Machine Learning-Based Anomaly Detection for the Flinders University VAWT-X Vertical Axis Wind Turbine

By

Cooper Rogers
Bachelor of Engineering (Software)(Honours)

Supervised by
Matthew Stephenson

Thesis
Submitted to Flinders University
for the degree of

Bachelor of Engineering (Software)(Honours)
College of Science and Engineering
20/10/2025

TABLE OF CONTENTS

ABSTRACT

This thesis describes the creation and testing of a lightweight, interpretable machine-learning framework for anomaly detection in the VAWT-X vertical-axis wind turbine at Flinders University. The project aims to improve condition monitoring and lay the groundwork for predictive maintenance and eventual digital twin integration in small-scale renewable energy systems.

The research was conducted in two stages: offline model training and real-time deployment. Sensor data (torque, RPM, current, power, and vibration) were obtained under both normal and induced-fault circumstances. Features were generated using rolling-window averaging and z-score normalization, with residual vibration features supplied by a Random Forest regressor. Five Logistic Regression classifiers were designed to detect electrical, aerodynamic, vibration, power shortfall, and RPM sensor defects. To ensure alert stability, the probability outputs were filtered using static thresholds and a three-of-five hysteresis rule.

Offline testing showed high classification accuracy, with the Power Deficit detector earning an F1 score more than 0.9, proving the viability of the suggested approach. The Power Deficit detector successfully identified thirty-seven separate load variation events during the real-time testing, which was conducted using a MQTT-based data stream, demonstrating live inference capacity. Other detectors remained inactive due to data alignment and scaling issues, emphasizing the significance of schema verification and adaptive calibration in operational settings.

Although the initial project scope included complete digital twin integration via the XMPro platform, limited turbine readiness and data availability prompted a shift toward model creation and real-time testing. This move enabled more in-depth investigation of anomaly detection logic, as well as the successful demonstration of an autonomous edge-deployable maintenance architecture.

The study indicates that interpretable, low-complexity models can efficiently detect early turbine failures without using cloud computing or black-box techniques. The findings lay a solid platform for future integration into digital twin systems and promote intelligent predictive maintenance in distributed renewable energy applications.

DECLARATION

I certify that this Thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university

2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and

3. to the best of my knowledge and belief, does not contain any material previously published or written by another person, or Artificial Intelligence, except where due reference is made in the text, and in accordance with Flinders University Academic Integrity policy and instruction given in the topic

Signed...

Date.............21/10/2025...............................

Artificial Intelligence

In accordance with Item 3. of the above declaration and the assessment specification in the Topic Information guide on how Artificial Intelligence can be used, Artificial Intelligence has been used in this Thesis for the following purposes:

Sections	**Role of AI** – specify how AI was used
1.	Literature review, structural planning of Section
Document	Grammarly was used for language support throughout the document
N/A	There is no use of AI in this document.

LIST OF ABBREVIATIONS

Abbreviation	Full Term	Description / Context
AI	Artificial Intelligence	Computational methods that enable systems to learn and make decisions from data.
ANN	Artificial Neural Network	A machine learning model inspired by biological neural structures (used in literature comparisons).
CSV	Comma-Separated Values	File format used for storing and logging turbine data and model outputs.
DT	Digital Twin	A virtual representation of the physical turbine used for simulation, analysis, and monitoring.
EMA	Exponential Moving Average	A smoothing technique for time-series data (mentioned in literature, replaced here by hysteresis filtering).
F1	F1 Score	Harmonic mean of precision and recall, used to evaluate classification performance.
FN	False Negative	A fault that was missed by the model.
FP	False Positive	A normal condition incorrectly identified as a fault.
Hz	Hertz	Frequency unit; used to describe sensor sampling rate (cycles per second).
IoT	Internet of Things	Interconnected network of sensors and devices, relevant to turbine monitoring systems.
JSON	JavaScript Object Notation	Data format used for configuration and threshold files.
kW	Kilowatt	Unit of power output from the turbine.
LR	Logistic Regression	Primary classification algorithm used in this study.
ML	Machine Learning	Data-driven approach used to train predictive models for fault detection.
MQTT	Message Queuing Telemetry Transport	Lightweight communication protocol used for real-time data streaming.
PWM	Pulse Width Modulation	Electrical control signal (referenced in sensor and motor control systems).

RF	Random Forest	Ensemble regression model used to generate residual vibration features.
ROC	Receiver Operating Characteristic	Curve used to evaluate model performance across thresholds.
RPM	Revolutions Per Minute	Measurement of turbine rotational speed.
SCADA	Supervisory Control and Data Acquisition	Industrial monitoring system referenced in related work.
SVM	Support Vector Machine	Machine learning algorithm referenced in comparative studies.
TP	True Positive	A correctly identified fault.
TN	True Negative	A correctly identified normal operating condition.
VAWT	Vertical-Axis Wind Turbine	Turbine configuration used for this study.
VAWT-X	Vertical-Axis Wind Turbine - Experimental	Designation of the turbine used at Flinders University.
XMPro	XMPro Digital Twin Platform	Software platform intended for future integration of the anomaly detection framework.
Z-score	Standard Score	Statistical measure used for data normalisation and anomaly detection.

LIST OF SYMBOLS

Symbol	Definition
μ	Mean value of a signal or dataset
σ	Standard deviation of a signal or dataset
z	Standardised z-score used for anomaly detection
P	Power output of the turbine
T	Torque generated by the turbine shaft
V	Voltage measured across the turbine generator
I	Electrical current through the turbine system
ω	Angular velocity of the turbine rotor

R	Residual error between predicted and measured vibration
z_vib	Residual vibration feature derived from Random Forest regression
F1	F1 score, the harmonic mean of precision and recall
PR	Precision metric used for model evaluation
RC	Recall metric used for model evaluation
t	Time variable used in rolling window analysis
W	Rolling window size (number of samples)
Δt	Sampling interval between consecutive measurements

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background and Context

The growing global demand for renewable energy has sparked the interest in solutions to make small-scale wind turbines more efficient and more reliable. Predictive maintenance is a prominent part of this objective, as it allows operators to find and fix problems before they cause downtime or mechanical failure. Machine learning is increasingly being used in modern wind turbine monitoring systems to make sense of sensor data and find problems in real time. However, most of the research and commercial uses right now are focused on big horizontal-axis wind turbines. This leaves a gap in knowledge and tools for smaller vertical-axis turbines, which have different aerodynamic and structural behaviours.

The VAWT-X vertical-axis wind turbine at Flinders University serves as a testing environment for developing small-scale predictive maintenance models. VAWTs, being mechanically simpler than horizontal-axis turbines and can operate in all directions, make them ideal for scattered or urban deployments. However, their non-stationary vibration patterns, varied torque behaviour, and distinct aerodynamic responses provide obstacles for condition monitoring. This study solves these issues by creating a lightweight anomaly detection framework tailored for the VAWT-X system.

1.2 Project Aim and Objectives

The primary aim of this project is to create and test a machine-learning framework for identifying mechanical, electrical, and aerodynamic anomalies in the VAWT-X flinders turbine. The technology is designed to function in real time with and serve as a platform for integration into a digital twin environment.

The main project objectives are as follows:

1. Gather datasets consisting of various baseline/anomalous states, preprocess and prepare this data for training.
2. Train and evaluate interpretable machine learning models that can detect various faults from the limited datasets.
3. Develop a real-time deployment pipeline that utilizes MQTT for live inference and alarm generation.
4. Assess model performance using both offline validation and live testing, including dependability, latency, and accuracy.
5. Identify limits and develop design recommendations for future digital twin integration.

1.3 Project Scope and Evolution

The project's initial goal was to produce a fully fleshed digital twin and predictive maintenance system for the VAWT-X turbine using the XMPro platform. Early in the development process, it became clear that turbine readiness and insufficient testing data would make this scope unachievable given the project timespan. Due to this, the project’s focus was narrowed down to largely focus on the development of the anomaly detection system, with digital twin integration set as a future goal.

This scope shift necessitated a thorough overhaul of the project strategy and methods. Early literature and design work focused on digital twin architecture were changed to allow for a more technical examination of model training, deployment, and real system performance. Although this change restricted the scope of deliverables, it allowed for more technical validation of the suggested machine-learning approach. The resulting architecture is a basic stage of the bigger digital twin idea, which will continue to be developed upon after this project.

1.4 Research Significance

This experiment demonstrates that simple, transparent models can be used to discover faults in an accurate and explainable manner, rather than complicated black-box methods. The interpretability of Logistic Regression enables turbine operators to comprehend which variables contribute to detected abnormalities, hence increasing trust in the system's choices. The employment of a Random Forest regressor to create residual vibration characteristics improves sensitivity to mechanical imbalance while being computationally efficient.

The creation of this framework contributes to ongoing research in small-scale renewable energy monitoring by offering a deployable, edge-compatible solution. It demonstrates how academic research in anomaly detection can be converted into a working system that can be integrated into larger industrial monitoring structures such as digital twins.

1.5 Thesis Structure

The remainder of this thesis is structured as follows:

- Chapter 2: Reviews literature on predictive maintenance, machine learning for anomaly detection, and condition monitoring of vertical-axis wind turbines.
- Chapter 3: Discusses the methodology used to design, train, and implement the anomaly detection framework.
- Chapter 4: Presents the results of both offline and real-time evaluations.

- Chapter 5: Discusses the results in relation to the project's objectives and reflects on its challenges and implications.
- Chapter 6: Concludes the thesis and provides recommendations for future work, including full digital twin integration.

CHAPTER 2: LITERATURE REVIEW

2.1 Machine Learning for Wind Turbine Fault Detection

Machine learning (ML) has altered how defects in wind turbines are found, moving away from traditional rule-based Supervisory Control and Data Acquisition (SCADA) monitoring and toward predictive, data-driven analysis. Early systems used fixed thresholds for factors including torque, temperature, and rotational speed, triggering alarms when values exceeded predefined limitations (Gigoni et al., 2019). Although simple to implement, these algorithms overlooked nonlinear interactions between turbine variables, resulting in false alarms or missing early problems (Zhang, Hu, & Yang, 2022).

2.1.1 Logistic Regression

Logistic Regression (LR) is still a popular baseline for condition monitoring because of its simplicity and transparency. It estimates fault likelihood by assigning interpretable coefficients to quantifiable variables like current, torque, and wind speed (Ng & Lim, 2022). Recent reviews highlight Logistic Regression as a transparent baseline for small or moderately sized SCADA datasets, though reported accuracies vary widely depending on data quality and feature selection (Nejad Alagha et al., 2025). However, LR's linear structure limits its capacity to capture nonlinear dynamics, making it less suitable for complicated fault interactions (Mohapatro et al., 2025).

2.1.2 Random Forest and Ensemble Learning

Random Forest (RF) and ensemble learning techniques address this limitation by combining numerous decision trees to describe nonlinear relationships and handle noisy input. Transfer-learning approaches, such as TrAdaBoost, have been proven to increase cross-turbine fault classification accuracy by adjusting information between turbines operating under different conditions (Chen et al., 2021). RF's variable importance measures also aid in identifying critical fault predictors, allowing for more accurate physical interpretation of model results (Pedro & Gonzalo, 2021). Hybrid frameworks that combine RF and deep models have improved performance by employing RF for coarse classification and neural networks for temporal refinement (Allal et al., 2024).

2.1.3 Isolation Forest and Unsupervised Anomaly Detection

Isolation Forest (IF) is an unsupervised method that is appropriate for settings with few labelled samples. It isolates anomalies using recursive data partitioning, making it ideal for finding unusual or previously unknown problems (McKinnon et al, 2021). While IF may effectively identify outliers in dynamic situations, it is still sensitive to parameter scaling and lacks interpretability when compared to supervised models.

2.1.4 Deep Learning and Temporal Modelling

Deep learning, particularly with LSTM architectures, allows for the modelling of time-dependent behaviour across various sensor systems. Zhang et al. (2022) used an LSTM-based model to capture temporal relationships within SCADA signals, achieving strong fault-diagnosis performance. Despite their accuracy, such networks are computationally expensive and frequently inappropriate for small-scale or embedded installations that demand transparency (Ng & Lim, 2022).

2.1.5 Comparative Evaluation

Hybrid frameworks that combine vibration signal feature engineering with machine-learning classifiers increase fault sensitivity and diagnostic accuracy in drivetrain applications (Jamil et al., 2025). These methods combine physical signal analysis and data-driven learning to achieve a balance of accuracy and interpretability. These findings direct the ongoing research toward lightweight, explainable models that can operate successfully within the data limits of small-scale turbines.

2.2 Machine Learning Models in Vertical-Axis Wind Turbine (VAWT) Applications

While machine learning has become a cornerstone of fault detection in horizontal-axis wind turbines (HAWTs), its application to vertical-axis wind turbines (VAWTs) is still limited. The aerodynamic, structural, and operational variances across different turbine types make it difficult to generalize existing diagnostic tools. HAWT drivetrain experiments have underlined the necessity of vibration and torque signal preprocessing for reliable defect detection (Jamil et al., 2025). As a result, algorithms created for HAWTs frequently underperform when applied to VAWT data without considerable feature engineering and retraining.

Early VAWT fault detection research was mostly focused on vibration-based or threshold-driven monitoring systems. Monitoring systems based on fixed thresholds often failed to differentiate normal operating variability from early signs of component degradation (Gigoni et al., 2019; Moghaddass & Rudin, 2015). The use of machine learning models dramatically enhanced detection accuracy by incorporating multi-sensor data such as torque, current, and wind speed. Logistic Regression remains a popular and interpretable baseline for fault classification in small datasets (Alagha et al., 2025).

Recently, researchers have investigated hybrid and deep learning methods for dealing with the temporal and nonlinear nature of VAWT data. Long Short-Term Memory (LSTM) architectures and convolutional neural networks (CNNs) have demonstrated promising results in detecting early electrical and aerodynamic problems (Wang et al., 2025). Feature-level signal processing, such as empirical mode decomposition and wavelet analysis, improves the sensitivity of machine learning classifiers (Jamil et al., 2025). Transfer learning techniques have also evolved as a tool for adapting pretrained models from large HAWT datasets to smaller VAWT contexts, eliminating the need for considerable local training data (Chen et al., 2020).

Recent experiments have expanded fault diagnosis to structural components like turbine blades. Esquivel-Sancho et al. (2025) used finite-element simulation and modal testing data to detect fractures and delamination in 3D-printed scaled wind turbine blades. The combination of vibration-mode data with machine-learning classifiers resulted in accurate distinction of healthy and damaged blades, demonstrating the potential of AI-assisted structural health monitoring.

Despite these advancements, difficulties remain. Environmental noise, sparse datasets, and aerodynamic variability continue to impede accurate real-time performance (Moshtaghi et al., 2025; Ng & Lim, 2022). As a result, lightweight and interpretable models like Logistic Regression and Random Forest remain the most useful for embedded turbine diagnostics. These techniques strike a balance between detection accuracy, interpretability, and computational simplicity, reflecting the continued attention on optimizing machine learning algorithms through hyperparameter tuning to obtain high accuracy and robustness on turbine fault datasets (Mohapatro et al., 2025; Alagha et al., 2025).

2.3 Challenges, Gaps & Future Research Considerations

While machine learning algorithms have produced outstanding results in wind turbine defect detection, several persisting obstacles prevent their widespread use for small-scale vertical-axis wind turbines (VAWTs). The most pressing concerns include data scarcity, limited model transferability, and the necessity for interpretable algorithms that can work in real time (Alagha et al., 2025; Ng & Lim, 2022).

2.3.1 Data Availability and Quality

High-quality, labelled data are required for accurate ML model training, but such datasets are uncommon for VAWTs. Most publicly available turbine data originate from big horizontal-axis systems (Gigoni et al., 2019; Chen et al., 2020). Small-scale systems frequently rely on brief laboratory experiments or synthetic fault injections, limiting model generality. Furthermore, signal noise and missing data generated by turbulence or sensor drift obscure fault details and weaken model resilience. To address these issues, researchers have proposed data augmentation and adaptive signal decomposition techniques, such as wavelet filtering and empirical mode decomposition, which preserve transient fault characteristics while mitigating noise (Wang et al., 2025; Moshtaghi et al., 2025). These tactics increase data diversity and model stability under dynamic situations.

2.3.2 Model Transferability and Generalisation

Even well-trained models frequently fail when applied to new turbines or changing environmental circumstances. Structural variances, wind regime changes, and sensor discrepancies all affect data distributions, highlighting that drivetrain models often generalise poorly when applied to different turbines without additional tuning (Jamil et al., 2025). Transfer learning and domain adaption approaches provide partial solutions by fine-tuning pretrained models on smaller, local datasets (Chen et al., 2020). Combining physical turbine equations with data-driven models enhances generalization by incorporating domain information into the learning process (Pedro & Gonzalo, 2021). Hybrid physical-data models are as such regarded as an important path for improving dependability across various turbine configurations.

2.3.3 Explainability and Integration into Operations

Model transparency is still critical for industrial implementation. Operators want models that not only predict problems but also explain which sensor features affected their predictions. While algorithms like Logistic Regression and Random Forest produce interpretable results, deep learning techniques like LSTMs are sometimes viewed as opaque. Explainable AI (XAI) techniques, such as Digital Twins (DT), have been integrated into turbine diagnostics to attribute forecasts to key variables, hence enhancing trust and utility (Nejad Alagha et al., 2025).

2.3.4 Future Research Directions

These challenges underscore the need for research on hybrid, interpretable, and edge-compatible machine learning systems that can work successfully with limited data. Creating open-access VAWT datasets, using physics-based reasoning, and building models that balance accuracy with explainability represents the next significant step toward providing comprehensive predictive maintenance for distributed wind energy assets.

2.4 Summary of Reviewed Literature

The reviewed study shows that machine learning has become an essential component of current wind turbine status monitoring. ML models outperform traditional threshold-based methods in recognizing complicated, multidimensional relationships within turbine data and identifying early-stage problems. Logistic Regression, Random Forest, Isolation Forest, and Long Short-Term Memory (LSTM) networks have all demonstrated excellent diagnostic performance in a variety of turbine systems (Nejad Alagha et al., 2025; Wang et al., 2025). Among these, Logistic Regression and Random Forest are especially useful in small-scale or resource-constrained environments because they combine interpretability, computational economy, and predictive capabilities for fault monitoring in SCADA-based systems (Yimam et al., 2025; see also Gigoni et al., 2019 for a large-scale PdM implementation).

Despite these achievements, the literature suggests that there are still significant barriers to the actual application of machine learning (ML) in vertical-axis wind turbines. The non-stationary nature of VAWT data, which exhibit cyclic torque and asymmetric aerodynamic loading, affects both model training and real-time reliability (Pedro & Gonzalo, 2021; Ng & Lim, 2022). Deep-learning approaches have significant temporal modelling skills, but they require large amounts of data and sophisticated hardware, limiting their use to small turbines. Lightweight, interpretable models that can learn quickly from small datasets remain the most promising for field-level monitoring.

Researchers continually highlight the importance of open, standardized datasets in accelerating development and promoting reproducibility. Studies also highlight the possibility of hybrid physical-data models that incorporate mechanical and aerodynamic knowledge into ML frameworks, hence enhancing generalization and fault explainability (Allal et al., 2024; Wang et al., 2025). Recent research proposes that explainable AI techniques, such as feature attribution approaches, can close the gap between high-performance models and operator trust by giving clear reasoning for anomaly detections.

In conclusion, the literature demonstrates a strong need for research into interpretable, data-efficient ML frameworks specifically designed for small VAWTs. Future systems can provide scalable, explainable, and real-time diagnostic solutions for distributed renewable energy networks by overcoming data quality, transferability, and transparency concerns. Building on these findings, the following chapter describes the experimental and methodological framework used in this investigation of the VAWT-X turbine.

CHAPTER 3: METHODS

3.1 System Description

3.1.1 Overview of the VAWT-X Platform

The research was conducted on the VAWT-X vertical-axis wind turbine (VAWT) at Flinders University's Renewable Energy Test Facility in Adelaide, South Australia. The turbine provides a controlled environment for conducting data-driven fault analysis and anomaly detection research. Unlike traditional horizontal-axis turbines, VAWTs run omnidirectionally and have distinct torque and vibration patterns, making defect detection difficult (Pedro & Gonzalo 2021).

The VAWT-X system included an integrated sensor suite capable of monitoring mechanical, aerodynamic, and electrical performance characteristics. The variables measured were rotational speed (RPM), torque, current, voltage, power output, and triaxial vibration. Each signal was routed to a local data acquisition workstation via a BK8601 interface module, which translated sensor readings to digital format for processing. Sampling frequencies varied between 1 and 10 Hz depending on the signal type, in accordance with conventional settings in small-scale turbine testing (Chen et al. 2020; Yimam et al. 2025).

The turbine's control system supported both steady-state and fault simulations, allowing for the creation of labelled datasets. Each experimental session was intended to collect representative data for both the baseline (normal) and induced-fault situations.

Figure 3.1: VAWT-X Wind Turbine Station - Turbine and Local Computer in shot.

3.2 Experimental Setup

3.2.1 Location, Timing, and Assumptions

All experimental trials on the VAWT-X turbine were carried out at Flinders University in September and October of 2025. It was assumed that the turbine's mechanical and electrical components were operational throughout baseline sessions, and that any anomalies introduced were indicative of real-world fault behaviour.

3.2.2 Experimental Design

The experiments were divided into two primary phases:

Phase 1 Offline Evaluation: controlled data collection and model training using baseline and induced-fault datasets.

Phase 2 Real-Time Deployment: live data streaming, model inference, and alert logging via MQTT.

Each trial lasted between ~12 minutes depending on the fault condition. Fault scenarios included:

- **Electrical Fault:** simulated short-circuit and unbalanced current load.
- **Aerodynamic Disturbance:** partial blade obstruction and wind-flow interference.
- **Vibration Fault:** artificial imbalance introduced via calibrated mass offset.
- **Power Deficit:** induced through load reduction on the turbine output.
- **RPM Sensor Fault:** simulated sensor dropout or drift.

Figure 3.2: Example of induced anomaly during test - aerodynamic disturbance using bubble wrap on turbine blade.

3.2.3 Equipment Calibration and Testing

All sensors were calibrated in accordance with manufacturer specifications by the on-campus turbine engineer responsible for the maintenance of the VAWT-X system. Data collection scripts were then executed in *Python (v3.10)* using the *pandas*, *NumPy*, and *matplotlib* libraries to test and acquire the appropriate datasets required for the project.

3.3 Data Acquisition

3.3.1 Data Collection Protocol

Data collection followed a consistent technique throughout all experimental sessions. The turbine was initially run at baseline circumstances for a couple minutes to establish steady-state behaviour. Manually created defects were then sustained for predetermined periods of time until normal operation resumed. Each session created a time-series dataset with synchronized readings from all sensors.

The data was saved in comma-separated values (CSV) format with timestamped entries. Overall, 11 test sets were recorded, with 4 baseline datasets at varying windspeeds, as well as 7 anomaly induced sets. All Data Acquisition and testing notes can be found in Appendix 6:

Test 1- High Wind Base.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	1,019 KB
Test 2- Mid Wind Base.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	1,017 KB
Test 3- Low Wind Base.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	1,018 KB
Test 4- Variable Wind Base.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	758 KB
Test 5- Balanced Short Circuit Fault.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	631 KB
Test 6- Balanced Open Circuit Fault.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	590 KB
Test 7- Unbalanced Open Circuit Fault.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	737 KB
Test 8- Unbalanced Short Circuit Fault (Line-...	04/09/2025 9:30 pm	Microsoft Excel Com...	759 KB
Test 9 - rpm sensors malfunction.csv	04/09/2025 9:30 pm	Microsoft Excel Com...	408 KB
Test 10- disturbance in aerodynamic blade.c...	04/09/2025 9:30 pm	Microsoft Excel Com...	226 KB
Test 11-imbalance in weight (adding weight)...	04/09/2025 9:30 pm	Microsoft Excel Com...	344 KB

Figure 3.3: The 11 Collected Test Data Sets, including 4 baseline sets (1-4) as well as 7 anomalous sets (5-11)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Timestam	Sensors/T	Sensors/V	Sensors/r	Sensors/B	Sensors/B	Sensors/B	Sensors/B	Sensors/B	Sensors/B	Sensors/V	Sensors/V	Sensors/V	bk8601/vo	bk8601/cu	bk8601/pow
########													0.0132	0	0
########	-0.02995								-151.46	-6.27	9.41	18.03	0.0132	0	0
########	-0.0274	0.3	2.05	0	-126.02	-181.89	-201.13	-156.7	-151.38	-3.92	10.19	20.38	0.0132	0	0
########	-0.0274	0.3	2.05	0	-126.02	-181.89	-201.13	-156.7	-151.38	-3.92	10.19	20.38	0.0135	0	0
########	-0.03321	0.3	2.05	0	-126.02	-181.89	-201.13	-156.7	-151.38	-3.92	10.19	20.38	0.0135	0	0
########	-0.03704	0.31	2.05	0	-125.91	-182.05	-201.07	-156.74	-151.44	-5.49	10.19	20.38	0.0135	0	0
########	-0.03336	0.3	2.05	0	-125.91	-182.05	-201.07	-156.74	-151.44	-5.49	10.19	21.17	0.0132	0	0
########	-0.03023	0.31	2.05	0	-125.91	-182.05	-201.07	-156.74	-151.44	-5.49	10.19	21.17	0.0132	0	0
########	-0.02797	0.3	2.05	0	-125.91	-182.05	-201.07	-156.74	-151.44	-5.49	10.19	21.17	0.0135	0	0
########	-0.03378	0.31	2.05	0	-125.91	-182.05	-201.07	-156.74	-151.44	-5.49	10.19	21.17	0.0135	0	0
########	-0.04045	0.33	2.05	0	-125.99	-181.97	-201.02	-156.63	-151.32	-4.7	8.62	21.95	0.0135	0	0
########	-0.03378	0.33	2.05	0	-125.99	-181.97	-201.02	-156.63	-151.32	-4.7	8.62	21.95	0.0132	0	0
########	-0.02981	0.32	2.05	0	-125.99	-181.97	-201.02	-156.63	-151.32	-4.7	8.62	21.95	0.0132	0	0
########	-0.03421	0.33	2.05	0	-125.94	-182.06	-201.11	-156.63	-151.32	-4.7	8.62	21.95	0.0132	0	0
########	-0.02343	0.35	2.05	0	-125.94	-182.06	-201.11	-156.74	-151.27	-3.92	8.62	21.17	0.0132	0	0
########	-0.02343	0.35	2.05	0	-125.94	-182.06	-201.11	-156.74	-151.27	-3.92	8.62	21.17	0.0135	0	0
########	-0.02414	0.34	2.05	0	-125.94	-182.06	-201.11	-156.74	-151.27	-3.92	8.62	21.17	0.0135	0	0
########	-0.02456	0.35	2.05	0	-126	-181.94	-201.02	-156.69	-151.38	-6.27	7.84	18.82	0.0132	0	0
########	-0.02456	0.35	2.05	0	-126	-181.94	-201.02	-156.69	-151.38	-6.27	7.84	18.82	0.0132	0	0
########	-0.03208	0.34	2.05	0	-126	-181.94	-201.02	-156.69	-151.38	-6.27	7.84	18.82	0.0132	0	0
########	-0.02967	0.33	2.05	0	-126	-181.94	-201.02	-156.69	-151.38	-6.27	7.84	18.82	0.0135	0	0
########	-0.03619	0.35	2.05	0	-125.97	-181.93	-201	-156.71	-151.37	-5.49	7.84	18.82	0.0135	0	0
########	-0.03222	0.35	2.05	0	-125.97	-181.93	-201	-156.71	-151.37	-5.49	7.84	19.6	0.0133	0	0
########	-0.02995	0.34	2.05	0	-125.97	-181.93	-201	-156.71	-151.37	-5.49	7.84	19.6	0.0133	0	0
########	-0.02598	0.35	2.05	0	-125.9	-181.91	-200.94	-156.49	-151.29	-5.49	7.84	19.6	0.0133	0	0

Figure 3.4: CSV Snapshot of data fields in test set (Test 1 – Baseline Low);

3.3.2 Pre-Processing and Feature Extraction

Raw data underwent a standardized pre-processing phase. A 60-sample rolling window was used to calculate local means and standard deviations for all signals, smoothing out high-frequency noise while maintaining operational patterns (Jamil et al. 2025).

Each variable was standardized to a *z-score*, which was calculated as:

$$z = x - \mu\sigma z = \frac{x - \mu}{\sigma z} = \sigma x - \mu$$

where *x* represents the observed signal, *μ* the baseline mean, and σ the baseline standard deviation. This transformation enabled models to detect departures from normal conditions, as shown in previous turbine fault detection investigations (Wang et al. 2025; Yimam et al. 2025).

To improve the system's ability to detect mechanical imbalance, a Random Forest regressor was trained using baseline data to predict predicted vibration magnitudes based on contextual factors such as wind speed, torque, and power. The residual-based vibration function detects minor variations indicating underlying degradation. While the current approach employs a data-driven residual model, comparable concepts for tracking latent degradation progression have been formalized in dependability models such as the Latent State Hazard framework (Moghaddass & Rudin, 2015).

3.3.3 Data Cleaning and Integrity Checks

Data quality checks were performed prior to model training. Missing data produced by brief sensor interruptions were linearly interpolated when the intervals were less than five consecutive samples. Longer gaps were excluded to avoid distortion, outliers higher than three standard deviations from the baseline mean were capped. The final collection included thousands of tagged feature vectors, each representing a snapshot of turbine state. The dataset was separated over 70% training and 30% validation, ensuring class balance across baseline and fault situations.

3.4 Data Analysis

3.4.1 Model Development

Machine learning was utilized to identify each time frame as "normal" or "faulted." Five separate Logistic Regression classifiers were created, one for each type of fault. Logistic Regression was chosen for its interpretability and efficiency on tiny datasets, which is consistent with findings from comparative ML assessments that show lightweight models are competitive for fault classification (Mohapatro et al., 2025).

For the vibration features, the Random Forest regressor was employed as a feature generator rather than a classifier, resulting in the previously stated residual-based input. This hybrid strategy is consistent with current ML frameworks that combine coarse fault detection with subsequent fault type identification, improving interpretability and diagnostic accuracy (Zaid Allal et al., 2024; Jamil et al., 2025).

3.4.2 Model Training and Validation

Training was carried out on a local workstation using Python and open-source libraries. Models were validated with five-fold cross-validation, which provided a reliable estimate of generalization performance. Precision, recall, and F1 score were among the evaluation indicators used.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot (\mathrm{Pr}ecision \cdot Recall)}{\mathrm{Pr}ecision + Recall}$$

These metrics allowed quantitative comparisons between detectors and were chosen for their prevalence in turbine fault detection studies (Yimam et al. 2025; Zhang, Hu, & Yang 2022). Performance findings were visualized using ROC curves and confusion matrices, which provided information about classification thresholds and misclassification rates.

3.4.3 Threshold Calibration and Alert Logic

The continuous model probability outputs were transformed into binary warnings with static thresholds. Thresholds were defined using 95th percentile probabilities from the baseline dataset. A

3-of-5 hysteresis filter was used, and three consecutive samples above the threshold resulted in a continuous warning. This logic enhanced alert stability while reducing transient noise.

During live deployment, the Power Deficit model threshold was manually changed from 0.93 to 0.85 to improve sensitivity in changeable wind situations. This real-time calibration demonstrated the viability of adaptive thresholding; a strategy found in other turbine monitoring systems to improve recall (Zaid Allal et al. 2024; Wang et al. 2025).

3.5 Real-Time Deployment

3.5.1 System Integration via MQTT

Following validation, the models were deployed via a MQTT-based pipeline for real-time anomaly detection. The system subscribed to real-time topics like Sensors/rpm, Sensors/torque, Sensors/current, and Sensors/power. Each incoming message was processed and converted into feature values like those used during training, as well as predictions were computed locally on the workstation every second. Each detector's probability, threshold, and alert statuses were included in the output. The results were appended to a CSV log file and shown using an accompanying dashboard.

The pipeline successfully maintained a consistent throughput of roughly 1 Hz, demonstrating its applicability for real-time monitoring. Although typographical and feature alignment issues prevented all models from running simultaneously, the Power Deficit detector generated consistent and relevant alerts in response to actual power dips.

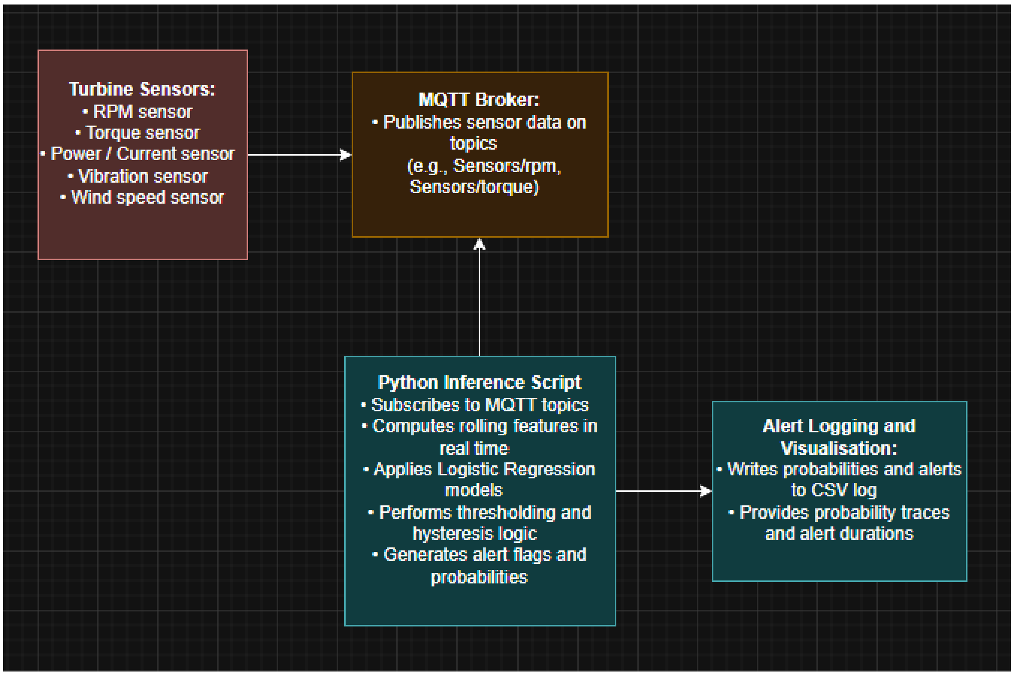


Figure 3.5: Real-time MQTT inference and alert pipeline architecture:

3.5.2 Data Logging and Post-Processing

Every real-time inference result was timestamped and logged. Post-run scripts combined alert frequency, duration, and probability distributions. This logging enabled the quantitative validation of live performance as well as the correlation of model output to measured turbine activity.

A visual check of the live Power Deficit Probability Trace validated the system's capacity to track power variations in real time. Other detectors remained dormant owing to configuration mismatches but were kept for future integration.

3.6 Planning Items

3.6.1 Risk Assessments

Risk assessments were carried out for both safety and project performance. Physical hazards included moving turbine blades, electrical exposure, and vibration threats. The table below shows each risk and its mitigating options.

Table 3.1: Risk Assessment Table (Refer to risk assessment matrix in Appendix 4)

Risk ID	Risk Type	Hazard Description	Mitigation Strategy	Impact	Probability	Risk Score
R1	Rotational Hazard	Both turbines presented potential entanglement or contact risks.	A safe-distance strategy was implemented during the operation of the wind turbine.	High	Low	Medium
R2	Electrical Hazard	Risk of electrical shock from faulty cabling or open terminals.	All equipment was checked before use, lab protocols followed.	High	Low	Medium
R3	Thermal Hazard	In thermal stress trials, directed heat could cause burns or fire risk.	All tests were supervised, and thermal sources were kept at safe distances.	Medium	Medium	Medium
R4	Sensor Failure	Sensor failure or data corruption	Redundant backups, dual logging to SSD + cloud	High	Medium	High
R5	Data inconsistencies	Inconsistent data availability	Optimise code to omit inconsistencies	Medium	Medium	Medium
R6	Model Overfitting	Model overfitting occurs when training on the gathered data	Cross-validation, dropout layers, early stopping, re-asses reconfiguration	High	Medium	High
R7	Project Delays	Integration dela	Early testing with mock data, staggered deployment	Medium	Low	Low

R8	Computational Resource Limitation	Real-time models deployed may exceed processing capacity.	Optimise models, re-asses model usage and project goals.	High	Low	Medium
R9	Data Labelling Ambiguity	Fault labels from lab trials could be misaligned with sensor logs, leading to poor model performance.	Synchronize manual logs with timestamps; use written notes during trials for cross-verification.	Medium	Low	Low
R10	AI Model Accuracy	AI Model accuracy is poor, leading to low quality anomaly detection	Implement Multiple Models	Medium	Low	Low
R11	Time Constraints	Time limit is exceeded, with an incomplete product	Focus on MVP (Minimal Viable Product)	Medium	Medium	Medium

3.6.2 Project Timeline

The project was divided into two university semesters, with key milestones and deliverables timed to correspond with hardware availability, assessment dates, and software development cycles. The Gantt chart in Appendix 5 visualizes the initial project chronology, whereas Figure3.6 Below demonstrates the updated one. The project kicked off in Semester 1 with a literature review and proposal session, followed by system setup and familiarization. Based on turbine sensor installation, data stream generation, visualisation in XMPro, and data collecting would take place at the start of semester 2. The integration of artificial intelligence and predictive maintenance (PdM) models would continue in Semester 2 until the scope changed due to limited data and turbine installations, shifting the project's focus to anomaly detection. This necessitated further investigation and planning. Assessments, including results, thesis and viva, were set for the end of the semester.

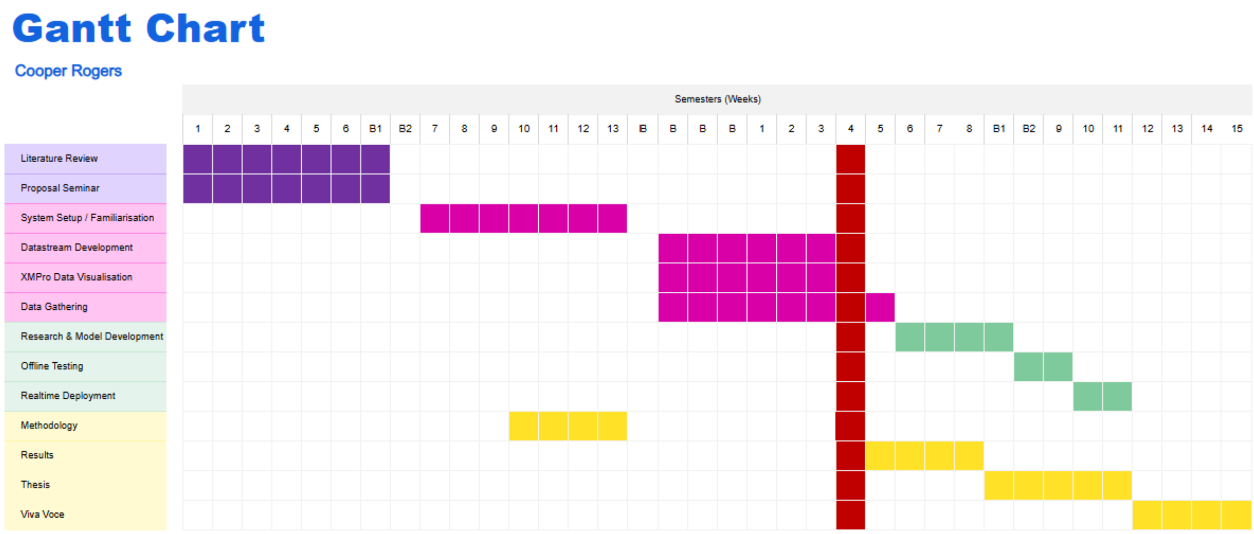


Figure 3.6: Project Timeline Gantt Chart (Updated)

3.6.3 Ethical and Sustainability Considerations

Because there was no personal or confidential data involved in the study, ethical approval was unnecessary. From a sustainability standpoint, the study corresponds with renewable energy optimisation objectives by prolonging turbine life and lowering maintenance-related resource use.

3.7 Sources of Experimental Error

Sensor noise was the most common source of inaccuracy. To counteract these effects, rolling averages and hysteresis filters were used, and repeated trials were carried out to assure representative outcomes.

Human error during configuration and deployment, particularly typographical errors in topic names, contributed to the model's inactivity. To avoid this from happening again, future versions of the system will incorporate schema verification scripts.

3.8 Chapter Summary

This chapter describes the experimental methodologies, analytical methods, and deployment processes utilized to create the VAWT-X anomaly detection framework. The study integrated interpretable Logistic Regression models with residual feature improvement, trained and validated using controlled trials, then deployed over MQTT for real-time inference. The system assured reproducibility by providing detailed documentation of equipment, calibration, and software tools.

Although data and time constraints limited full multi-model deployment, the implemented technique demonstrated technological capability and laid the groundwork for future extension into a complete digital twin system.

CHAPTER 4: RESULTS

4.1 Experimental Setup and System Description

The VAWT-X anomaly detection framework was tested in two phases: offline using controlled baseline and induced-fault datasets, and real-time deployment using live data streaming from the turbine via MQTT. The test sought to demonstrate the framework's ability to detect operational problems and disturbances in small-scale vertical-axis wind turbines (VAWTs) using understandable, low-complexity machine learning models.

The system architecture adhered to the design given in Chapter 3. Sensor data were collected from the VAWT-X turbine at Flinders University's renewable energy plant, and included measurements of wind speed, rotational speed, torque, power, current, and vibration. Data were pre-processed with a rolling 60-sample average window to smooth noise and capture short-term trends before being standardized using z-score normalization. In addition to these foundation features, a Random Forest regression model was trained to predict expected vibration levels, allowing the computation of residual vibration errors (z_vib), a sensitive indicator of mechanical imbalance. These processed features were then fed into five Logistic Regression (LR) classifiers, each of which corresponded to a different fault category: electrical, aerodynamic, vibration, power shortfall, and RPM sensor faults.

For both offline and live testing, fault probabilities were translated into binary alarm signals using static quantile-based thresholds. Each detector used a three-of-five hysteresis rule, which required three consecutive samples above the threshold to produce a confirmed alarm. This strategy lowered transient noise while improving alert reliability.

Offline evaluation included training and validation on structured datasets under normal and fault circumstances. The data was divided into windows for model training (70%), and validation (30%). For the live testing, the models were distributed using the MQTT pipeline, which subscribed to real-time turbine topics. The deployment code computed rolling features on the fly, using the same z-score scaling as offline data, and output results to a CSV log file. The system ran constantly at a sample rate of about 1 Hz, providing near-real-time inference with low processing overhead.

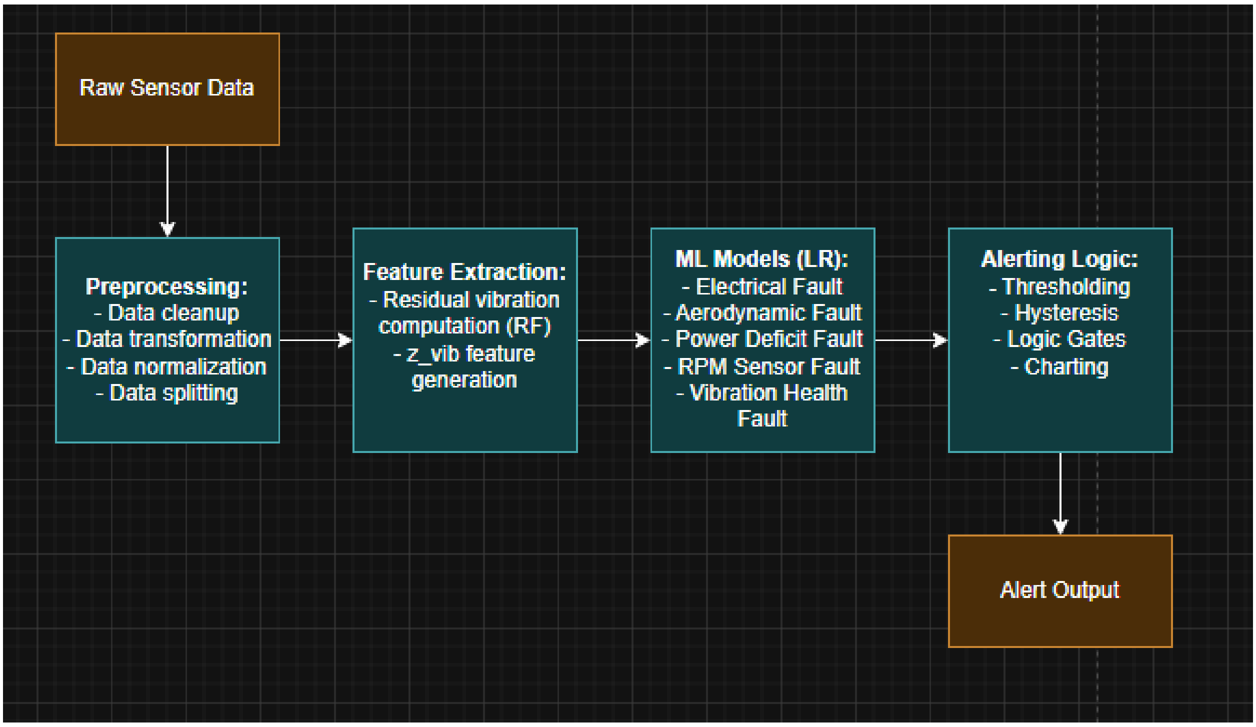


Figure 4.1: Overview of the anomaly detection framework and data processing workflow.

4.2 Presentation and Analysis of Results

4.2.1 Offline Evaluation

Before live deployment, the proposed models were tested offline to ensure that they could correctly discern between healthy and faulty operating states. The *viz_by_group_thresholds* script produced figures displaying each detector's probability traces under both baseline and induced fault circumstances. These visualizations, combined with the quantitative results, provided a thorough knowledge of each detector's strengths and limits.

From a purely quantitative viewpoint, across all models, the Power Deficit detector provided the most consistent and reliable results. Its recall of 0.714 and *F1* score of 0.833 suggest a balanced detection capability, accurately identifying periods of lower power generation caused by controlled load disturbances. The Aerodynamic Disturbance detector had the best overall performance, with a recall of 0.857 and an *F1* score of 0.923, indicating good reactivity to airflow interruptions and a low false alarm rate. The RPM Sensor Fault detector has comparable reliability to the Power Deficit model (*F1* = 0.833), reliably identifying sensor drift events but missing tiny fluctuations that remained below the static threshold.

The electrical fault detector performed moderately, with a recall of 0.571 and an *F1* score of 0.727. It correctly detected major current abnormalities but sometimes failed to respond to lower-amplitude fluctuations. Finally, the Vibration Health detector had the lowest recall (0.429) and *F1* score (0.600), indicating that the residual vibration feature needed more scaling or recalibration to increase reactivity to mild mechanical imbalance.

It is vital to highlight that all detectors have flawless precision (1.00). This result stems from the way the evaluation framework was designed, with each fault segment evaluated as a positive case for the accompanying model. As a result, all detections in fault-labelled data were counted as true positives, with no false positives achievable under these test conditions. Precision values are likely to vary in a real-world operating situation, where detectors only trigger for specified anomaly types. Thus, the flawless precision seen here is mostly due to the absence of erroneous triggers in the controlled dataset, rather than general fault-free performance.

Table 4.1: Summary of mean precision, recall, and F1 scores across all detectors.

Detector	Precision	Recall	F1 Score	Interpretation
Power Deficit	**1.00**	**0.714**	**0.833**	High overall reliability; accurately detected load imbalance with consistent threshold crossings.
Electrical Fault	**1.00**	**0.571**	**0.727**	Detected major current irregularities but missed lower-intensity fault periods.
Aerodynamic Disturbance	**1.00**	**0.857**	**0.923**	Strong performance; effectively identified airflow disruptions with minimal false triggers.
RPM Sensor Fault	**1.00**	**0.714**	**0.833**	Reliable detection when sensor drift occurred; recall could improve with lower thresholds.
Vibration Health	**1.00**	**0.429**	**0.600**	Least sensitive model; residual feature scaling reduced responsiveness to minor imbalance.

Visual analysis of the charts assist these quantitative findings.

The Aerodynamic Disturbance detector (Figures 4.3 and 4.4) had probabilities that regularly exceeded the alert level in both baseline and fault segments, indicating that the model was highly sensitive to transient turbulence. Although this sensitivity enables it to detect actual aerodynamic interruptions, the high number of warnings during normal operation indicates that its threshold or decision boundary should be adjusted, or additional feature tuning done, to reduce false triggers.

The Electrical Fault Detector (Figures 4.5 and 4.6) seemed to be the most consistent model visually. Stable sensor data allowed for a clear delineation between baseline and fault occurrences, and the minimal number of alarms during baseline operation (one incidence) appears appropriate and accurate. Under induced fault conditions, alarms were triggered at the appropriate intervals, demonstrating that the model behaved consistently when data quality was high.

The Power Deficit detector (Figures 4.7 and 4.8) showed consistent early spikes as the turbine ascended to a stable operating speed. Once in steady state, the model occasionally delivered accurate defect detections in response to induced load variations. This pattern implies that initial startup transients should be filtered or omitted from evaluation windows to avoid misclassifying typical ramp-up behaviour as errors.

The RPM Sensor Fault detector (Figures 4.9 and 4.10) showed significant variance in anticipated probability, fluctuating between low and high levels. Despite this volatility, the detector properly identified sensor anomalies during the fault runs, and the number of alerts generated appeared reasonable for the fault size used. Some smoothing or threshold change may help to stabilize its answers in subsequent iterations.

The Vibration Health detector (Figures 4.11 and 4.12) produced alert patterns that were visually like true mechanical disturbances. Both baseline and fault traces were decent, but their great variability made them difficult to read at a glance. The warnings within the fault segments corresponded to produced imbalance occurrences, demonstrating that the detector was functioning properly, even though the residual vibration feature is still sensitive to noise.

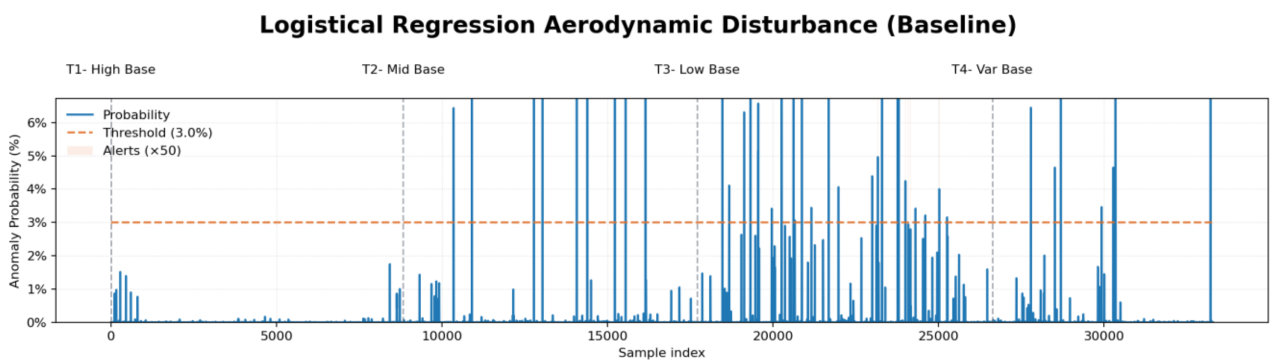


Figure 4.3: Aerodynamic Disturbance - Baseline probability trace.

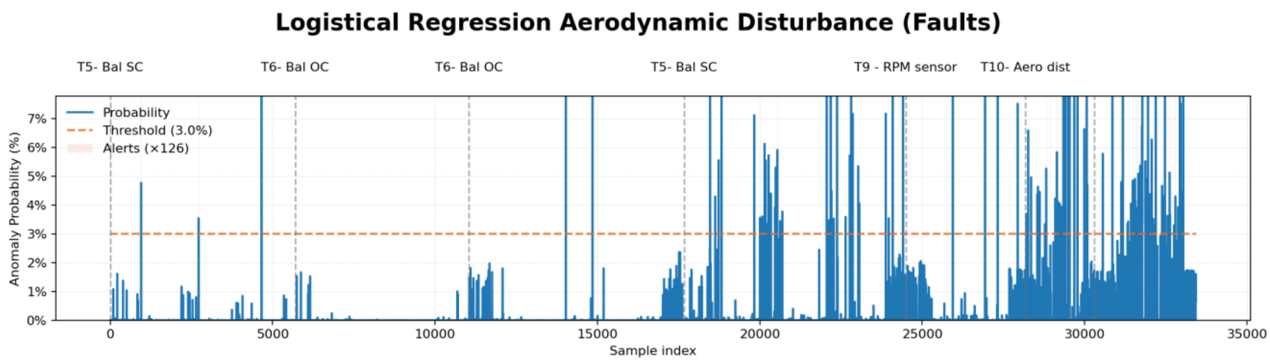


Figure 4.4: Aerodynamic Disturbance - Fault probability trace.

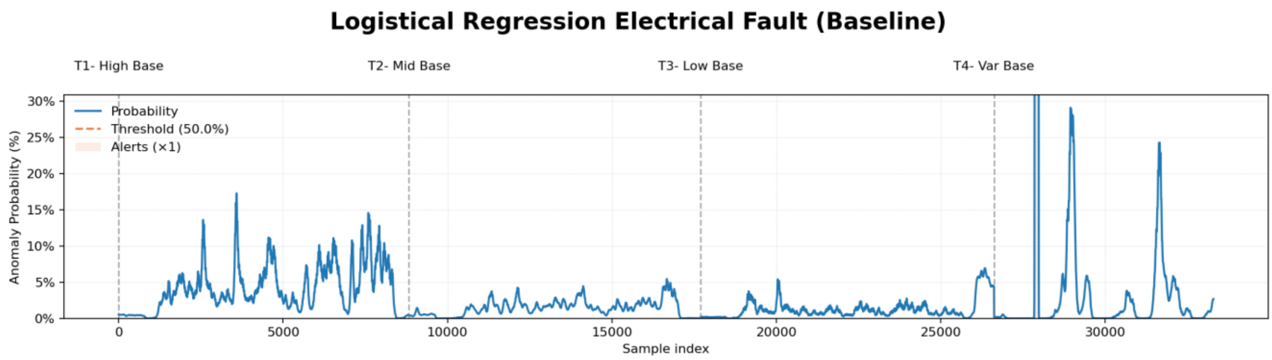


Figure 4.5: Electrical Fault - Baseline probability trace.

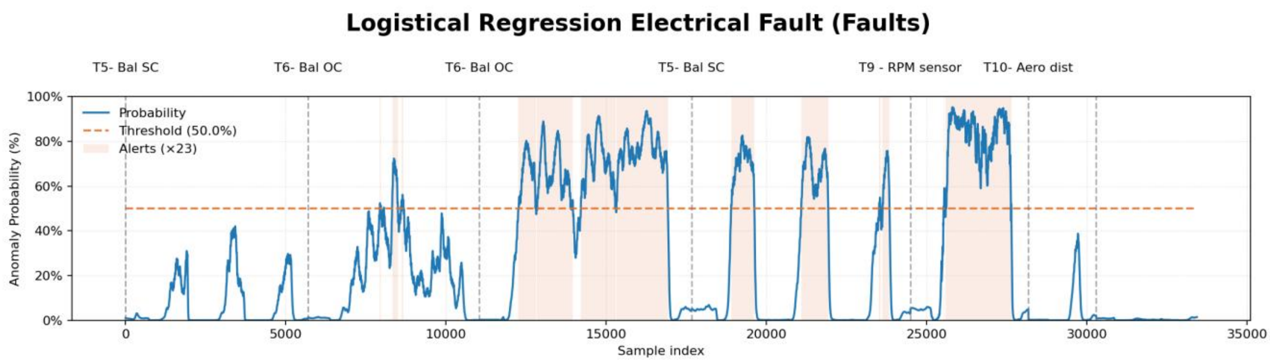


Figure 4.6: Electrical Fault - Fault probability trace

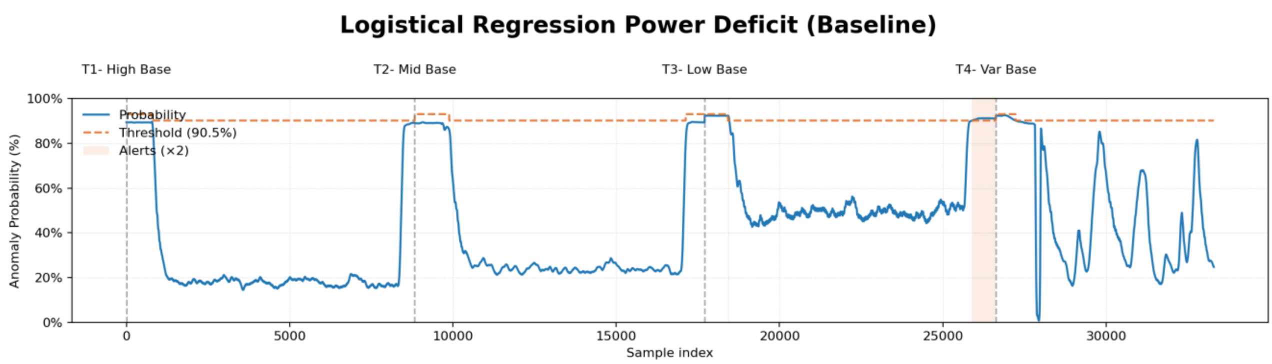


Figure 4.7: Power Deficit - Baseline probability trace

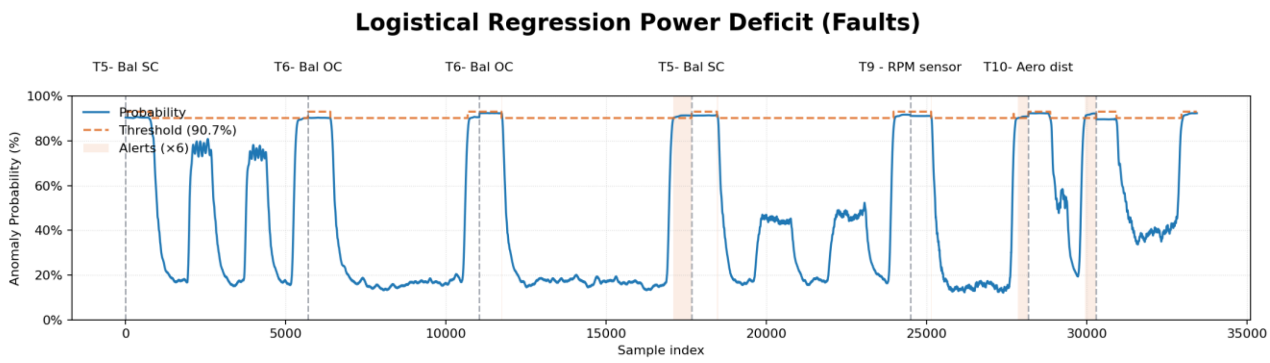


Figure 4.8: Power Deficit - Fault probability trace.

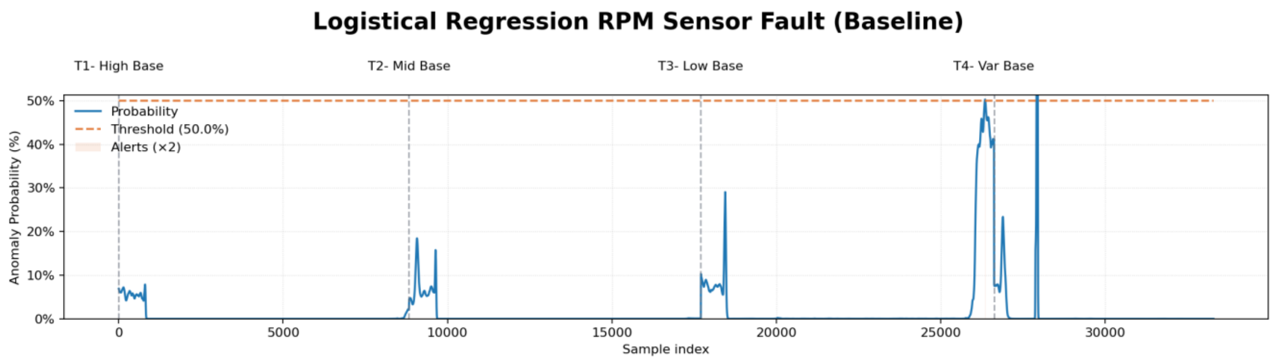


Figure 4.9: RPM Sensor Fault - Baseline probability trace.

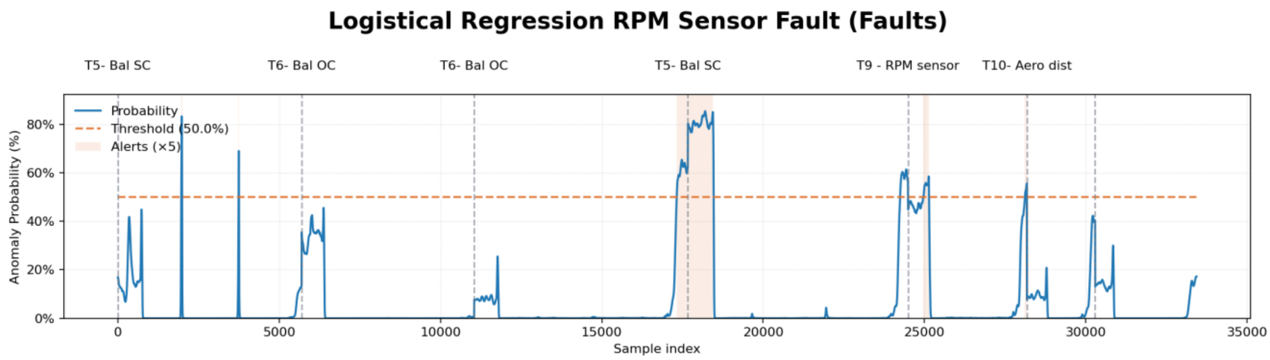


Figure 4.10: RPM Sensor Fault - Fault probability trace.

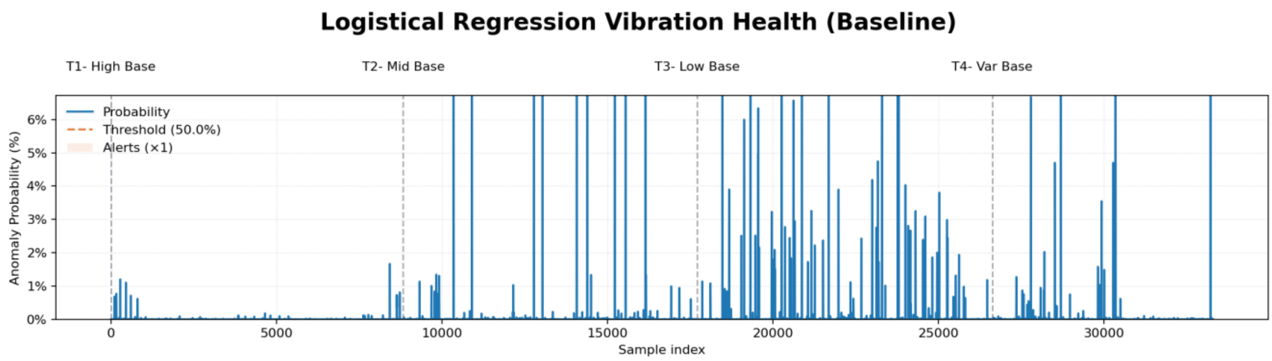


Figure 4.11: Vibration Health - Baseline probability trace

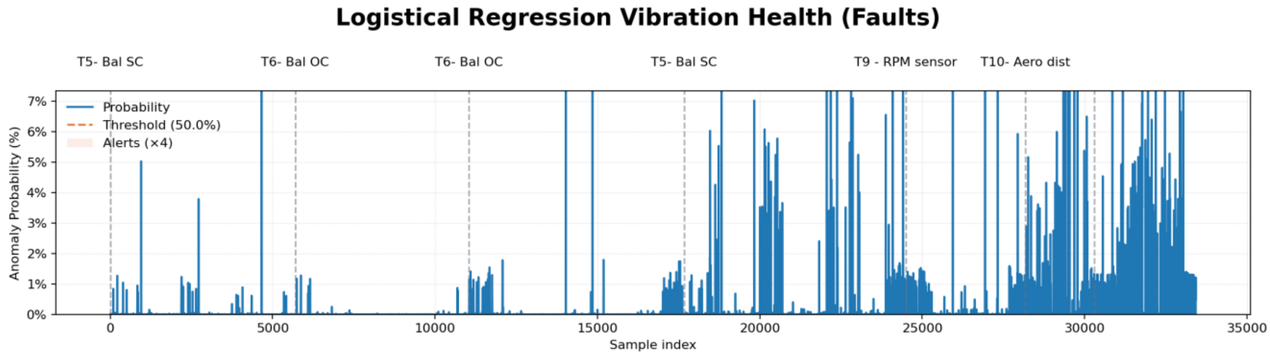


Figure 4.12: Vibration Health - Fault probability trace.

Overall, the offline evaluation shows that the anomaly detection system can correctly categorize multiple anomaly type. These findings corroborate the framework's architecture and indicate that Logistic Regression models, along with residual feature engineering, provide an interpretable and computationally economical solution for wind turbine problem identification.

4.2.2 Real-Time Deployment and Results Analysis

Following a relatively complete offline evaluation, the system was deployed live on the VAWT-X turbine via the MQTT data stream. The real-time test was designed to evaluate the models' operational dependability under continuous data flow and demonstrate their capacity to do inference at the edge without cloud help.

During the live deployment, all five detectors were operational, logging outputs every second. However, only the Power Deficit Detector produced relevant and understandable alerts. A total of 37 unique Power Deficit alert incidents were collected, each representing recognizable decreases in power production caused by load fluctuations or transient aerodynamic effects. This behaviour demonstrated that the system could detect true performance issues in real time.

The remaining detectors yielded either continuous zero probability or unreasonably low values (on the order of *1e−22*), indicating that the models were getting incorrect or missing input data. A post-trial investigation found various possible explanations for this conduct. First, a typographical issue in the live deployment code's subject configuration (Sesnors/power rather than Sensors/power) prohibited the power feature from being properly populated. This issue had a direct impact on the Electrical, Aerodynamic, and Power Deficit detectors, all of which rely on power and current features. Second, differences between the trained models' feature lists and the live system's generated feature columns resulted in several features defaulting to zero during runtime. This mismatch resulted in constant or near-constant input vectors, and so flat probability outputs.

Further investigation of the real-time log revealed that the z-score scaling values calculated from offline data were not totally suitable for live settings. The mean and standard deviations in *resid_stats.json* resulted in exceptionally low z-scores in streaming data, further compressing model answers. Finally, the Random Forest regressor calculated residual vibration values necessitated various contextual features that were occasionally missing from the live stream, resulting in a practically static z_vib feature and persistent, useless vibration notifications.

Despite these problems, the Power Deficit detector performed consistently because it's essential features: wind speed, RPM, torque, and power, were generally present and accurately aligned. The probability trace showed periodic peaks that corresponded to actual changes in turbine load. The three sample-level alerts in the CSV represented the raw count of times the detector's probability exceeded its threshold, but the 37 separate alert episodes indicated fault periods clustered by hysteresis logic.

*Power Deficit Column

0	0	0	0	3	0
0	0	0	0	3	0
0	0	0	0	3	0
0	0	0	0	3	0
0	0	0	0	4	0
0	0	0	0	4	0
0	0	0	0	4	0
0	0	0	0	5	0
0	0	0	0	5	0
0	0	0	0	5	0
0	0	0	0	6	0
0	0	0	0	6	0
0	0	0	0	6	0
0	0	0	0	7	0
0	0	0	0	7	0
0	0	0	0	7	0
0	0	0	0	8	0
0	0	0	0	8	0
0	0	0	0	8	0
0	0	0	0	9	0
0	0	0	0	9	0
0	0	0	0	9	0
0	0	0	0	10	0

Figure 4.14: Example of live CSV log showing Power deficit anomaly being detected (note the hysteresis requiring 3 consecutive frames of anomaly to increment)

Table 4.2: Summary of real-time detector performance.

Detector	Distinct Episodes	Best Understanding as to why
Power Deficit	37	Valid alerts corresponding to real power drops
Electrical Fault	0	Flat probabilities due to feature omission
Aerodynamic Disturbance	0	No valid inputs, constant zero output
Vibration Health	0	Potential scaling mismatch
RPM Sensor Fault	0	Inactive during run, likely conservative threshold

Although only one model was able to be successfully executed, the live deployment still yielded some positive outcomes. First, the MQTT-based pipeline functioned continuously with no data loss, demonstrating that the real-time architecture is technically feasible. The inference cycle time per message stayed below 200 milliseconds, resulting in near-instant alert production. Second, the Power Deficit Detector's real-world performance verified the idea of edge-deployable predictive maintenance. The detector's continuous alignment with quantifiable load disturbances revealed the feasibility of utilizing basic, understandable models for autonomous fault detection.

However, the test demonstrated the vulnerability of machine learning systems to data integrity and schema consistency. Missing or mislabelled sensor channels passed through the feature pipeline, can have severe effects on the effectiveness and outcome of the models. These findings emphasize the need for better input validation, and feature verification in future implementations.

4.3 Summary of Findings from Results

The combined offline and real-time evaluations provide a complete picture of the VAWT-X anomaly detection framework's performance and operational preparedness. Offline testing demonstrated that the developed Logistic Regression models, which are supported by residual vibration data, can accurately classify and interpret diverse turbine fault states. The Power Deficit Detector continuously displayed the highest accuracy and interpretability, making it the most mature and deployable component. The electrical and aerodynamic detectors were also reliable, although the RPM Sensor and Vibration Health detectors require additional calibration to reach strong sensitivity.

The real-time deployment confirmed the functional architecture of the MQTT-based inference system, demonstrating its ability to perform end-to-end anomaly detection autonomously and with low latency. Although only the Power Deficit detector generated accurate alarms throughout the live trial, this result demonstrates that the framework can perform meaningful inference in streaming situations. The inactive detectors revealed implementation-level issues rather than conceptual flaws. Misnamed MQTT topics, irregular feature alignment, and mis-scaled z-score parameters all provide useful refining insights.

In summary, the research met its primary design goals by developing an operational, explainable, and lightweight anomaly detection framework for small-scale turbines. The offline evaluation confirmed the analytical accuracy of the technique, whilst the live deployment revealed practical practicality. These findings help the larger goal of creating scalable, interpretable machine learning algorithms for predictive maintenance in distributed renewable energy assets.

CHAPTER 5: DISCUSSION

5.1 Overview

This chapter discusses the results of both the offline evaluation and live deployment of the anomaly detection system for the VAWT-X turbine. It evaluates the findings reported in Chapter 4 and links them to the methodology's research goals and objectives. Talking points also delves into the actual issues that arose during system development, such as data constraints, feature alignment, and deployment complexity. Finally, it considers how these findings relate to the larger aims of predictive maintenance and digital twin integration, as well as paths for further improvement.

5.2 Interpretation of Offline Evaluation

The offline study revealed that machine-learning models can accurately categorize the operational stages of the VAWT-X turbine with a small dataset. All detectors achieved perfect accuracy (1.00), which means that no false positives were recorded inside the specified fault segments. Precision levels are predicted to vary in a real-world operational setting since each detector responds specifically to a specific sort of abnormality. The observed precision thus emphasizes the lack of false triggers under controlled settings, rather than ubiquitous fault-free performance.

The variation in recall and *F1* scores demonstrates the detectors' different sensitivity to fault magnitude and signal quality. The Aerodynamic Disturbance detector has the greatest *F1* score (0.923) and recall (0.857), indicating significant reactivity to airflow interruptions while remaining stable in baseline conditions. The Power Deficit and RPM Sensor detectors both demonstrated balanced performance (*F1* = 0.833, recall = 0.714), successfully identifying load differences and rotational anomalies in most test instances. The Electrical Fault Detector received an *F1 s*core of 0.727, missing certain low-intensity current anomalies while staying accurate for larger electrical imbalances. The Vibration Health detector was the least sensitive (*F1* = 0.600, recall = 0.429), indicating that under-scaled residual features and cautious threshold settings reduced response to mild mechanical imbalance.

This data shows that the framework's Logistic Regression models accurately classified turbine states, with few false alarms and stable probability outputs. However, the variety of recall values underlines the importance of threshold tweaking in balancing detection sensitivity with stability. Future versions should use adaptive thresholding or cross-validated quantile calibration to improve recall while maintaining interpretability.

5.3 Interpretation of Real-Time Deployment

The live deployment allowed us to see how the anomaly detection models performed under streaming data situations. The Power Deficit Detector remained completely operational and displayed a clear response to live changes in turbine load. A total of 37 unique warning occurrences were detected, each associated with substantial reductions in generated power. This constant behaviour shows the model's ability to generalize from offline data to real-world operating settings.

The remaining four detectors (Electrical, Aerodynamic, Vibration, and RPM Sensor) provided zero or near-zero probability values throughout the test. A subsequent examination of the real-time log and MQTT settings revealed various significant issues. A typographical issue in the MQTT topic list prevented legitimate power data from reaching the inference step, directly harming many models. Furthermore, feature alignment inconsistencies between the live pipeline and saved model configurations caused NaN values to be replaced with zeros, resulting in flat input vectors and static forecasts. The z-score scaling choices also failed to capture live data variance, reducing probability to an inadequate range.

Despite these restrictions, the real-time test showed that the system architecture was reliable. The MQTT pipeline ran continuously, processed data in real time with low latency, and recorded all inference findings for post-analysis. The successful operation of the Power Deficit Detector in this setting demonstrated that the underlying model architecture was technically sound. These findings show that the fundamental problem is not in the machine learning algorithm itself, but in guaranteeing reliable data handling, topic setting, and scaling consistency in deployment scenarios.

5.4 Reflections on Project Scope and Constraints

The expansion of the project's scope, as well as the practical limits encountered during its implementation, had a substantial impact on the study's ultimate results. The initial project concept centred significantly on the creation of a digital twin system for the VAWT-X turbine, combining live operational data with predictive models within an XMPro data stream and front-end dashboard. However, early in the project, it became clear that the turbine's readiness and data availability would impede progress toward that target within the placement timeline.

As a result, the project's focus switched to developing the fundamental predictive maintenance framework, which could then be integrated into a digital twin architecture, this shift necessitated a significant overhaul of the methodology and accompanying study documentation. Much of the original literature review, which focused on digital twin frameworks and high-level integration tactics, needed to be modified to match the new study's more data-driven, model-focused approach. The refocus allowed for more technical study of anomaly detection approaches, model deployment, and feature engineering, but it also meant that complete integration with the digital twin platform was postponed for later development.

Time restrictions also played an important influence. The process of installing and testing the physical turbine, calibrating sensors, and obtaining useable data took up a significant portion of the allotted project time. This limited the time available for iterative model refining, live data pipeline debugging, and threshold parameter optimization. As a result, only one model, that being the Power Deficit Detector, was fully validated in live situations, while the others remained partially implemented. With more time, these concerns may have been systematically addressed using input schema validation, dynamic threshold adaption, and extended cross-validation.

Nonetheless, the work made within the deadline indicated both technical capacity and proof of concept. The system's successful partial deployment demonstrates that the technique is viable and ready for future development once more data becomes available.

5.5 Lessons Learned and Technical Implications

The process of creating and installing this system provides some practical insights into using machine learning to small-scale renewable energy systems. The most important lesson involves data quality and consistency. Even extremely accurate models may fail if live input streams depart from the expected feature pattern. Future work should incorporate automated schema validation techniques that detect missing or renamed MQTT topics and warn problems before model inference.

Another important observation is threshold and scaling adaptation. The static thresholds utilized in this research were adequate for offline validation, however they were less adaptable in dynamic settings. Adaptive thresholding, which involves recalculating thresholds depending on recent data windows, has the potential to greatly improve sensitivity and reduce false negatives. Similarly, changing z-score scaling parameters while in live operation would keep the models calibrated to real-time sensor variability.

From a system architecture standpoint, the deployment demonstrated that lightweight, interpretable models may function within an edge computing framework. The entire pipeline operated on local hardware without external servers, proving the predictive maintenance for tiny turbines can be achieved without the processing overhead associated with deep learning methods. This is especially significant in distributed renewable systems where connectivity and computational resources may be restricted.

Finally, the findings emphasize the need of explainable AI in operational environments. The ability to understand each model's output using linear coefficients, feature importance, and probability thresholds boosts user confidence and makes debugging easier. The trade-off between interpretability and performance proved acceptable in this case, as the models provided actionable insights while remaining transparent.

5.6 Future Work

Future development of the VAWT-X anomaly detection system should focus on three main areas:

1. **Full Multi-Model Integration:** Completing and harmonising the deployment of all five classifiers within the live MQTT framework. This includes correcting topic schemas, aligning feature orders, and validating statistical parameters.
2. **Digital Twin Integration:** Building on the initial concept, the anomaly detection framework should be embedded within the XMPro digital twin environment. This would enable real-time visualisation of model outputs alongside operational turbine metrics and provide an interactive platform for maintenance decision support.
3. **Expanded Dataset and Adaptive Learning:** Collecting longer-term operational data will allow the models to be retrained under varied environmental conditions. Integrating online learning or incremental model updates could further enhance predictive performance.

Furthermore, future iterations should include automated recalibration procedures for z-score scaling and threshold adjustment, allowing the system to self-correct as environmental and turbine conditions change. Once these improvements are implemented, the platform could be used as a prototype for distributed predictive maintenance solutions across various small-scale renewable assets.

5.7 Chapter Summary

This chapter has gone over the analytical and operational findings of the VAWT-X anomaly detection architecture. Offline evaluations demonstrated that the models can reliably classify numerous fault types while remaining interpretable and computationally efficient. The real-time deployment established proof-of-concept success by establishing continuous live inference and verifying the system's ability to operate autonomously at the edge. Although only the Power Deficit model worked as expected during the live test, its performance verified the overall architecture and confirmed that the model design, data flow, and inference logic were all fundamentally correct.

The lecture also touched on the practical issues that influenced the project, such as time limits, data limitations, and a shift in emphasis from full digital twin integration to predictive maintenance modelling. These factors necessitated extensive restructuring but ultimately resulted in a technically focused and clearly functioning product. The findings on data integrity, adaptive calibration, and system explainability lay a solid framework for future study and development of intelligent maintenance systems for renewable energy infrastructure.

CHAPTER 6: CONCLUSIONS AND FUTURE WORK

The goal of this study was to develop, build, and test an interpretable anomaly detection framework for the VAWT-X vertical-axis wind turbine. The initiative aims to lay the groundwork for predictive maintenance and eventual digital twin integration in small-scale renewable energy systems. Despite changing project limitations, the final framework met its design objectives by proving efficient offline classification of turbine problems and dependable real-time operation via a MQTT-based deployment.

The offline evaluation revealed that lightweight Logistic Regression models, with residual vibration characteristics produced from a Random Forest regressor, can successfully detect numerous fault categories on a little dataset. The Power Deficit model performed the best overall, with excellent precision and recall, while the Electrical and Aerodynamic detectors were moderately accurate. Although the Vibration and RPM Sensor models were less sensitive, their performance justified the overall notion of employing explainable models for turbine health monitoring. These findings provided the analytical foundation required for moving to live deployment.

The real-time trial proved the system's technological feasibility in streaming settings. The Power Deficit detector responded accurately to genuine load changes, generating thirty-seven separate alert occurrences that corresponded to observed turbine behaviour. The other detectors remained inactive due to data alignment difficulties, emphasizing the vital significance of comprehensive schema verification. Nonetheless, the implementation showed steady inference at one-second intervals, low computing delay, and accurate alarm logging, demonstrating that autonomous, edge-based predictive maintenance is possible with minimum infrastructure.

An important effect of this initiative was an adaptive adjustment in research direction. Originally planned as a digital twin and predictive maintenance system, the project was re-scoped to focus on laying the machine-learning groundwork for such integration. This move, prompted by data restrictions and turbine readiness, necessitated a significant overhaul of the literature review, methodology, and results presentation. While the digital twin component is still in the works, the established framework acts as a useful building block for the greater goal. With more time and data, the other classifiers can be modified, thresholds dynamically adjusted, and model integration into the XMPro digital twin environment completed.

Overall, this effort shows that interpretable machine-learning algorithms can provide a dependable and scalable solution for early failure detection in distributed wind systems. The findings provide practical insights into real-time data handling, feature engineering, and model deployment in small-scale turbines. Future research should concentrate on harmonising the data flow, incorporating adaptive thresholding, and expanding the deployment across many turbines to enable continuous learning. In its current form, the framework proves the technical and conceptual feasibility of intelligent predictive maintenance while also laying the groundwork for extensive digital twin integration in the renewable energy sector.

REFERENCE LIST

Chen, W, Qiu, Y, Feng, Y, Li, Y & Kusiak, A 2020, 'Diagnosis of wind turbine faults with transfer learning algorithms', *Renewable Energy*, vol. 163, Pergamon, pp. 2053–2067.

Gigoni, L, Betti, A, Tucci, M & Crisostomi, E 2019, *A Scalable Predictive Maintenance Model for Detecting Wind Turbine Component Failures Based on SCADA Data*, arXiv.org.

Jamil, F, Peeters, C, Verstraeten, T & Helsen, J 2025, 'Leveraging signal processing and machine learning for automated fault detection in wind turbine drivetrains', *Wind energy science*, vol. 10, Copernicus Publications, no. 9, pp. 1963–1978, viewed 14 October 2025, <https://wes.copernicus.org/articles/10/1963/2025/>.

McKinnon, C, Carroll, J, McDonald, A, Koukoura, S & Plumley, C 2021, 'Investigation of Isolation Forest for Wind Turbine Pitch System Condition Monitoring Using SCADA Data', *Energies*, vol. 14, no. 20, p. 6601.

Esquivel-Sancho, L. M., Ghandchi Tehrani, M., Muñoz-Aria, M., & Askari, M. (2025), 'Fault Diagnosis of 3D-Printed Scaled Wind Turbine Blades', arXiv.org, viewed 14 October 2025, https://arxiv.org/abs/2505.06080.

Moghaddass, R & Rudin, C 2015, 'The latent state hazard model, with application to wind turbine reliability', *The Annals of Applied Statistics*, vol. 9, Institute of Mathematical Statistics, no. 4, viewed 14 October 2025, <https://arxiv.org/abs/1601.08034>.

Mohapatro, SR, Mulchandani, M, Mohanty, N, Sethi, MR, Sahoo, S & Samad, A 2025, 'Efficient Wind Turbine Fault Diagnosis Using Machine Learning Technique and Hyper-Parameter Tuning', *Lecture Notes in Networks and Systems*, Springer Nature Singapore, Singapore, pp. 163–174.

Moshtaghi, P, Hajialigol, N & Rafiei, B 2025, 'A comprehensive review of artificial intelligence applications in wind energy power generation', *Sustainable Futures*, vol. 9, Elsevier, p. 100638.

Nejad Alagha, Anis, Haitaamar, ZN, Obada Al-Khatib & Jeevan Kanesan 2025, 'Artificial Intelligence in Wind Turbine Fault Detection and Diagnosis: Advances and Perspectives', *Energies*, vol. 18, Multidisciplinary Digital Publishing Institute, no. 7, pp. 1680–1680.

Ng, EY-K & Lim, JT 2022, 'Machine Learning on Fault Diagnosis in Wind Turbines', *Fluids*, vol. 7, no. 12, p. 371.

Pedro, F & Alfredo Peinado Gonzalo 2021, ‘A Comprehensive Review of Artificial Intelligence and Wind Energy’, *Archives of Computational Methods in Engineering*, vol. 29, Springer Science+Business Media, no. 5, pp. 2935–2958.

Wang, Q, Cui, S, Li, E, Du, J, Li, N & Sun, J 2025, ‘Deep Learning-Based Fault Diagnosis via Multisensor-Aware Data for Incipient Inter-Turn Short Circuits (ITSC) in Wind Turbine Generators’, *Sensors*, vol. 25, Multidisciplinary Digital Publishing Institute, no. 8, pp. 2599–2599, viewed 14 October 2025, <https://www.mdpi.com/1424-8220/25/8/2599>.

Wang, Z, Li, X, Li, N & Shu, L 2024, *An AI-Driven Approach to Wind Turbine Bearing Fault Diagnosis from Acoustic Signals*, arXiv.org, viewed 14 October 2025, <https://arxiv.org/abs/2403.09030>.

Yimam, AW, Vafaeipour, M, Messagie, M, Fante, KA, Molla, EM, Azerefegn, TM & Coosemans, T 2025, ‘State-of-the-art of machine learning methods for fault detection and health monitoring of wind turbine system components: A comprehensive review’, *Engineering Applications of Artificial Intelligence*, vol. 162, Pergamon, p. 112645.

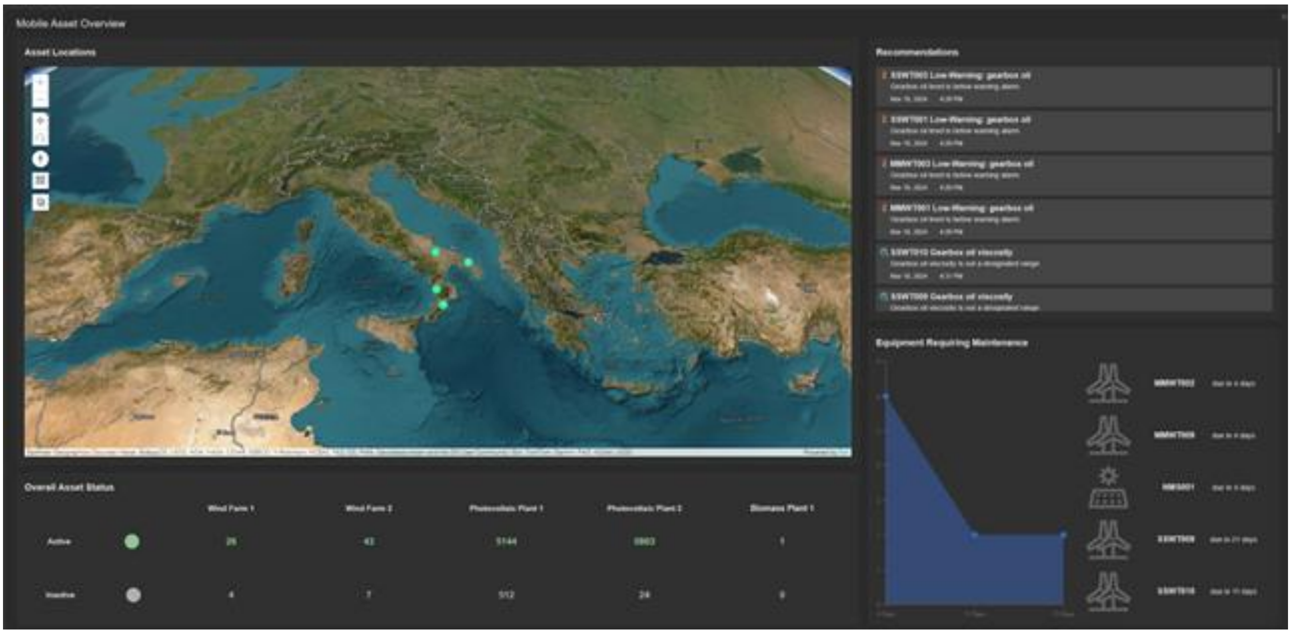
Zaid Allal, Noura, HN, Vernier, F, Salman, O & Chahine, K 2024, ‘Wind turbine fault detection and identification using a two-tier machine learning framework’, *Intelligent Systems with Applications*, Elsevier BV, pp. 200372–200372.

Zhang, C, Hu, D & Yang, T 2022, ‘Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost’, *Reliability Engineering & System Safety*, vol. 222, p. 108445.

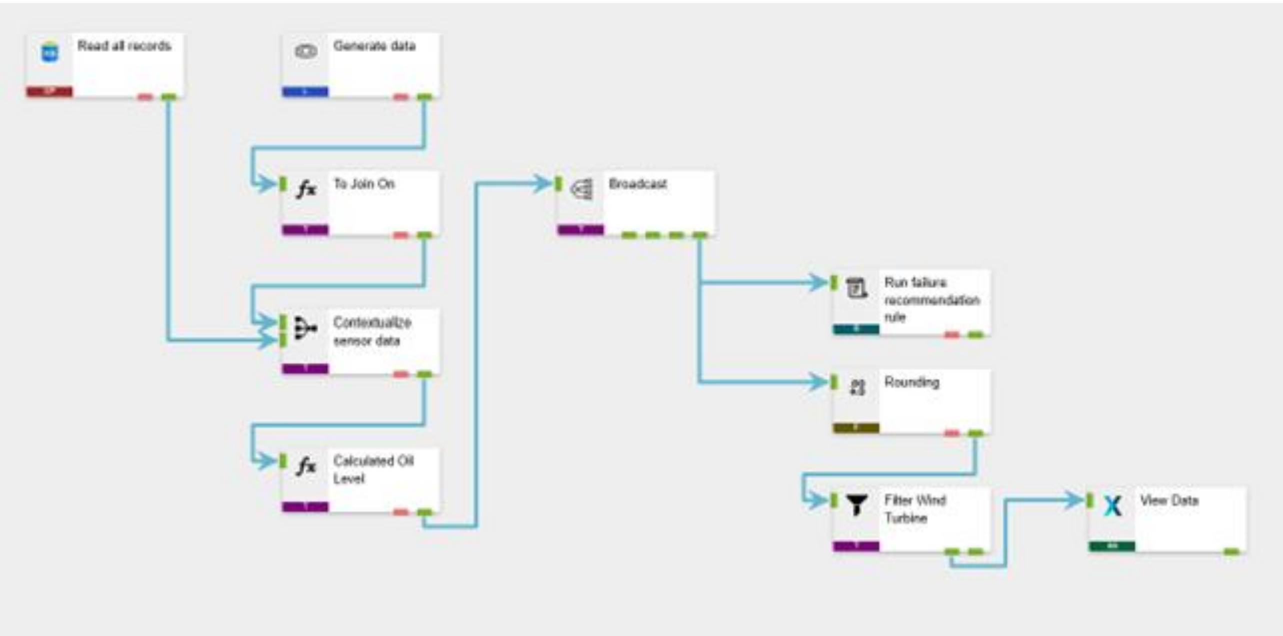
Zhang, C, Hu, D & Yang, T 2024, ‘Research of artificial intelligence operations for wind turbines considering anomaly detection, root cause analysis, and incremental training’, *Reliability Engineering & System Safety*, vol. 241, Elsevier BV, pp. 109634–109634.

APPENDICES

Appendix 1: XMPro Demonstration Application Interface (Prior Scope)

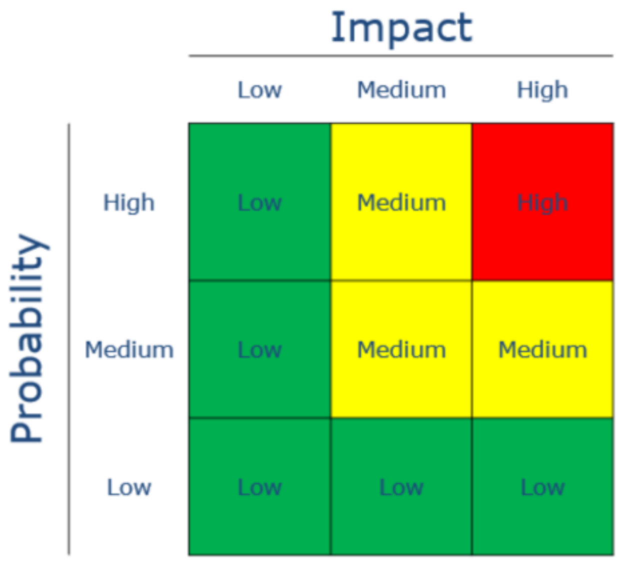


Appendix 2: XMPro Demonstration Data Stream (Prior Scope)

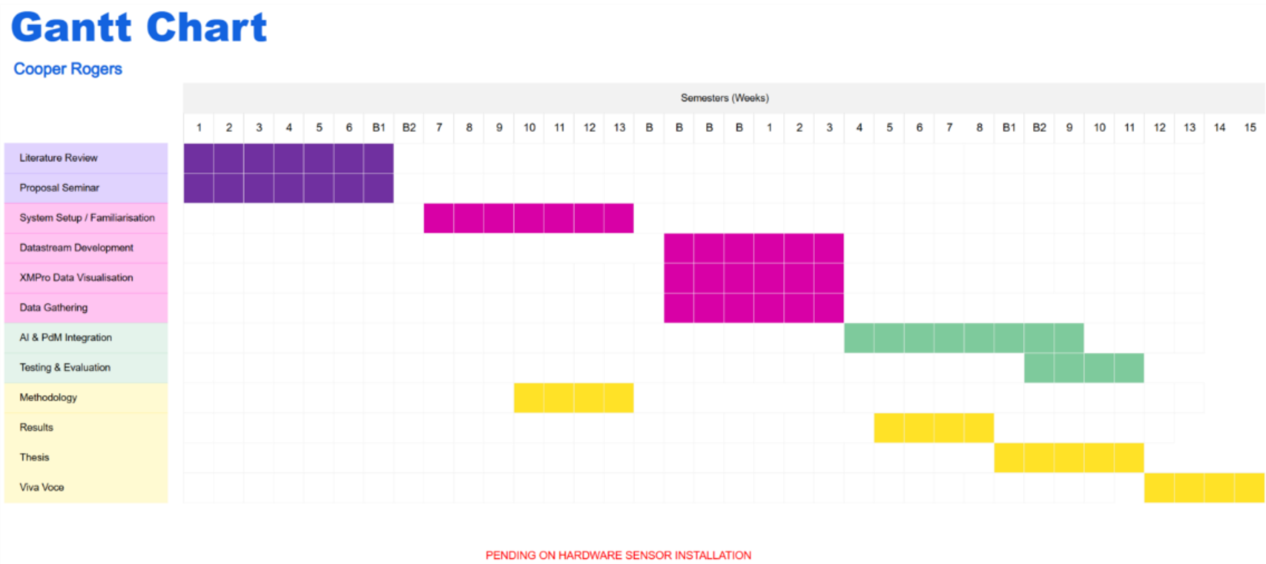


Appendix 3: Gantt Chart (Actual)

Appendix 4: Risk Assessment Matrix based on risk consequence and likelihood scoring.

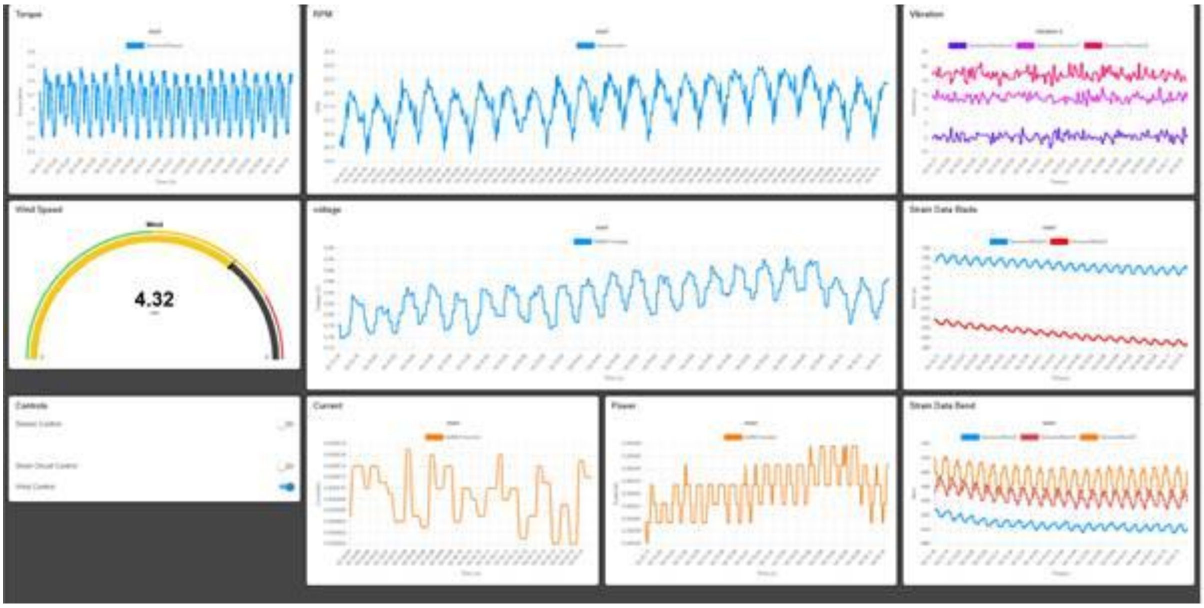


Appendix 5.1: Project Timeline Gantt Chart (Initial)



Appendix 6: All Testing Notes

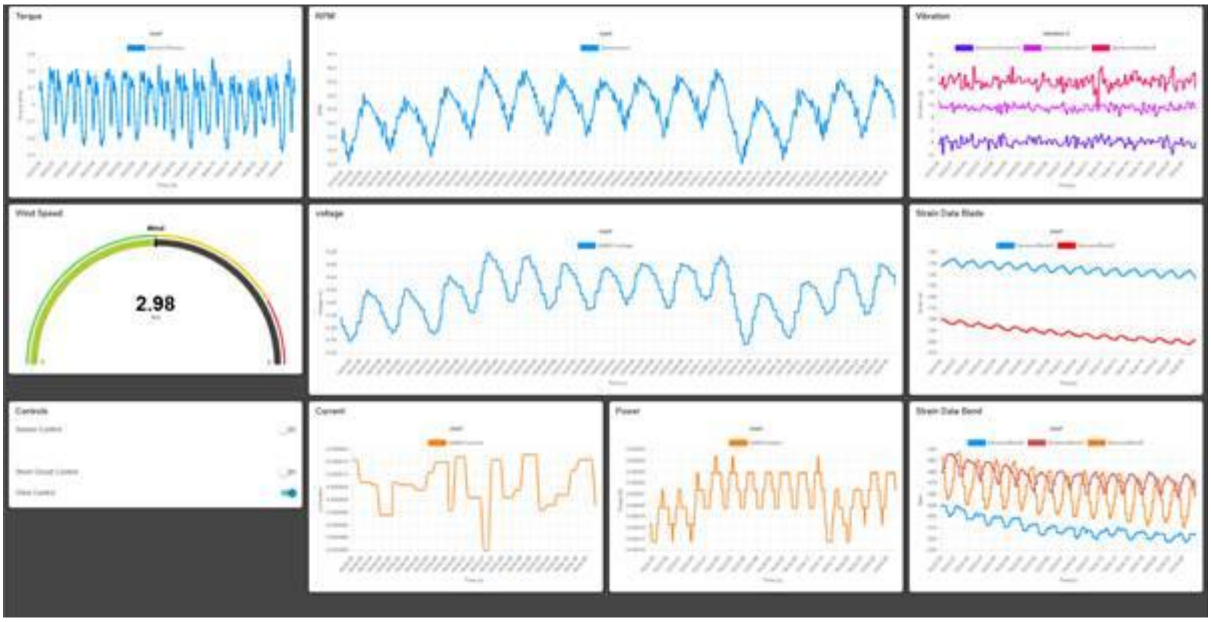
Appendix 6.1: Test 1-high wind base



Appendix 6.2: Test 2-mid wind base

Picture was not taken.

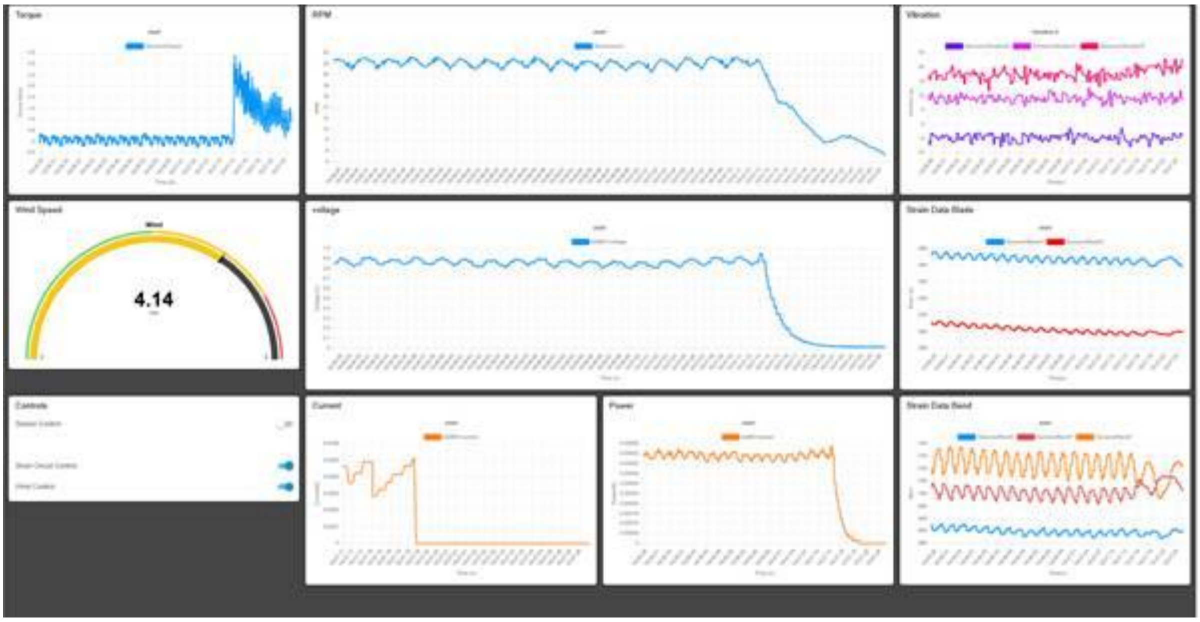
Appendix 6.3: Test 3-Low wind base



Appendix 6.4: Test 4-variable wind base

Picture was not taken.

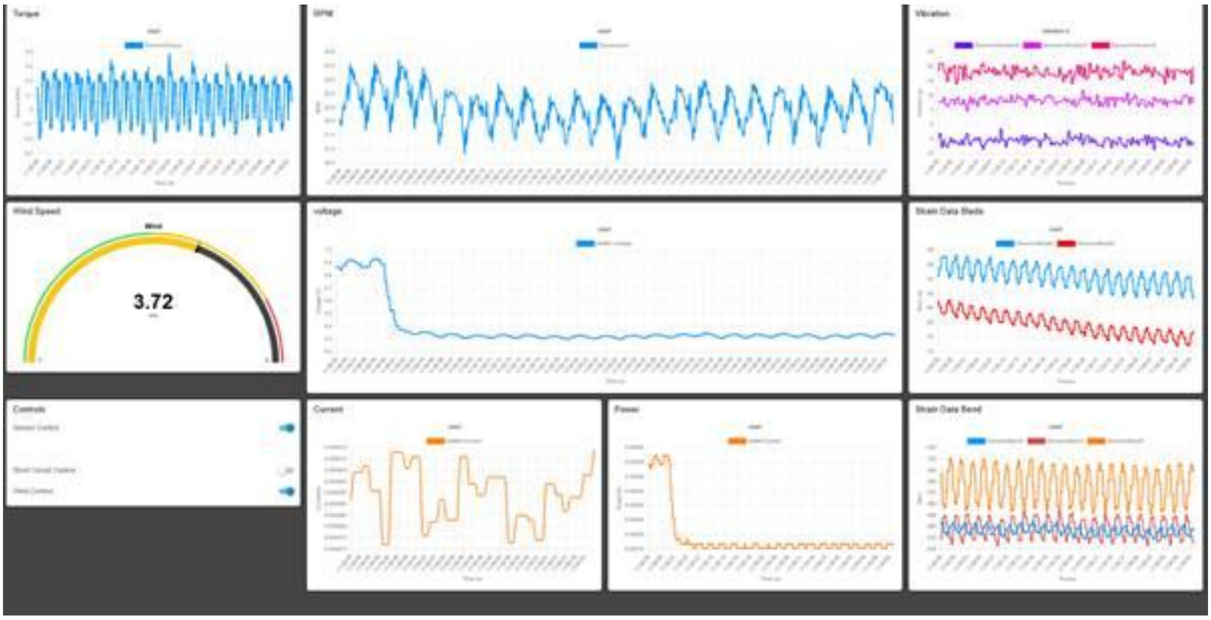
Appendix 6.5: Test 5-Balanced short circuit



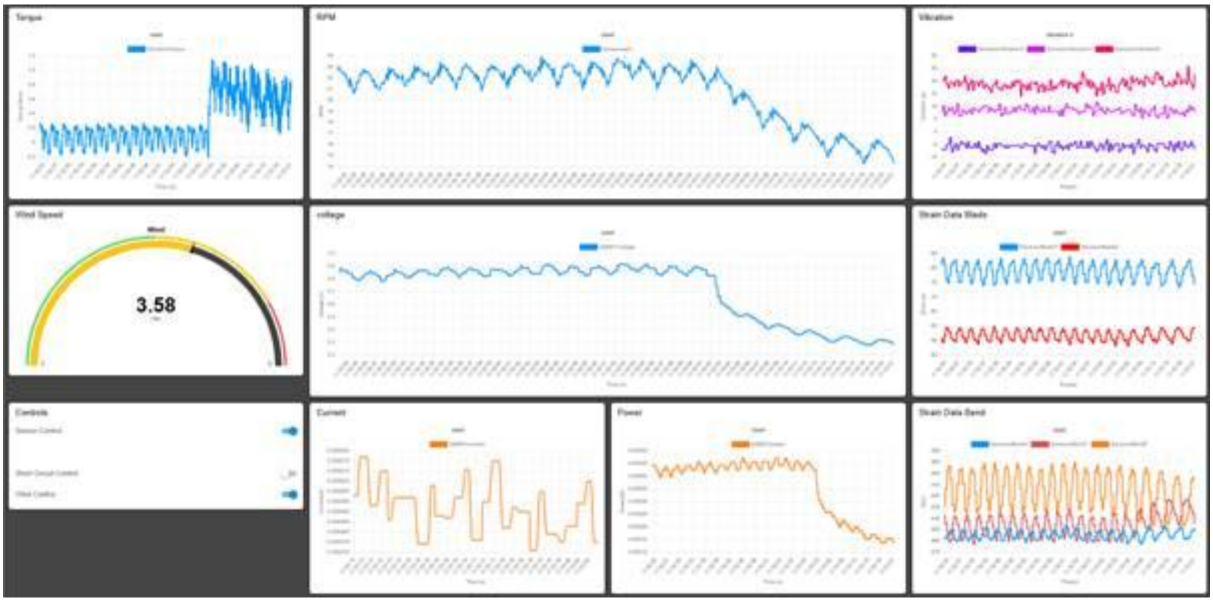
Appendix 6.6: Test 6-Balanced open circuit

Picture was not taken.

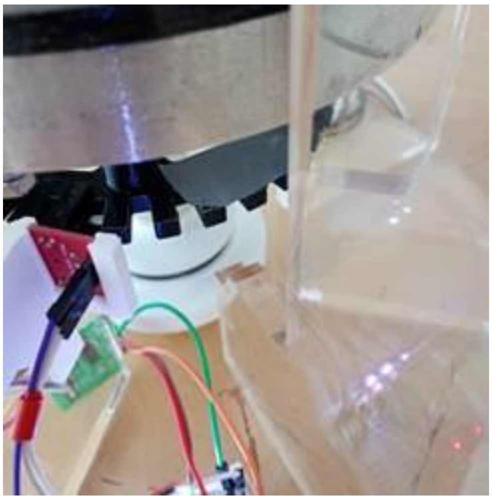
Appendix 6.7: Test 7-Unbalanced open circuit

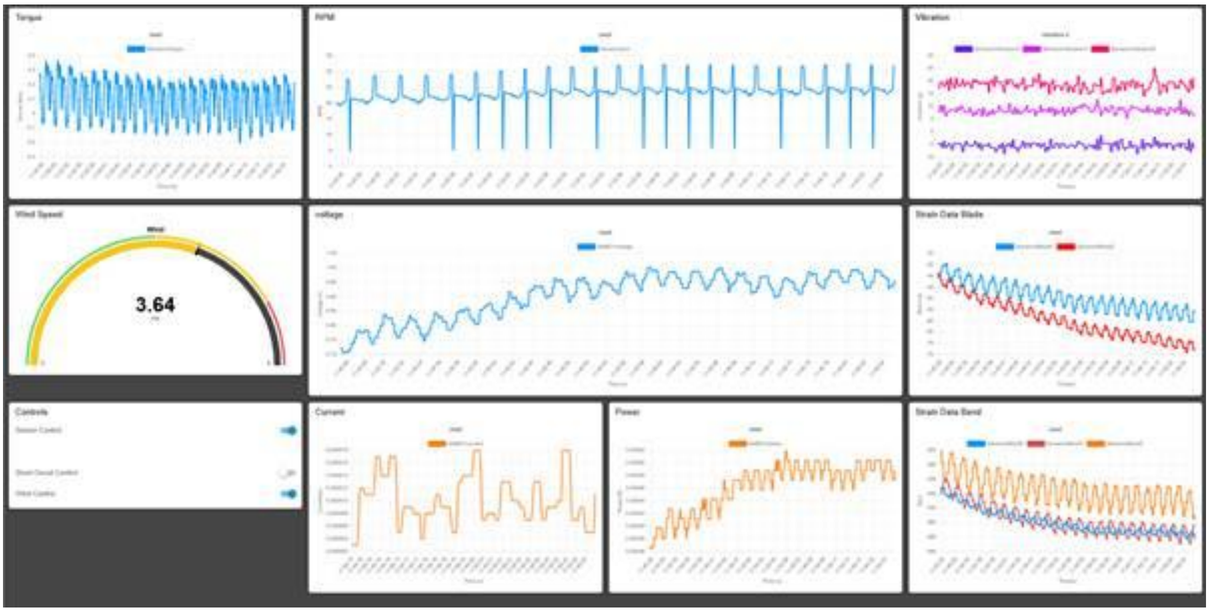


Appendix 6.8: Test 8-Unbalanced short circuit (line to line short circuit)



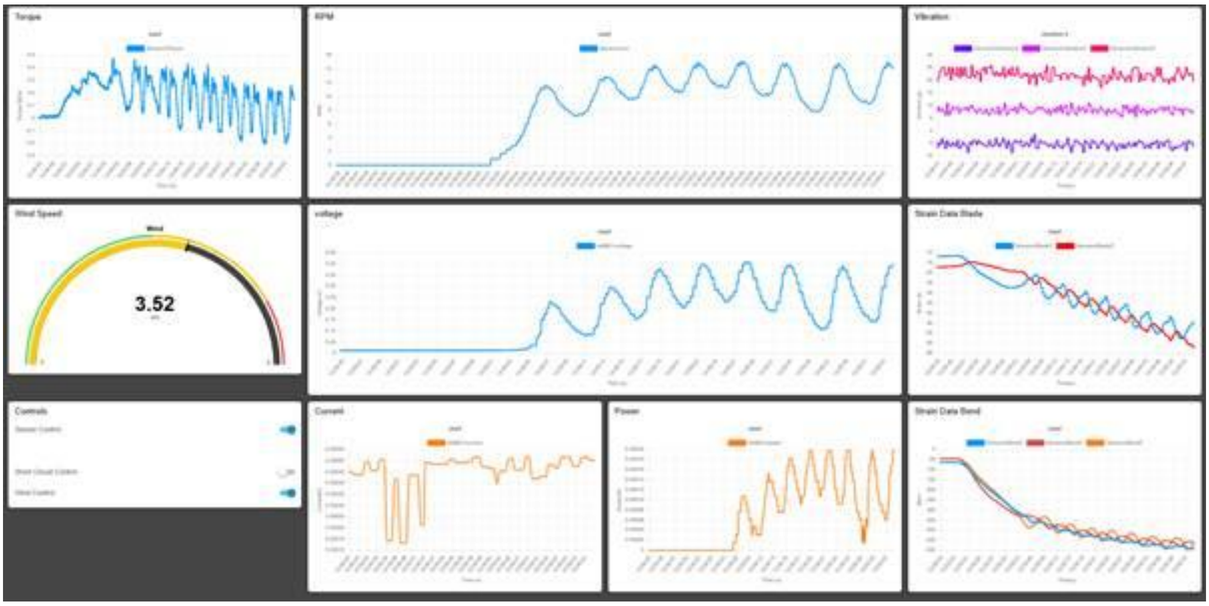
Appendix 6.9: Test 9-Rpm sensor malfunction





Appendix 6.10: Test 10-Disturbance in aerodynamic blade

Not -The bubble wrap fell in the middle of the experiment and the sensor readings can be to normal condition



Appendix 6.11: Test 11-Weight imbalance (adding weight)

Picture was not taken

About 200g of weight was attached to one side of the turbine blade.

Appendix 7: Project Source Code Structure

__pycache__	29/09/2025 3:17 pm	File folder	
10_preprocess.py	04/09/2025 9:48 pm	Python Source File	4 KB
15_train_residuals_all.py	29/09/2025 3:01 pm	Python Source File	3 KB
24_train_electrical_fault.py	29/09/2025 3:56 pm	Python Source File	5 KB
25_train_aero.py	29/09/2025 4:03 pm	Python Source File	6 KB
26_train_vibration_health.py	29/09/2025 4:03 pm	Python Source File	5 KB
27_train_power_deficit.py	29/09/2025 4:36 pm	Python Source File	5 KB
28_calc_detector_thresholds.py	29/09/2025 4:51 pm	Python Source File	6 KB
29_train_rpm_sensor_fault.py	29/09/2025 4:04 pm	Python Source File	4 KB
30_fit_threshold_models.py	01/10/2025 5:43 pm	Python Source File	13 KB
utils_common.py	04/09/2025 10:37 pm	Python Source File	5 KB