Machine learning meets PVsyst: A novel framework to detect, classify, and forecast faults in utility-scale PV

Sijin Wang[1], Brendan Wright[1], Ali Shakiba[1], Abhnil Prasad[1], Marcus Vengas[2], Felix Daddo[2], Georgia Tovich[2], Remi Coni[2], Deyang Su[3], John Rodriguez[1], Ziv Hameiri[1]

[1]The University of New South Wales, Australia, [2]Spark Renewables, Australia, [3]Jinko Solar, Australia

53rd IEEE Photovoltaic Specialists Conference, Montreal, QC, Canada · June 8–13, 2025

Introduction

- **Utility-scale photovoltaic** (PV) farms are growing, making reactive operation and maintenance (O&M) more costly and challenging [1]
- We propose a **machine learning (ML)** based method to **detect faults** in real time, **classify** the faults, and potentially **forecast** near future faults

Methodology

- The test site is a 120 MWp PV farm with north-south single-axis tracking
- An ML model was built as an interpolated PVsyst with a flexible timestamp

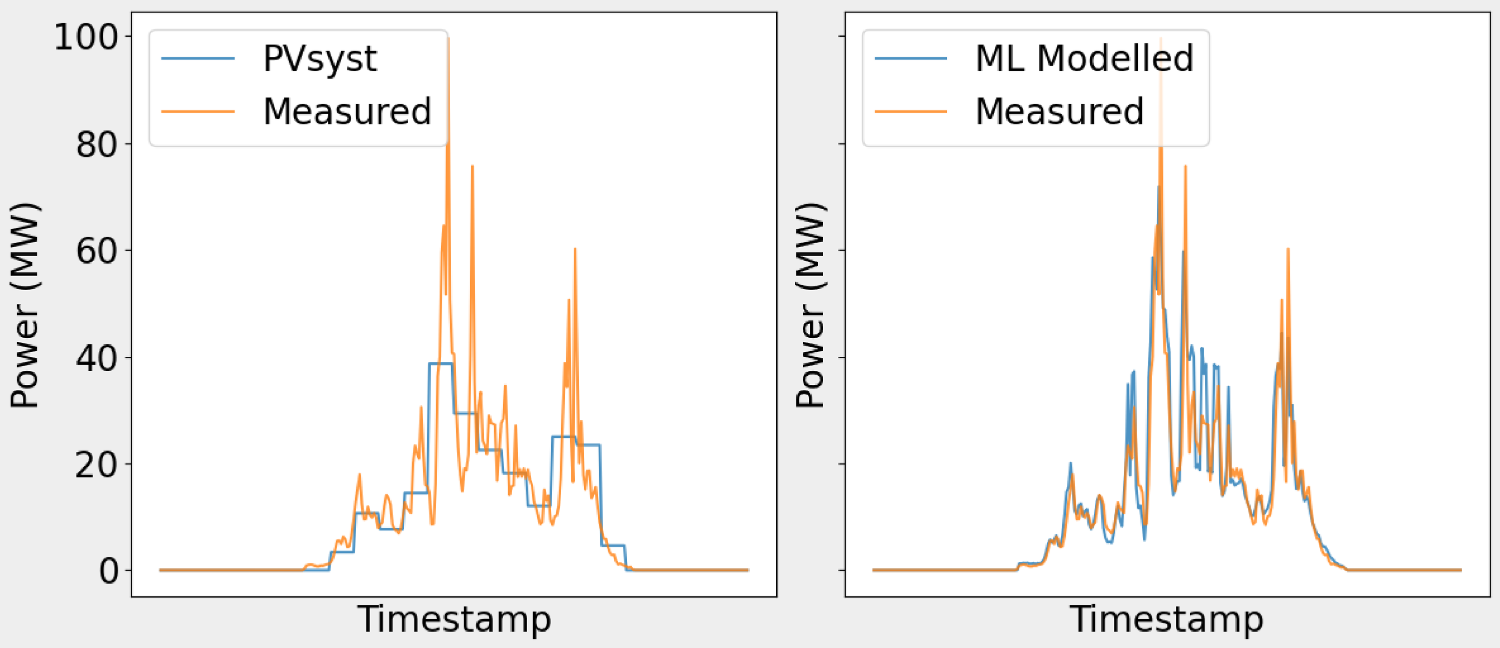

Fig. 1 – A comparison between the measured daily power profile and the power predicted by the (a) PVsyst and (b) ML models

- The ML **input** is the on-site measured weather data
- The **output** data includes the PVsyst-simulated power from each inverter

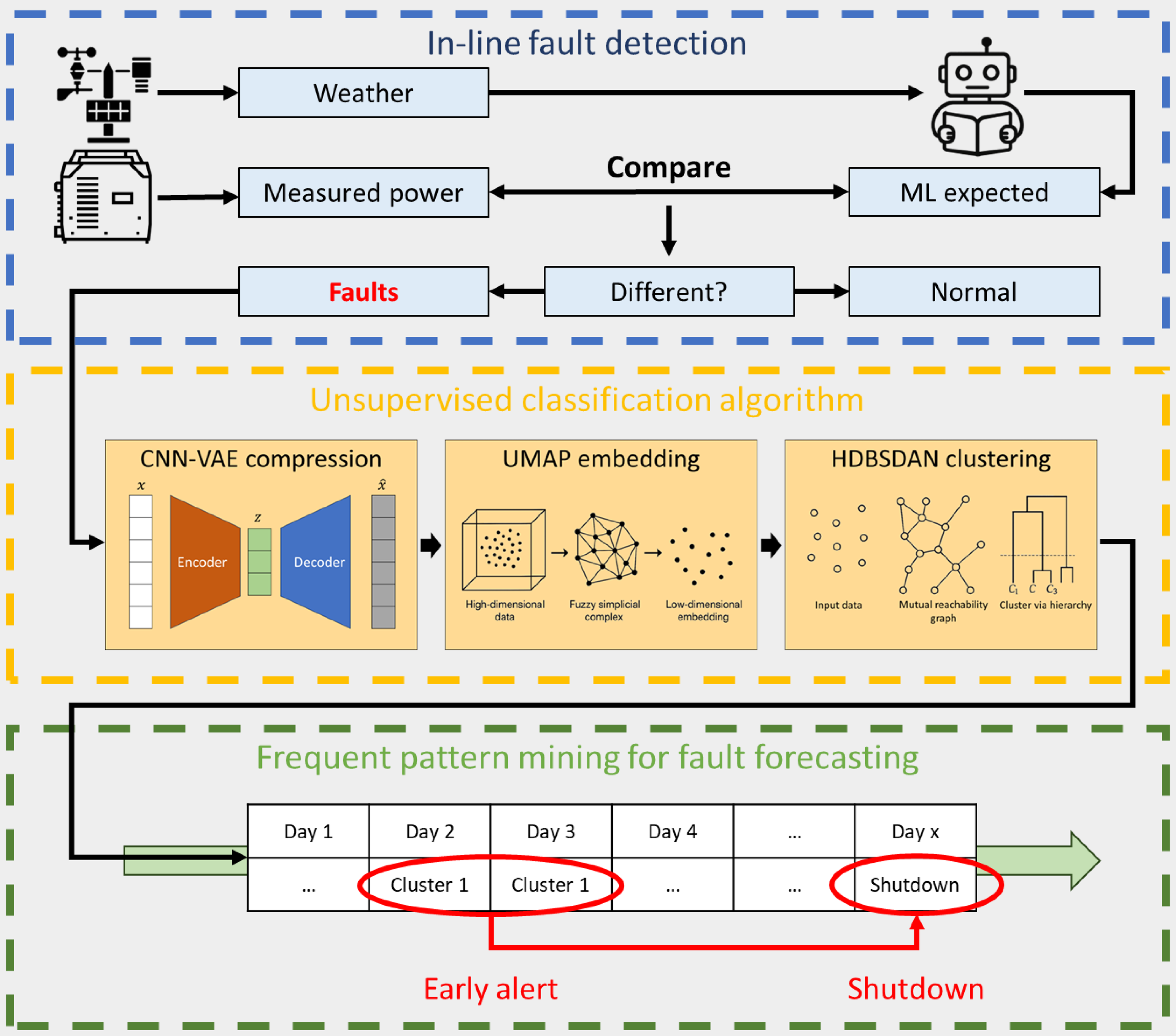

Fig. 2 – Workflow of the proposed ML pipeline

- The classification algorithm groups together the days with **similar features**

In-line detection

- When running the model **in line** with the measurements, red regions indicate underperformance
- The **flexible timestep** enables the sample-by-sample comparison between the measured and ML-predicted power

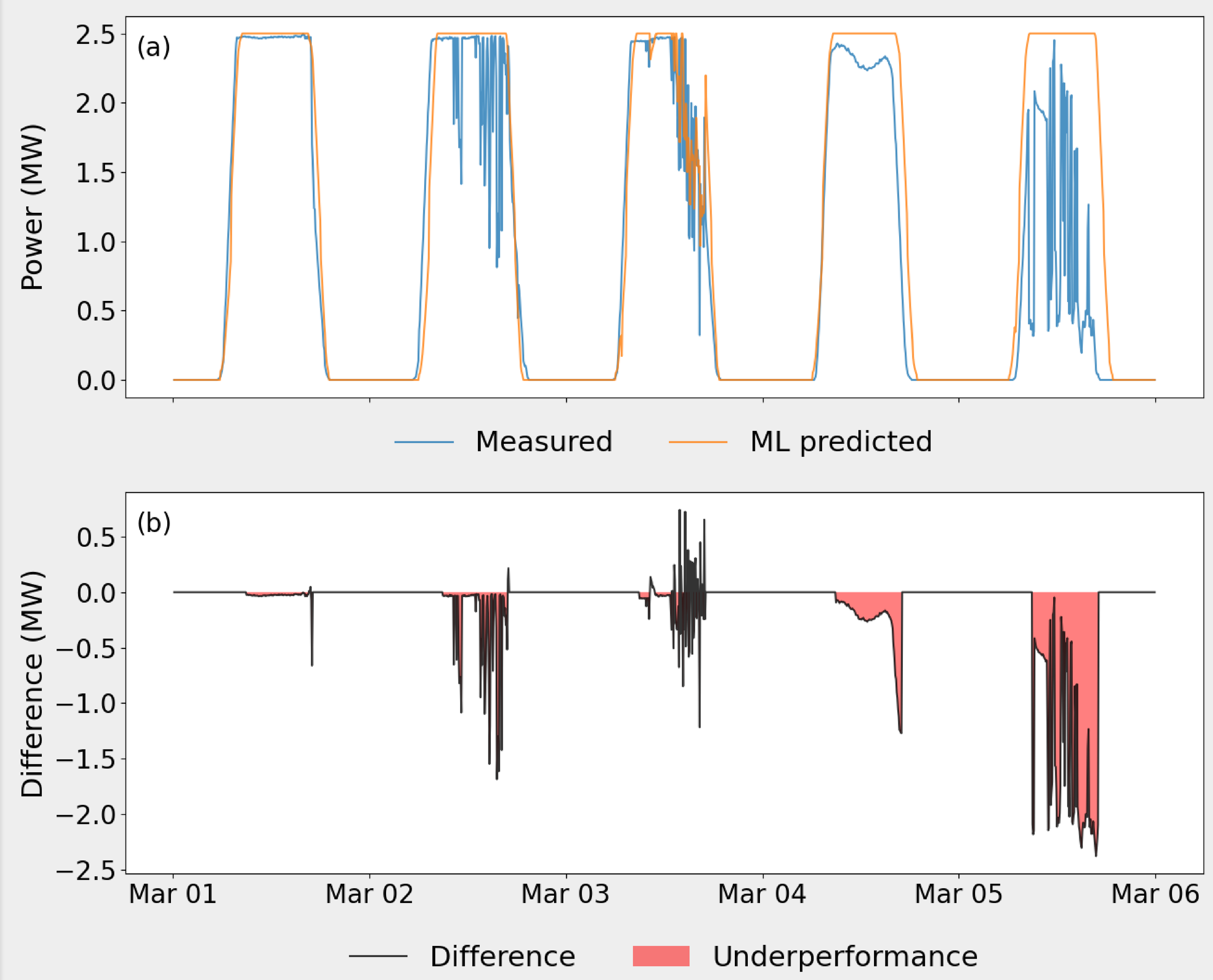

Fig. 3 – A comparison between the measured and ML-predicted power

Fault classification

- Cluster 2 shows pronounced underperformance in both morning and afternoon, likely due to tracker issues
- Cluster 5 could be an early alert for ground fault, which will be discussed in the next session

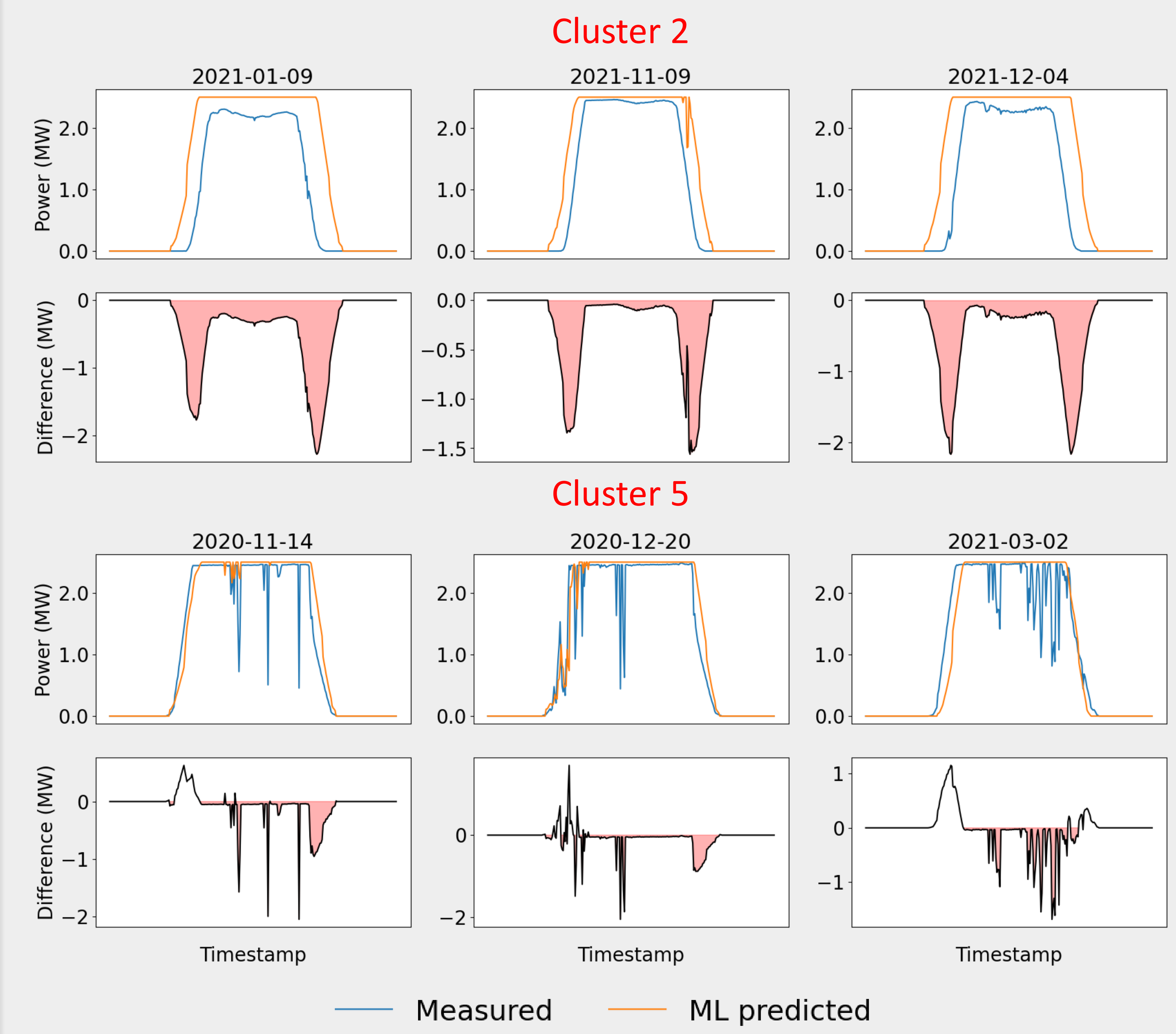

Fig. 4 – Two example fault classification clusters

Fault forecasting

- Forecast faults using the **frequent patterns** seen in the **historical data** to trigger an early alert before a fault
- For example, **Cluster 5** occurred on 28th Feb, which could serve as an early alert for a ground fault (12 days later, the inverter was shut down, and the logbook recorded an underground cable problem)

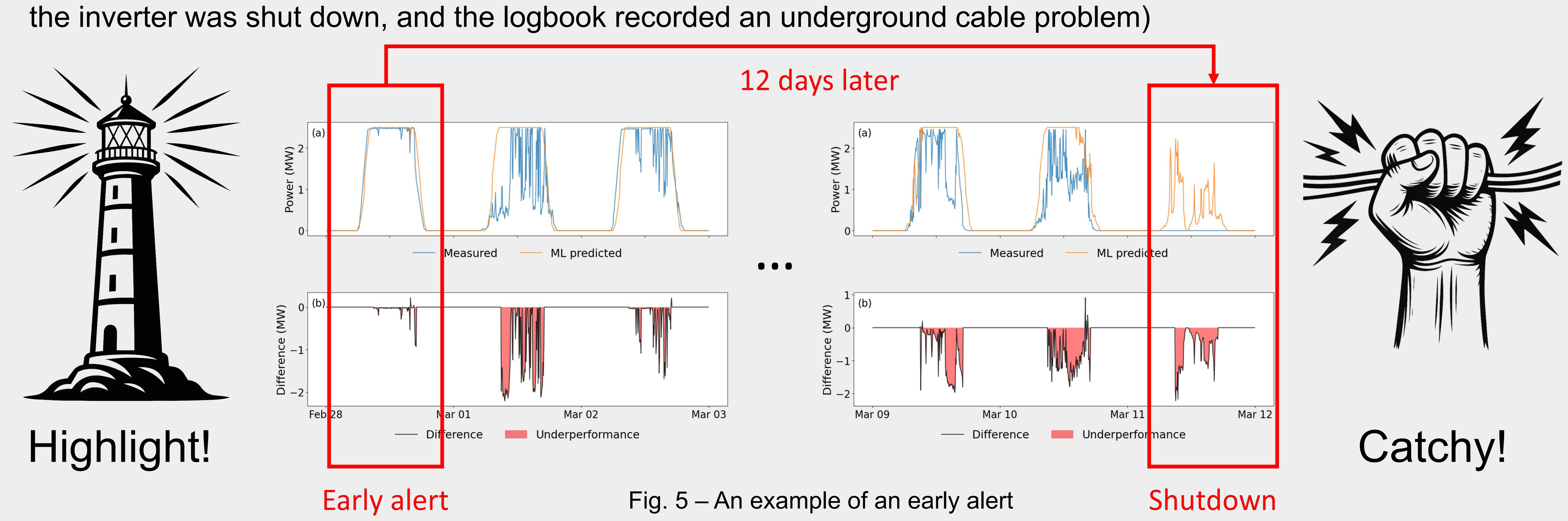

Fig. 5 – An example of an early alert

Conclusions

- We developed a novel, end-to-end pipeline to **detect, classify,** and **forecast** faults in utility-scale PV farms
- An ML model was deployed in **real time** with a **flexible time step** to compare with the measured data
- Another model was developed to classify the faults; some clusters could provide an early alert for forecasting faults
- While rule-based models require manual rule updates for each PV system, this pipeline is fully **autonomous**

Acknowledgements

This work was supported by the Australian Government through the Australian Renewable Energy Agency (ARENA, Grant 2022/ TRAC02).

References

[1] A. Hind *et al.*, *Renew. Sustain. Energy Rev*, vol. 195, pp.114342, 2024.
[2] S. Xu *et al.*, *J Supercomput.*, vol. 79, no. 11, pp. 12266-12291, 2023.
[3] B. Etienne *et al.*, *Nat Biotechnol.*, vol. 37, no. 1, pp. 38–44, 2019.
[4] S. Geoffrey *et al. Appl.Sci.*, vol. 12, no. 5, pp. 2405, 2022.

Contact Details

Sijin Wang
Sijin.Wang@student.unsw.edu.au
www.acdc-pv-unsw.com/