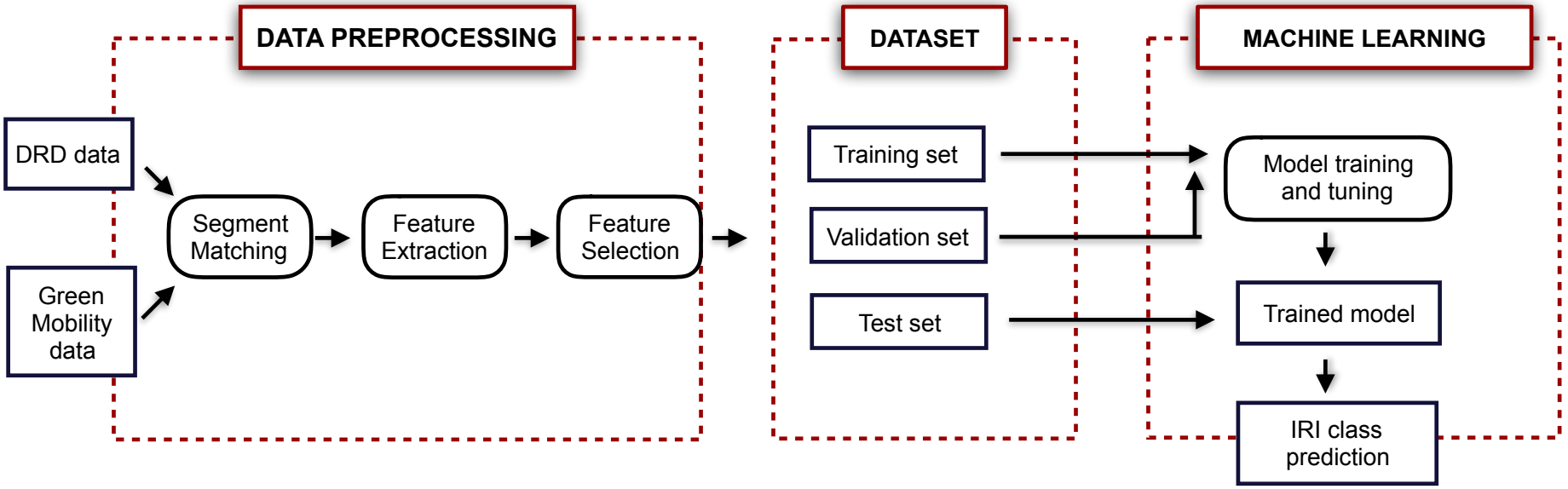
MACHINE LEARNING FOR ROAD CONDITION DATA

Milena Bajic, Postdoc
Tommy Sønne Alstrøm, Associate Professor

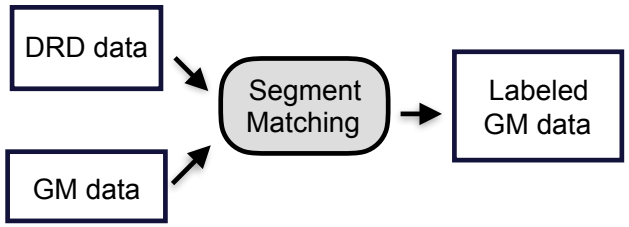
DTU Compute
Department of Applied Mathematics and Computer Science

Machine Learning Flowchart



Segment Matching

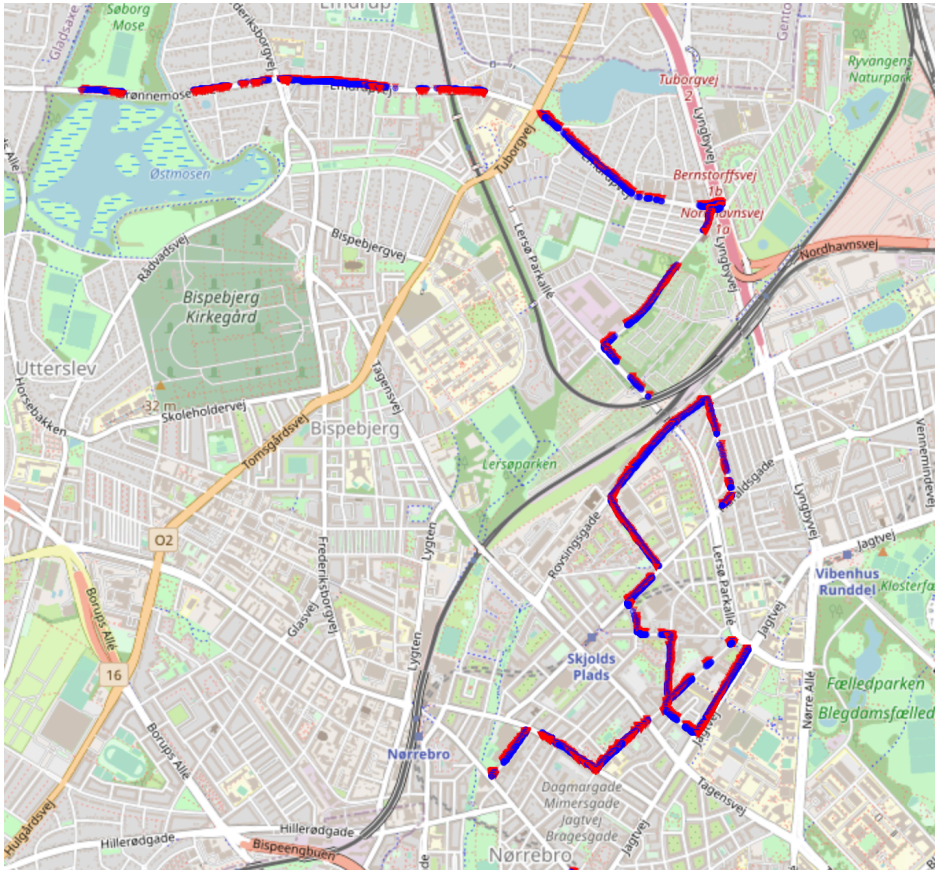
- **kNN Algorithm**
  1) Compute distances (GM point to DRD)
  2) Find the closest GPS point (<5m)



- **Matched data (with acc+speed):**
  1105 10-meter segments ~ 11km
- **IRI Class**

IRI Range	(0,2)	(2-6)	(6-15)
Class	0	1	2

- **Stratified data split:**
  Train (80%), Test (20%)



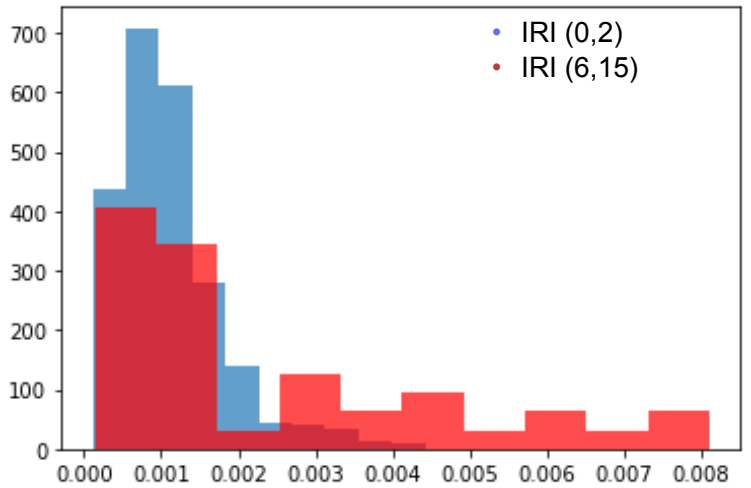
Feature Extraction

* *Barandas, Marília and Folgado, Duarte, et al. "TSFEL: Time Series Feature Extraction Library." SoftwareX 11 (2020).*

- **Inputs:**
  Acceleration and speed measurements in matched segments

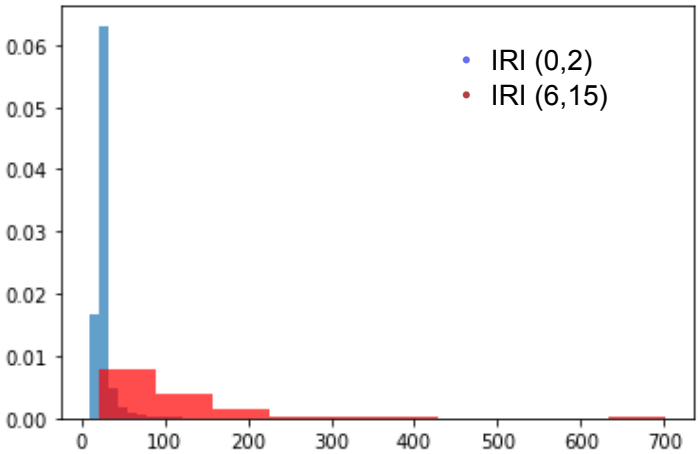
- **Idea:**
  Compute relevant features which help discrimination between different IRI classes

- **TSFEL**: Feature extraction library *
  - Time domain
  - Statistical domain

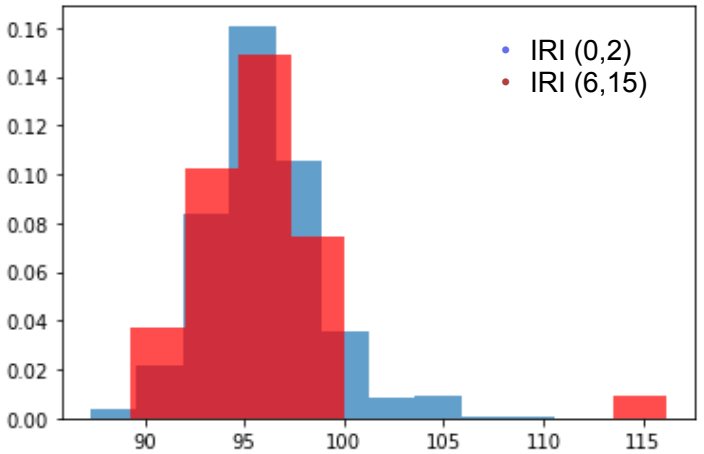


Variance (GM_acceleration_z)

Feature Extraction



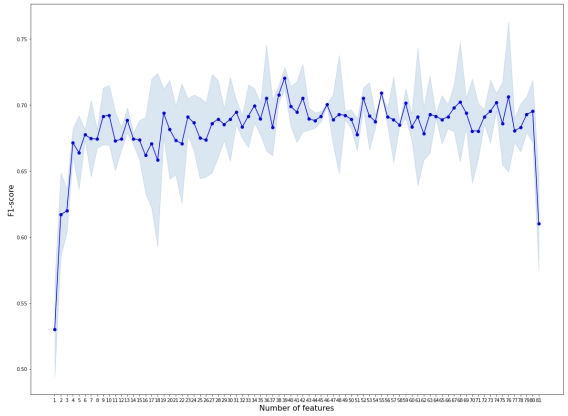
Autocorrelation (GM_acceleration_z)



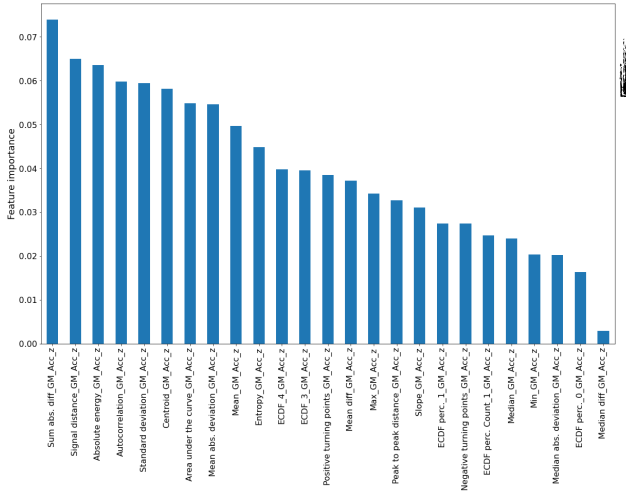
Total Energy (GM_acceleration_z)

Feature Selection

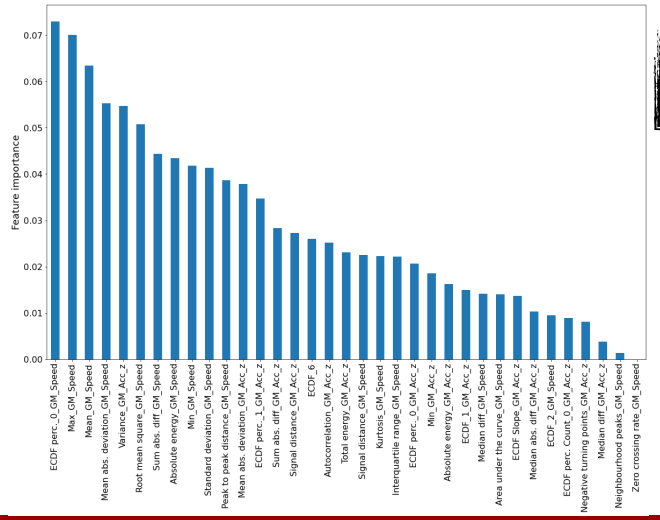
- **Procedure:**
  1. Removal of constant features
  2. Sequential Feature Selector



- **Selected features:**
  - Acceleration: 25 out of 39 features
  - Acceleration + speed: 35 out of 81 feature



Acceleration



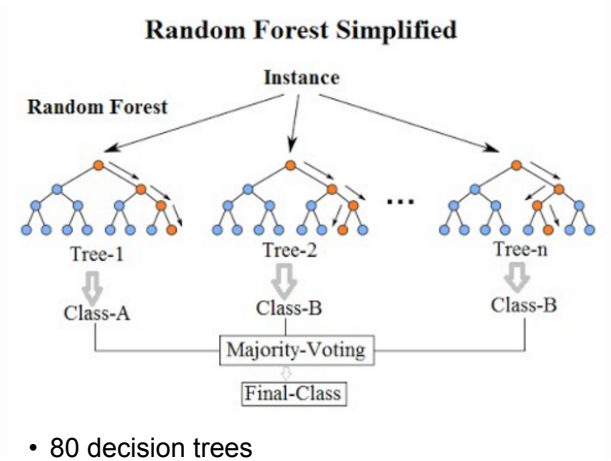
Acceleration + Speed

Random Forest Model

- **Idea:**

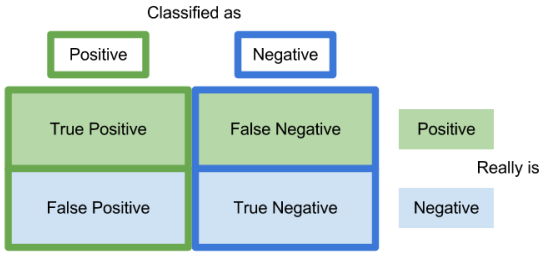
  Zhang, Zhiming, et al. "Application of a machine learning method to evaluate road roughness from connected vehicles." Journal of Transportation Engineering(2018)

- **Random Forest Algorithm**



- 80 decision trees
- Class imbalance -> class weights

- **Performance measurement:**



$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

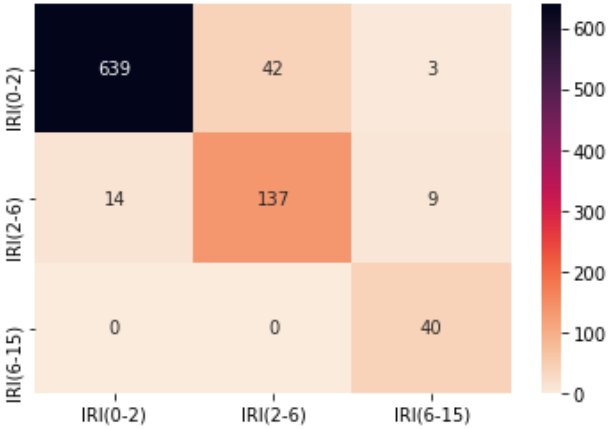
$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

```
F1 score = 2 * (recall * accuracy) / (recall + accuracy)
```

Random Forest Model: Predictions (I)

Acceleration

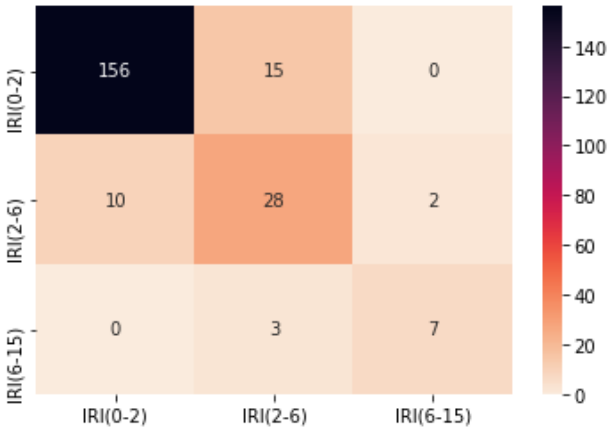
Train data



Train set	IRI (0-2)	IRI (2-6)	IRI (6-15)
Precision	0.98	0.77	0.77
Sensitivity	0.93	0.86	1.00
F1-score	0.96	0.81	0.87

884 segments

Test data



Test set	IRI (0-2)	IRI (2-6)	IRI (6-15)
Precision	0.94	0.61	0.78
Sensitivity	0.91	0.70	0.70
F1-score	0.93	0.65	0.74

221 segments

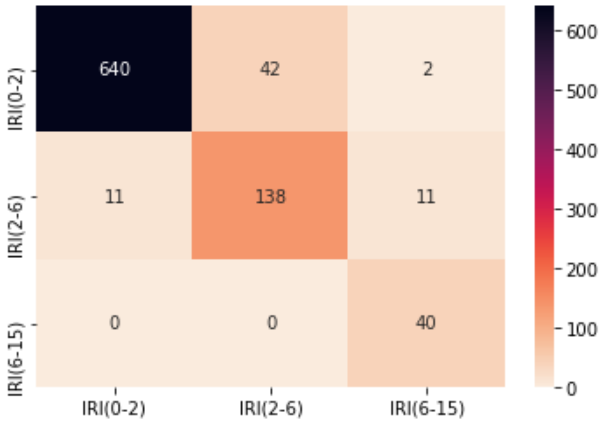
** rows: true label*

** columns: predicted label*

Random Forest Model: Predictions (II)

Acceleration + Speed

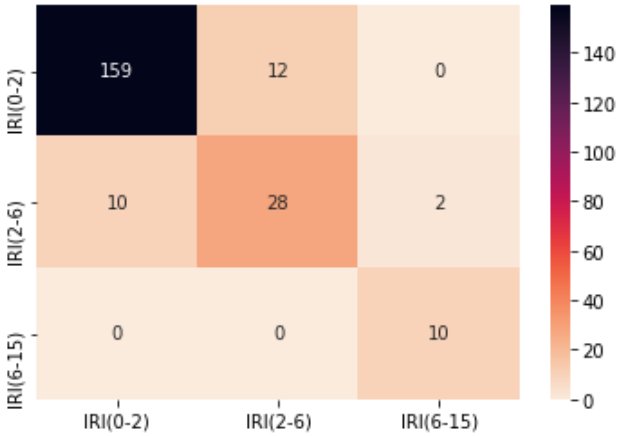
Train data



Train set	IRI (0-2)	IRI (2-6)	IRI (6-15)
Precision	0.98	0.77	0.75
Sensitivity	0.94	0.86	1.00
F1-score	0.96	0.81	0.86

884 segments

Test data



Test set	IRI (0-2)	IRI (2-6)	IRI (6-15)
Precision	0.94	0.70	0.83
Sensitivity	0.93	0.70	1.00
F1-score	0.94	0.70	0.91

221 segments

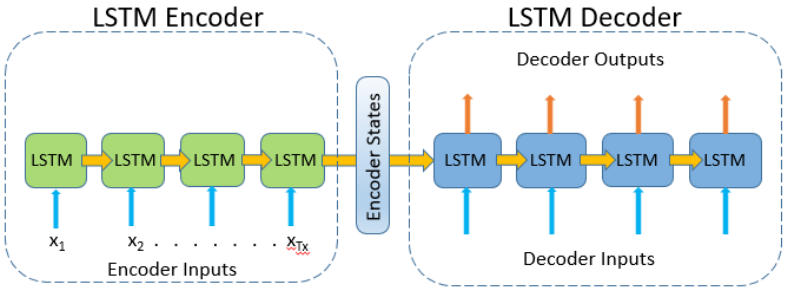
** rows: true label*

** columns: predicted label*

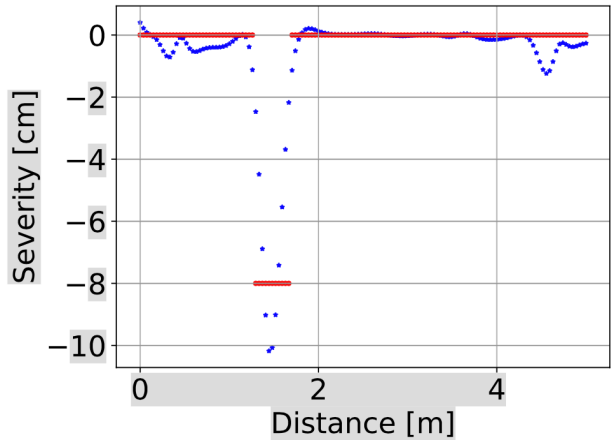
Synthetic Data Study

- **Simulation:**
  Quarter Car Model
  (DTU Civil Engineering)
- **Problem:**
  Map acceleration sequence to road severity sequence
- **Model:**
  Encoder-Decoder with attention
  (inspired by machine translation)

Encoder-Decoder Architecture



Preliminary results



Towards Hybrid Modelling

- **Next goal:**
  Combine theoretical prediction and data-driven approach

- **Ideas: ***
  - Add a penalty term to the loss function for predictions far away from the theoretical prediction
  - Pretrain the model on simulation, fine tune on measured data
  - Use the theoretical prediction as additional input

**Jared Willard and et.:"Integrating Physics-Based Modeling with Machine Learning: A Survey" , arXiv:2003.04919 (2020)*

Summary

- **Machine Learning flowchart:**
  Matching of vehicles trajectories
  Feature extraction and selection
  Prediction of IRI class using GM cars with variable speeds

- **Synthetic data study:**
  Prediction of road severity *sequence* from the acceleration sequence

- **Towards Hybrid modelling:**
  Combination of simulation and the data driven approach