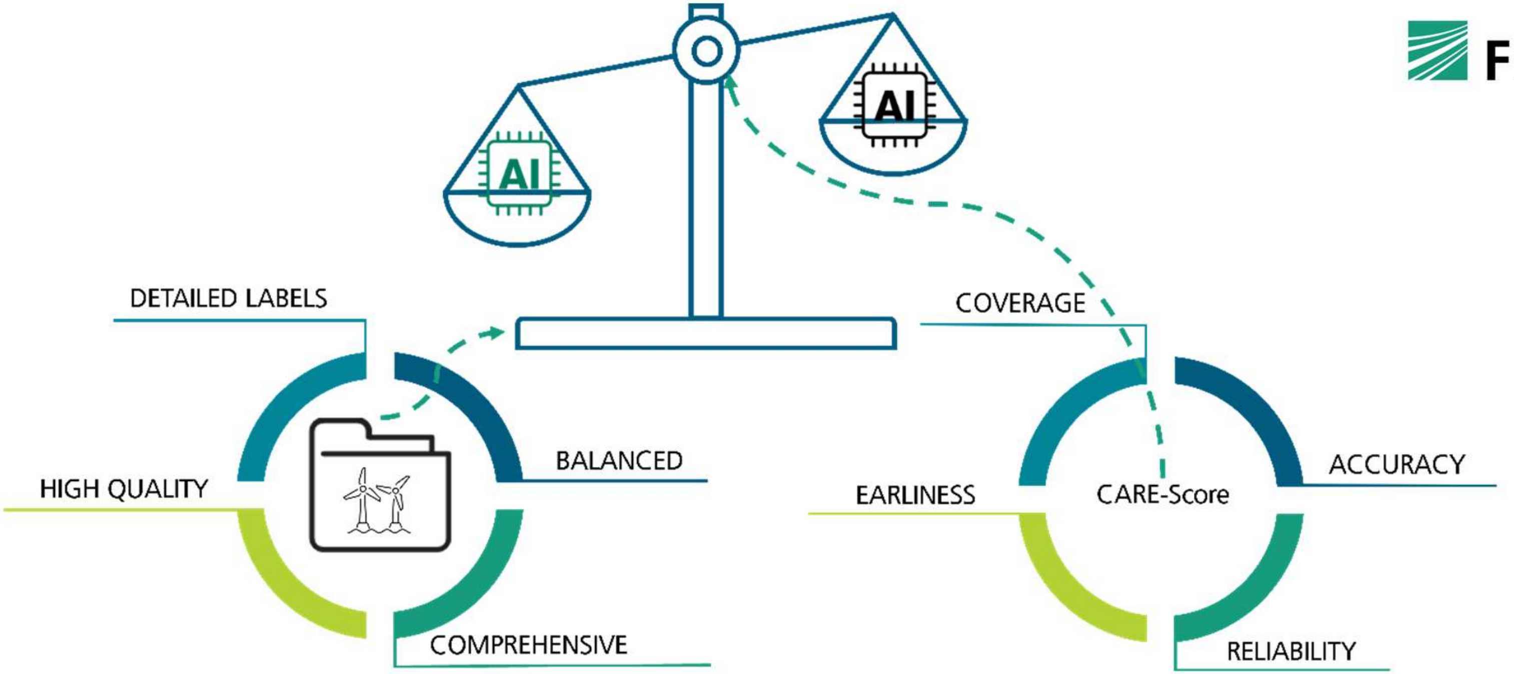


Fraunhofer
IEE

Care to Compare – a real-world dataset for early fault detection in wind turbine data

Cyriana Roelofs | 19.09.2024

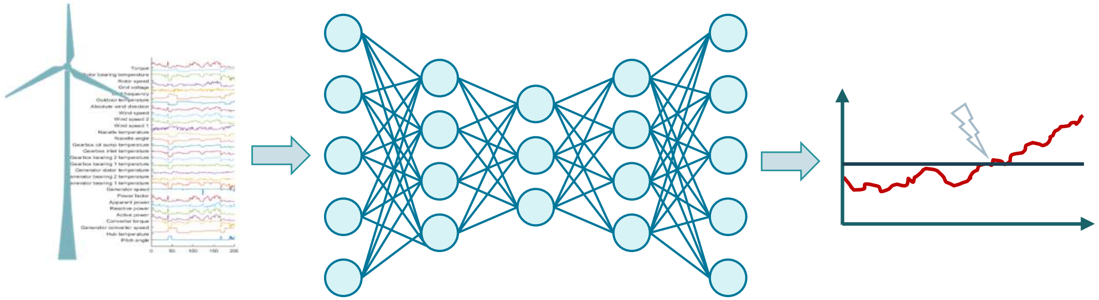
KonKIS 2024

Care to Compare – a real-world dataset for early fault detection in wind turbines

Motivation

Early fault detection in wind turbines

- Reduce downtime of wind turbines
- Corrective → predictive maintenance
- Usually anomaly detection in SCADA-data



Why do we need more public wind turbine SCADA data?

- Many publication use inaccessible datasets → Results are not reproducible or comparable
- Almost all public datasets with wind turbine SCADA data lack information about anomalies or faults.

To enable meaningful comparisons between early fault detection algorithms in the wind energy domain, more public datasets are necessary.

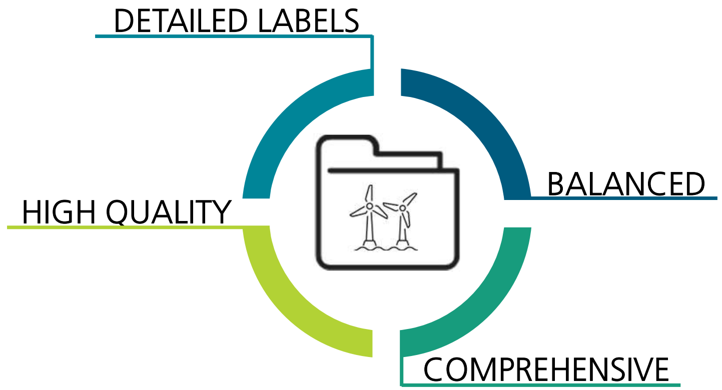
 © Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

Motivation

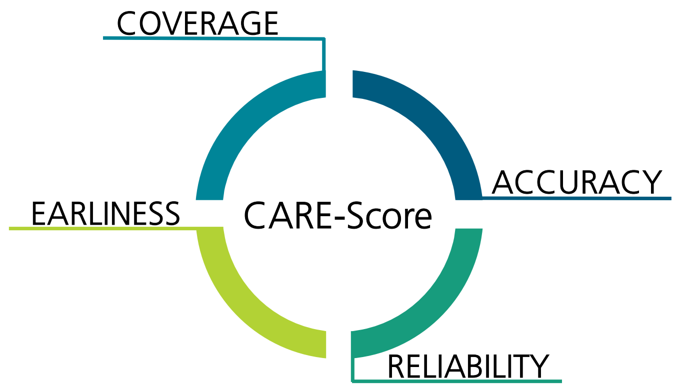
Care to Compare introduces a high-quality dataset

- Detailed information on faults and turbines status
- 44 labeled time frames for anomalies that led up to faults
  - 51 representing normal behaviour
- A total of 89 years of operating data from 36 wind turbines across 3 wind farms



Scoring method

- Uses all information in the dataset to identify good early fault detection models
- Perfect early fault detection model:
  - Detects all anomalies, as early as possible, without generating false alarms.



 © Fraunhofer IEE

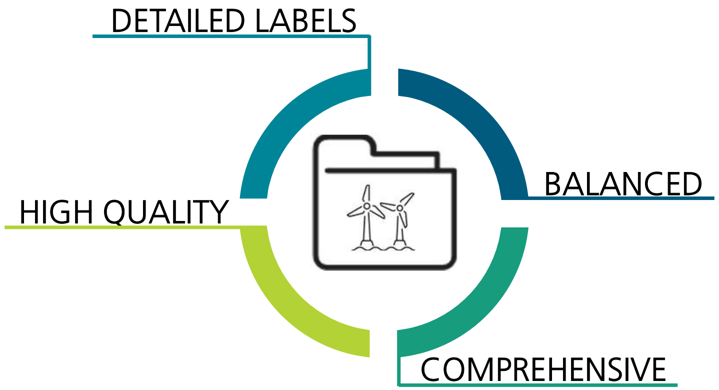
Care to Compare – a real-world dataset for early fault detection in wind turbines

Data collection and labeling

Dataset collection

- Data of 3 wind farms available
  - Combines data of 1 open dataset (EDP) and data of 2 new wind farms

- Data selection requirements
  - Different wind turbines types
  - Different wind farms
  - Different fault types
  - Balanced, anomalies and normal behaviour
  - Good quality training data available before fault
  - Labeled anomaly time frames



Labeling

- Wind turbines status
- Maintenance information from service reports
- Operator feedback

© Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

Dataset overview

- Each dataset contains
  - 10Min avg/min/max/std SCADA-data
  - 1 year training data
  - At least 2 weeks prediction data
  - Fault information (if anomaly dataset), including anomaly time frame

	wind farm A	wind farm B	wind farm C	total
# Wind turbines	5	9	22	36
# Datasets	22	15	58	95
# Anomaly datasets	11	6	27	44
# Normal behaviour datasets	11	9	31	51
# Features	86	257	957	-
# Sensors	54	63	238	-

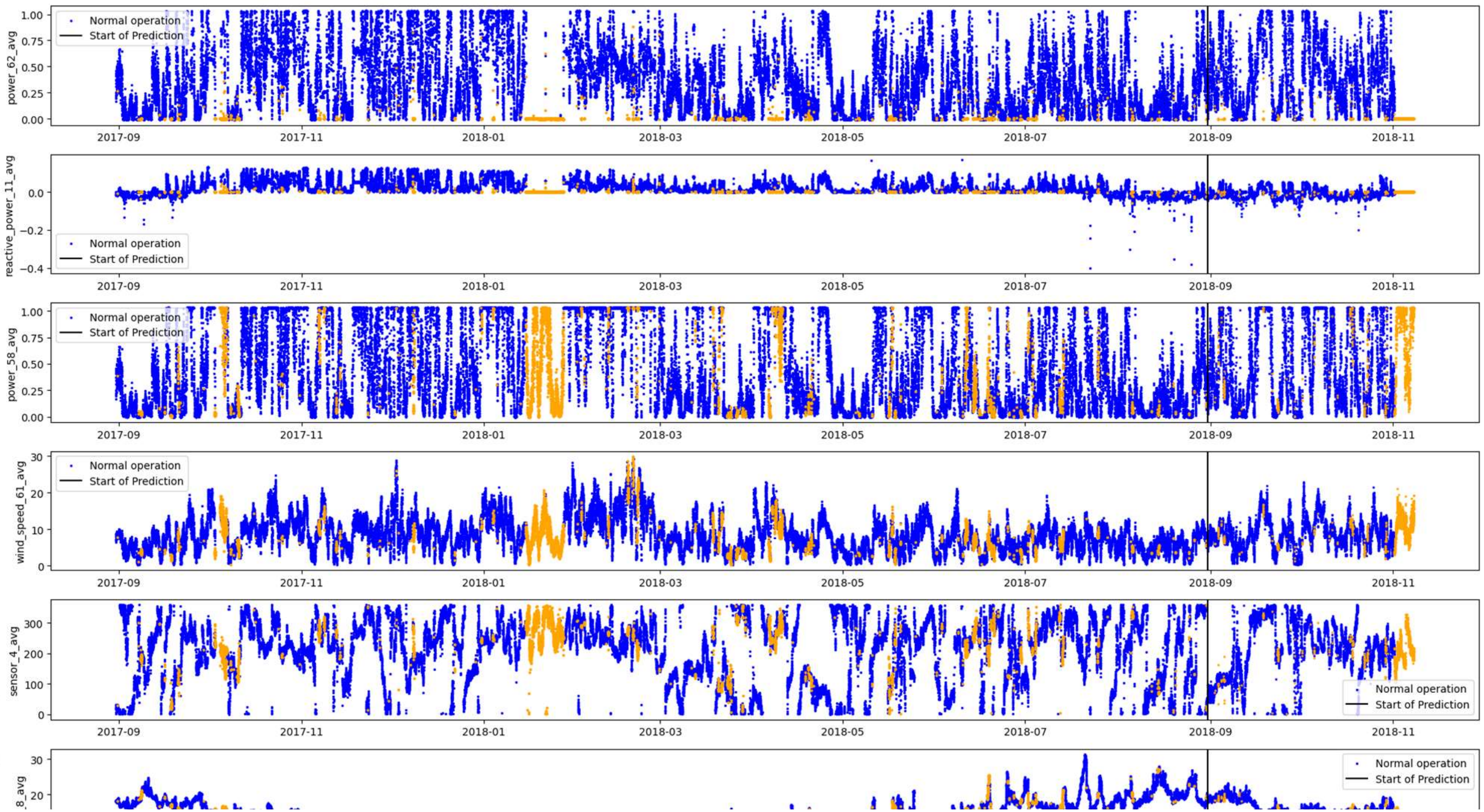
 © Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

Dataset overview

Example dataset

Rotor bearing damage



 © Fraun

Care to Compare – a real-world dataset for early fault detection in wind turbines

CARE Score - motivation

Perfect AD for early fault detection

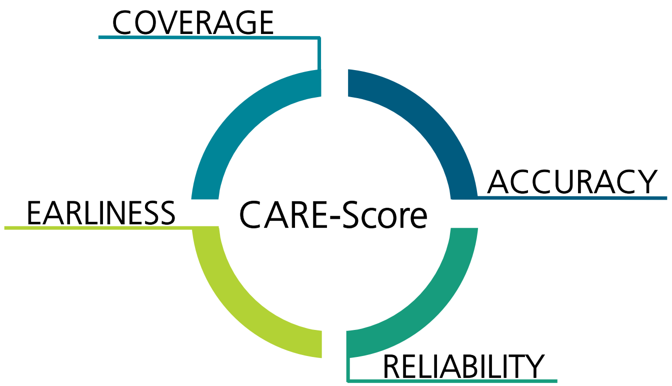
- Recognizes all faults correctly, without generating false alarms, as early as possible.

CARE-Score:

- Coverage – detect many anomalies
- Accuracy – recognise normal behaviour accurately
- Reliability – without generating false alarms
- Earliness – detect fault early

Why CARE?

- Cover all key aspects of a good early fault detection model
- Emphasizes the detection of anomalies - while minimizing false alarms.
- Rewards recognition of normal behaviour and anomalies.
- Adaptable



 © Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

CARE Score – breakdown

COVERAGE

- Detect as many anomalies as possible.
- Evaluated by F_β with $\beta = 0{,}5$

ACCURACY

- Recognise normal behaviour correctly
- Evaluated by Accuracy-Score

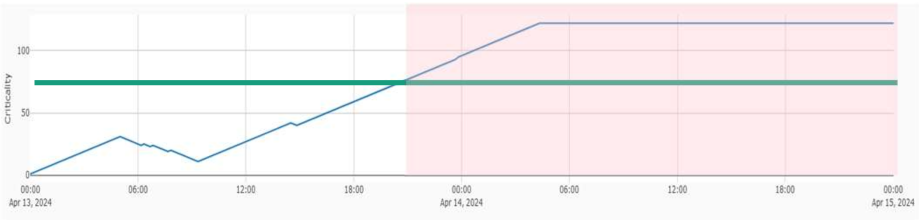
 © Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

CARE Score – breakdown

RELIABILITY

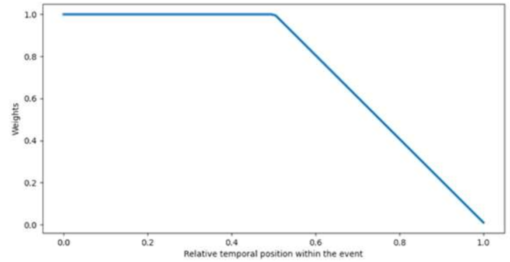
- No false alarms



- Criticality threshold > 72 → Anomaly event

EARLINESS

- Detect as early as possible



- Weighted score WS for correctly detected anomalies.

© Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

CARE Score – breakdown

Combining the scores through a weighted average:

- $WA := \frac{1}{\Sigma_{i=1}^{4} \omega_i}\left(\omega_1 \overline{F_\beta} + \omega_2 \overline{WS} + \omega_3 \overline{EF_\beta} + \omega_4 \overline{Acc}\right)$
- $\omega_1 = \omega_2 = \omega_3 = 1,\ \omega_4 = 2$

Special cases

- Do not reward models that detect no anomalies
- Do not reward random strategies or models that detect too many anomalies

$$CARE := \begin{cases} 0, & \text{if no anomalies were detected} \\ \overline{Acc}, & \text{if } \overline{Acc} < 0.5 \\ WA, & \text{else.} \end{cases}$$

 © Fraunhofer IEE

Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

Conclusions and future work

Care to Compare

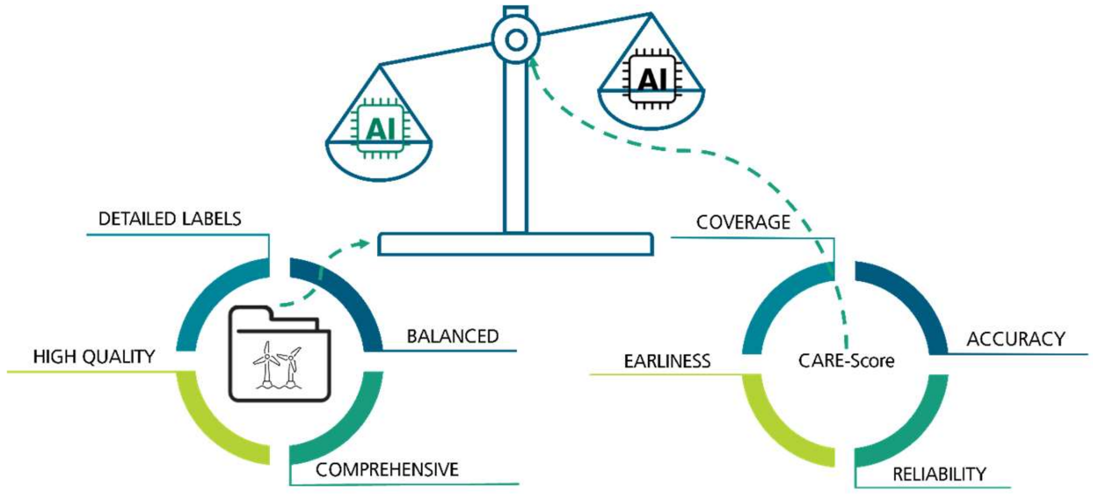
- A high-quality, comprehensive, balanced real-world dataset with detailed fault labels for early fault detection in wind turbines
- CARE scoring method

Future work

- Compare wide range of algorithms
- Extending the dataset with more fault types, etc.

- Adding Root Cause Analysis evaluation (XAI)
  - Dataset with root cause information
  - Extending CARE

- Early fault detection for other renewables, such as PV…



 © Fraunhofer IEE

Care to Compare – a real-world dataset for early fault detection in wind turbines

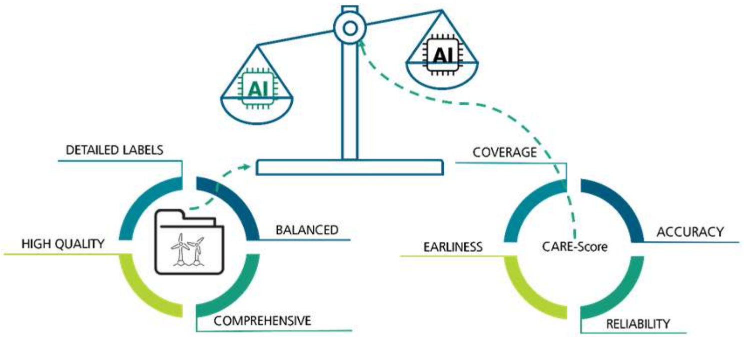
Conclusions and future work

Care to Compare

- A high-quality, comprehensive, balanced real-world dataset with detailed fault labels for early fault detection in wind turbines
- CARE scoring method

Dataset available on Zenodo:

- https://zenodo.org/records/10958775
  - Or search "Care2Compare"
- Paper preprint on arXiv - https://arxiv.org/abs/2404.10320



ADWENTURE
Anomaly Detection
For Wind Turbine Efficiency

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag

 © Fraunhofer IEE

Fraunhofer IEE

Thank you for your attention

Fraunhofer IEE