TU Clausthal

Evaluating the Impact of Data Sampling Rates on Event Detection Accuracy in Load Signatures Using a Shapelet based Approach

1st August 2019

Abdul Rahman Hakmeh

First assessor: Dr.-Ing. habil. Andreas Reinhardt

Second assessor: Prof. Dr. Thorsten Grosch

Roadmap and Goals

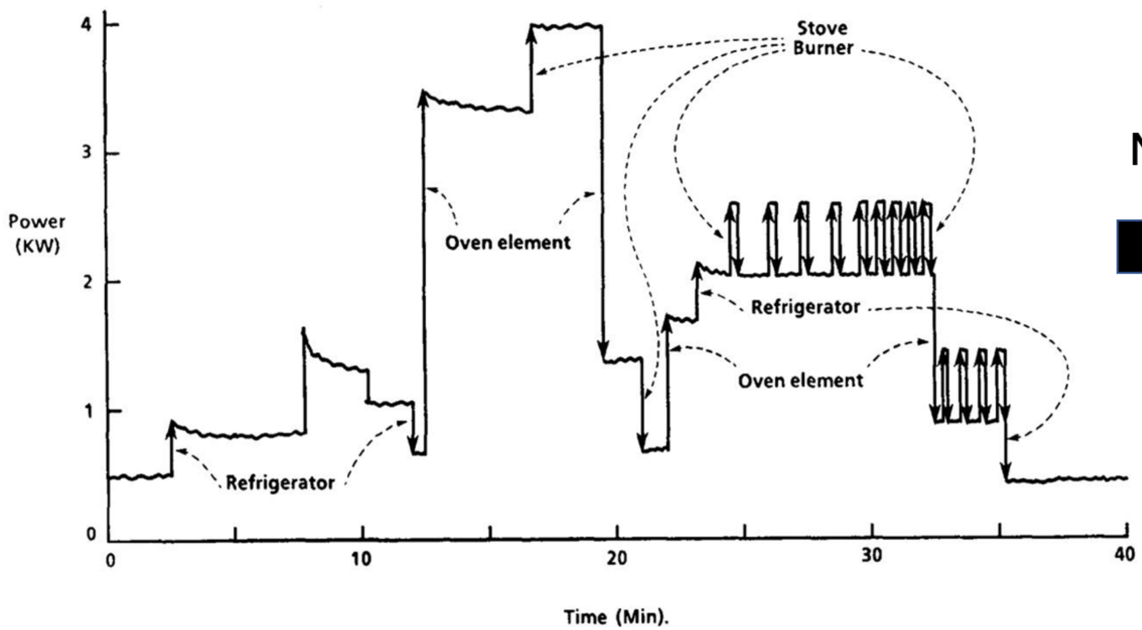
- Introduction
  - Non-intrusive Load Monitoring (NILM)
  - Research Question
- Methodology
  - Fast Shapelet Approach
  - Data Pre-processing
  - Model Training
- Experiments Results
  - Results Phase A
  - Results Phase B
- Conclusion
  - Future Work

TU Clausthal

INTRODUCTION

Non-intrusive Load Monitoring (NILM)

- Recording and analyzing the load measurements in the power system.
- Recognizing the power consumption and the electrical device’s state.
- Extracting the signatures of the appliances.



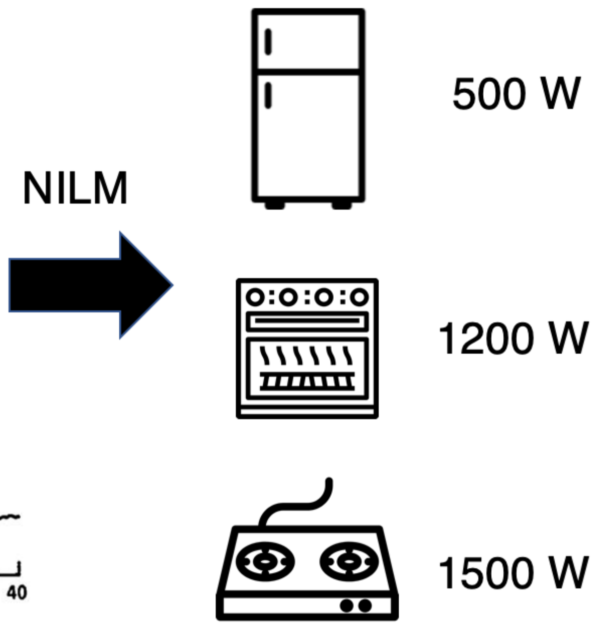


Image source: HART, George William. Nonintrusive appliance load monitoring. Proceedings of the IEEE, 1992.

Non-intrusive Load Monitoring

- Data Acquisition (DQ).
- Appliance Event Detection.
- Appliance Classification.
- Energy Disaggregation.

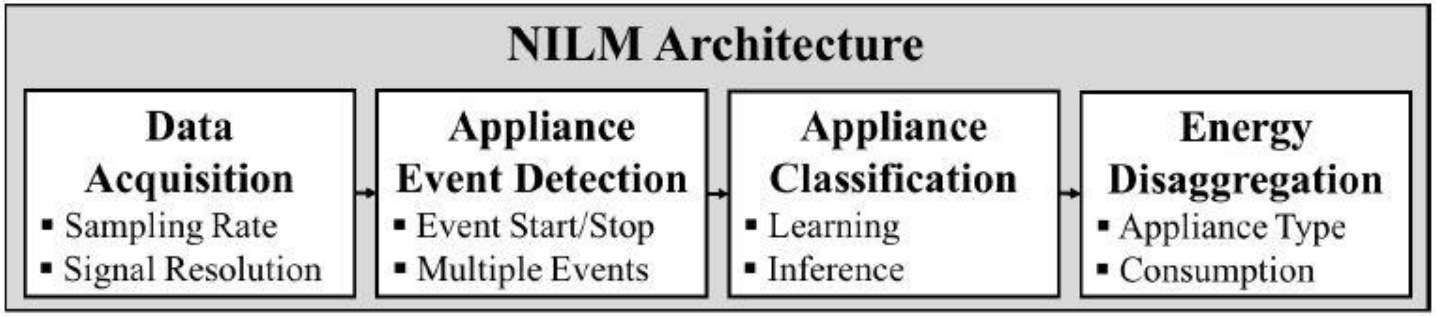


Image source: Armel, K. Carrie, et al. "Is disaggregation the holy grail of energy efficiency? The case of electricity." *Energy Policy* 52 (2013).

Research Question

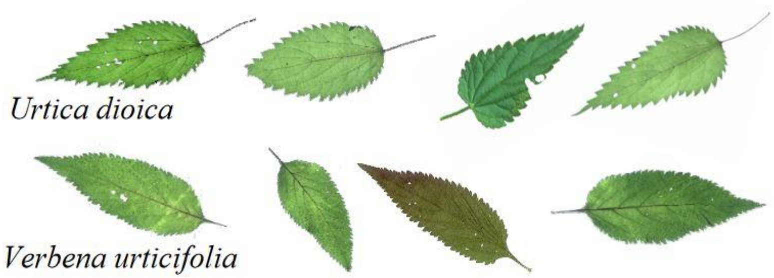
The specific research question is:

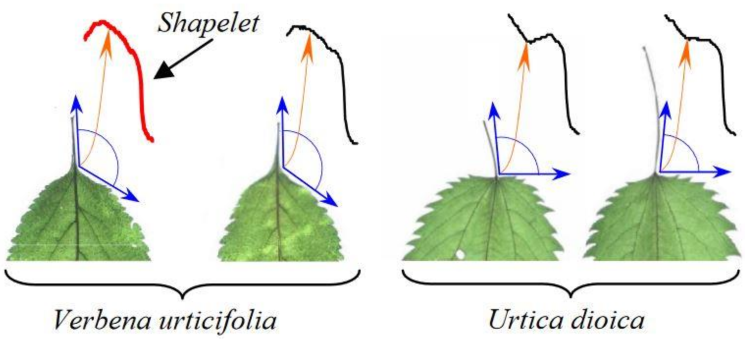
" Is it achievable to detect appliance switching on or off events in downsampled data without sacrificing the detection accuracy ? And how does the low sampling rate affect the event detection accuracy ? “

METHODOLOGY

Fast Shapelet Approach

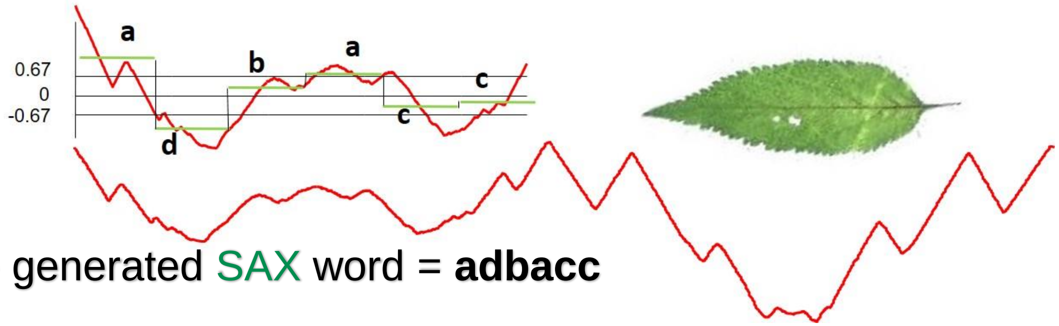
- Shapelet is a subsequence of a time series that has high discriminative power.
- Fast Shapelet makes the search more efficient through reducing the dimensional space of the time series by using Symbolic Aggregate Approximation (SAX) representation $O(n^2m^3) \rightarrow O(nm^2)$.
- A Decision Tree Classifier is built to classify the subsequences of the unseen data.
- **Work method based on an exempel:**
    - The task is to classify two very similar types of leaves.





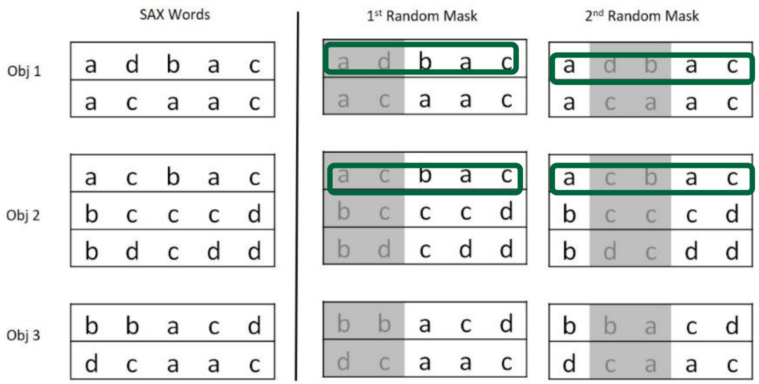
Fast Shapelet Approach

- **Generating SAX words:** By using Symbolic Aggregate Approximation.



The generated SAX word = **adbacc**

- **Random masking:** Using the multi masking resolves the problem of false dismissals:



The random masking makes **a**<u>d</u>**bac** and **a**<u>c</u>**bac** identical.

Image source: T. Rakthanmanon and E. Keogh. “Fast shapelets: A scalable algorithm for discovering time series shapelets.” In: proceedings of the 2013 International Conference on Data Mining.

Fast Shapelet Approach

- **Counting similar objects & Finding the best candidates:** By calculating the distinguishing score of every object from the collision table:
  - A is the final collision table.
  - B is the sum of objects-based counts.
  - C represents the complementary data of table B.
  - D is the distinguishing score of each SAX word.

A)

	Class1		Class2	
	Obj 1	Obj 2	Obj 3	Obj 4
1	5	5		
2	5	1	1	1
3	5	3		
4		5	1	1
5		5		5
6	1		5	3
7	3		5	2

B) Close to Ref

	Class1	Class2
1	10	0
2	6	2
3	8	0
4	5	2
5	5	5
6	1	8
7	3	7

C) Far from Ref

	Class1	Class2
1	0	10
2	4	8
3	2	10
4	5	8
5	5	5
6	9	2
7	7	3

D) Distinguishing Power

(10-0)+(10-0) = 20
(6-4)+(8-2)=8
(8-2)+(10-0)=16
(5-5)+(8-2)=6
(5-5)+(5-5)=0
(9-1)+(8-2)=14
(7-3)+(7-3)=8

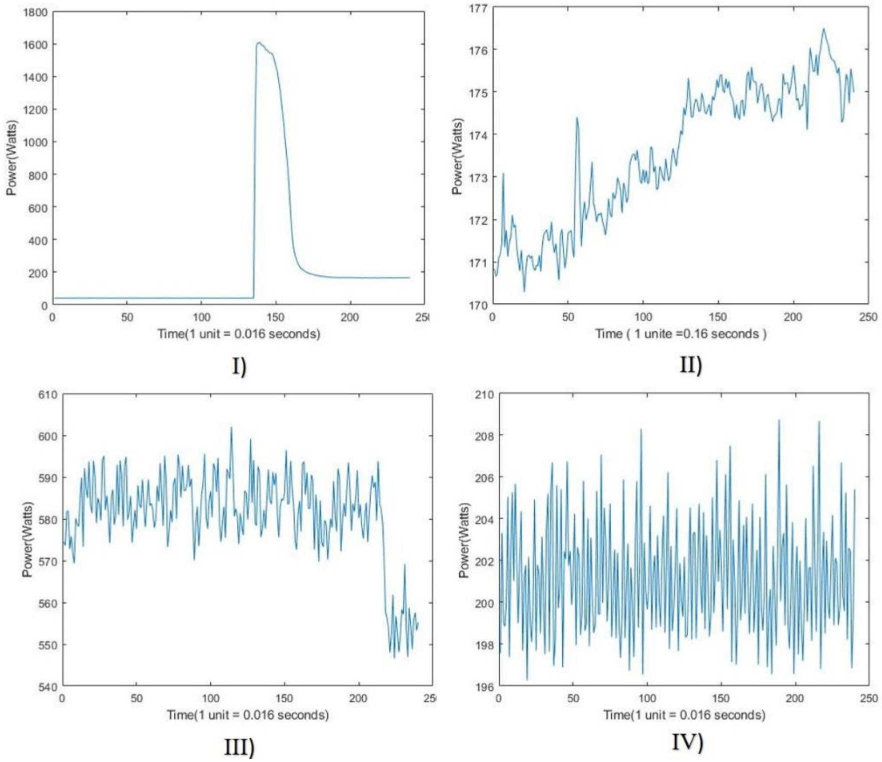
Image source: T. Rakthanmanon and E. Keogh. "Fast shapelets: A scalable algorithm for discovering time series shapelets." In: proceedings of the 2013 International Conference on Data Mining.

Data Pre-processing

- **Data Source**: Building-Level fUlly-labeled dataset for Electricity Disaggregation (BLUED).
  - BLUED includes one week of the current and the voltage measurements for a family house in Pennsylvania, USA.
  - BLUED is the de facto dataset for the NILM benchmarks.
  - BLUED is annotated with the events (switching on/off) time stamp and the triggering appliance.
  - Due to the electrical wiring system in USA, BLUED has two phases A and B.
  - Every phase will be divided into training set and testing set.

Data Pre-processing

- **Fixed length window technique** is used to extract sequences from data which represent events or non-events.

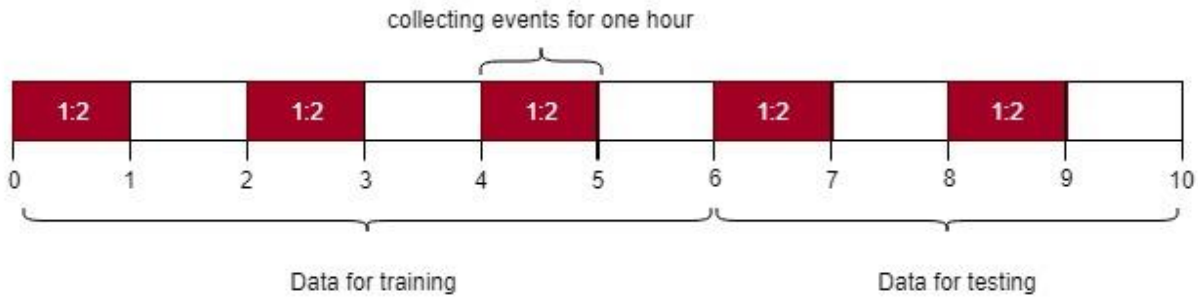


Four samples of the training instances; figure I and II represent respectively event and non-event segments from phase A, while III and IV draw event and non-event segments from phase B.

Data Pre-processing

- Processing 56 GB of raw data increases the computational complexity.

- Reducing the collected data to improve the performance by applying 3 filters on the segments extracting process:



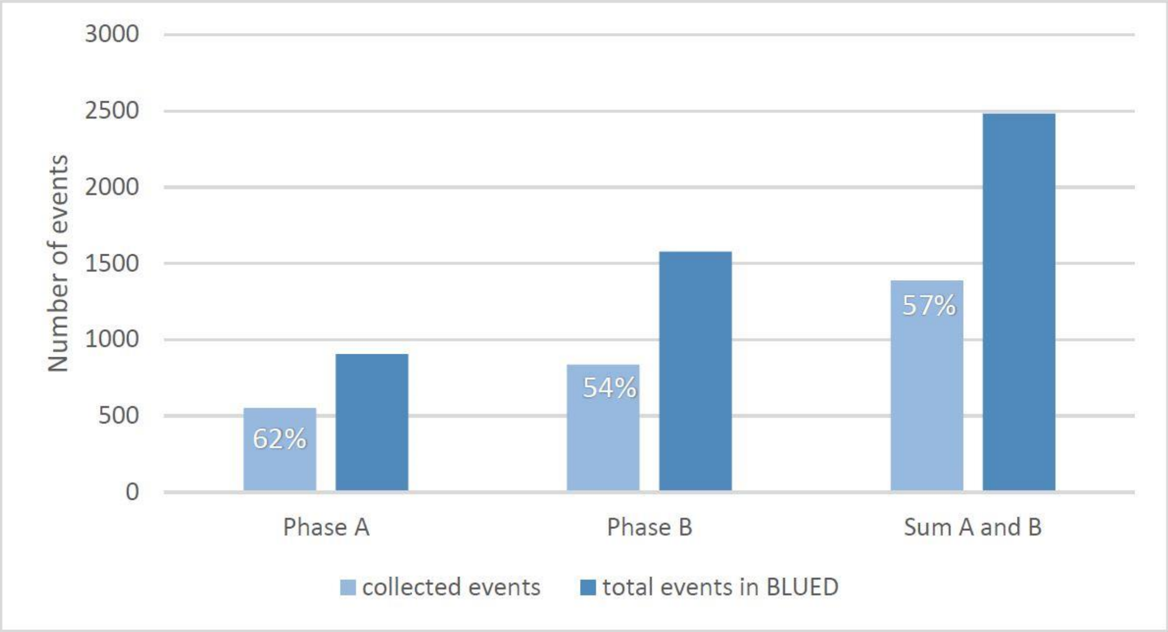
  - Splitting the data into training set and testing set according to 60/40 rule.

  - Skipping every other hour during the collecting process.

  - Collecting data according to the manner 1:2 ratio of events: non-events.

Data Pre-processing

- **Raw data covering percent:**



The covering percent of the collected events by the fixed length window technique in comparison to overall events in the BLUED.

Model Training

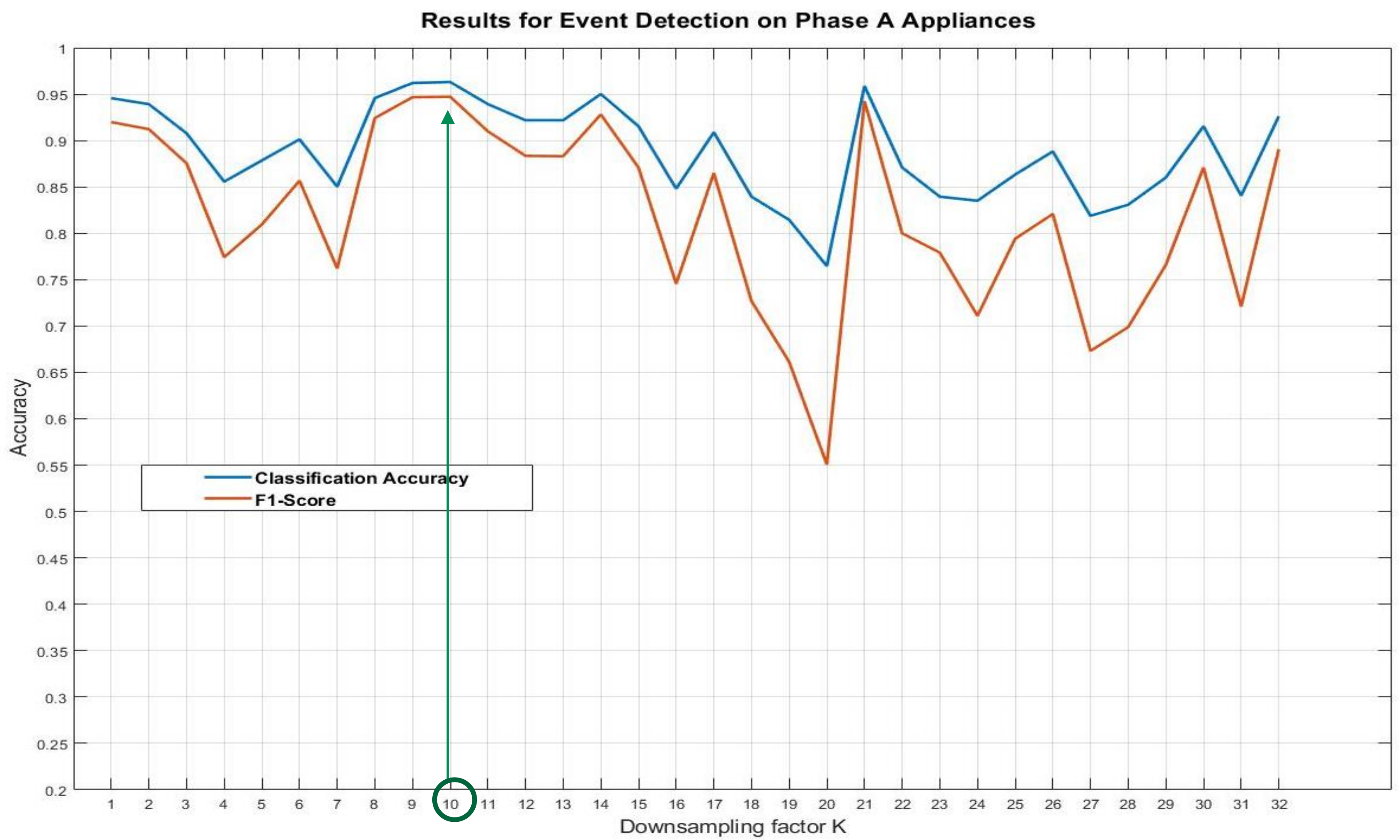
- The proposed stages to train the model und collect the results after testing with unseen data:

- Extracting segments for the training and the testing.
- Downsampling the segments by factor $K \in [1, 32]$.
- Training the Shapelet based classifier with the training data.
- Feeding the trained classifier with the unseen data (the testing set).

EXPERIMENTS RESULTS

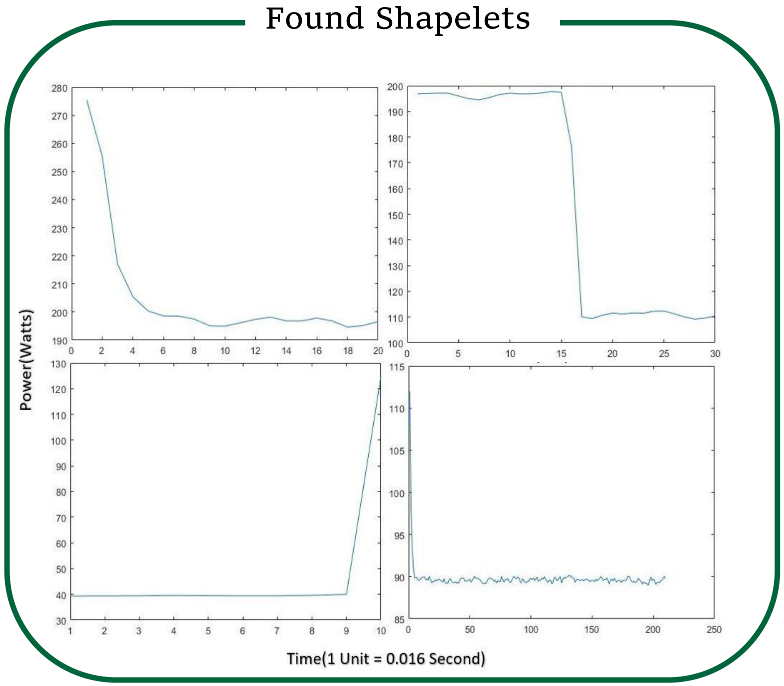
TU Clausthal

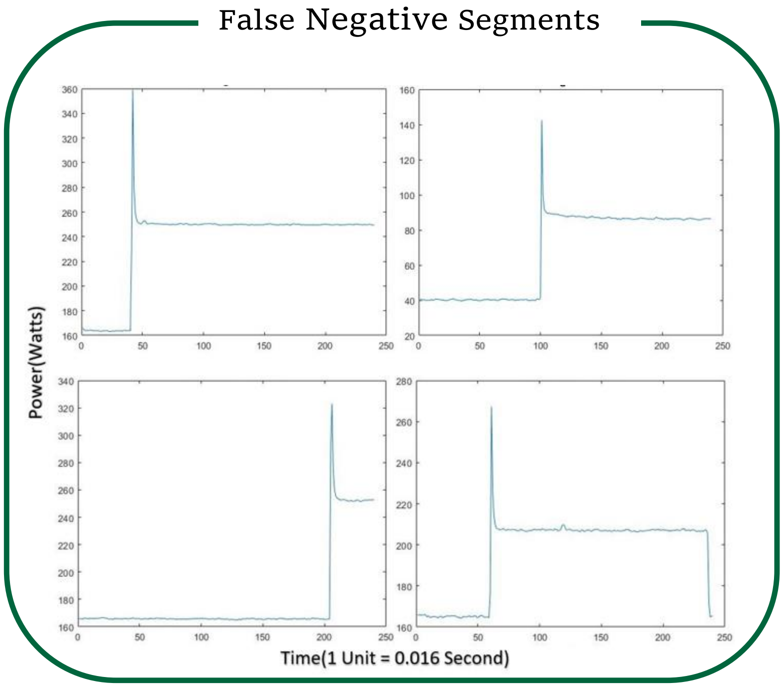
Results Phase A



Results Phase A

- **Results of non-downsampled instances:**



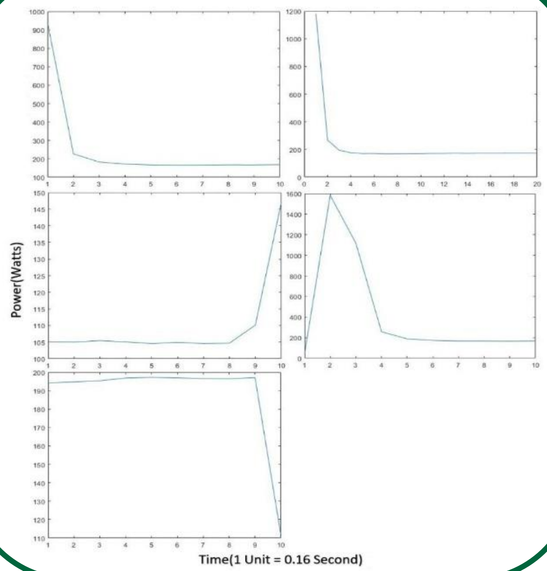


Accuracy	F1-Score	Confusion Matrix			
94.58 %	0.92	real		event	non-event
		pred	event	278	0
		pred	non-event	50	585

Results Phase A

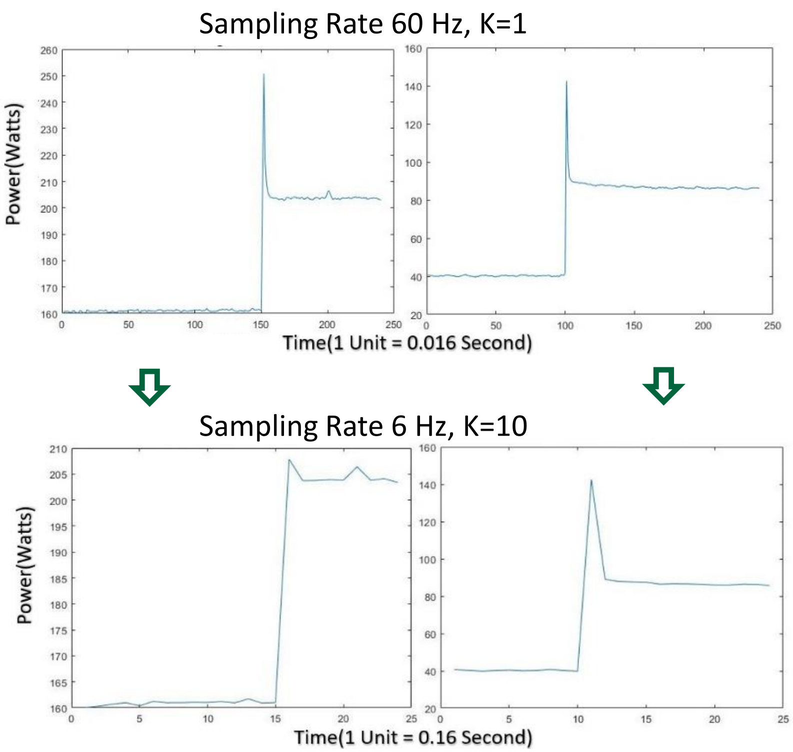
- **Results of downsampled instances by factor 10:**

Found Shapelets

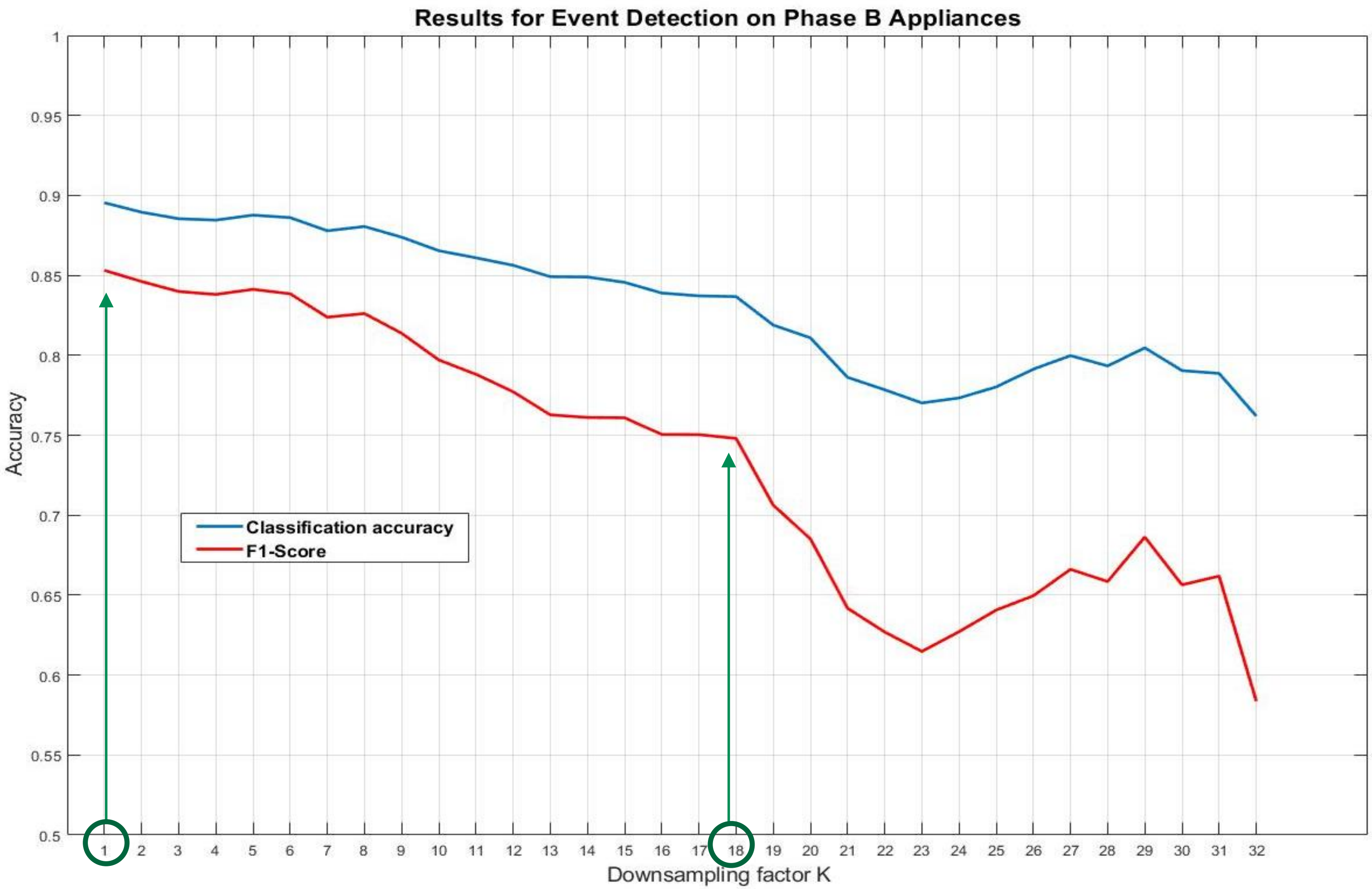


Accuracy	F1-Score	Confusion Matrix		
		real	event	non-event
96.31 %	0.94	pred event	305	2
		pred non-event	32	583

False Negative to True Positive



Results Phase B

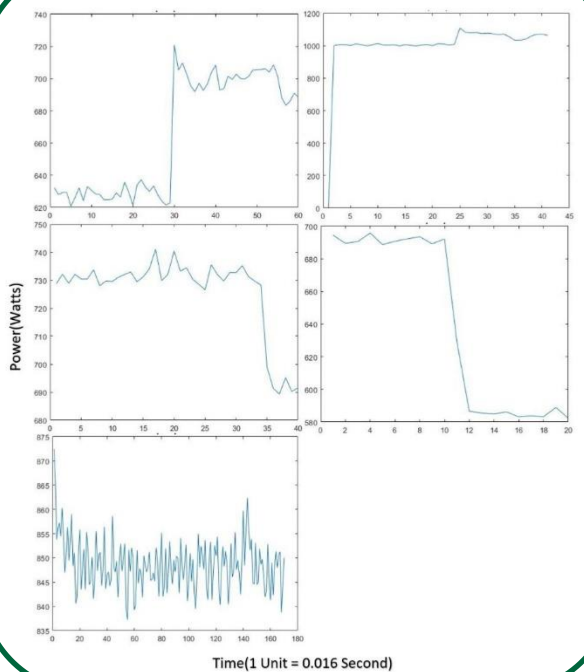


TU Clausthal

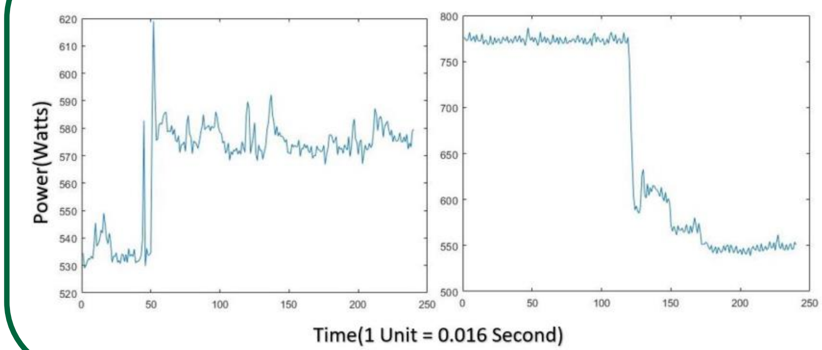
Results Phase B

- **Results of non-downsampled instances:**

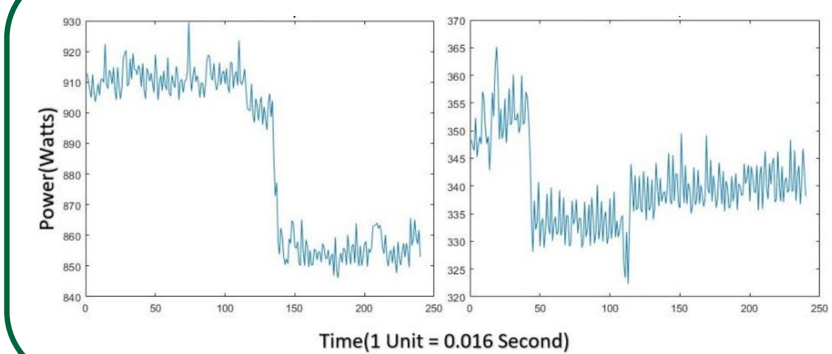
Found Shapelet



True Positive Segments



False Negative Segments

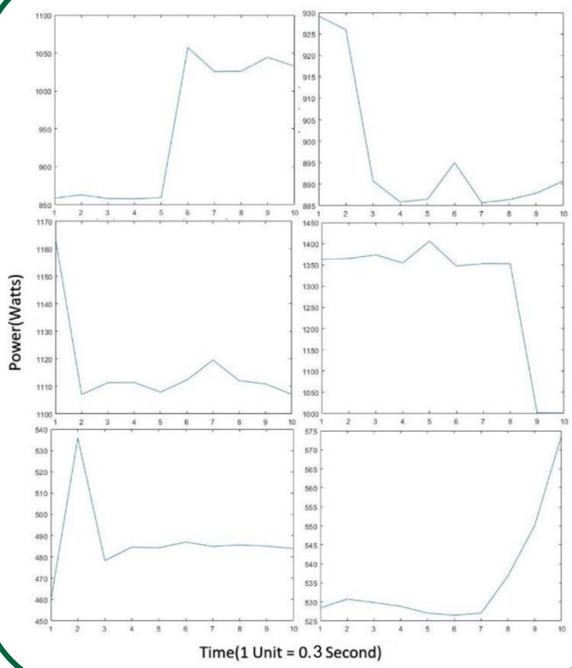


Accuracy	F1-Score	Confusion Matrix			
		real		event	non-event
89.54 %	0.85	pred	event	273	28
		pred	non-event	66	532

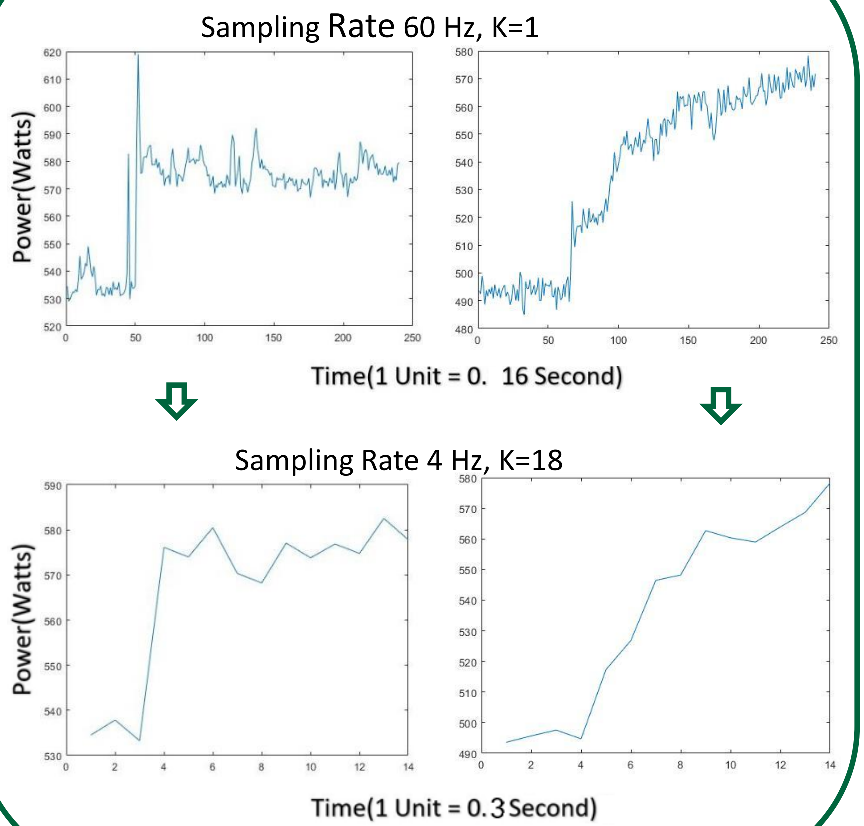
Results Phase B

- **Results of downsampled instances by factor 18:**

Found Shapelet



True Positive to False Negative



Accuracy	F1-Score	Confusion Matrix		
		real	event	non-event
82.75 %	0.74	pred event	215	31
		pred non-event	124	529

CONCLUSION

Conclusion

- The research question:

" Is it achievable to detect appliance switching on or off events in down sampled data without sacrificing the detection accuracy ? and how does the low sampling rate affect the event detection accuracy ? "

- **It is possible to detect events in downsampled data without sacrificing the detection accuracy.**
- The fast Shapelet approach generally performs better at the low sampling frequency with the clean data than the noisy data.
- The sampling rate is linearly to the model accuracy on the noisy data. At low sampling frequency the detection accuracy decreases.
- **Positive impact at the forefront of the NILM field, as collecting the data at low sample can save the recording, storage, and financial costs of the data acquisition phase.**

Future Work

- **Short Term Research:**
  - Additional training method to ensure the predictive performance of the model e.g. the cross-validation.
  - Another real-world dataset could be chosen, that replaces the BLUED and has an acceptable degree of noise.
- **Long Term Research:**
  - Another event detection approaches could be researched with considering the impact of utilizing different models.

Any questions?

Abd Hakmeh
ahak15@tu-clausthal.de

Thank you for your attention !!