75 years
of the Silesian University
of Technology

Silesian
University
of Technology

RESEARCH
UNIVERSITY
EXCELLENCE INITIATIVE

FEDERATED LEARNING AND DEEP LEARNING FOR AGV ANOMALY DETECTION

Daniel Kostrzewa, Bohdan Shubyn, Paweł Benecki,
Piotr Grzesik, Dariusz Mrozek

18.05.2022

Table of Content

75 years
of the Silesian University
of Technology

Silesian
University
of Technology

RESEARCH
UNIVERSITY
EXCELLENCE INITIATIVE

DEEP LEARNING FOR AGV ANOMALY DETECTION

Daniel Kostrzewa, Bohdan Shubyn, Paweł Benecki,
Piotr Grzesik, Dariusz Mrozek

18.05.2022

Forecasting of Energy Consumption

1. Some problems may be detected during scheduled checks
2. Others have to be found by different means
3. Telemetry is used for this purpose
4. Forecasting of energy consumption is the first step for the anomaly detection

Forecasting of Energy Consumption

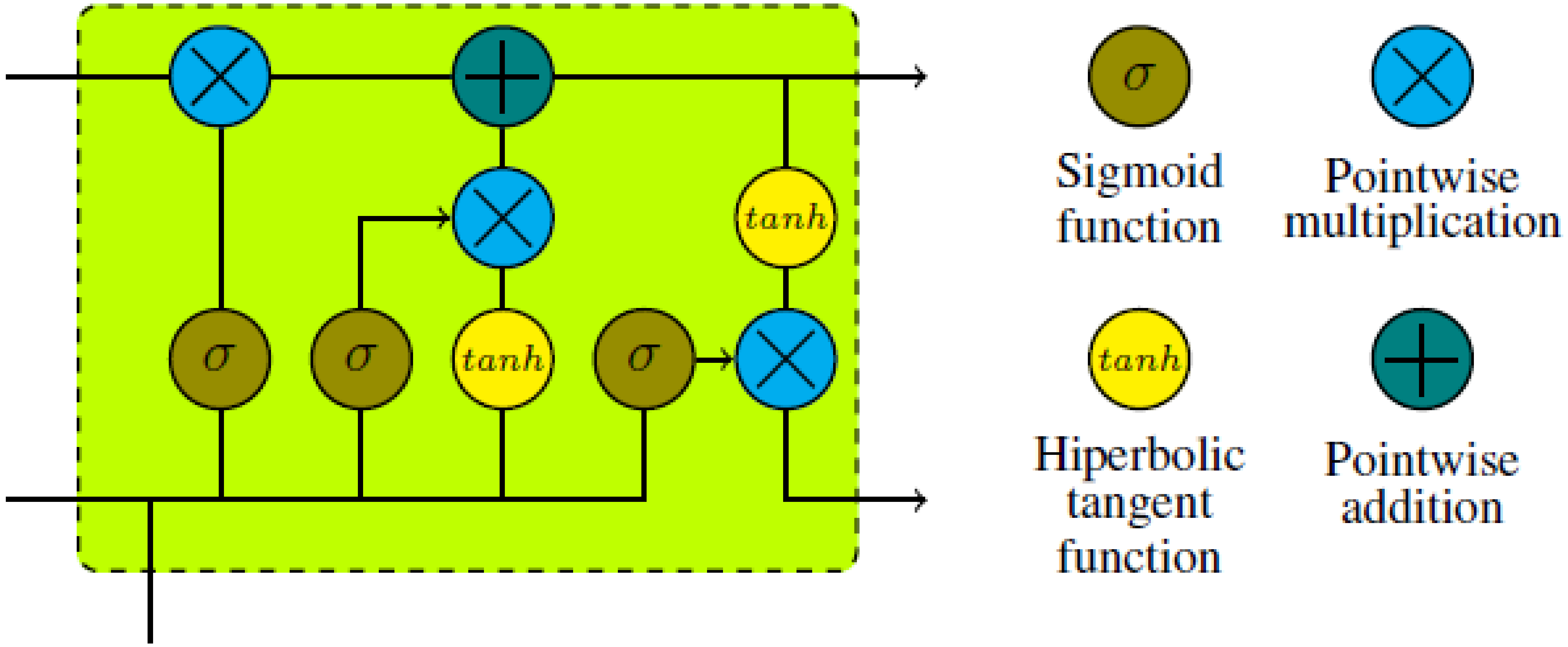
1. Anomaly detection can be formulated as a one-class classification
2. Machine Learning models are learned with data considered as normal
3. Models are evaluating unseen data and compute an anomaly score
4. The evaluation is often the difference between forecasted and actual values

Data Sets

1. Formica-1 (2022): 9 trials,
   avg. seq. length 1600 pts, 34 features
2. Husky A200 (2021): 92 trials (113 sequences),
   avg. seq. length 4200 pts, 22 features
3. IEEE Battery (2020): 72 sequences,
   4k-56k seq. length, 13 features

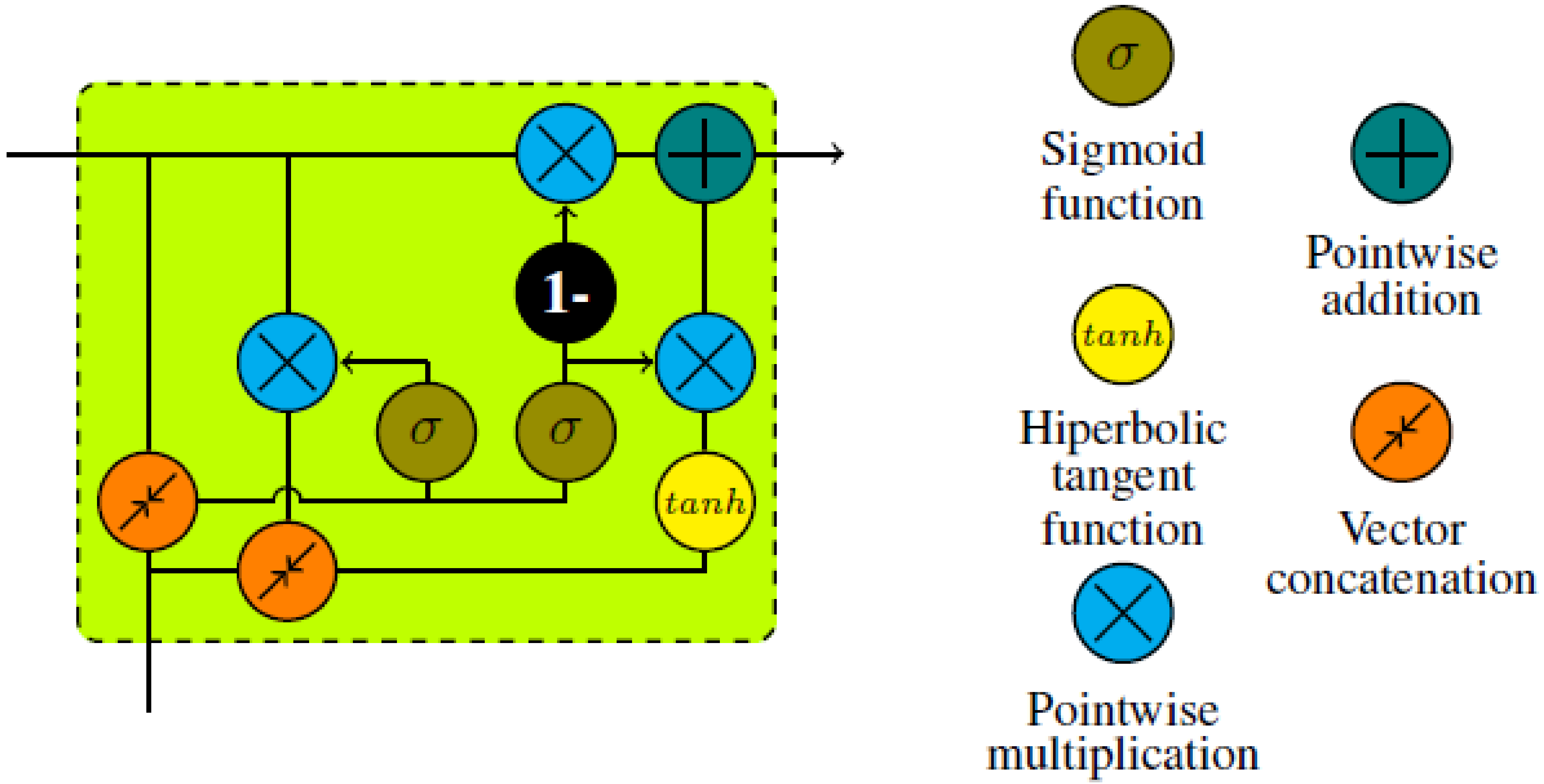
Machine Learning Models

1. LSTM cells (1- and 2-layer architectures)



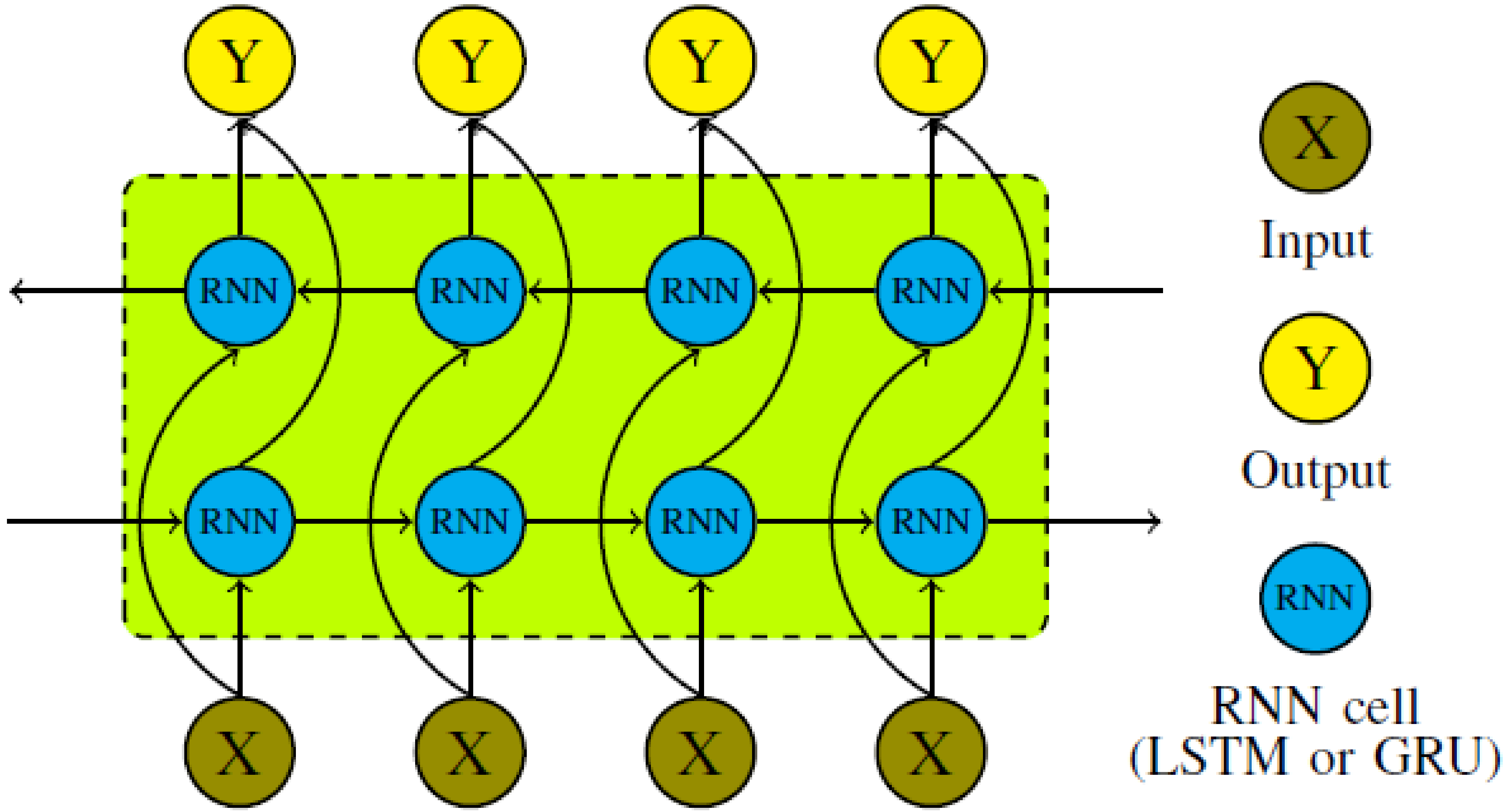
Machine Learning Models

2. GRU units (1- and 2-layer architectures)



Machine Learning Models

3. and 4. BiLSTM and BiGRU (1- and 2-layer architectures)



Numerical Experiments

1. Goal – forecast of power consumption signal based on other telemetry signals
2. Evaluation metrics
    1. Mean Square Error (MSE)
    2. Mean Absolute Error (MAE)
    3. Mean Absolute Percentage Error (MAPE)

Numerical Experiments

Step 1 – Identification of model and training scheme

1. 8 different architectures (LSTM, GRU, BiLSTM, BiGRU; 1- and 2-layers)
2. Different history length (10-190 points, step 20)
3. 3 data sets
4. 240 trainings in total
5. Result – best architectures and history length for each data set

Numerical Experiments

Step 2 – Input signals correlation

1. Power consumption singal depends on the other signals to a very different level
2. Common approach – well-built deep network can learn all dependencies…
3. … with huge cost: increased data, computing power, and time requirements
4. Pearson’s correlation coefficients between power consumption and other signals
5. Result – thresholds set (0.05, 0.1, 0.2, 0.3, 0.4) to reduce number of input signals

Numerical Experiments

Step 3 – Feature selection research

1. Subsets of features created based on identified thresholds (Formica-1 – 4 subsets, Husky A200 – 4 subsets, IEEE Battery – 3 subsets)
2. 3 best models from step 1 for each data set were selected
3. 33 additional trainings performed

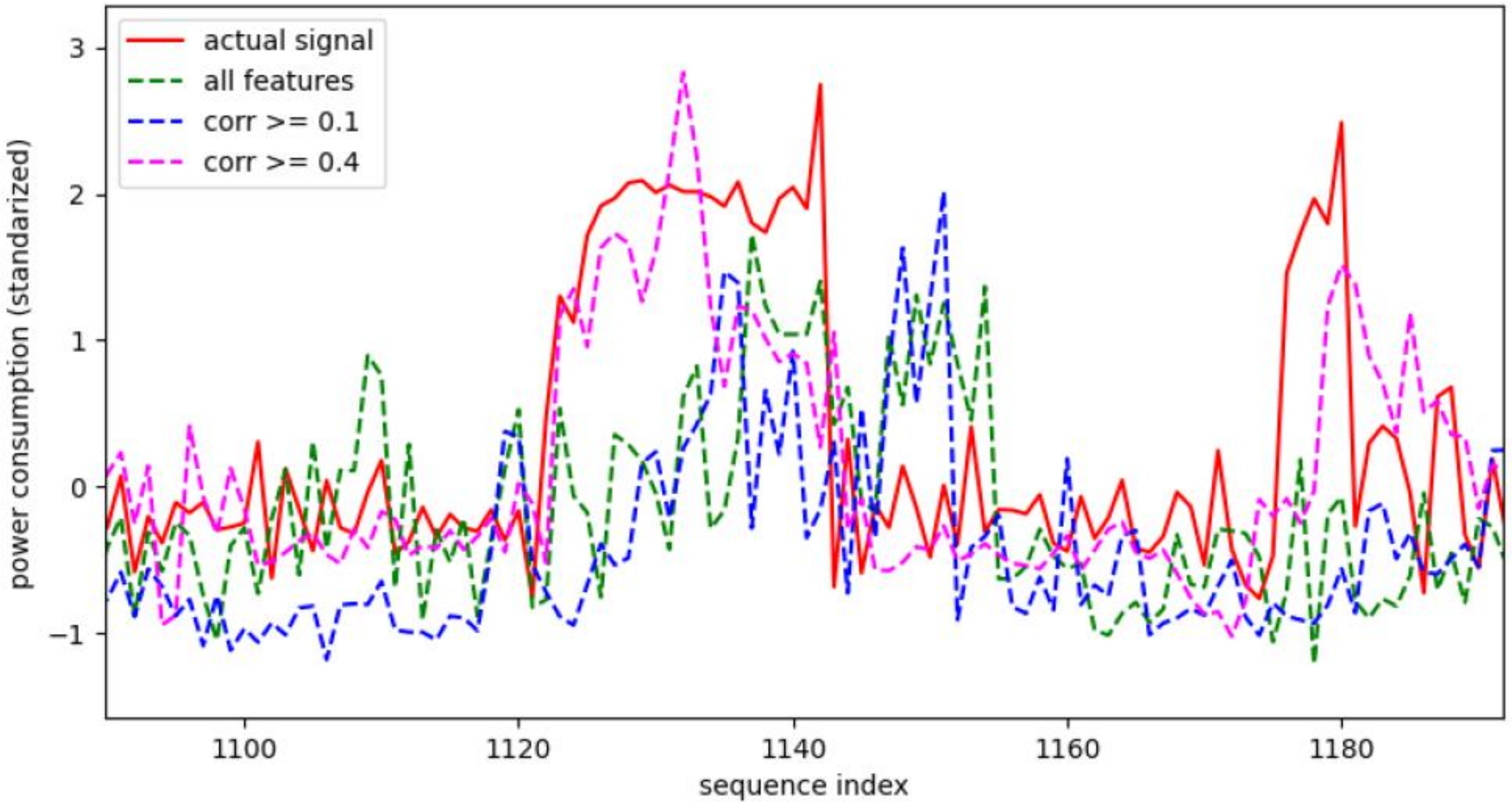
Numerical Experiments

Final results…

Model	$L_{hist.}$	Features	$\mu_{MSE}(Val)$	$\mu_{MSE}(T)$	$\mu_{MAE}(T)$	$\mu_{MAPE}(T)$	Pred. time	No. of param.
Formica-1								
LSTM 2-layers	50	all	0.0209	1.0086	0.6846	2.3460	0.0426	92330
		corr $\geq$ 0.05	0.0304	1.0162	**0.6362**	2.6156	**0.0407**	87210
		corr $\geq$ 0.1	0.0278	1.1113	0.7108	4.1627	0.0434	86890
		corr $\geq$ 0.2	0.0613	1.0511	**0.6632**	**1.7265**	**0.0400**	85290
		corr $\geq$ 0.4	0.1536	1.0481	0.7208	**2.0023**	**0.0416**	83690
LSTM 2-layers	150	all	0.0294	1.0891	0.7430	2.4750	0.0885	92330
		corr $\geq$ 0.05	0.0366	**1.0030**	**0.6237**	**2.4240**	**0.0843**	87210
		corr $\geq$ 0.1	0.0425	**1.0186**	**0.6689**	**1.9178**	**0.0849**	86890
		corr $\geq$ 0.2	0.0884	**1.0830**	**0.7341**	**1.9567**	**0.0833**	85290
		corr $\geq$ 0.4	0.1172	**1.0866**	**0.7369**	**2.0022**	**0.0825**	83690
LSTM 2-layers	170	all	0.0299	1.0799	0.7434	2.4378	0.1000	92330
		corr $\geq$ 0.05	0.0424	**1.0447**	**0.6354**	2.8822	**0.0961**	87210
		corr $\geq$ 0.1	0.0419	**0.9887**	**0.6615**	2.5382	**0.0957**	86890
		corr $\geq$ 0.2	0.0798	1.0813	**0.7381**	**2.0794**	**0.0917**	85290
		corr $\geq$ 0.4	0.1475	1.0814	**0.7371**	**2.0403**	**0.0917**	83690
Husky A200								
LSTM 2-layers	10	all	0.0711	1.1849	0.8166	6.7158	0.0546	85610
		corr $\geq$ 0.05	0.1071	**1.0667**	**0.7401**	**4.8691**	0.0554	83690
		corr $\geq$ 0.1	0.1126	**1.0491**	**0.7632**	**5.5597**	**0.0538**	81770
		corr $\geq$ 0.2	0.1450	**0.7214**	**0.6033**	**5.3452**	**0.0537**	80170
		corr $\geq$ 0.4	0.3772	**0.7142**	**0.6245**	**5.4322**	**0.0540**	79210
LSTM 2-layers	50	all	0.0749	1.4002	0.8915	7.8652	0.0993	85610
		corr $\geq$ 0.05	0.0980	**1.3614**	**0.8816**	**5.3055**	**0.0985**	83690
		corr $\geq$ 0.1	0.1016	**1.2507**	**0.8295**	**6.6086**	0.1002	81770
		corr $\geq$ 0.2	0.1243	**1.0034**	**0.7182**	**5.7255**	**0.0969**	80170
		corr $\geq$ 0.4	0.2352	**0.7015**	**0.6166**	**4.1202**	**0.0924**	79210
LSTM 2-layers	30	all	0.0880	1.5979	0.9469	7.6308	0.0767	85610
		corr $\geq$ 0.05	0.1176	**1.0802**	**0.7777**	**4.6433**	0.0772	83690
		corr $\geq$ 0.1	0.1087	**1.0650**	**0.7526**	**4.3236**	0.0783	81770
		corr $\geq$ 0.2	0.1355	**0.8857**	**0.6691**	**5.6918**	**0.0743**	80170
		corr $\geq$ 0.4	0.2176	**0.6681**	**0.5876**	**3.9296**	**0.0748**	79210
IEEE Battery								
BiLSTM 1-layer	70	all	0.0102	0.2609	0.3348	2.8338	0.4482	66890
		corr $\geq$ 0.05	0.0238	**0.2387**	**0.3003**	**2.4243**	**0.4474**	59850
		corr $\geq$ 0.3	0.0321	**0.2411**	**0.2794**	**2.3744**	**0.4327**	56650
		corr $\geq$ 0.4	0.0245	0.5294	0.3711	3.0120	**0.4395**	56010
BiLSTM 2-layers	10	all	0.0109	0.2367	0.3130	2.5488	0.2607	221770
		corr $\geq$ 0.05	0.0116	0.2883	0.3405	**2.4717**	**0.2605**	214730
		corr $\geq$ 0.3	0.0292	**0.2311**	**0.2983**	**2.3042**	**0.2595**	211530
		corr $\geq$ 0.4	0.0529	0.3973	0.3573	2.5904	**0.2582**	210890
LSTM 1-layer	70	all	0.0114	0.2328	0.3373	3.2991	0.3181	33450
		corr $\geq$ 0.05	0.0228	**0.2268**	**0.3218**	**2.4583**	**0.3069**	29930
		corr $\geq$ 0.3	0.0279	0.2642	**0.3056**	**2.3422**	**0.3005**	28330
		corr $\geq$ 0.4	0.0272	0.4957	0.3730	**3.0981**	**0.3053**	28010

Numerical Experiments

Final results…



Improvement	Formica-1	Husky A200	IEEE Battery
MSE	2%	44%	13%
MAE	9%	28%	17%
MAPE	26%	41%	19%
Prediction time	8%	7%	6%
No. of model params.	9%	7%	16%

Conclusions and future work

1. Usage of LSTM, GRU, BiLSTM, BiGRU models for energy consumption prediction
2. Forecasting works well for all data sets
3. Feature selection – generalization, processing time, and size of the final model

1. Increase forecasting results (with historical energy consumption – optional)
2. Enlarge Formica-1 data set (additional signals and scenarios)
3. Further reduce of processing time and size of the model (knowledge distillation)
4. Research on anomaly detection based on previous research

75 years
of the Silesian University
of Technology

Silesian
University
of Technology

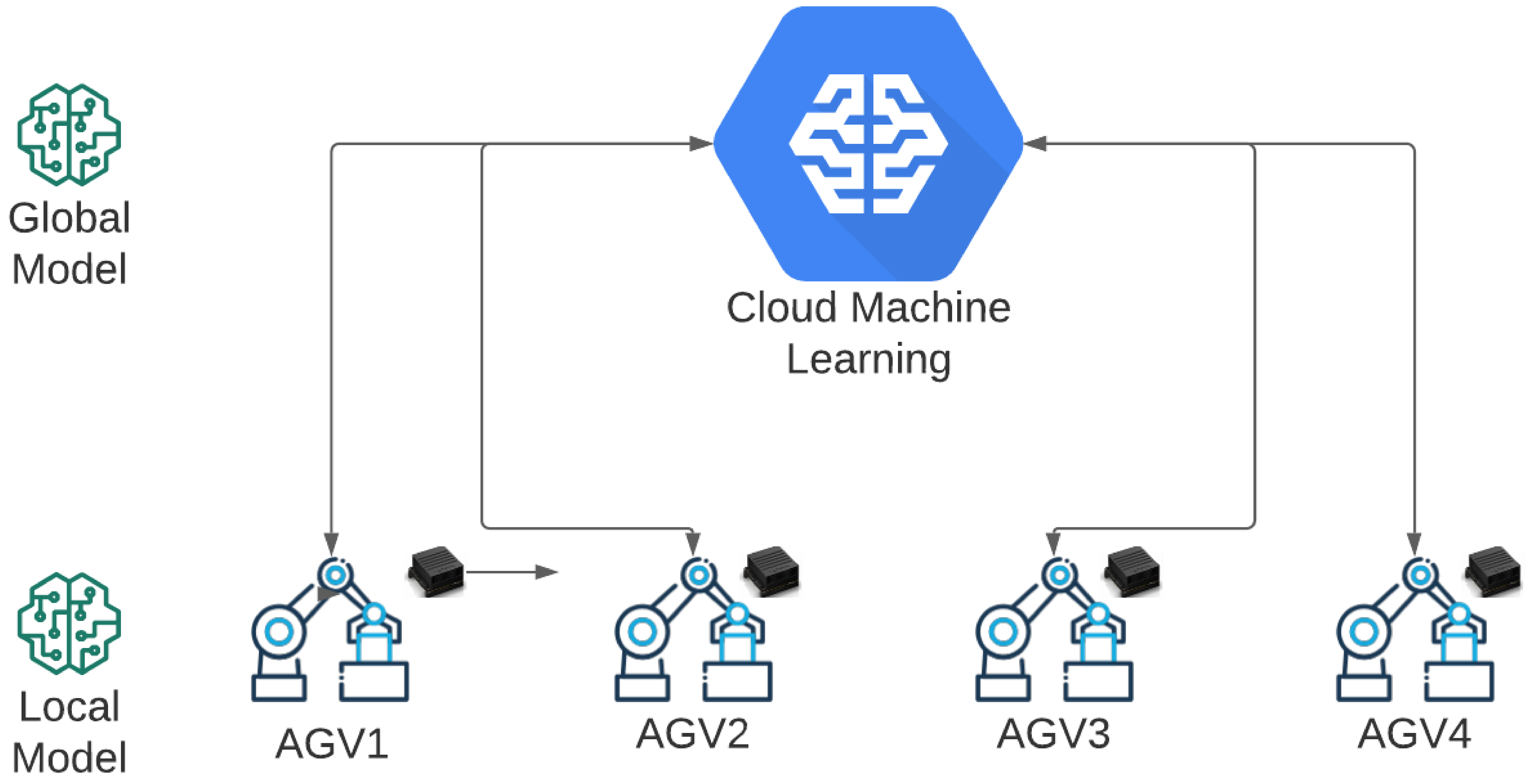
RESEARCH
UNIVERSITY
EXCELLENCE INITIATIVE

FEDERATED LEARNING FOR AGV ANOMALY DETECTION

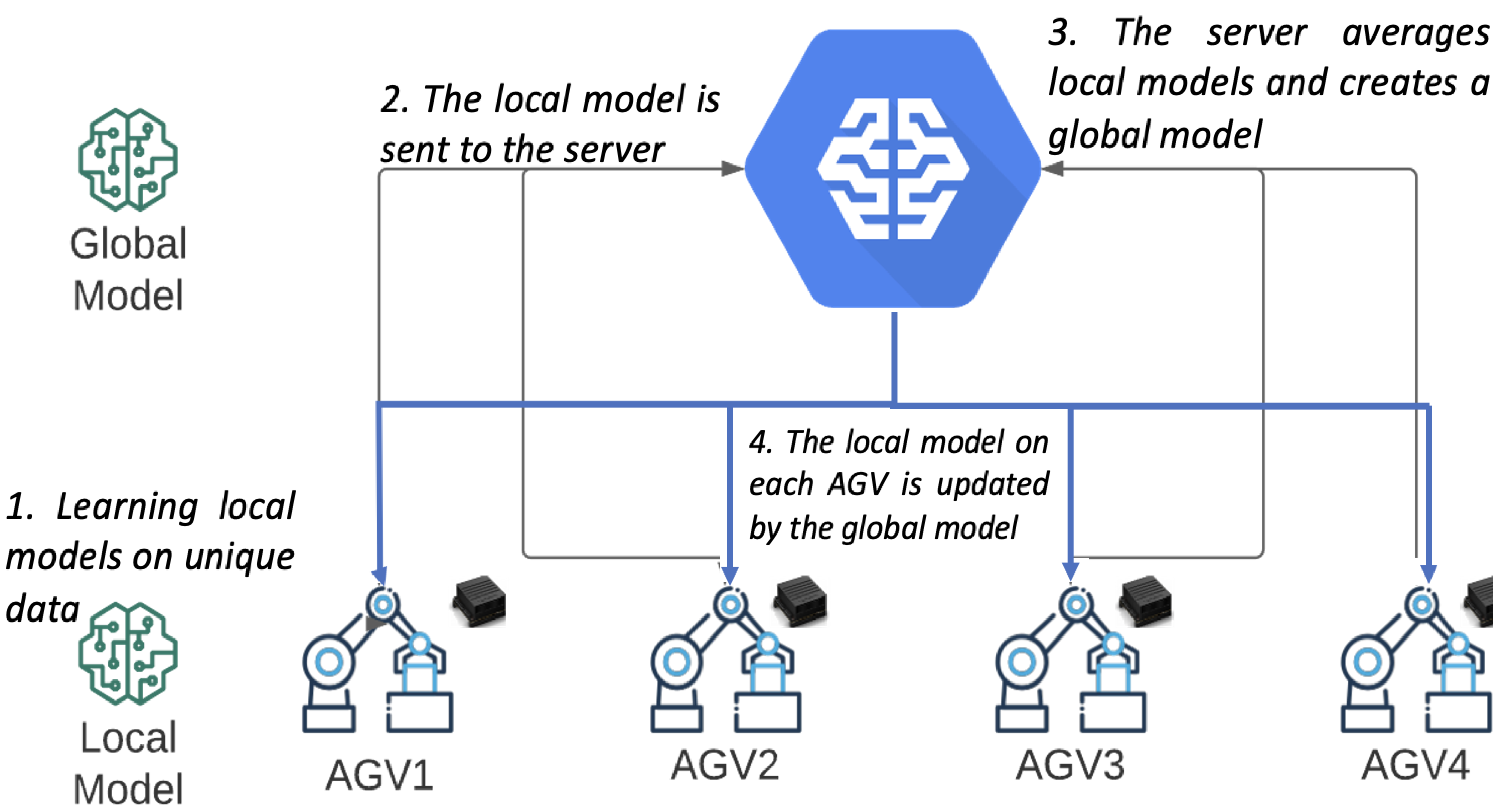
Daniel Kostrzewa, Bohdan Shubyn, Paweł Benecki,
Piotr Grzesik, Dariusz Mrozek

18.05.2022

Federated Learning for AGV Anomaly Detection



Distributed architecture with AI on the on-board IoT devices



Algorithm 1: Algorithm of the round

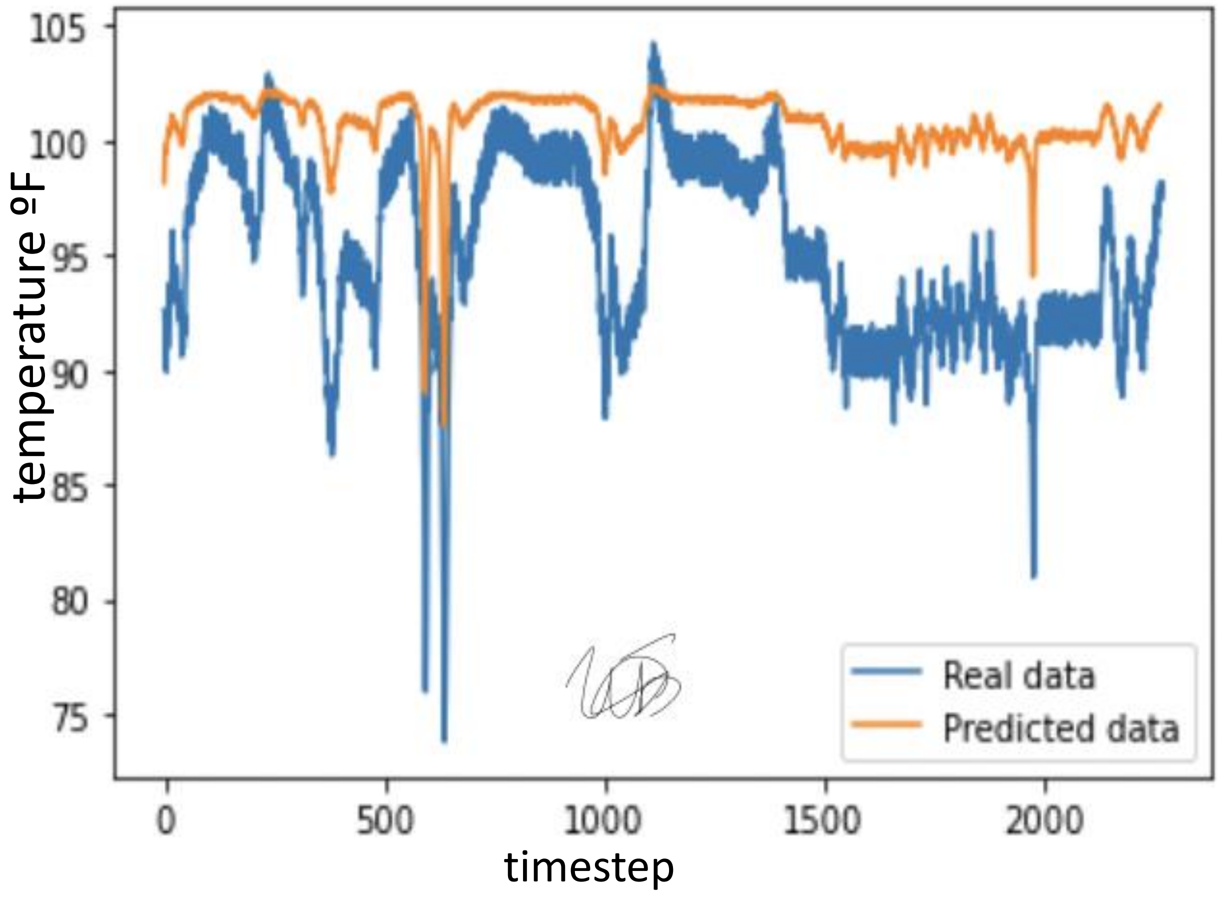
Data: lm (Local models on AGVs), gm (Global model), $AGVs$ (the fleet of AGVs), N (the number of AGVs), sgm (Server with a global model)

Result: $upAGVs$ (AGVs updated by global model)

```
for i ← 1 to AGVs do
    Train the RNN of AGV_i locally on unique, AGV-specific data;
    lm ← weights of the local RNN;
end
foreach lm ∈ AGVs do
    Send lm to the sgm;
end
Build the gm by averaging lms on the sgm;
foreach lm ∈ AGVs do
    Update lm by the sgm;
end
```

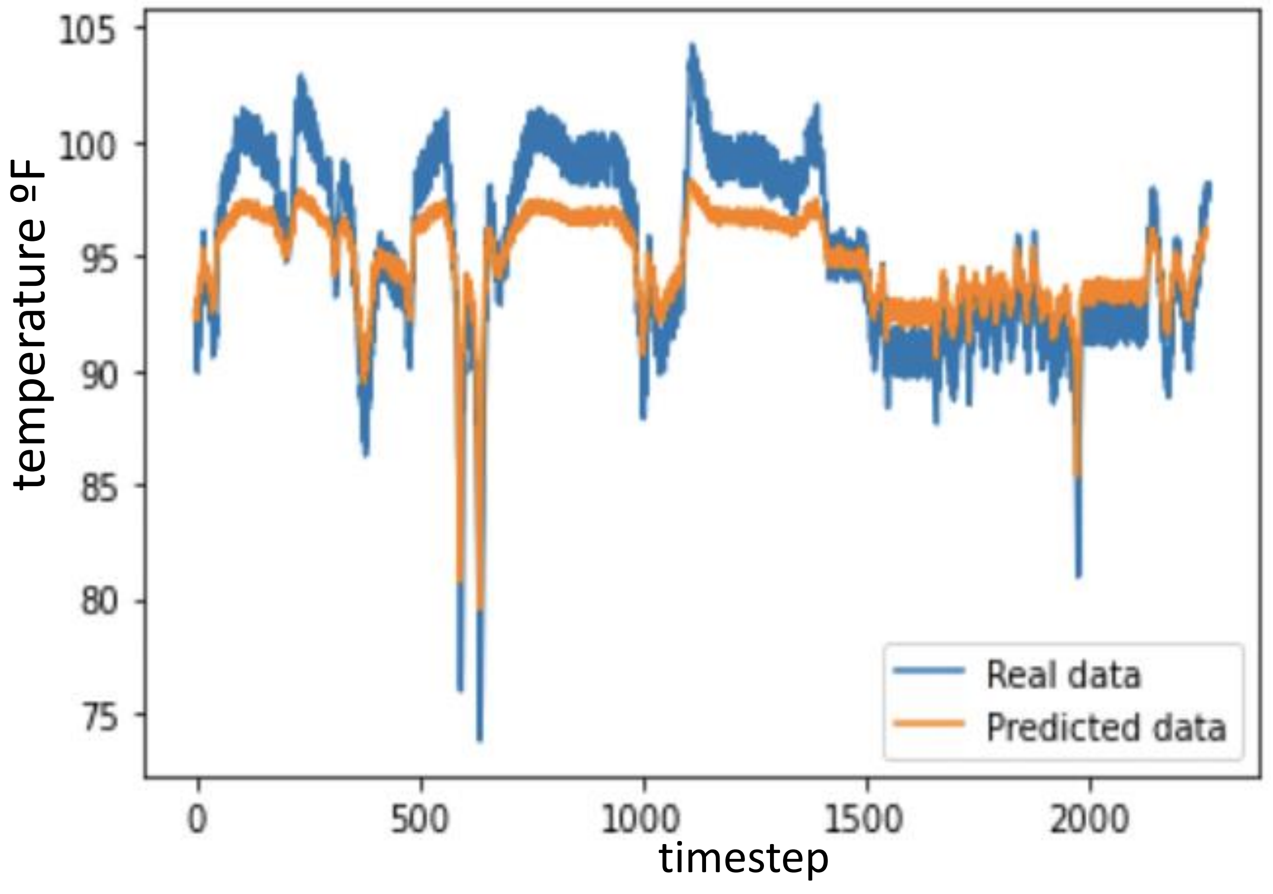
Experiment

Mean Squared Error (MSE) = 39.97
Mean absolute percentage error (MAPE) = 6.02 %
Root Mean Squared Error (RMSE) = 6.32



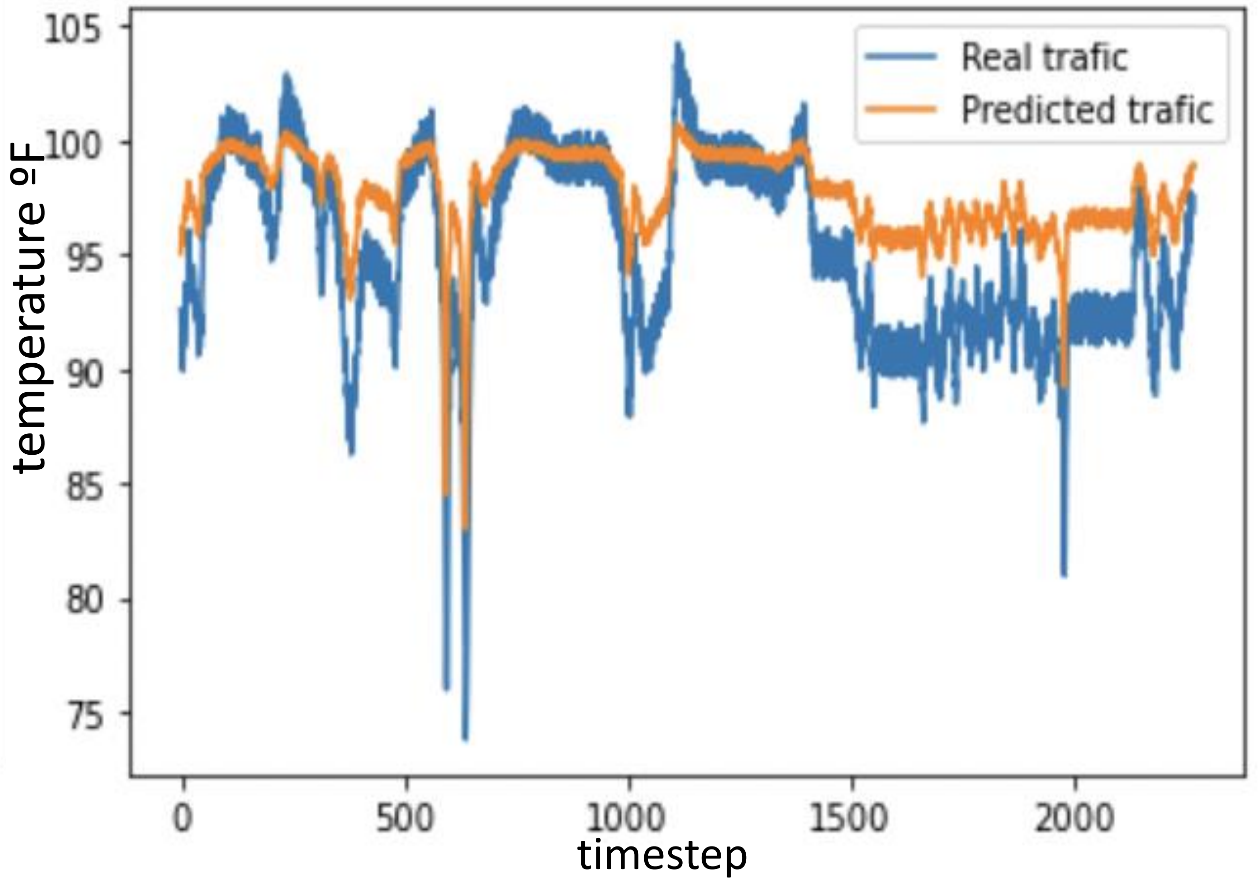
Effectiveness of the local model for the virtual client 1

Mean Squared Error (MSE) = 4.74
Mean absolute percentage error (MAPE) = 1.91
Root Mean Squared Error (RMSE) = 2.18



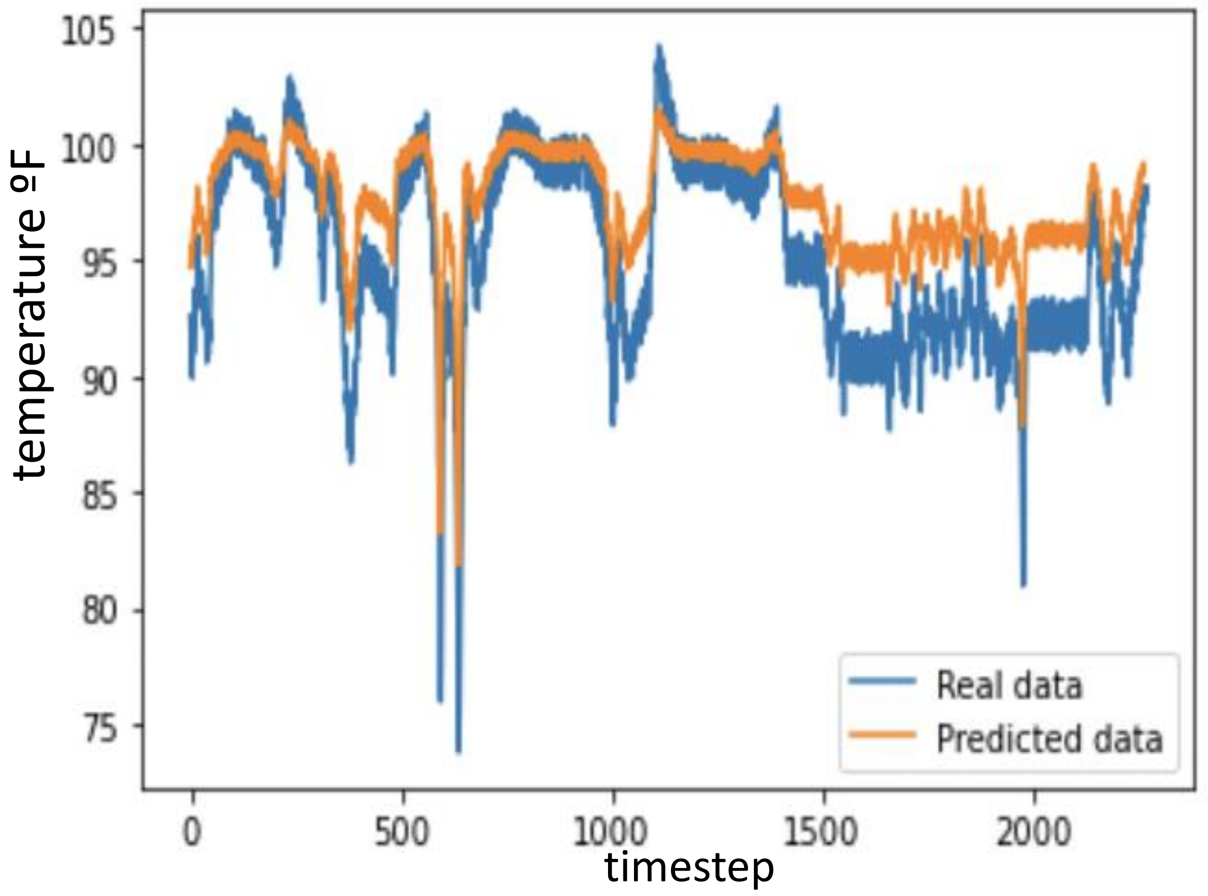
Effectiveness of the local model for the virtual client 2

Mean Squared Error (MSE) = 11.81
Mean absolute percentage error (MAPE) = 3.02
Root Mean Squared Error (RMSE) = 3.44



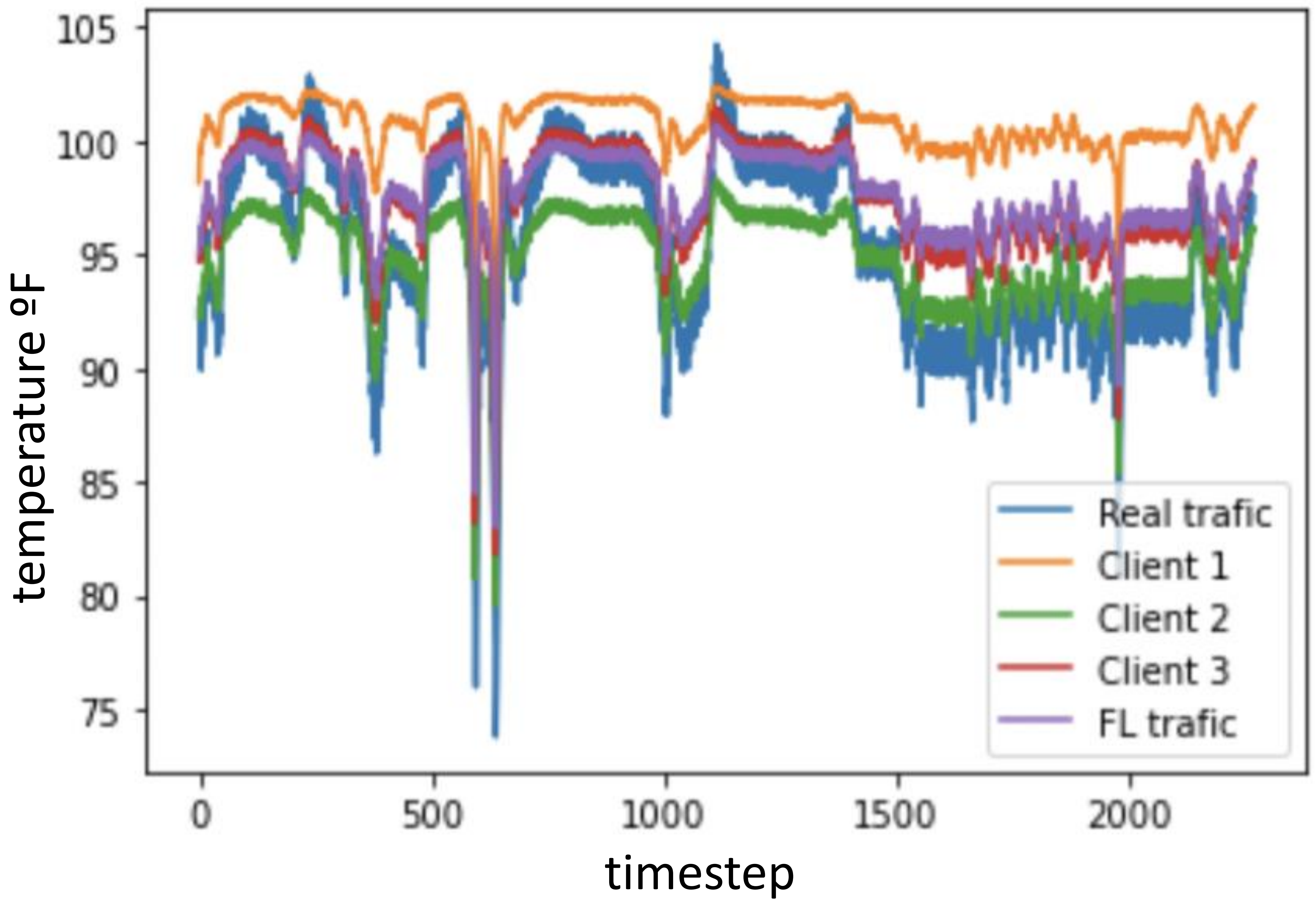
Effectiveness of the local model for the virtual client 3

Mean Squared Error (MSE) 9.43
Mean absolute percentage error (MAPE) = 2.74 %
Root Mean Squared Error (RMSE) = 3.07



Effectiveness of the global model

Experiment



Client/Metric	MSE	MAPE	RMSE
Local Model Client 1	**39,97**	6.02%	6,32
Local Model Client 2	**4,74**	1.91%	2,18
Local Model Client 3	**11,81**	3.02%	3,44
Global model (FL)	**9,43**	2.4%	3,07

75 years
of the Silesian University
of Technology

Silesian
University
of Technology

RESEARCH
UNIVERSITY
EXCELLENCE INITIATIVE

FEDERATED LEARNING AND DEEP LEARNING FOR AGV ANOMALY DETECTION

THANK YOU FOR YOUR ATTENTION

18.05.2022