Valentin GUIEN

Early-Career Researchers Day

DATA Program

2024 / 12 / 18

Characterization and anomaly detection in daily cow activities using wavelet-based features

Introduction

- Digital revolution : Using sensors to control many parameters
- Precision Livestock Farming
- Helping the farmer to anticipate situations requiring action on animals, buildings,…
- Automated systems to reduce the workload of monitoring and improve efficiency
- Tools to improve animal welfare by detecting specific conditions based on behavior?

Introduction

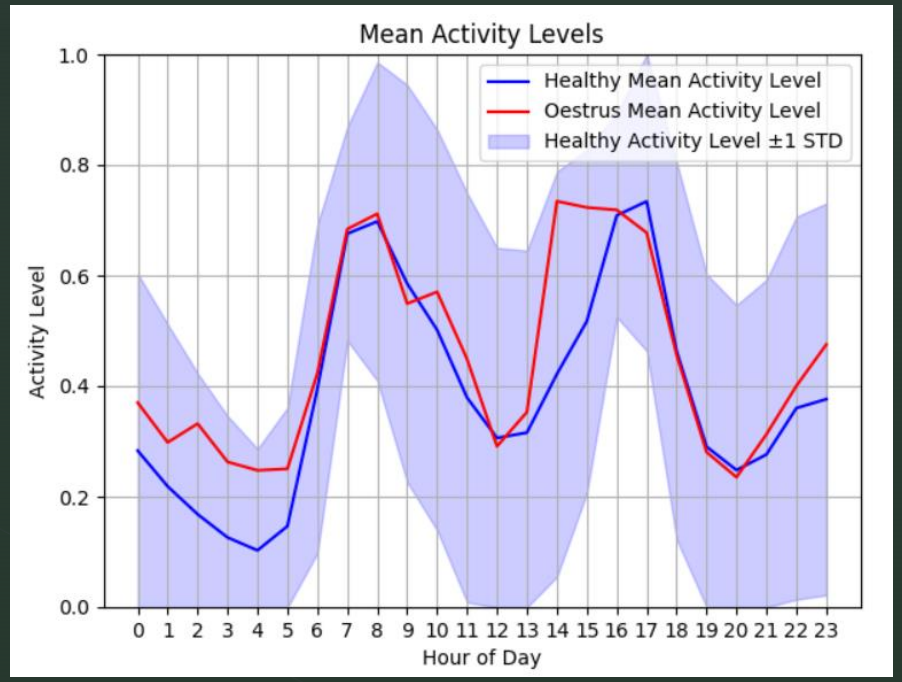
- Data collected with CowView : sensor placed on the cow's collar which provides the position and therefore the activity
  - Alleys → Walking/Standing
  - Cubicles → Resting
  - Feeding table → Eating

The activity level (AL) is then computed per cow and per hour by a weighted sum of the time spent in each activity:

AL = 0.16*(time spent standing) + 0.42*(time spent eating) – 0.23*(time spent resting)

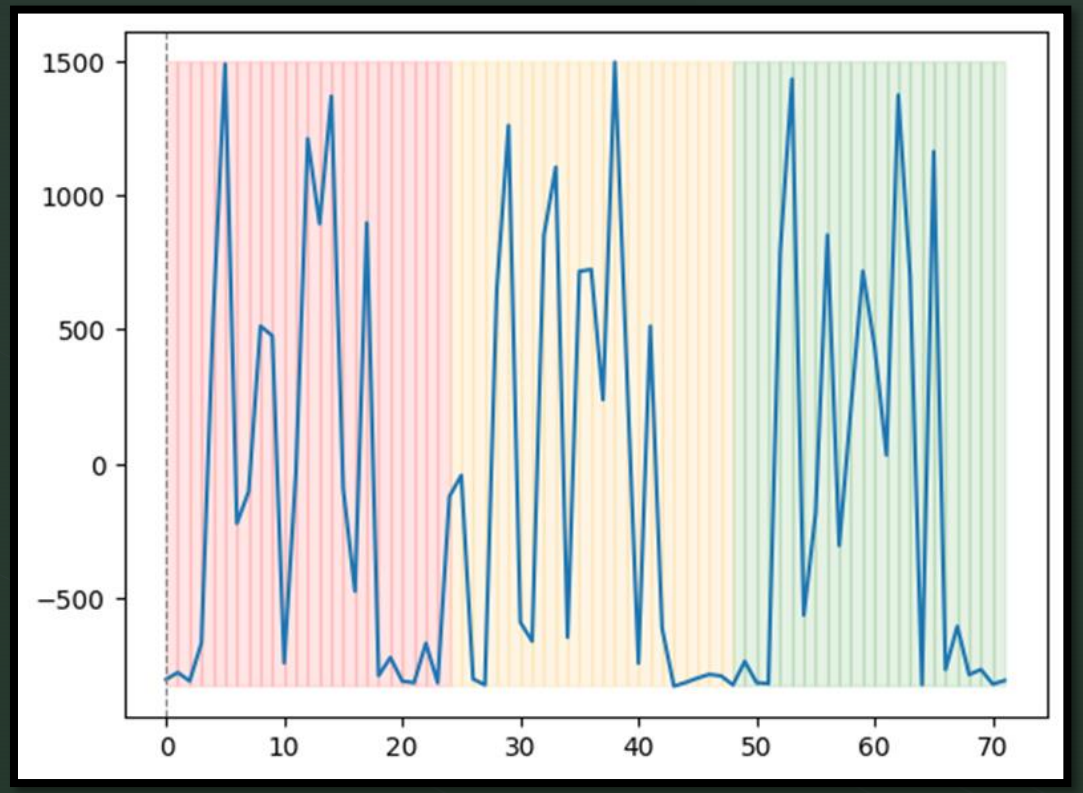
Introduction

- Events that can change the behavior of the cow:
  - Diseases: lameness, mastitis,…
  - Reproductive status: oestrus, calving
  - Stress
  - Change of environment



Introduction

- Significant inter-individual and intra-individual variations
- Wide range of possible perturbations
- Possible errors in the annotation of the internal state by the farmer

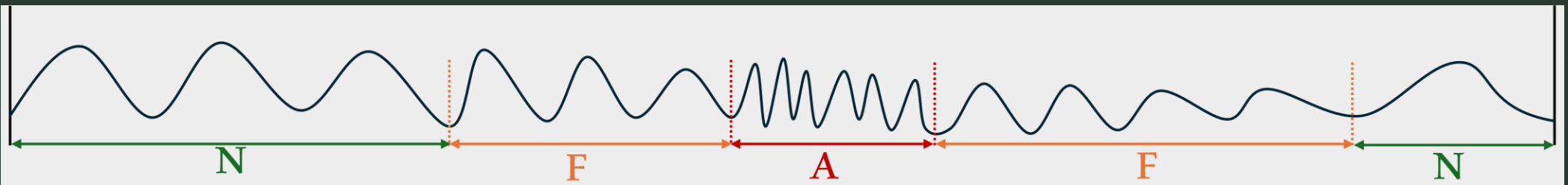


Introduction

- Anomaly detection methods preprocess the data by applying projections and/or transformations before using similarity threshold or one-class classifier
- Recent techniques (GANs, autoencoders, transformers) lack of explainability.
- Objective of this work: capture the characteristics essential for anomaly detection in dairy cows activity

Data windowing

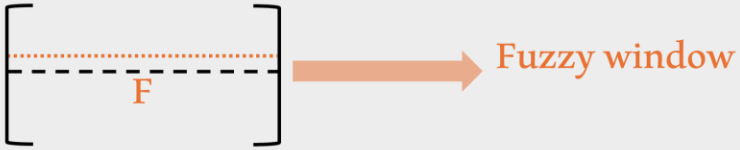
- Windows of size q running through the time series : W_i
- Labels on W_i based on the worst-case scenario: Y_i
  - If at least one point in W_i has label A, $Y_i = Abnormal$
  - If no A label point in W_i, but at least one point has label F, $Y_i = Fuzzy$
  - If all points have label N, $Y_i = Normal$
- Two windows are consecutives if they represent the same sequence in the time series with one unit of time in interval

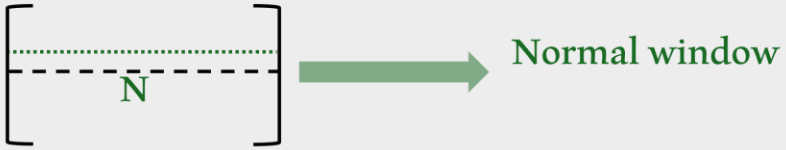


N, Normal sequence

F, Fuzzy sequence

A, Abnormal sequence





Feature extraction

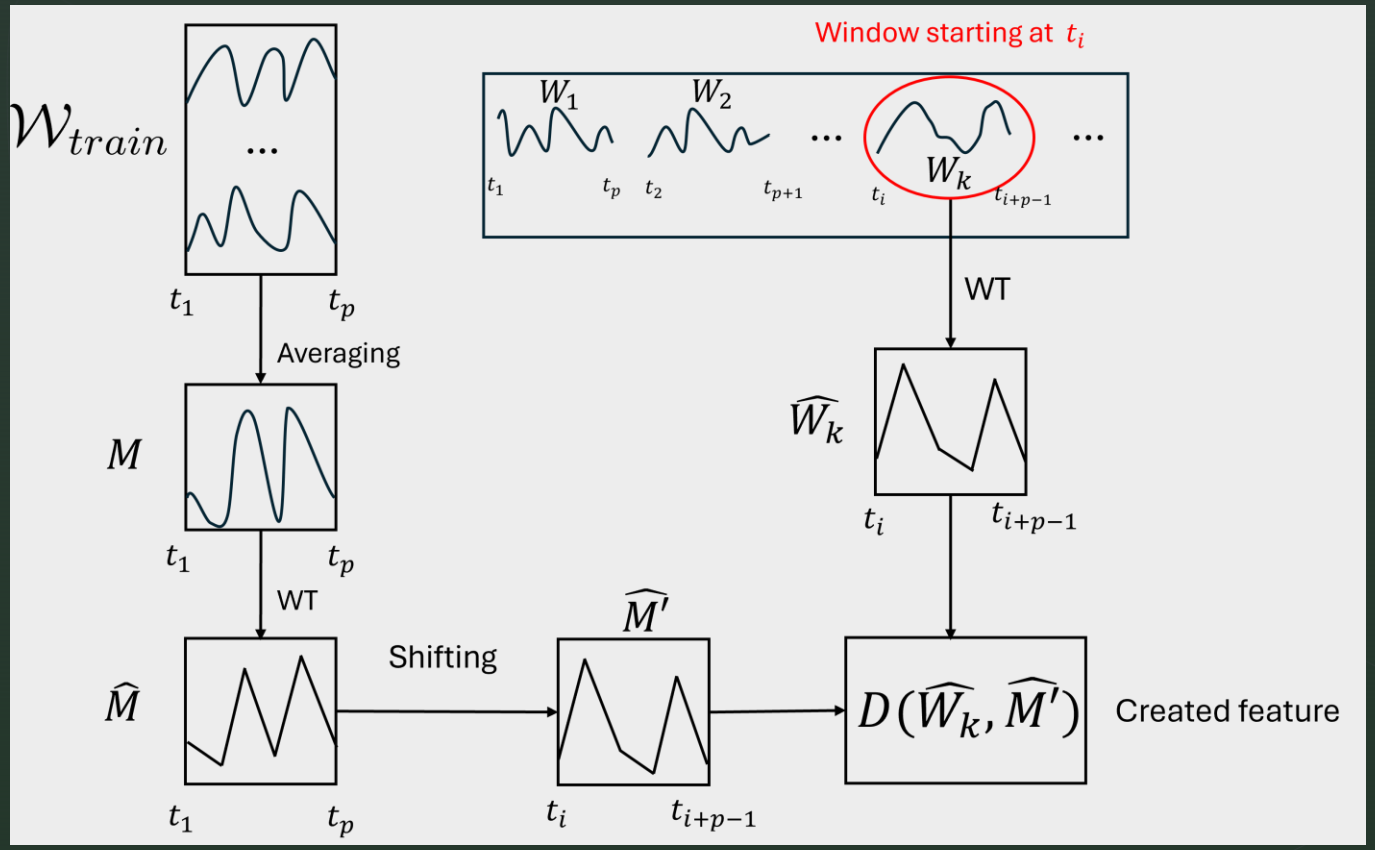
- 26 basic time series features to capture fundamental properties of the activity level over 24 hours (minimum, maximum, mean, STD,…)
- 23 wavelet-based features using wavelets with compact support
  - Orthogonal wavelets: Haar, Daubechies, Coiflet
  - Bi-orthogonal wavelets

Wavelet based features

- Size of windows = the size of the period: $q = p$
- M: Average period of the times series of the training set
- $\widehat{M}$: Discrete Wavelet Transform (DWT) of M

- The DWT-based features of W_i are computed as follow:
  - $\widehat{W_i}$: DWT of W_i
  - $\widehat{M'}$: Shifting of $\widehat{M}$ based of the time where W_i starts
  - The Euclidean distance is computed between $\widehat{W_i}$ and $\widehat{M'}$

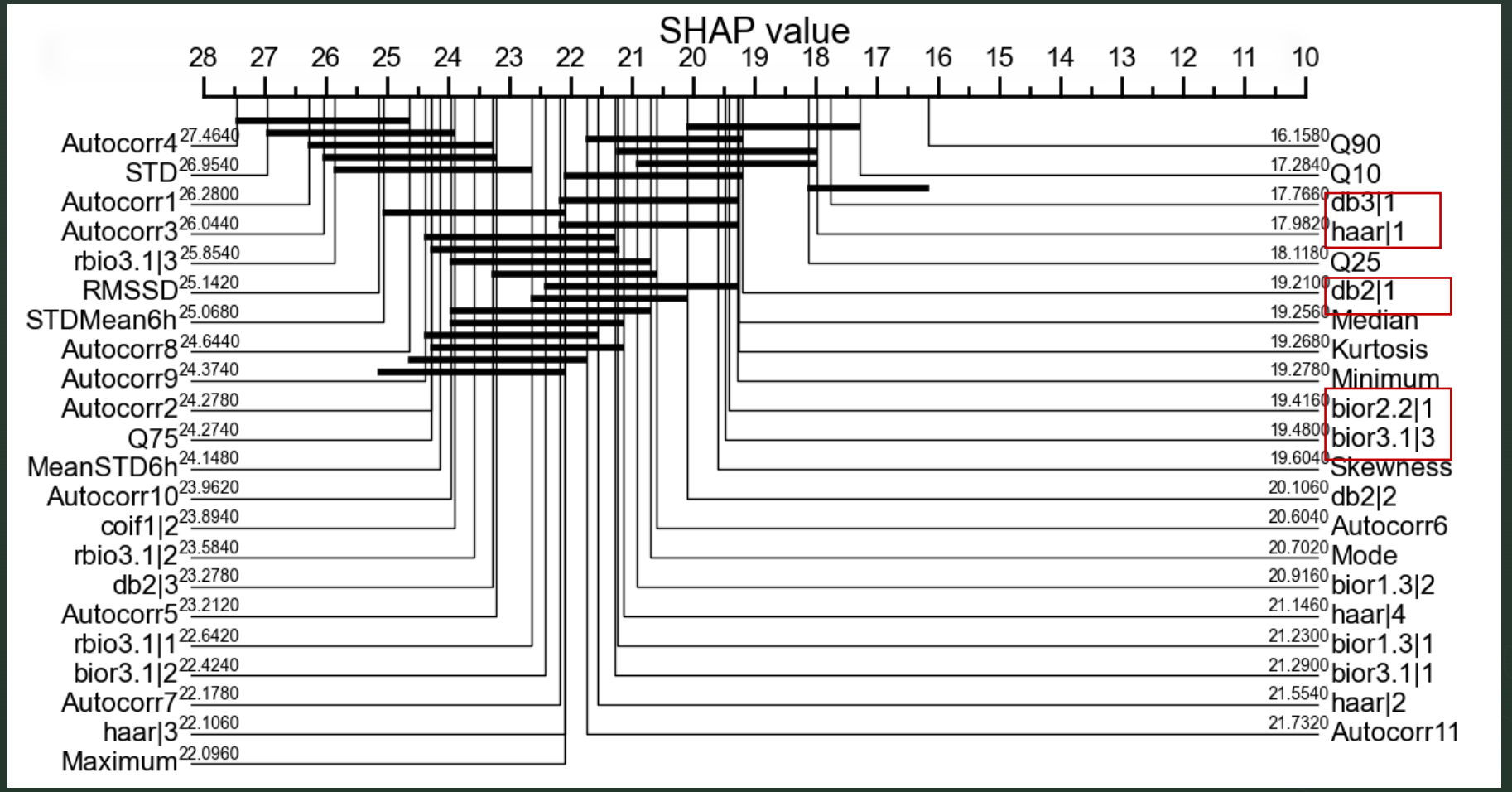
Wavelet based features



SHAP values to evaluate the features contribution

- The SHAP (SHapley Additive exPlanations) of a feature indicates the significance of a feature for one instance of algorithm detection
- It varies depending how the training and the testing sets are build, and also due to the use of Isolation Forest, so we evaluate among multiples executions.
- 10 separations of training and testing sets:
  - For each separation, 50 iterations of IF are done
  - The mean SHAP value of each features is extracted
- Critical Difference Diagram generated for the “performance” of the features

Results


SHAP value
28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10
Autocorr4 27.4640
STD 26.9540
Autocorr1 26.2800
Autocorr3 26.0440
rbio3.1|3 25.8540
RMSSD 25.1420
STDMean6h 25.0680
Autocorr8 24.6440
Autocorr9 24.3740
Autocorr2 24.2780
Q75 24.2740
MeanSTD6h 24.1480
Autocorr10 23.9620
coif1|2 23.8940
rbio3.1|2 23.5840
db2|3 23.2780
Autocorr5 23.2120
rbio3.1|1 22.6420
bior3.1|2 22.4240
Autocorr7 22.1780
haar|3 22.1060
Maximum 22.0960
16.1580 Q90
17.2840 Q10
17.7660 db3|1
17.9820 haar|1
18.1180 Q25
19.2100 db2|1
19.2560 Median
19.2680 Kurtosis
19.2780 Minimum
19.4160 bior2.2|1
19.4800 bior3.1|3
19.6040 Skewness
20.1060 db2|2
20.6040 Autocorr6
20.7020 Mode
20.9160 bior1.3|2
21.1460 haar|4
21.2300 bior1.3|1
21.2900 bior3.1|1
21.5540 haar|2
21.7320 Autocorr11

Results

- 70 separations of training and testing sets:
  - For each separation, 20 iterations of IF
- First, Wavelet-Based Features (WBF) are not used. Afterward, they are used among the other features
- If two features are correlated more than 0.9, one of them is dropped
- The scores are computed on the normal and abnormals windows, the fuzzy one are not considered

	Balanced training set	
	Statistical feat.	Statistical and wavelet feat.
Accuracy	0,51±0,01	**0,54±0,01**
Recall	0,12±0,02	**0,13±0,01**
Precision	0,54±0,07	**0,67±0,07**

Conclusion

- We created features that slightly improve results on complex periodic datasets with a classical anomaly detection algorithm
- Core idea to extract the endogenous rhythm of a day to detect an anomalous signal
- Highly interpretable approach