morning glycemia

Dicatil+ Dataset

- 12 patients (data quality & quantity varies)
  - dirty, outliers, data mixed from multiple sources
- different glycemia sensor → irregular & longer intervals
- no insulin, little sleep, no stress data
- richer nutrition data

Glycemia

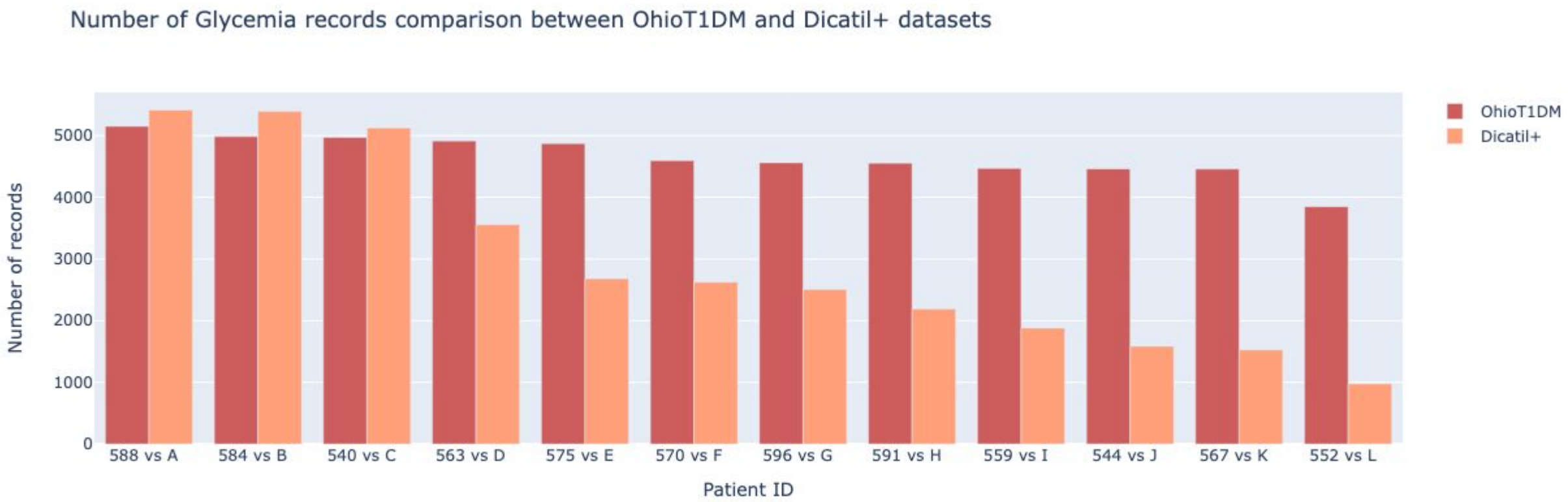


Nutrition

Physical Activity

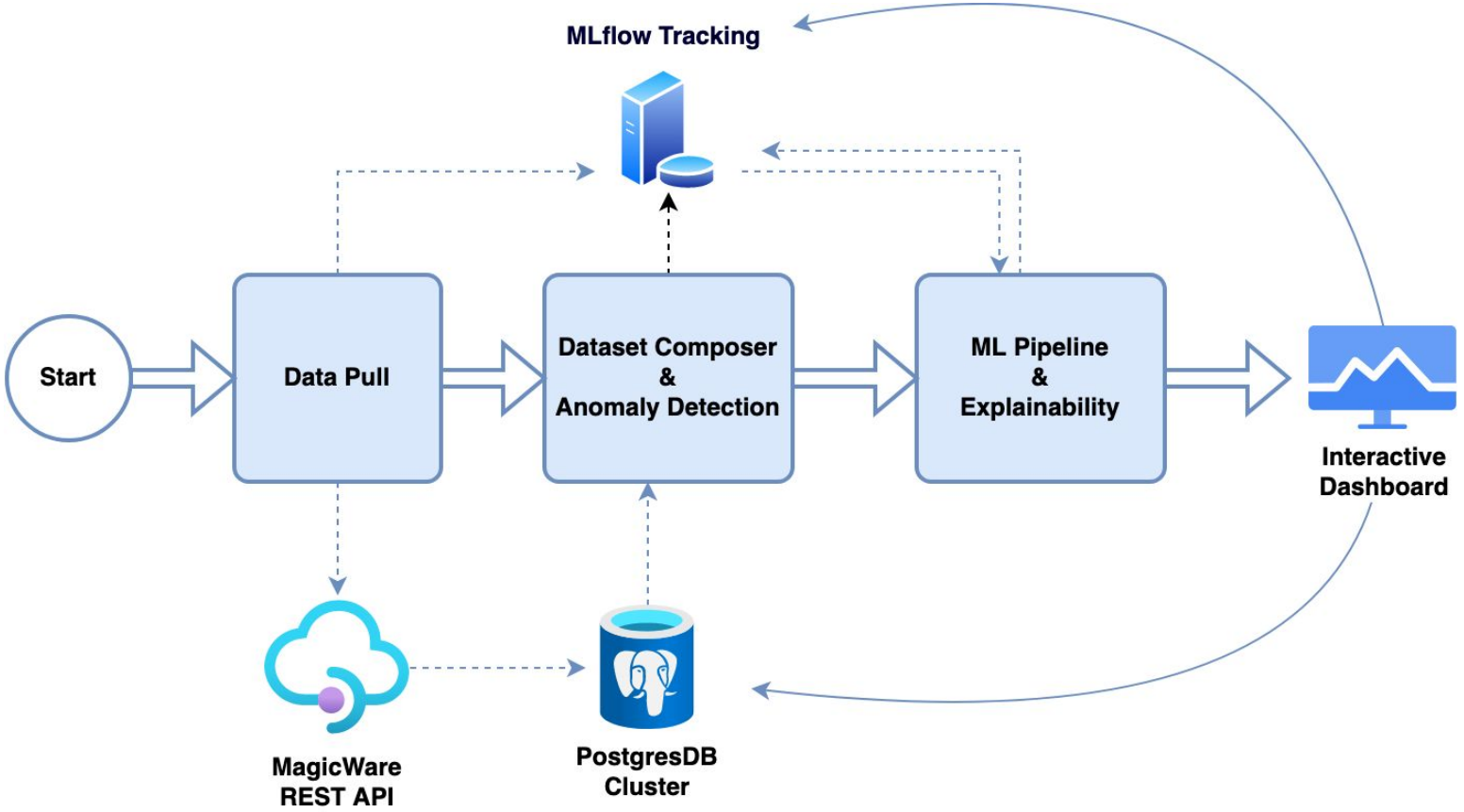
MUNI FI

Dicatil+ Glycemia Data Volume



Infrastructure & Time Series Framework

- all data & compute in Kubernetes



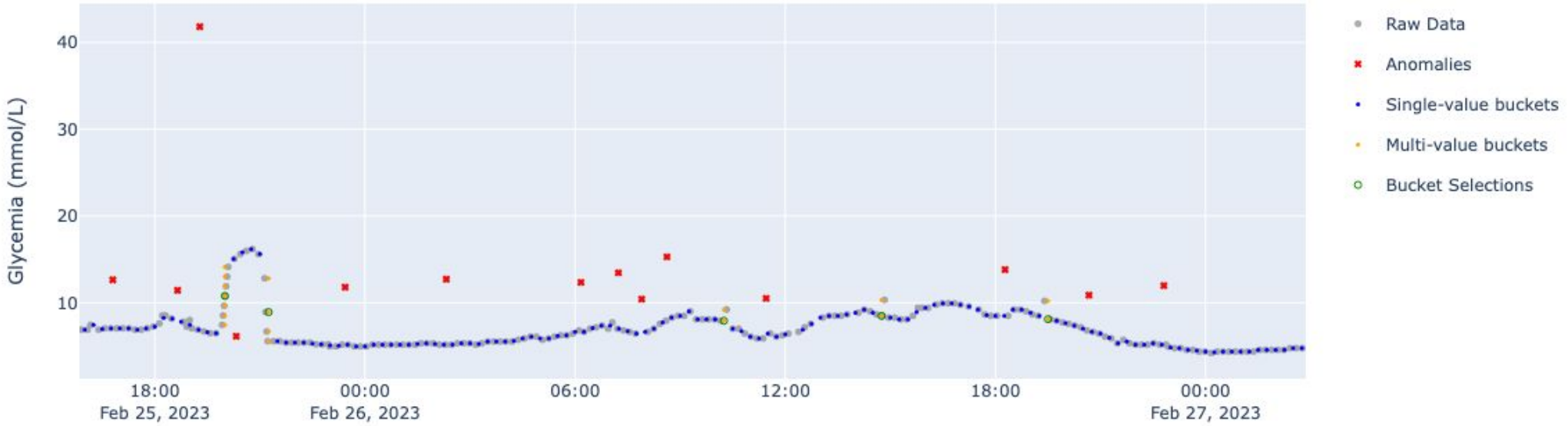
MUNI
FI

Dataset Pipeline

1. raw data extraction from DB
2. anomaly detection
3. resampling & aggregation
4. feature engineering
   - moving averages, moving sums, datetime features
5. dataset file & HTML report

Dataset Pipeline: Anomaly Detection

1. smooth data using a filter
2. fit spline & compute distances to raw data points
3. fit IsolationForest & predict outliers



MUNI FI

ML Pipeline

1. train-validation-test split
2. method-specific data preprocessing & windowing
   - missing values, standardization / normalization / scaling, …
3. training (k times)
   - transfer learning, sampling strategy
4. evaluation
5. explainability

Dataset Predictability

OhioT1DM

	RMSE_0:30:00		RMSE_1:00:00	
	mean	std	mean	std
subject				
540	21.6292	0.1012	39.1055	0.1324
544	17.9436	0.0486	31.9047	0.0503
552	15.9397	0.0455	29.0406	0.0368
559	18.0900	0.0880	32.7803	0.1140
563	16.3294	0.0398	28.0885	0.0391
567	21.4610	0.0571	37.4724	0.1114
570	16.3247	0.2337	29.3317	0.2894
575	22.3754	0.1159	35.8142	0.0974
584	23.0457	0.0856	38.3496	0.1075
588	18.9031	0.0381	32.3290	0.0871
591	20.1655	0.0555	32.3255	0.1328
596	16.4016	0.0234	28.5205	0.0097

Dicatil+

	RMSE_0:30:00		RMSE_1:00:00	
	mean	std	mean	std
subject				
364	12.9189	0.2410	24.1848	0.4957
1046	13.1582	0.2097	24.0213	0.2851
2265	14.4160	0.0764	26.3580	0.1219
2746	14.4317	0.1158	26.1781	0.1558
4366	15.7669	0.0965	24.9327	0.0926
4948	14.4208	0.0950	27.0548	0.0955
5275	9.9755	0.0383	17.0041	0.1277
5549	13.3153	0.0529	23.3762	0.0963
6537	11.6444	0.0420	18.6090	0.0843
20736	11.8561	0.0869	20.1658	0.2187
20762	10.7348	0.0785	19.1377	0.2044

MUNI FI

CRNN Experiments

	RMSE @ 30	RMSE @ 60	MAE @ 30	MAE @ 60
Glyc	16.017315	26.881805	10.864706	18.933036
Glyc Steps	14.009471	24.139170	9.209684	16.132882
Glyc Carbs	15.690608	27.035870	10.314846	18.068828
Glyc Steps Carbs	12.662501	21.367086	8.578651	13.936472
Glyc Carbs Steps HR	12.613626	22.449807	7.998876	14.053213

MUNI FI

Single vs Multi Horizon Models

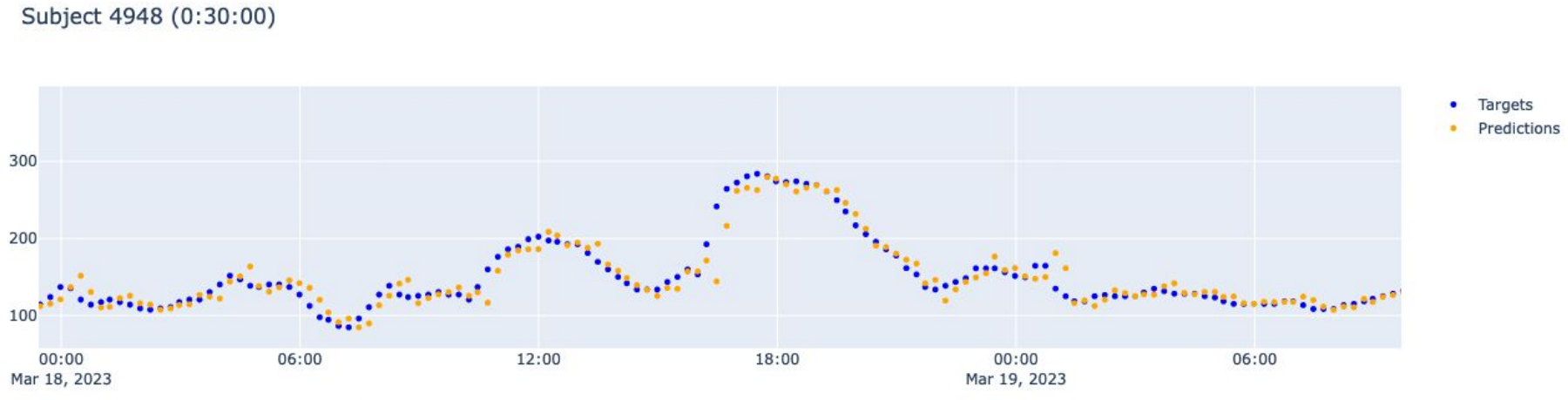
- one model per horizon
- slightly better performance?

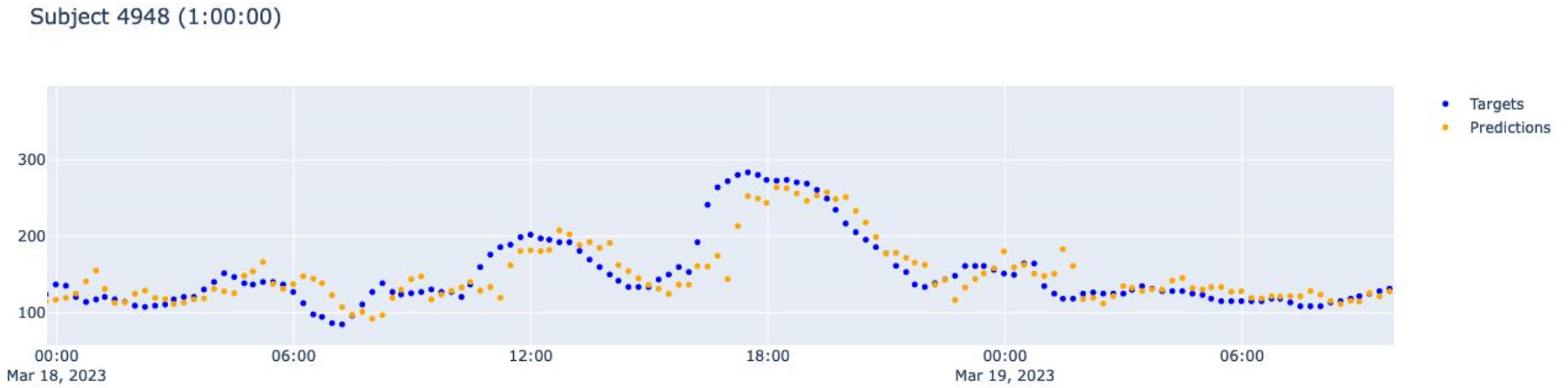
- model capacity focused
- classic approach & metrics

- single model
- more coherent forecast

- model capacity divided
- specialized metrics

Predictions Example





Model Personalization

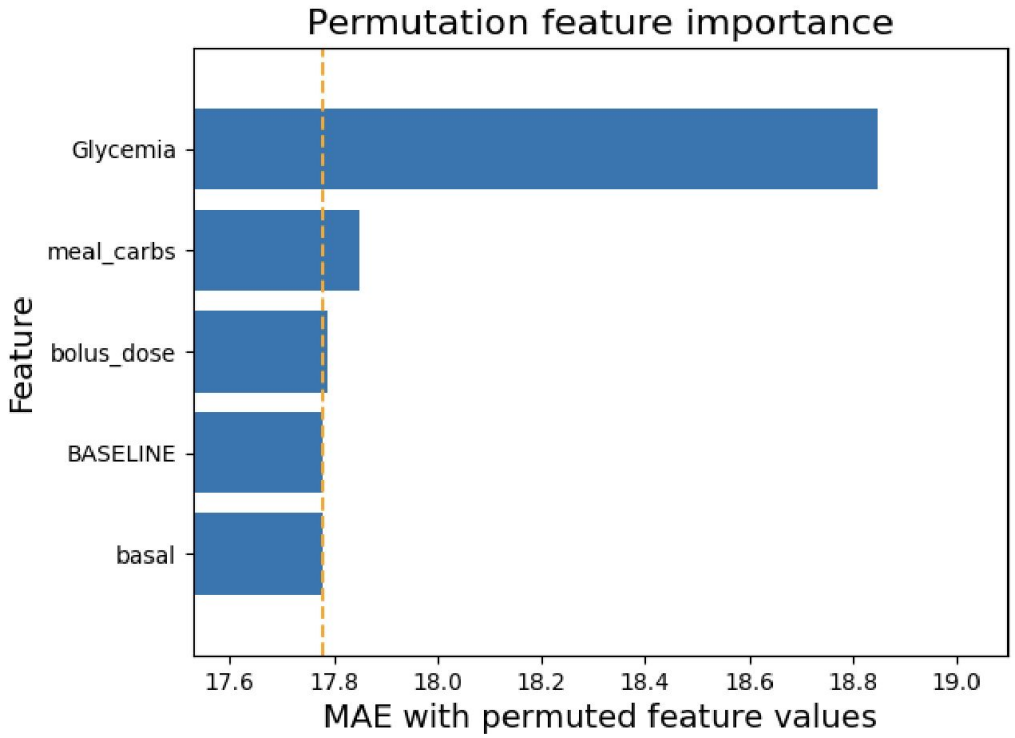
- improves performance in general
- short-term patterns ~ more general
- long-term patterns ~ more personalized

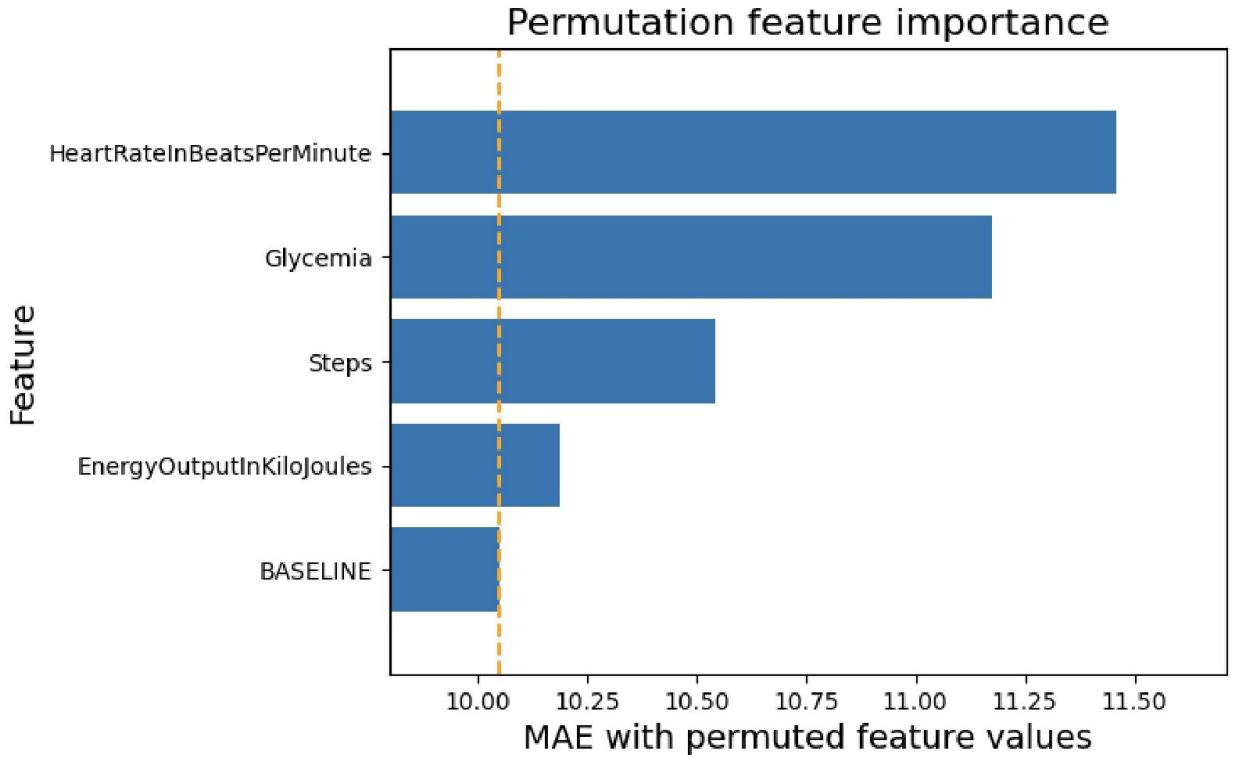
subject	RMSE_0:30:00	RMSE_0:45:00	RMSE_1:00:00	RMSE_1:15:00	RMSE_1:30:00
364	0.490700	-0.217000	-1.038500	-1.902600	-2.425700
2265	-0.178200	-0.831000	-1.835200	-2.905800	-3.842900
2746	-0.660900	-2.415400	-3.473000	-3.520500	-5.120900
4349	-0.113600	-0.656000	-1.430900	-1.980100	-2.398700
4948	-0.469800	-0.869100	-1.016900	-1.261400	-1.554100

MUNI FI

XAI: Permutation Feature Importance

- different patients react to different features





MUNI FI

Future plans

- ensembles
- window sampling strategies
- Time2Vec
- additional xAI methods (e.g. Shap)
- morning Glycemia forecasts (need more data)
- better feature engineering
- XGBoost, LightGBM

Sources

- https://cgi.csc.liv.ac.uk/~frans/PostScriptFiles/bglp_final_2020.pdf
- https://arxiv.org/pdf/1807.03043.pdf
- http://smarthealth.cs.ohio.edu/bglp/bglp-results.html
- https://is.muni.cz/auth/th/j0gda/Thesis.pdf

MUNI

FI