Interpretable Operational Regime Classification for a Wind Turbine PHM Digital Twin Architecture

Gard Indrekvam[1], Maneesh Singh[1], Anne-Lena Kampen[1], Mayank Shekhar Jha[2]

[1,2]*Western Norway University of Applied Sciences, Bergen, Inndalsveien 28, 5063, Norway*
gard.indrekvam@icloud.com
maneesh.singh@hvl.no
anne-lena.kampen@hvl.no

[2]*CRAN, CNRS, University of Lorraine, France*
mayank-shekhar.jha@univ.lorraine.fr

ABSTRACT

Wind turbines operate under varying environmental and control conditions that make reliable interpretation of SCADA data challenging. Without contextual information about the current operating state, normal variability may be mistaken for abnormal behaviour, reducing the reliability of diagnostic and prognostic analysis. This study develops a Digital Twin framework for wind turbines using SCADA data, where periodically updated models represent turbine structure, behaviour, and operating context. To address this, physics-informed operational zones are defined based on power curve characteristics and control logic, providing structured labels. These are used to train interpretable rule-based algorithms, Repeated Incremental Pruning to Produce Error Reduction (RIPPER) and First-Order Inductive Learner (FOIL), which generate explicit IF-THEN rules linking measured variables to operational states. Evaluation using overall accuracy, macro-F1 score, and per-class precision and recall shows that both methods achieve classification while producing compact, physically interpretable rule sets aligned with known turbine behaviour. The study demonstrates that rule-based learning enables transparent and effective operational regime classification, forming a critical contextual layer for prognostics and health management (PHM) oriented Digital Twins, with applicability beyond wind turbines to other industrial assets.

1. INTRODUCTION

1.1. Background

Wind turbines operate under highly variable and uncertain environmental conditions, where wind speed, air density, and turbulence continuously fluctuate. These variations lead to significant changes in turbine behaviour and power production, making it challenging to distinguish between normal operational variability, expected aging performance degradation and early indicators of faults.

Modern wind turbines are equipped with Supervisory Control and Data Acquisition (SCADA) systems that continuously record key operational (position, temperature, RPM, hydraulic and electrical) and meteorological (wind speed, humidity, temperature, direction) parameters. These measurements are used for performance monitoring, operational analysis, condition assessment and maintenance decision-making in wind farms (Dai et al., 2016; Hong et al., 2025).

1.2. Condition Monitoring of Wind Turbines Using SCADA

While SCADA systems are primarily designed for monitoring and control, they are closely linked to condition monitoring systems, which focus on diagnostic and prognostic analysis. Consequently, SCADA data collected from wind turbines, including temperature, RPM, wind speed, and power generation, can also be effectively utilised for fault detection and health assessment.

The time interval between fault initiation and eventual failure varies significantly across different failure modes, ranging from seconds (e.g., generator earth faults) to several weeks or longer (e.g., progressive gear wear). Given that SCADA data are typically recorded at relatively low temporal resolution, they are particularly suited for detecting faults that evolve over longer time scales. As a result, many modern SCADA-based monitoring systems incorporate real-time analytics using statistical and artificial intelligence methods to identify abnormal behaviour (Manwell et al., 2009; Pandit & Wang, 2024; Tavner, 2012; Yang et al., 2018).

One of the widely used approaches for analysing wind turbine performance using SCADA data is to examine power generation as a function of key variables, particularly wind speed. The power curve of a wind turbine represents the relationship between generated power and the environmental and operational conditions under which the turbine operates. Power production depends on a combination of technical factors (e.g., rotor radius), environmental conditions (e.g., wind speed and air density), and operational variables (e.g., blade pitch angle and yaw misalignment between wind direction and nacelle orientation) (Manwell et al., 2009).

A discrepancy between rotor power (P_{Rotor}, predicted using models), and electrical power ($P_{Electrical}$, measured at the generator output), indicates additional energy losses within the system. These losses are typically associated with increased mechanical friction, vibration, or thermal dissipation in drivetrain components, and may reflect degradation in mechanical or electrical subsystems. Consequently, power-based analysis provides a valuable basis for condition assessment and fault detection (Duguid, 2018). While such analysis may not always identify the exact fault mechanism, significant deviations can serve as early indicators of abnormal behaviour, prompting further diagnostic investigation. In addition, power curve analysis can be used to detect suboptimal performance, for example due to control inefficiencies or misalignment. Comparative analysis across multiple turbines within a wind farm can further help identify units that are underperforming relative to their peers.

1.3. Methodologies for Predicting Produced Power

A substantial body of literature has focused on modelling wind turbine behaviour using physical, statistical, and machine learning approaches (Lydia et al. 2014; Pandit, Infield & Kolios, 2019; Saint-Drenan et al., 2020; Pandit & Wang, 2024).

1.3.1. Parametric Models

Parametric models describe wind turbine power generation using predefined mathematical functions that establish relationships between key variables. These models are typically derived from the fundamental aerodynamic power equation, which is based on Betz's law and expresses power as a function of air density, rotor swept area, wind speed, and the power coefficient (Saint-Drenan et al., 2020, Manwell et al., 2009). Common formulations include linearized segmented models, polynomial regressions, and sigmoid-type functions such as 4- or 5-parameter logistic models. These models aim to approximate the characteristic S-shaped power curve of a wind turbine, capturing key regions such as cut-in, partial load, rated power, and cut-out behaviour.

Parametric models offer several advantages. They are computationally efficient, relatively easy to interpret, and require only a limited number of parameters to describe turbine performance. This makes them suitable for applications such as performance benchmarking, anomaly detection, and large-scale wind farm analysis. However, their accuracy depends on the appropriateness of the chosen functional form, and they may struggle to capture complex nonlinearities arising from changing environmental conditions, control strategies, or turbine-specific characteristics.

A fundamental limitation of parametric models is the implicit assumption of a consistent relationship between environmental variables, such as wind speed and air density, and power output. In practice, wind turbines operate as closed-loop systems, where the overall operating strategy is determined by continuously balancing multiple, and often competing, objectives, including ensuring safe operation, maximising power generation, minimising structural loads and vibrations, protecting components from damage, and maintaining grid compliance. These objectives are achieved through coordinated control actions that dynamically adjust system behaviour in response to changing environmental and operational conditions. Key control mechanisms include blade pitch regulation, yaw orientation control, brake torque management, and generator torque control. Through these interacting control loops, the turbine transitions between different operating states, each characterised by distinct relationships between environmental inputs and power output (Manwell et al., 2009).

1.3.2. Data Driven Models

Data-driven models typically offer strong predictive capabilities and can capture complex nonlinear relationships that are difficult to represent using parametric formulations. These models include machine-learning techniques such as Support Vector Machines, Gaussian Processes, Random Forests, and Artificial Neural Networks (Ouyang et al., 2017; Pandit, Infield & Kolios, 2019). Once trained, these models can be used to predict turbine performance under varying environmental and operational conditions.

Data-driven models suffer due to their reliance on black-box approach that lacks interpretability. This presents a significant limitation in engineering applications, where model outputs must be transparent, explainable, and verifiable by domain experts. The need for interpretable models has therefore become increasingly critical, particularly in applications related to condition monitoring, maintenance decision-making, and digital twin systems (Nössig et al., 2024).

1.3.3. Hybrid Models

Hybrid approaches attempt to resolve the limitations of parametric and data-driven models by embedding physical knowledge within data-driven frameworks or by augmenting physics-based models with data-driven corrections. Thus, these approaches aim to leverage the strengths of both paradigms: the interpretability and theoretical grounding of physics-based models, and the flexibility of machine learning methods to capture complex, nonlinear behaviour from data.

Hybrid models are particularly relevant in wind energy applications, where turbine behaviour is governed by well-established aerodynamic and mechanical principles, but is also influenced by site-specific conditions, control strategies, and component degradation. For example, hybrid models, combine analytical representations of the power curve with non-parametric or data-driven components that adapt to observed data. This allows local deviations from the idealised power curve to be captured without losing the overall physical structure of the model. It also enables use of residual learning techniques where data-driven models capture the discrepancy between measured and physics-predicted outputs. Such approaches ensure that predictions remain physically plausible while improving representation of turbine performance under varying environmental and operational conditions (Saint-Drenan et al., 2020).

Despite their advantages, hybrid models still face challenges related to data quality, model calibration, and the influence of operating regimes. Since turbine behaviour varies significantly across different control states, models that do not explicitly account for the active operating regime may produce inconsistent results. While hybrid approaches can incorporate mechanisms to infer the current regime, explicit operational regime classification remains a useful preprocessing step that improves consistency across all modelling approaches, including hybrid frameworks.

1.3.4. Rule-based Models

Rule-based learning methods offer an alternative to black-box modelling approaches. Unlike opaque models, where the internal reasoning is difficult to assess, rule-based models expose the relationships between input variables and system behaviour in a form that is accessible to domain experts. By generating explicit IF-THEN rules, these methods provide transparent, traceable, and auditable decision logic that can be directly interpreted and validated against established engineering knowledge.

A key advantage of rule-based approaches is their ability to represent decision rules in a structured and modular manner. Each rule corresponds to a specific region of the operating space, allowing complex system behaviour to be decomposed into a set of interpretable conditions. This is particularly relevant for wind turbines, where different operating regimes are governed by distinct physical mechanisms and control strategies. Rule-based models can naturally align with these regime-dependent behaviours, making them well suited for context-aware analysis.

Classical algorithms such as Repeated Incremental Pruning to Produce Error Reduction (RIPPER) and First-Order Inductive Learner (FOIL) were specifically developed to learn human-readable rule sets from data (Quinlan, 1990; Cohen, 1995). When applied to discretized SCADA variables, they can produce concise rules that map measurable parameters, such as wind speed, rotor speed, and blade pitch angle, to specific operational states.

RIPPER is a propositional rule induction algorithm designed to generate compact IF–THEN classification rules. The algorithm follows a sequential covering strategy, where rules are learned iteratively for one class at a time. After rule construction, a pruning phase based on minimum description length principles is applied to reduce overfitting and limit unnecessary rule complexity. This typically results in relatively small and interpretable rule sets (Cohen, 1995). RIPPER is often selected due to its balance between predictive performance and interpretability, which is important in operational monitoring contexts where transparent decision logic is required.

FOIL is a rule-learning algorithm that constructs logical definitions by iteratively adding literals that improve class separation. In contrast to tree-based methods, FOIL directly produces sets of logical rules that describe a target class as conjunctions of feature conditions (Quinlan, 1990). When applied to discretized SCADA variables, FOIL generates multiple candidate rules per operational regime. These rules can overlap in feature space, which makes an explicit arbitration strategy necessary during prediction.

In addition to interpretability, rule-based models offer practical advantages in engineering applications. The learned rules can be easily inspected, modified, and integrated into existing monitoring frameworks or digital twin architectures. They also facilitate validation against physical principles, enabling inconsistencies or implausible behaviours to be identified and corrected.

Despite these strengths, rule-based learning has received relatively limited attention in recent wind turbine research compared to more complex machine learning approaches. This study addresses this gap by systematically evaluating rule-induction methods for operational regime classification, demonstrating their potential to provide both high predictive performance and the level of transparency required for reliable PHM applications.

1.4. Digital Twin in an Industrial PHM Context

The concept of a *digital twin* has received considerable attention across many industrial sectors. As a result, a wide range of software platforms and definitions has emerged, each describing a digital twin in slightly different ways. While this reflects strong interest, it also creates confusion - especially for PHM researchers and practitioners who need a clear and practical understanding of what a digital twin should provide for specific operational and maintenance tasks.

For this work, the definition from the National Academies of Sciences, Engineering, and Medicine (2024), has been accepted:

"A set of virtual information constructs that mimics the structure, context, and behaviour of a natural, engineered, or social system; is dynamically updated with data from its physical counterpart; possesses predictive capabilities; and informs decisions to realize value."

This definition emphasises that a digital twin:

1. mirrors the structure, context, and behaviour of a real system,
2. is continuously updated using data from the physical asset,
3. has predictive capabilities, and

4. supports decision-making that creates practical value.

1.5. Model Requirements for Interpretable and Adaptive Digital Twins

The long-term objective of this research is the development of a physics-consistent digital twin for wind turbines that can evolve over time using operational data. The models embedded within this digital twin should have the following features:

1.5.1. Contextual Dependency of Models

For effective PHM, SCADA data cannot be analysed in isolation. Measurements must be interpreted in the context of the turbine's current operating conditions. A digital twin designed for PHM must therefore explicitly account for this contextual dependency in order to reliably distinguish normal operational variability from early signs of faults. Operational regime classification is thus a key enabling step, as it provides the contextual information required by diagnostic and prognostic algorithms.

Wind turbines operate in distinct control regimes that are defined by aerodynamic behaviour and control actions (e.g., pitch regulation and generator torque control), such as below-rated operation, transition regions, and above-rated power regulation. Without explicit identification of the active regime, condition-monitoring algorithms may misinterpret normal control-driven variations as abnormal behaviour, leading to false alarms or missed fault detections (Xie et al., 2023).

From a digital twin perspective, operational regime classification plays several important roles. First, it enables context-aware interpretation of SCADA data, ensuring that diagnostic models are applied under appropriate operating conditions. Second, it supports model adaptability by allowing different physics-based or data-driven models to be selected or parameterised according to the current regime. Third, it improves transparency, since regime definitions can be directly linked to known turbine control logic and aerodynamic principles, making model behaviour easier to interpret.

A key challenge is that SCADA data are generated by a closed-loop controlled system, not by an unconstrained physical process. Measured signals reflect not only aerodynamic behaviour, but also grid limitations, safety logic, and control actions. As a result, simple handcrafted IF-THEN-ELSE rules with fixed thresholds are often inadequate. Such rules impose sharp decision boundaries and fail to represent the smooth, nonlinear structure of the turbine power curve. Unlike manually defined IF-THEN-ELSE logic, algorithms such as RIPPER and FOIL can infer flexible, nonlinear rule structures that better follow the complex shape of the power curve. Importantly, they can identify data points that lie outside physically meaningful operating regions, cases that fixed-threshold rules would typically fail to detect, while still providing transparent and auditable decision logic.

1.5.2. Explainability / Interpretability of Models

The performance of data-driven models is also highly dependent on the quality of the input data. Preprocessing techniques applied to SCADA data, including filtering and outlier removal, can improve predictive accuracy (Hong et al., 2025). However, the identification and removal of outliers is not trivial. Depending on the criteria used, such preprocessing steps may inadvertently discard data that correspond to meaningful abnormal behaviour, including transition states and early fault signatures. Hence, in this study, unsupervised outlier detection methods such as density-based clustering (e.g. DBSCAN) are deliberately not adopted. Although these methods can identify anomalous data points based on distance measures, they provide limited insight into why a particular data point is classified as an outlier.

1.5.3. Dynamic Update of Models

In a PHM-oriented digital twin, the digital representation is not a single model, but a set of interacting models with different roles. These include models for identifying the current operating state, monitoring condition, diagnosing faults, and estimating future degradation. All of these models must be updated using real-time data from the physical turbine and must remain consistent with known physical behaviour and turbine control logic.

At the time a wind turbine is commissioned, only parametric and physics-based models are typically available. These models are built to satisfy fundamental aerodynamic and thermodynamic laws, but they rarely match real operational behaviour exactly. Differences arise due to manufacturing tolerances, site-specific wind conditions, control settings, and gradual aging of components. As operational data from SCADA systems become available, the digital twin must therefore evolve from a purely parametric description towards a data-informed representation that reflects the actual performance of the turbine.

One practical approach is to update the parameters of the physical power model using operational data, in particular the coefficients that define the power coefficient $C_P(\lambda, \beta)$. Here, λ is the tip speed ratio, defined as the ratio of the blade tip speed to wind speed, and β is the blade pitch angle. By estimating these parameters (e.g., C_1 to C_{10}) from SCADA measurements and tracking how they change over time, it becomes possible to identify performance shifts linked to degradation, aging, or changes in control behaviour (Saint-Drenan et al., 2020).

In practice, the wind turbine power curve is more accurately represented by a set of models, each corresponding to a distinct operating region, such as start-up, partial-load, rated-power, and shutdown regimes. Consequently, it is essential to identify the current operating segment from incoming

SCADA data so that only the relevant model parameters are updated. This regime-specific updating strategy improves model accuracy and enables more meaningful interpretation of parameter changes within the digital twin framework.

2. MOTIVATION AND AIM

2.1. Motivation

To improve the operation and maintenance of wind turbines, a digital twin provides a practical and effective approach. In this framework, sensor data from the physical turbine, collected from multiple sources, are continuously transmitted to a digital environment, allowing the current operating state, structural condition, and environmental influences to be tracked in near-real time.

Physics-consistent and interpretable models within the digital twin form the core of the system. These models are regularly updated to reflect changes in operating conditions and component aging, and they are used to support both operational optimisation and maintenance-related analysis.

For operational purposes, the digital twin uses these models to generate recommendations that can either be applied directly through control actions or presented to operators as clear, actionable guidance via user interfaces. This closed-loop setup ensures that operational decisions are continuously informed by the current health and operating condition of the turbine.

For maintenance planning, whether scheduled or condition-based, the models accumulate degradation history and health indicators over time. This supports diagnostic and prognostic analysis, enabling the system to distinguish between normal operational variability and early signs of faults. As a result, maintenance actions can be planned in a timely and risk-informed manner, consistent with PHM principles.

2.2. Aim of the Project

The aim of this study is to develop an interpretable and physically consistent method for identifying wind turbine operating regimes, enabling reliable SCADA data interpretation and preprocessing for PHM applications.

To achieve this, the study employs rule-based machine learning methods, specifically RIPPER and FOIL, trained using physics-informed operational labels. These approaches enable the extraction of transparent IF-THEN rules that remain consistent with known aerodynamic behaviour and turbine control logic, while still adapting to patterns observed in operational data.

Once operating regimes are correctly identified, SCADA data can be utilised more reliably to support:

- robust diagnostics and prognostics with a reduced likelihood of false alarms,
- separation of long-term performance degradation from short-term operational or environmental disturbances, and
- continuous model updating to ensure that digital representations remain aligned with the actual behaviour of the turbine.

2.3. Contributions of this Research

While previous studies have explored SCADA-based modelling, anomaly detection, and explainable machine learning, relatively limited attention has been given to the automatic generation of interpretable operational regimes for wind turbine analysis. In particular, there is a lack of approaches that identify physics-aligned operational zones using rule-based machine learning, while maintaining both interpretability and consistency with turbine control behaviour. This study addresses this gap by combining physics-informed operational regime definitions with rule-induction algorithms, specifically FOIL and RIPPER.

A key distinguishing aspect of this work is that no data points are discarded as outliers solely based on deviation from an expected power curve. Instead, each observation is assigned to an interpretable operational regime, ensuring that all data are contextualised, thereby, contributing to the understanding of turbine behaviour. This approach reduces the risk of removing fault-relevant information and improves traceability in the analysis.

3. METHODS

3.1. Data Used for the Research

To demonstrate the feasibility of the proposed approach, this study utilises SCADA data provided by EDP from two of the four horizontal-axis offshore wind turbines (Turbines 7 and 11) located off the west coast of Africa. The dataset spans a two-year period (2016–2017) and consists of 10-minute averaged measurements. The dataset was subject to standard quality filtering, including removal of records with missing values and exclusion of measurements outside physically plausible operating range prior to the analysis.

In total, 76 SCADA variables are available, encompassing mechanical, electrical, and environmental variables. For the purposes of this study, a subset of seven key variables is selected, focusing on measurements most relevant to aerodynamic performance, turbine control behaviour, and operational regime identification.

3.2. Operational Regime Classification of SCADA Data in Closed-Loop Operation

Since the SCADA data are influenced by closed-loop operation, explicit operational regime classification is required. Rule-based regime classification allows SCADA data to be partitioned according to these control regimes, ensuring that each data point is interpreted within the correct operational context. Based on the analysed SCADA data, the following operational zones are defined:

3.2.1. Lockdown and Transition Regimes

Zone 1 - Safety Lockdown (Blades_Pitch_Angle >= 86°)
In the lockdown regime, the turbine is intentionally prevented from producing power, regardless of wind availability. This state is typically triggered by external constraints such as grid saturation or transformer limitations rather than turbine aerodynamics. Blade pitch angles are at or above approximately 86°, rotor speed is zero, and power production ceases even under favourable wind conditions. SCADA data from this regime appear as a horizontal line defining Grid_Prod_Power = 0, Rtr_RPM = 0, Blades_Pitch_Angle >= 86° for all values of Amb_Wind_Speed. This data does not reflect turbine response to wind input.

Zone 2 - Transition to or from Safety Lockdown (Blades_Pitch_Angle < 86°)
Blade pitch is actively adjusted during these transitions while the turbine moves between operational zones and lockdown states. Transitions into or out of lockdown occur over time intervals longer than the SCADA sampling period (10 minutes), resulting in recorded data during intermediate states. These data points are characterised by scattered data points between the operating curve and horizontal line defining Grid_Prod_Power = 0, Rtr_RPM = 0, Blades_Pitch_Angle >= 86° for all values of Amb_Wind_Speed. Although measurements are valid, they correspond to transient control actions rather than steady operational behaviour and need to be treated as such.

3.2.2. Standby and Transition Below Cut-In Wind Speed Regimes

When ambient wind speed falls below approximately 4 m/s, power production is not economically viable.

Zone 3 – Standby (Low-Wind Shutdown) (Amb_Wind_Speed < 4 m/s; Blades_Pitch_Angle = 24°; Rtr_RPM = 0 rpm)
When the Ambient Wind Speed is below the *Cut-In Wind Speed* (4 m/s), the turbine remains fully braked. Blade pitch angle is fixed at approximately 24°, rotor speed is zero, and no electrical power is produced. This regime is characterised by a horizontal straight line in the Grid_Prod_Power = 0, Rtr_RPM = 0, for all values of Amb_Wind_Speed brought about by fixed Blades_Pitch_Angle = 24°.

Zone 4 – Transition to or from Standby (Amb_Wind_Speed < 4 m/s; Blades_Pitch_Angle = 24°; 0 rpm < Rtr_RPM)
During startup or shutdown under low wind conditions, transient states are observed in which the brakes are released and the blade pitch angle remains near 24°. These data correspond to control-driven transitions rather than steady operation. In this zone, while there is a fixed Blades_Pitch_Angle = 24°, Amb_Wind_Speed = *Cut-In Wind Speed* (4 m/s), Grid_Prod_Power = 0. Due to the release of brakes, the rotor speed (Rtr_RPM) becomes non-zero.

3.2.3. Transitions to Power Production Regimes

Zone 5 – Free Wheeling (Below Cut-in) (Amb_Wind_Speed < 4 m/s; 0° < Blades_Pitch_Angle < 24°; 0 rpm < Rtr_RPM < 11 rpm)
Below the *Cut-In Wind Speed* (4 m/s), at low rotor angular speed (0 rpm < Rtr_RPM < 11 rpm), the rotor rotates freely without generator engagement. During the increasing wind speeds, blade pitch angle decreases toward 0°, rotor speed increases slowly; and during the decreasing wind speeds the pitch angle increases toward 24° with the rotor decreasing slowly. In either case, there is no electrical power generation because the generator is not engaged. Data in this regime appear scattered between 0° < Blades_Pitch_Angle < 24°, 0 rpm < Rtr_RPM < 11 rpm for all values of Amb_Wind_Speed below *Cut-In Wind Speed* (4 m/s).

Zone 6 – Generator Engagement (Below Cut-in) (Amb_Wind_Speed < 4 m/s; Blades_Pitch_Angle = 0°; 11.0 < Rtr_RPM = 11.2)
Below the *Cut-In Wind Speed* (4 m/s), at rotor speeds of approximately 11–12 RPM, the generator becomes engaged. Rotor speed remains nearly constant while torque increases with the increase in wind speed, resulting in low but increasing power production. In this zone, with the increase in wind speed (Amb_Wind_Speed), while the blades pitch angle (Blades_Pitch_Angle) is maintained at about 0° and the rotor speed (Rtr_RPM) remains constant at 11.0 - 11.2 rpm, the power generation (Grid_Prod_Power) slowly increases

Zone 7 – Partial Load Operation (Above Cut-in but Below Rated) (4 <= Amb_Wind_Speed <= 12; Blades_Pitch_Angle = 0°; 11.2 < Rtr_RPM < 14.8)
Between *Cut-In Wind Speed* (4 m/s) and *Rated Wind Speed* (12 m/s), the blade pitch angle is maintained at 0° to optimise aerodynamics. This results in an increase in rotor angular speed and corresponding electric power generation. With increase in wind speed, power generation follows the cubic power curve. Here the turbine operates in a torque-controlled region aimed at maximising aerodynamic efficiency. SCADA data shows that with increase in wind speed (Amb_Wind_Speed), while blade pitch angle (Blades_Pitch_Angle) remains near 0°, rotor speed (Rtr_RPM) increases from approximately 11.2 rpm to about 14.8 rpm, and power output follows the characteristic cubic power curve with respect to wind speed.

3.2.4. Full Power Production Regime

Zone 8 – Full Load Operation (Above Rated) (12 < Amb_Wind_Speed < 25; 0° < Blades_Pitch_Angle; 11.8 <= Rtr_RPM < 15)
Beyond rated wind speed, the turbine intentionally reduces aerodynamic efficiency preventing overspeed and overloading. Blade pitch angle is actively increased, while rotor speed and power output are regulated to near-constant values (approximately 15 rpm and 2000 kW respectively). SCADA data shows that with an increase in wind speed (Amb_Wind_Speed), there is an increase in blade pitch angle

(Blades_Pitch_Angle), while maintaining stable power (Grid_Prod_Power = 2000 kW) and rotational speed (Rtr_RPM = 14.8 rpm).

Zone 9 – High-Wind Safety Shutdown (25 < Amb_Wind_Speed)

At wind speeds exceeding approximately 25 m/s, the turbine enters a safety shutdown. Blades are pitched toward 90°, rotor speed drops to zero, and power generation stops. The shutdown is initiated automatically by the turbine control system to prevent structural overload. No observations corresponding to this regime were present in the available SCADA dataset.

3.3. Discretization of SCADA Variables

To train the RIPPER and FOIL algorithms, the continuous SCADA variables were discretized into a finite set of categories. This step was undertaken to enhance the interpretability of the extracted rules and to reduce sensitivity to small measurement fluctuations and noise. Importantly, the discretization directly enables the formulation of rule-based models.

Both RIPPER and FOIL construct classification models as sets of logical rules of the form:

IF (Condition$_1$ ∧ Condition$_2$ ∧ …) THEN (Operational Regime)

Where each condition corresponds to a discretized variable interval. For instance, a rule may take the form:

IF Rtr_RPM_Bin = Stationary_RPM AND Blades_Pitch_Angle_Bin = Standby_Angle THEN Predict Zone_3

In this way, discretization serves as a critical bridge between raw SCADA measurements and interpretable rule-based representations. It ensures that the learned rules remain both human-readable and physically meaningful, while preserving the ability of the model to distinguish between different operational regimes.

Table 1: Discretization of Wind Speed (Amb_Wind_Speed)

Category	Interval (m/s)	Interpretation
Very_Low_Wind_Speed	0 – 3.0	
Low_Wind_Speed	3.1 – 4.0	
Low_Gen_Engagement	4.1 – 5.0	Above Cut-in wind speed
Medium_Wind_Speed	5.1 – 11.0	
High_Wind_Speed	11.1 – 15.0	Above Rated wind speed
Very_High_Wind_Speed	15.1 – 25.0	

Table 2: Discretization of Rotor Speed (Rtr_RPM)

Category	Interval (rpm)	Interpretation
Stationary_RPM	0.0 – 0.2	No rotation
Free_Wheeling_RPM	0.3 – 10.8	Free rotation
Gen_Engagement_RPM	10.9 – 11.2	Generator engagement
Torque_Controlled_RPM	11.3 – 14.8	Increasing torque
Rated_RPM	14.9 – 15.0	Rated rotational speed

Table 3: Discretization of Blades Pitch Angle (Blades_Pitch_Angle)

Category	Interval (°)	Interpretation
Full_Efficiency_Angle	-3.0 – 2	Aerodynamic operation
Control_Angle	2.1 – 23.8	Transition region during to and from standby and control during Rated power generation
Standby_Angle	23.9 – 24.2	Standby (Low-Wind Shutdown)
Lockdown_Transition_Angle	24.3 – 85.8	Transition to or from Safety Lockdown
Lockdown_Angle	85.9 – 90.0	Safety Lockdown

Table 4: Discretization of Power Output (Grid_Prod_Power)

Category	Interval (kW)	Interpretation
Loss_or_Negligible_Power	-30 – 5	No production
Very_Low_Power	6 – 500	Initial output
Low_Power	501 – 1000	Increasing production
Medium_Power	1001 – 1500	Near-rated growth
High_Power	1501 – 1990.0	High production
Rated_Power	1991 – 2010	Rated production

The discretization boundaries were defined based on turbine operating behaviour and remained fixed throughout the modelling process. These boundaries were initially derived from the physics-informed operational regime definitions introduced in the previous section. Subsequently, minor adjustments were applied during exploratory analysis to mitigate classification gaps near regime boundaries arising from measurement uncertainty.

3.4. Rule-Based Learning Framework

While physics-based rules can describe most steady operating zones reasonably well, transition zones such as Zone 2 and Zone 4 are difficult to capture with fixed thresholds, because their boundaries are gradual and sensitive to small variations in the measurements.

The operational regime classification problem was formulated as a supervised multi-class task covering Zones 1 to 8. Rather than training a single multi-class classifier, both RIPPER and FOIL were implemented using a one-versus-rest strategy. In this approach, a separate binary classifier is constructed for each zone, where the target zone is treated as the positive class and all remaining zones as the negative class. This formulation facilitates the extraction of explicit, interpretable rule sets tailored to each operational regime.

3.4.1. Training and Validation Procedure

The 2016 dataset for Wind Turbine 7 was first labelled using the physics-informed regime definitions described in Section 3.2. After labelling, a visual inspection step was conducted to verify the consistency of the regime assignments, particularly in transitional regions where boundary conditions are sensitive to small measurement variations. Manual corrections were made to clearly inconsistent data points to ensure that the training labels aligned with physically meaningful operating states.

For RIPPER, the following hyperparameters were used: prune_size = 0.1, k = 2, and dl_allowance = 64. These values were manually identified to allow moderate pruning while retaining sufficient rule expressiveness. A sensitivity check with $k \in \{1, 2, 3\}$ showed less than 0.5 percentage points variation in overall accuracy, confirming that the results are not sensitive to this choice. The parameters were fixed throughout the experiments and were not tuned on the validation set in order to avoid introducing bias into the reported performance.

3.4.2. Rule Arbitration Strategy

FOIL may generate multiple rules that match a given observation, necessitating a conflict resolution mechanism. In this study, predictions were assigned based on the matching rule with the highest number of literals. This strategy prioritises rules that are more specific over more general ones, thereby reducing ambiguity in overlapping regions of the feature space.

The arbitration strategy was defined prior to evaluation and applied consistently across both internal validation and external testing.

3.4.3. Internal Validation Protocol

Following labelling of the 2016 dataset for Wind Turbine 7, a chronological 80/20 split was applied. The first 80% of the time series was used for training, while the remaining 20% was used for validation. A chronological split was chosen instead of random sampling to preserve the temporal structure of the SCADA data. Since turbine operation is influenced by seasonal and environmental variations, random shuffling was not considered.

3.4.4. External Validation Protocol

To evaluate temporal generalisation, the trained models were applied, without retraining, rebalancing, or parameter adjustments, to an independent and entire dataset from 2017 for Wind Turbine 11. This external evaluation therefore represents a test of model performance under a full year of operating conditions not encountered during training. In addition, using data from a different turbine enabled evaluation of the models' ability to generalise not only across time but also across turbines with potentially different operating characteristics and degradation histories.

3.4.5. Evaluation of the Classification

Model performance on the 2017 dataset was evaluated using overall accuracy and macro-F1 score to account for class imbalance across operational regimes. In addition, confusion matrices were analysed to identify systematic misclassifications, particularly between neighbouring regimes where transitions are inherently gradual.

4. RESULTS

4.1. Internal Validation

4.1.1. RIPPER (Internal Validation)

RIPPER achieved a validation accuracy of 94%, with a macro-F1 score of 0.92.

The confusion matrix for RIPPER during internal validation is shown in Figure 1. The confusion matrix shows strong diagonal dominance, indicating that most operational regimes are clearly separated. A closer look reveals that the majority of errors occur between neighbouring regimes rather than between physically unrelated states.

In particular, misclassifications occur mainly between Zones 5 and 6, which are separated by a narrow rotor speed interval corresponding to generator engagement. Since these regimes share similar wind and pitch conditions and the primary difference lies within a small RPM band, boundary ambiguity is expected when measurements fluctuate around this region. Importantly there is little confusion between distant regimes, such as shutdown states and full-load operations, which suggests that the classifier preserves the overall operational structure of the turbine across regimes.

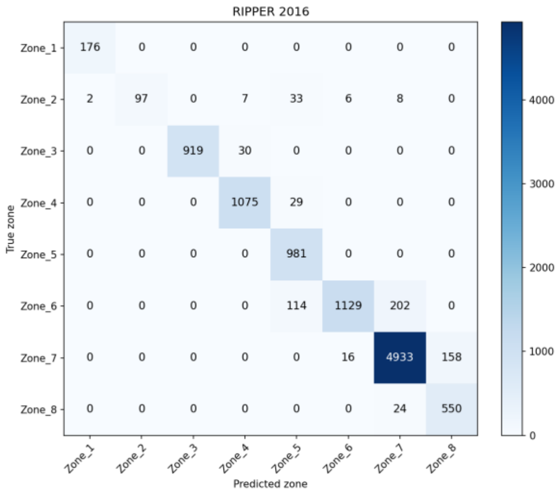

Figure 1. Confusion matrix for RIPPER during internal validation.

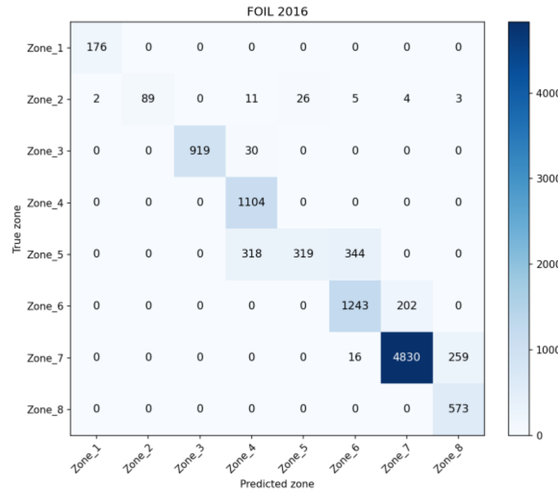

Figure 2. Confusion matrix for FOIL during internal validation.

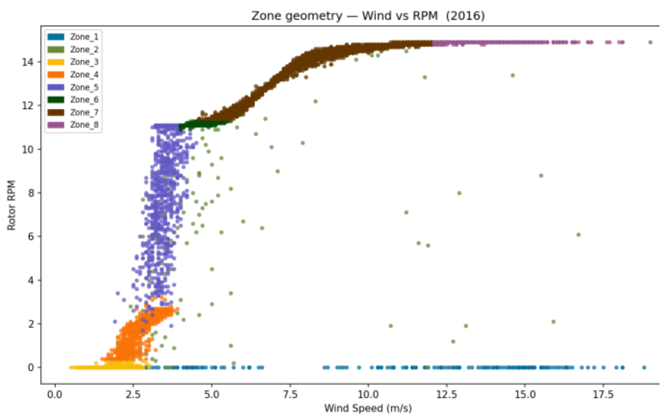

Figure 3 Wind speed versus rotor speed for internal validation.

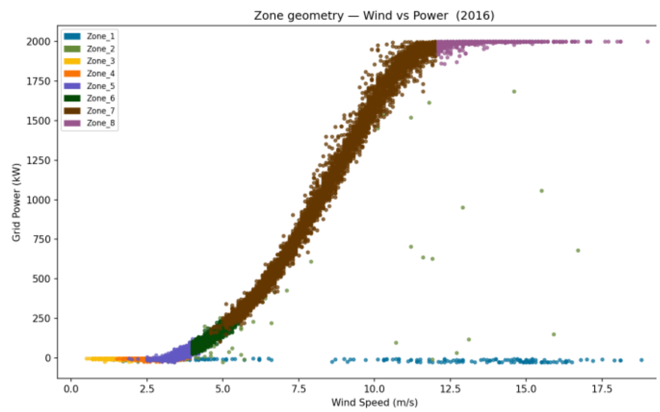

Figure 4. Wind speed versus power output for internal validation.

Steady operating regimes such as standby (Zone 3) and partial load operation (Zone 7) are classified with consistently high precision and recall, reflecting their clearer separation in feature space.

The RPM versus wind speed plot (Figure 3) shows that Zone 6 occupies a very narrow band around the generator engagement speed. Small variations in rotor speed measurements may therefore shift samples between Zone 5 and 6, explaining the observed confusion between the zones.

The power generated versus wind speed plot (Figure 4) shows that partial load operation (Zone 7) follows a stable cubic relationship between wind speed and power production. This consistent aerodynamic behaviour aligns with the strong classification performance observed for this regime. In contrast, transitional regimes such as Zone 2 appear more diffuse in both plots, reflecting their control-driven nature rather than aerodynamic behaviour.

4.1.2. FOIL (Internal Validation)

FOIL achieved a validation accuracy of 88%, with a macro-F1 score of 0.83.

The confusion matrix for FOIL during internal validation is shown in Figure 2. Compared to RIPPER, FOIL shows slightly more confusion in transitional regimes. The primary overlap again appears between Zones 5 and 6, indicating sensitivity to small variations near the generator engagement boundary.

However, similar to RIPPER, the errors remain structured and largely confined to neighbouring regimes. There is no evidence of widespread cross-regime confusion, which suggests that FOIL also captures the main operational structure, but with reduced stability in boundary regions.

The geometric patterns observed in Figures 3 and 4 provide further context for this behaviour. The narrow RPM band associated with Zone 6 increases sensitivity to small variations in measurements, which again contributes to the confusion between Zone 5 and 6. At the same time, the stable

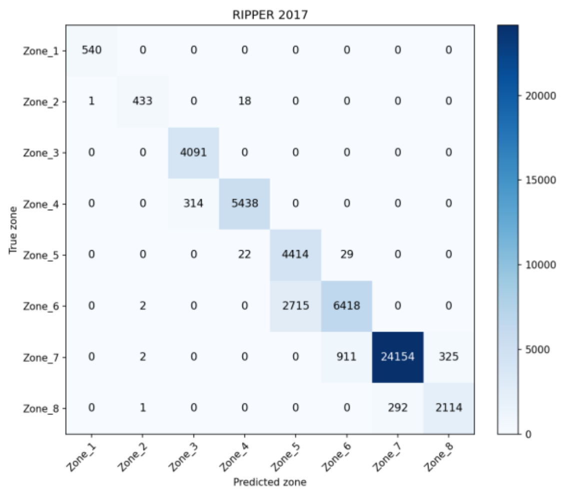

Figure 5 Confusion matrix for RIPPER during external evaluation.

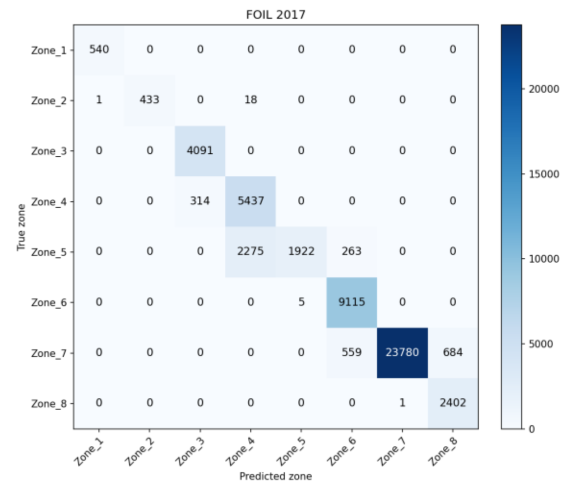

Figure 6. Confusion matrix for FOIL during external evaluation.

wind-power relationship in Zone 7 supports consistent classification performance, although FOIL shows slightly more variability compared to RIPPER in these regions.

Table 6: Per-class classification results - Internal Validation

	RIPPER Prec.	RIPPER Rec.	RIPPER F1	FOIL Prec.	FOIL Rec.	FOIL F1
Zone_1	0.989	1.000	0.994	0.989	1.000	0.994
Zone_2	1.000	0.634	0.776	1.000	0.582	0.736
Zone_3	1.000	0.968	0.984	1.000	0.968	0.984
Zone_4	0.967	0.974	0.970	0.755	1.000	0.860
Zone_5	0.848	1.000	0.918	0.925	0.325	0.481
Zone_6	0.981	0.781	0.870	0.773	0.860	0.814
Zone_7	0.955	0.966	0.960	0.959	0.946	0.952
Zone_8	0.777	0.958	0.858	0.686	0.998	0.813

4.2. External Validation

4.2.1. RIPPER (External Evaluation)

On the external dataset, RIPPER achieved an accuracy of 91%, with a macro-F1 score of 0.91.

The confusion matrix for RIPPER during external validation is shown in Figure 5. The confusion matrix indicates that most operational regimes remain well separated when applied to the 2017 dataset. As in the internal validation case, the majority of errors are concentrated between neighbouring regimes rather than being randomly distributed across classes.

The partial load operation regime (Zone 7) continues to show strong classification performance, reflecting the stable aerodynamic relationship between wind speed, rotor speed, and power in this region.

The largest performance reduction occurs in transitional regimes. In particular, some instances of zone 6 (generator engagement) are classified as Zone 5 (freewheeling), which is consistent with the narrow rotor speed interval defining the engagement boundary. Small variations in RPM measurements around this threshold naturally increase ambiguity. However, there is no evidence of widespread confusion between distant regimes, indicating that the learned rule structure remains physically consistent across both years.

Overall, the results indicate that the learned rules retain their physical meaning when applied to new data, and that the classifier generalises well despite the small shifts in the distribution of the data.

4.2.2. FOIL (External Evaluation)

FOIL achieved an external accuracy of 92%, with a macro-F1 score of 0.89.

The confusion matrix for FOIL during external validation is shown in Figure 6. Compared to internal validation, FOIL shows improved performance in the external evaluation, with both accuracy and macro-F1 increasing relative to internal results. Similar to RIPPER, the primary source of misclassification remains the boundary between Zone 5 and 6, where Zone 5 instances show lower recall. The concentration of errors in transitional regions suggests that FOIL captures the overall regime structure well, though boundary regions remain the main source of uncertainty

Importantly, even in the external setting, the confusion remains structured and limited to adjacent regimes rather than spreading across unrelated operating states. This indicates that the learned rules retain their physical meaning when applied to new data.

Table 7: Per-class classification results - External Validation

	RIPPER Prec.	RIPPER Rec.	RIPPER F1	FOIL Prec.	FOIL Rec.	FOIL F1
Zone_1	0.998	1.000	0.999	0.998	1.000	0.999
Zone_2	0.921	0.958	0.939	1.000	0.958	0.979
Zone_3	0.929	1.000	0.963	0.929	1.000	0.963
Zone_4	0.993	0.945	0.968	0.703	0.945	0.807
Zone_5	0.613	0.989	0.756	0.959	0.430	0.594
Zone_6	0.871	0.703	0.778	0.917	0.998	0.956
Zone_7	0.988	0.951	0.969	1.000	0.937	0.967
Zone_8	0.867	0.878	0.872	0.778	0.998	0.875

4.3. Comparative Analysis

A comparison of the internal validation results (Table 6) with the external validation results (Table 7) shows that the RIPPER classifier maintains relatively stable performance between the two years, with only a slight reduction in macro-F1 (0.92 to 0.91) alongside a moderate decrease in overall accuracy (94% to 91%). This suggests that the extracted rule sets capture the underlying operational logic of the wind turbine rather than patterns specific to the 2016 training dataset.

In contrast, FOIL shows a more notable improvement in external validation, with accuracy increasing from 88% to 92% and macro-F1 from 0.83 to 0.89. The performance gap between the two classifiers, which was six percentage points in internal validation, narrows to less than one percentage point in external validation. This suggests that the FOIL rule set, while less compact, captures decision boundaries that generalise well to data from a different year and turbine.

Overall, both classifiers generalise well to unseen data. RIPPER maintains a small macro-F1 advantage in external validation (0.91 vs. 0.89), while FOIL achieves marginally higher overall accuracy (92% vs. 91%). The compactness and interpretability of RIPPER rule sets, combined with its consistent macro-F1 performance, make it a practical choice within a digital twin framework. FOIL's strong external performance indicates it is also a viable option, particularly where overall classification rate is prioritised over rule compactness.

4.4. Structural Comparison of RIPPER and FOIL Rule Sets and Their Interpretability

Beyond overall classification accuracy, the internal structure of the learned rule sets is important for understanding how the models behave. Both RIPPER and FOIL generate explicit IF-THEN rules based on the discretized SCADA variables. However, their rule sets differ in how compact and how specific these rules are.

To illustrate these differences, Zone 7 (partial load operation) is used as a representative example.

For Zone 7, RIPPER produced the following set of rules:

[RPM_bin=Torque_Controlled_RPM ∧
Wind_bin=Medium_Wind_Speed ∧
Power_bin=Low_Power] ∨
[RPM_bin=Torque_Controlled_RPM ∧
Wind_bin=Medium_Wind_Speed ∧
Power_bin=Medium_Power ∧
Pitch_bin=Full_Efficiency_Angle] ∨
[RPM_bin=Torque_Controlled_RPM ∧
Wind_bin=Medium_Wind_Speed ∧
Power_bin=High_Power] ∨
[RPM_bin=Torque_Controlled_RPM ∧
Wind_bin=Medium_Wind_Speed] ∨
[Power_bin=High_Power ∧
Pitch_bin=Full_Efficiency_Angle] ∨
[Power_bin=High_Power ∧
RPM_bin=Torque_Controlled_RPM ∧
Wind_bin=High_Wind_Speed]

These rules show that Zone 7 is primarily characterised by torque-controlled rotor speed in combination with medium wind speed, which appears in four of the six clauses. The fourth clause acts as a general catch-all for this combination regardless of power level. Two additional clauses extend coverage to high-power and high-wind conditions, using pitch angle as a secondary discriminator. RIPPER therefore captures this regime mainly through rotor speed and wind speed, adding power and pitch only where necessary to avoid overlap with adjacent zones.

FOIL, in contrast, generated eleven separate rules for the same zone:

R1: IF RPM_bin=Torque_Controlled_RPM AND Wind_bin=Medium_Wind_Speed AND Power_bin=Low_Power AND Pitch_bin=Full_Efficiency_Angle THEN Zone_7
R2: IF RPM_bin=Torque_Controlled_RPM AND Wind_bin=Medium_Wind_Speed AND Power_bin=Medium_Power AND Pitch_bin=Full_Efficiency_Angle THEN Zone_7
R3: IF RPM_bin=Torque_Controlled_RPM AND Wind_bin=Medium_Wind_Speed AND Power_bin=High_Power AND Pitch_bin=Control_Angle THEN Zone_7
R4: IF RPM_bin=Torque_Controlled_RPM AND Wind_bin=Medium_Wind_Speed AND Power_bin=High_Power THEN Zone_7
R5: IF RPM_bin=Torque_Controlled_RPM AND Wind_bin=Medium_Wind_Speed AND Power_bin=Very_Low_Power THEN Zone_7
R6: IF Power_bin=High_Power AND Pitch_bin=Full_Efficiency_Angle AND RPM_bin=Torque_Controlled_RPM AND Wind_bin=High_Wind_Speed THEN Zone_7
R7: IF Power_bin=High_Power AND Pitch_bin=Full_Efficiency_Angle AND RPM_bin=Rated_RPM AND Wind_bin=Medium_Wind_Speed THEN Zone_7
R8: IF Power_bin=High_Power AND Pitch_bin=Full_Efficiency_Angle AND RPM_bin=Rated_RPM THEN Zone_7

R9: IF Power_bin=High_Power AND RPM_bin=Torque_Controlled_RPM AND Wind_bin=High_Wind_Speed AND Pitch_bin=Control_Angle THEN Zone_7
R10: IF Power_bin=High_Power AND RPM_bin=Rated_RPM AND Wind_bin=High_Wind_Speed AND Pitch_bin=Control_Angle THEN Zone_7
R11: IF RPM_bin=Torque_Controlled_RPM AND Wind_bin=Low_Gen_Engagement THEN Zone_7

FOIL enumerates every combination of power level and pitch angle individually. The first five rules cover the core medium-wind, torque-controlled regime, separating Very_Low, Low, Medium, and High power levels and splitting High_Power by pitch angle (full efficiency vs control angle). Rules R6-R10 extend the zone to high-power and high-wind conditions, further distinguishing Torque_Controlled_RPM from Rated_RPM and the two pitch angle bins across separate rules. R11 covers the torque controlled RPM at low generator-engagement wind conditions. Every condition that RIPPER leaves as an implicit default is made explicit in a separate FOIL rule.
This pattern is also observed in the other operating zones. In general, RIPPER produces fewer rules per zone and each rule covers a larger part of the operating region. FOIL, on the other hand, often generates several closely related rules that differ only in one condition. In this study, RIPPER generated between one and eight clauses per zone, while FOIL generated between one and eleven rules per zone.
The more compact structure of the RIPPER rules may help explain its consistent macro-F1 performance across both evaluation periods. Since the rules cover a broader part of the operating region, they are less sensitive to small changes in the binned SCADA variables. In contrast, the more specific FOIL rules show stronger improvement in external validation, suggesting that their narrower coverage adapts well to the 2017 data distribution despite covering smaller regions of the feature space.

5. LIMITATIONS

Although the proposed rule-based framework demonstrates encouraging performance, several limitations should be acknowledged.
First, the study is based on data from only two turbines (Wind Turbines 7 and 11) using datasets from two years (2016 and 2017). While the external evaluation across two years indicates temporal stability, the generalisability of the learned rules to turbines of different designs, control strategies, or environmental conditions has not been assessed. Future work should therefore include cross-turbine validation across multiple turbines and sites.
Second, the regime labeling relies on physics-informed threshold definitions. Although these thresholds are aligned with turbine control behaviour, small deviations in SCADA measurements around regime boundaries, especially near the generator engagement region, may lead to ambiguous classifications. This limitation is inherent to the operational characteristics of the turbine rather than the learning algorithm itself.
Third, the discretization of continuous variables introduces simplifications in the representation of the operating space. While this improves the interpretability and rule stability, it may reduce sensitivity to subtle variations that could be relevant for early fault detection or degradation monitoring.
Fourth, the framework treats each SCADA observation independently, without incorporating temporal information such as lagged variables or signal derivatives. While such features could improve identification of transition zones, rules over lagged variables are considerably harder to interpret physically. Maintaining the instantaneous feature space was therefore a deliberate choice to preserve interpretability, and extending the framework with temporal context remains a direction for future work.
Finally, this study does not include a systematic comparison with other interpretable classifiers such as decision trees. Adding a full baseline evaluation is beyond the scope of the current paper and is proposed as a direction for future work.

6. CONCLUSION

This study presented an interpretable, rule-based framework for operational regime classification of wind turbine SCADA data. Physics-informed regime definitions were first established based on turbine control behaviour and aerodynamic principles. These labels were then used to train two rule-induction algorithms, RIPPER and FOIL, using discretized representations of key operating variables.
Internal validation on a hold-out subset of the 2016 dataset demonstrated classification potential for both models. RIPPER achieved relatively higher accuracy and macro-F1 score, while FOIL provided slightly lower but still competitive performance. The confusion matrices indicated that most regimes are clearly separated, with misclassifications primarily occurring near narrow transitional boundaries.
To assess temporal robustness, the trained models were evaluated on the full 2017 dataset without retraining or parameter adjustment. RIPPER maintained stable performance across both years, with only a slight reduction in macro-F1, suggesting that the learned rules capture consistent operational behaviour rather than year-specific patterns. FOIL showed improved performance in external validation, with both accuracy and macro-F1 increasing, and the gap between the two classifiers narrowing substantially.
An analysis of the extracted rule sets showed that the learned rules follow the known turbine control logic. Simple and compact rules were obtained for steady operating regimes, while transition regimes required more detailed rule combinations because the boundaries between these states are close to each other in the measurement space.

Overall, the results show that rule-based learning can provide reliable operational regime classification while remaining transparent and easy to interpret. These classifiers can be used as a supporting component in digital twin systems, ensuring that condition monitoring models operate within clearly defined and physically meaningful operating states.

Future work should include validation across multiple turbines, integration with performance tracking models, and investigation of how regime definitions may need to be adjusted as turbine behaviour changes over time.

DECLARATIONS

Declaration on the Use of AI Tools in Manuscript Preparation - During the preparation of this paper, an AI tool (specifically ChatGPT) was used to assist with grammar correction, clarity improvement and general language refinement. The tool was applied conscientiously and in line with ethical guidelines. No AI was used to generate original ideas, analyses, or references. All research, analysis, interpretations, and conclusions are entirely the authors' own, and the authors take full responsibility for the content.

REFERENCES

Cohen, W. W. (1995). *Fast effective rule induction*. In Proceedings of the 12th International Conference on Machine Learning (ICML) (pp. 115-123), Morgan Kaufmann

Dai, J., Liu, D., Wen, L., Long, X. (2016*). Research on power coefficient of wind turbines based on SCADA data*. Renewable Energy, 86, 206-215. https://www.sciencedirect.com/science/article/pii/S0960148115302238?via%3Dihub.

Duguid, L. (2018), *Data analytics in the offshore wind industry – pilot case study outcomes*, CATAPULT - Offshore Renewable Energy Report No. PN000229-RPT-001. https://ore.catapult.org.uk/wp-content/uploads/2018/05/Data-Analytics-in-Offshore-Wind-Pilot-Case-Study-Outcomes.pdf.

Hong, H. S., Hue, N. T., Ninh, N. T., Thuan, N. D., Huong, N. T. L. (2025). *Physics-constrained scheme for outlier removal in wind turbine SCADA data for power curve modeling*. Electric Power Systems Research, 248, 111953. https://www.sciencedirect.com/science/article/pii/S0378779625005449?via%3Dihub

ISO/IEC (2023). *ISO/IEC 30173:2023 – Digital twin – Concepts and terminology*. Geneva: International Organization for Standardization.

Lydia, M., Kumar, Suresh Kumar, S., Selvakumar, A. I., Prem Kumar, G. E. (2014). *A comprehensive review on wind turbine power curve modeling techniques*, Renewable & Sustainable Energy Reviews, Vol. 30, pp.452-460

Manwell, J. F., McGowan, J.G. and Rogers, A.L. (2009). *Wind energy explained – Theory, design and application (2nd ed.)*, John Wiley & Sons Ltd., ISBN 978-0-470-01500-1.

National Academies of Sciences, Engineering, and Medicine (2024). *Foundational Research Gaps and Future Directions for Digital Twins*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26894.

Nössig, A., Hell, T., Moser, G. (2024). *Rule learning by modularity*. Machine Learning, 113, 7479–7508. https://link.springer.com/article/10.1007/s10994-024-06556-5.

Ouyang, T., Kusiak, A., He, Y. (2017). *Modeling wind-turbine power curve: A data partitioning and mining approach*, Renewable Energy, Vol. 102, pp. 1-8

Pandit, R. K., Infield, D. and Kolios, A. (2019). *Comparison of advanced non-parametric models for wind turbine power curves*, IET Renewable Power Generation, Vol. 13(9), pp. 1503-1510

Pandit, R. and Wang, J. (2024). *A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights*, IET Renewable Power Generation, Vol. 18, pp. 722-742

Quinlan, J. R. (1990). *Learning logical definitions from relations.* Machine Learning, 5(3), 239-266. https://link.springer.com/article/10.1007/BF00117105

Saint-Drenan, Y.-M., Besseau, R., Jansen, M., Staffell, I., Troccoli, A., Dubus, L., Schmidt, J., Gruber, K., Simoes, S. (2020). *A parametric model for wind turbine power curves incorporating environmental conditions*. Renewable Energy, 157, 754–768. https://www.sciencedirect.com/science/article/pii/S0960148120306613?via%3Dihub

Singh, M. (2024). *A hybrid – machine learning and possibilistic – methodology for predicting produced power using wind turbine SCADA data*, *Proc. of the 8th European Conference of the PHM Society*, July 3-5, Prague, Czech Republic, 2024, https://papers.phmsociety.org/index.php/phme/article/view/4006.

Singh, M., Kampen, A.-L., Mishra, R., Jha, M. (2025). *Development of an integrated – condition monitoring, SCADA and digital twins – human cognition inspired condition management system for wind turbines*, Proc. of the 8th International Congress and Workshop on Industrial AI and eMaintenance, Luleå, Sweden, May 13-15, 2025.

Tavner, P. (2012). *Offshore wind turbines – Reliability, availability and maintenance*, The Institution of Engineering and Technology, IET Renewable Energy Series 13, ISBN 978-1-84919-230-9.

Xie, J., Dong, H., Zhao, X. (2023). *Data-driven torque and pitch control of wind turbines via reinforcement learning*. Renewable Energy, 215, 118893. https://www.sciencedirect.com/science/article/pii/S0960148123007905?via%3Dihub

Yang, W., Wei, K., Peng, Z. and Hu, W. (2018). *Chapter 7, Advanced health condition monitoring of wind turbines*, W. Hu (ed.), Advanced Wind Turbine Technology, Springer International Publishing AG