ANOMALY DETECTION IN BIOSENSOR WAVEFORMS

DATA 599 | MIDTERM UPDATE

JUSTINE FILION | NEETHU GOPALAKRISHNA | SAISREE GR | SARA HALL

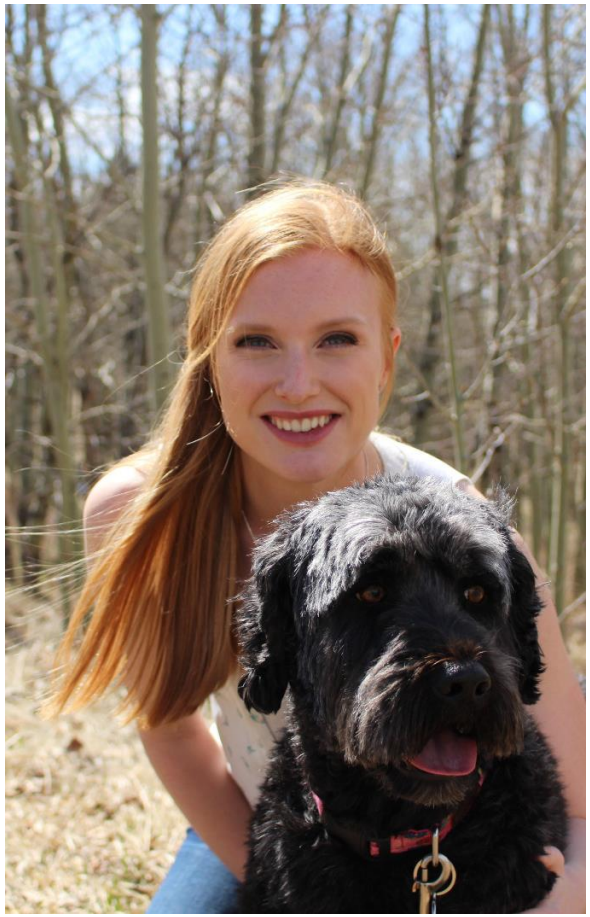
AGENDA

1. About Us
2. About Siemens Healthineers
3. Project Details
4. Our Approach
5. Results and Roadblocks

NEETHU GOPALAKRISHNA

- B.E. in Computer Science
- Grew up in India.
- Loves to read, cook and dance.

SARA HALL



- BSc in Neuroscience
- Grew up in Calgary
- Loves running and cycling…especially uphill

SAISREE GR

- B.E. in Electronics and Communication.
- Worked as a Senior Systems Engineer
- Loves hot chocolate and singing!

JUSTINE FILION

- BSc in Actuarial Science
- Grew up in northern Quebec
- Loves skiing and painting.

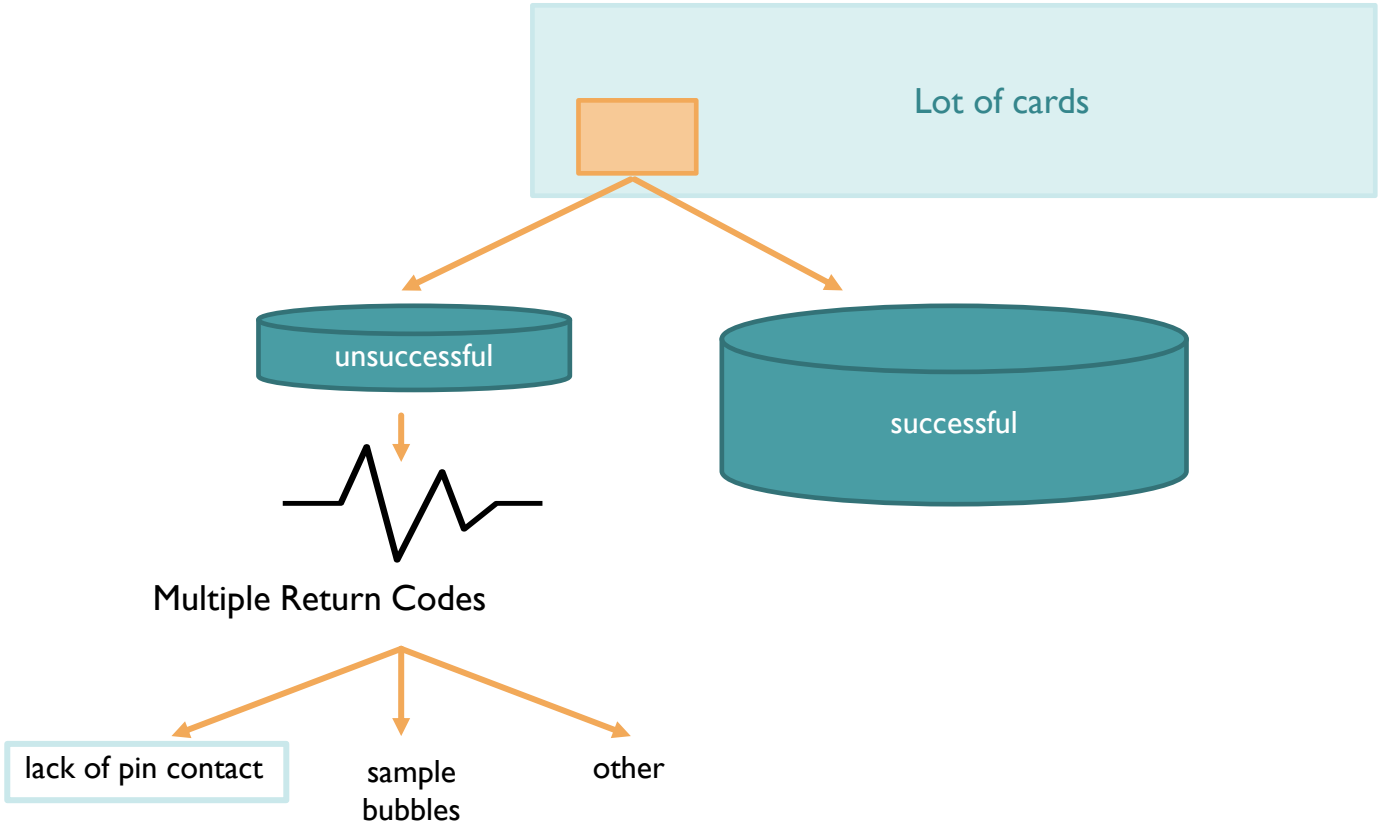
SIEMENS HEALTHINEERS

Medical device company based in Germany

Create devices that help inform clinical decision making

EPOC BLOOD ANALYSIS SYSTEM

QUALITY CONTROL PROCESS

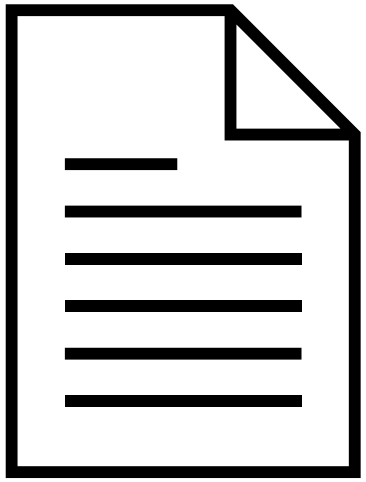


AIMS AND OBJECTIVES

DEVELOP MACHINE LEARNING PIPELINES TO CLUSTER READINGS

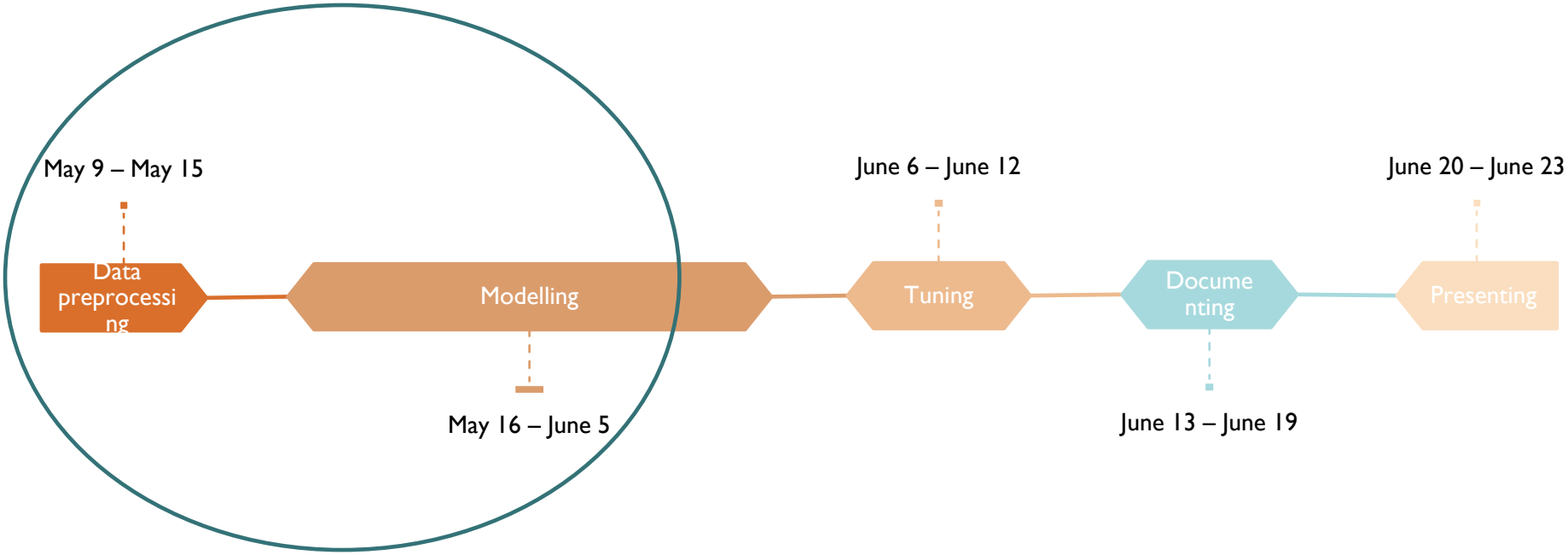
DETERMINE WHICH METHODS ARE EFFECTIVE AND WHICH ARE NOT FOR IDENTIFYING ANOMALIES IN BIOSENSOR READINGS

DELIVERABLES



- Well commented Python code for everything we have tried
- A final report detailing our attempts

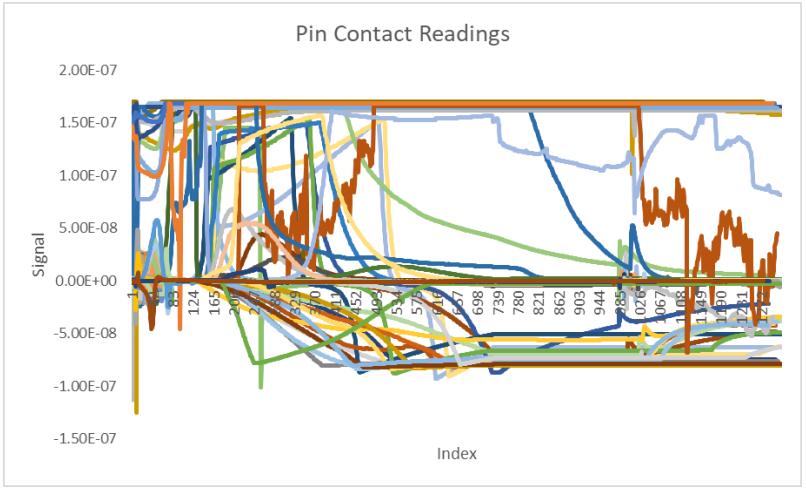
SCHEDULE

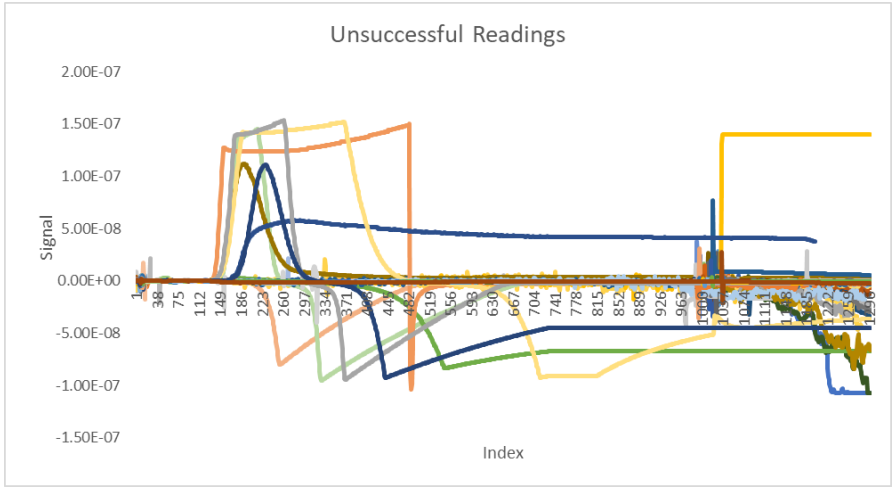

May 9 – May 15
June 6 – June 12
June 20 – June 23
Data preprocessing
Modelling
Tuning
Documenting
Presenting
May 16 – June 5
June 13 – June 19

OUR DATA

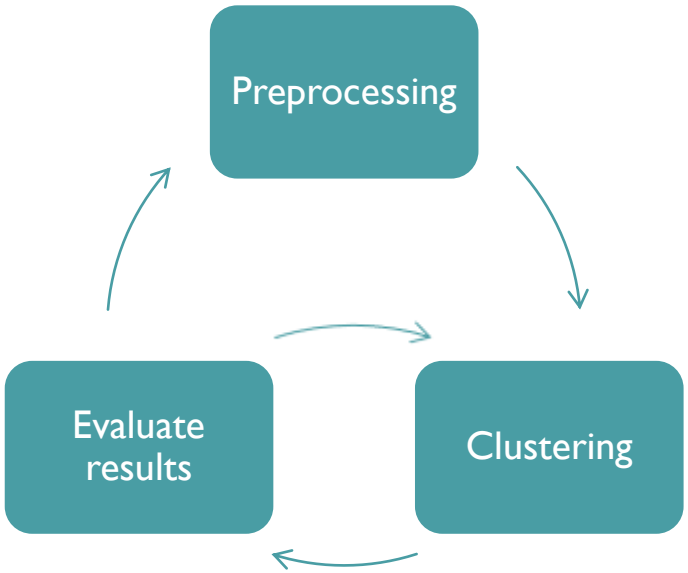

Successful Readings
4.00E-08
3.00E-08
2.00E-08
1.00E-08
0.00E+00
-1.00E-08
-2.00E-08
-3.00E-08
-4.00E-08
Signal
1
50
99
148
197
246
295
344
393
442
491
540
589
638
687
736
785
834
883
932
1079
1128
1177
1226
1275
Index

OUR DATA

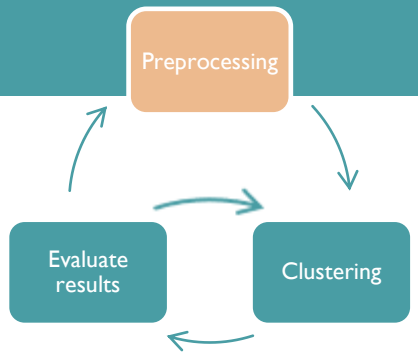

Pin Contact Readings
2.00E-07
1.50E-07
1.00E-07
5.00E-08
0.00E+00
-5.00E-08
-1.00E-07
-1.50E-07
Signal
Index


Unsuccessful Readings
2.00E-07
1.50E-07
1.00E-07
5.00E-08
0.00E+00
-5.00E-08
-1.00E-07
-1.50E-07
Signal
Index

METHODS OVERVIEW

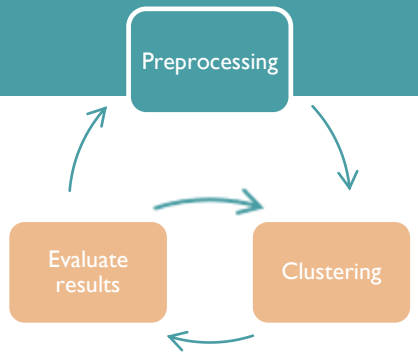


METHODS – ITERATION 1



1. Removed uninformative window
2. Standardized the waveform
3. Split the waveform into different windows

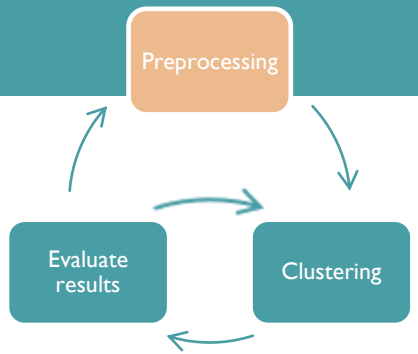
METHODS – ITERATION 1

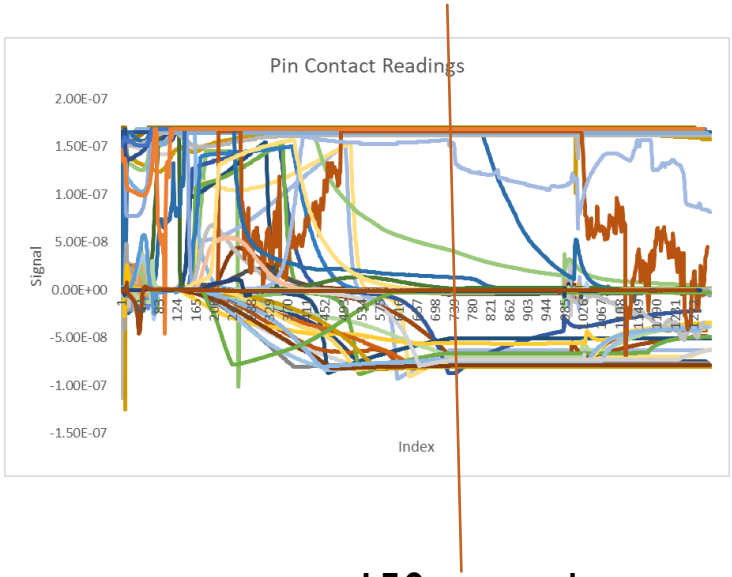


1. Principal Component Analysis
2. Autoencoder
3. Self-Organizing Map
4. Feature extraction using TSFRESH

METHODS – ITERATION 2

Removed uninformative window → Normalized → Smoothed

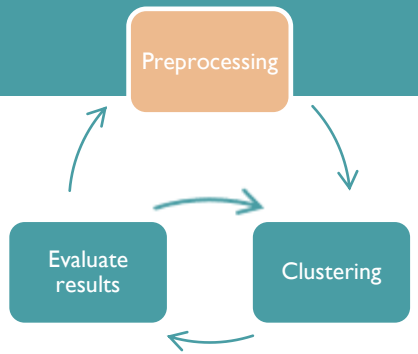


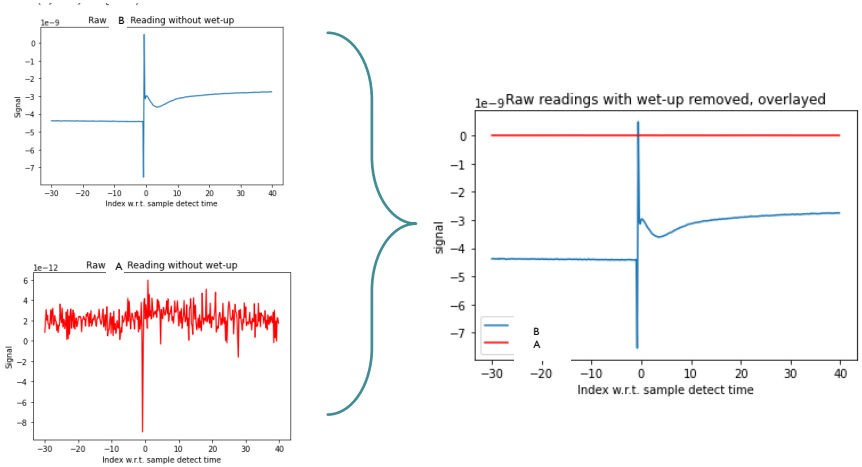


150 seconds

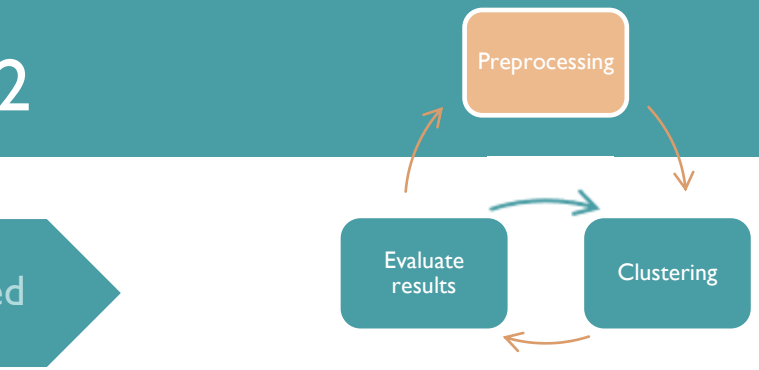
METHODS – ITERATION 2

Removed uninformative window → Normalized → Smoothed

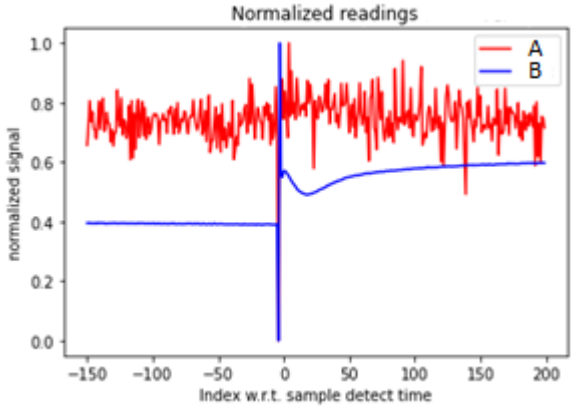




METHODS – ITERATION 2



Removed uninformative window → Normalized → Smoothed



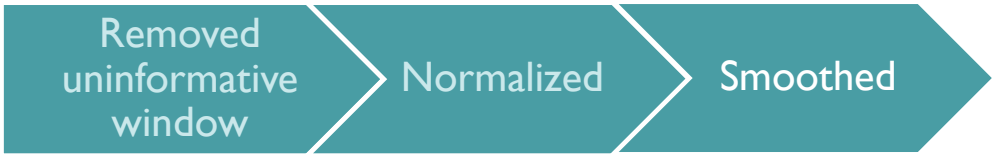
Let T be a time series consisting of n data points: $\{t_0, t_1, \ldots t_n\}$

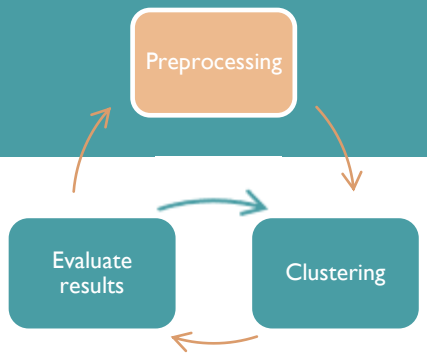
Then

$T_{norm} = \{(t_i - \min(T))/(\max(T) - \min(T)), \text{for } t_i \text{ in } T\}$

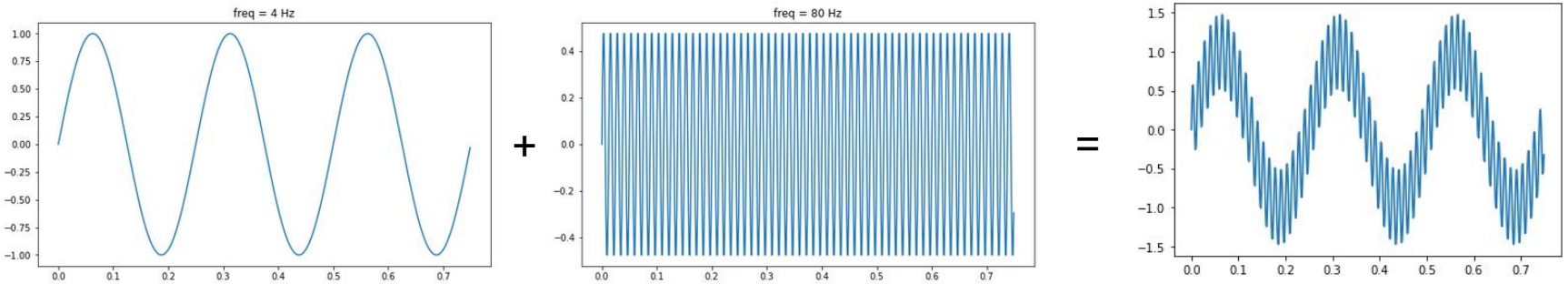
Waveform keeps its shape but the values are scaled from 0 to 1

METHODS – ITERATION 2

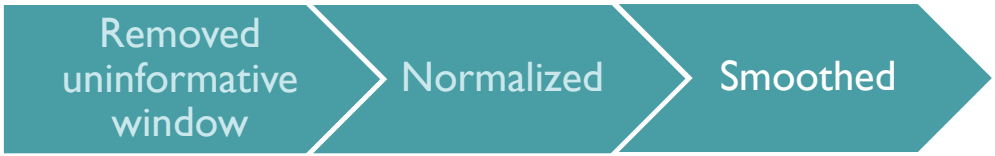


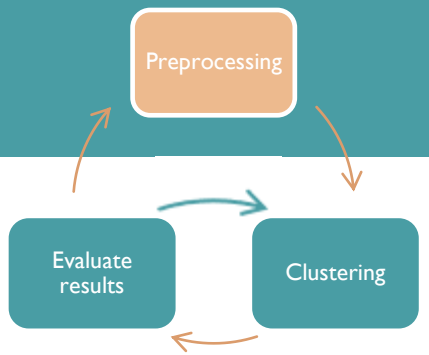


Fourier transformations and Power Spectral Density

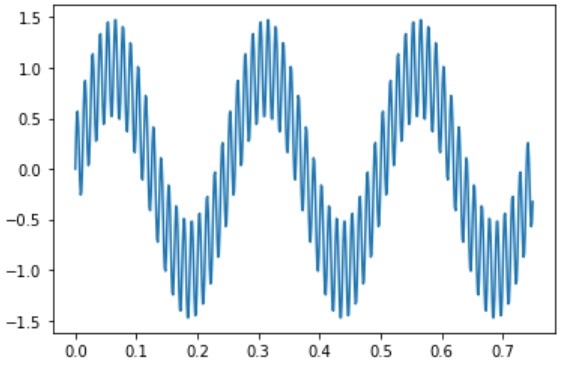


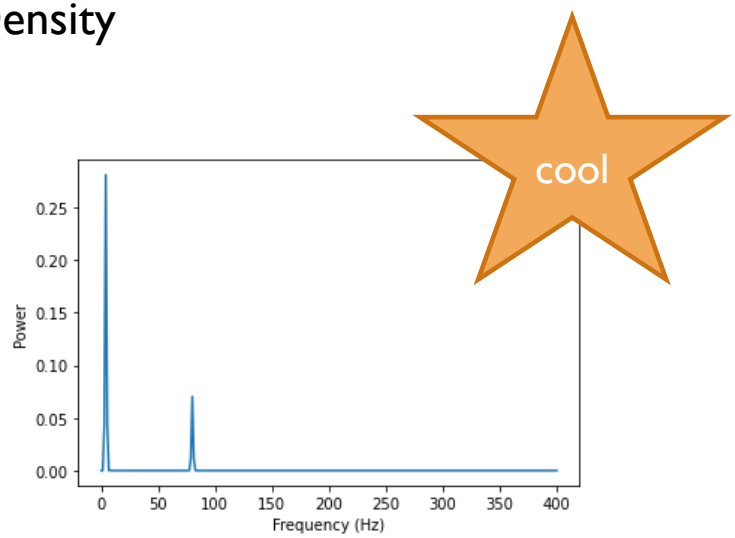
METHODS – ITERATION 2



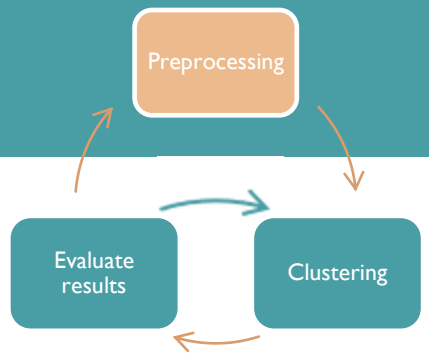


Fourier transformations and Power Spectral Density



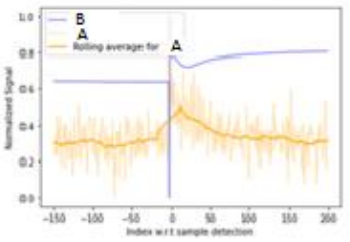


METHODS – ITERATION 2

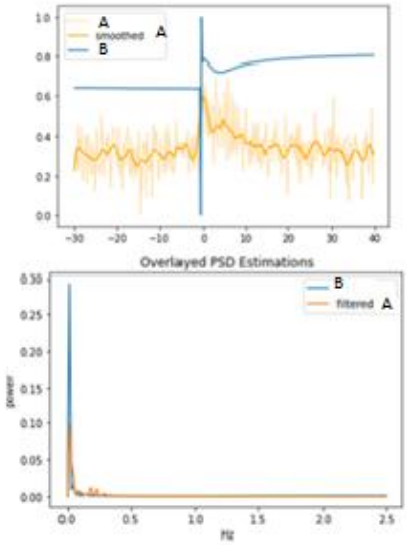


Wet-up removed → normalized → Smoothed

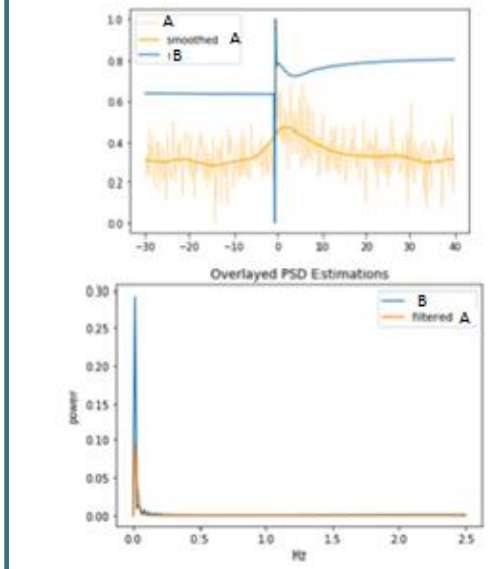
Moving Average



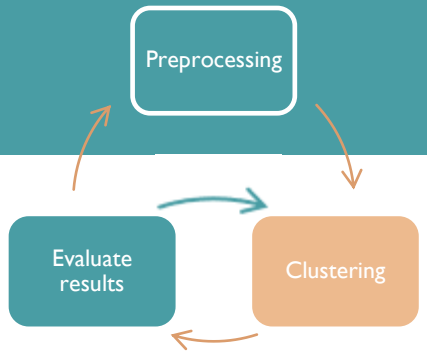
Savgol



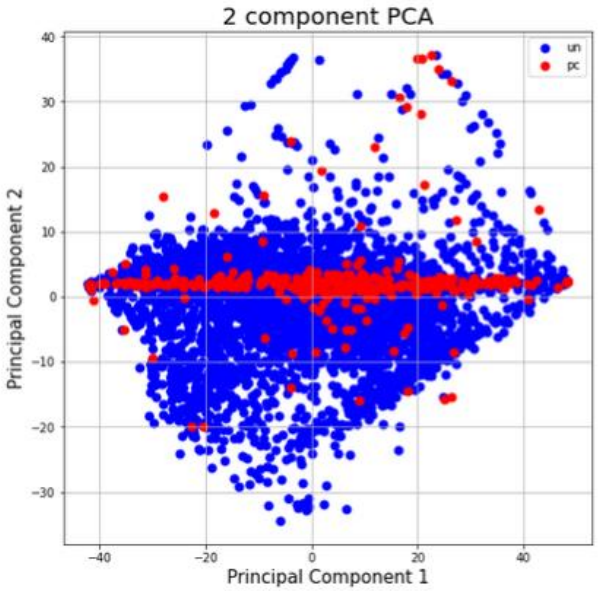
Convolution

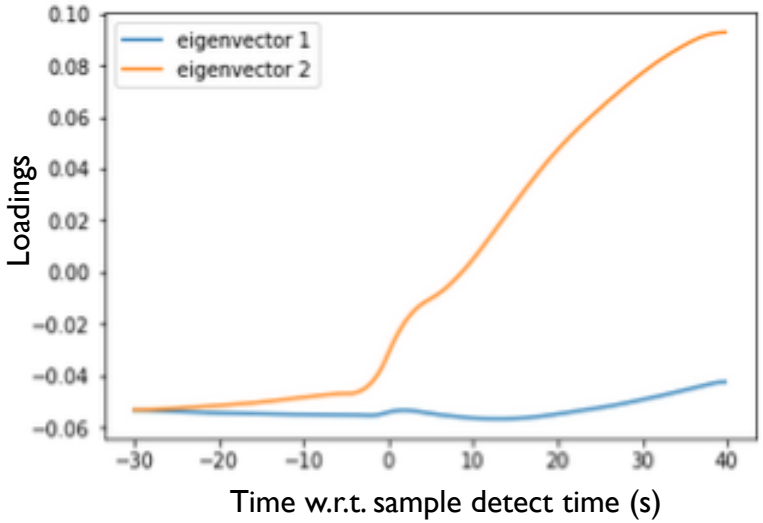


METHODS – ITERATION 2

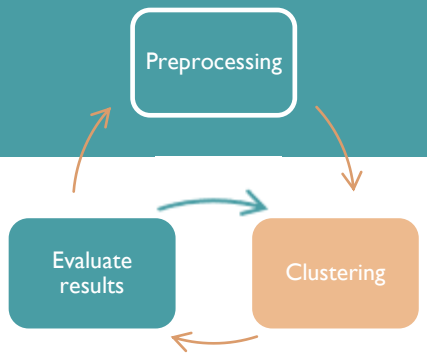


Principle Component Analysis (PCA)



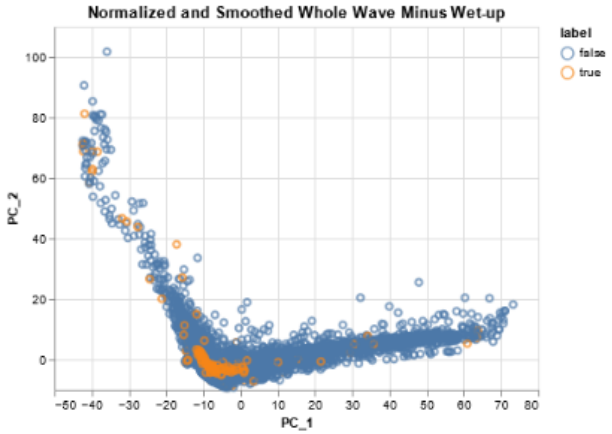


METHODS – ITERATION 2

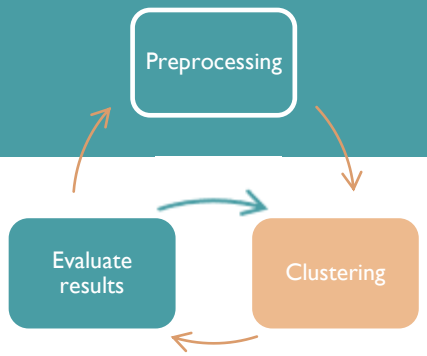


TSFRESH Predictors

- Phase 1: Feature Extraction (~ 770 features)
- Phase 2: Feature Significance Testing (~550 features)
  - Only the features that are significant with respect to classifying the readings are kept.
- Phase 3: PCA for dimension reduction (~ 30 components)
  - 95 % accumulated amount of variance explained



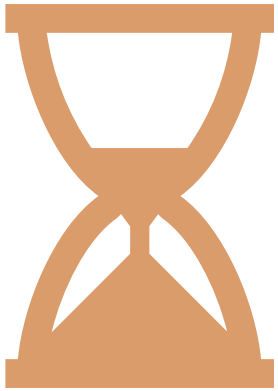
METHODS – ITERATION 2



TSFRESH Predictors

- Phase 4: Clustering the components
  - Create clusters using various algorithms (Gaussian Mixture Model/Agglomerative Clustering)
    - Try to get a cluster with most of the pins and a small amount of total readings
  - Cluster the subcluster that contains most of the pins

ROADBLOCKS



SUMMARY

- Project Goal: Group readings into categories to describe ways in which failure occurs as seen in the waveforms
- Our process is iterative:
  - Preprocessing → clustering → evaluation
- First iteration:
  - Standardization (mean 0, stdev 1) → Windows
  - Clustering on separate windows – disappointing
- Second iteration:
  - Normalization (scale between 0 and 1) → Smoothing (noise reduction)
  - Clustering on the whole wave (and features from it) – still in progress

NEXT STEPS

1. Look at other methods of feature extraction
2. Look at other clustering algorithms
3. Find metrics to describe readings that fall into clusters together
4. Compare results from different smoothing and filtering methods

QUESTIONS OR FEEDBACK?

THANK YOU FOR YOUR TIME!