https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

A Real-Time IoT, LLM, AI-Supported Wind Turbine Failure Prediction System

İlhan Gamze Saygı[1], Sara Mahyanbakhshayesh[1], Berdan Ak[1], Mehmet Hilal Özcanhan[2], Mehmet Süleyman Ünlütürk[3]

[1]Department of Computer Engineering, Dokuz Eylul University, Izmir, 35390, Turkey
[2]Department of Computer Engineering, Dokuz Eylul University, Izmir, 35390, Turkey
[3]Department of Software Engineering, Yaşar University, Izmir 35390, Turkey
Correspondence to: Sara Mahyanbakhshayesh (saramahyan.smb@gmail.com), İlhan Gamze Saygı (ilhangamzesaygi@gmail.com)

Abstract. Wind and solar energy are two popular alternative energy sources. However, wind turbines are larger and more complex than solar panels. Accordingly, wind turbines are more exposed to environmental factors and therefore more prone to mechanical failures. Our work improves the reliability and efficiency of wind energy systems by presenting an artificial intelligence (AI)-based system that predicts mechanical gearbox and electrical failures. Historical data are aggregated with real-time sensor data to train the prediction model. Apart from related mechanical and environmental data, sensors provide real-time vibration, internal nacelle temperature, wind speed, noise and smoke levels. The developed system integrates AI and Large Language Model (LLM)-based interfaces for real-time interactive monitoring of turbines. The user interface of the developed system allows users to receive informative responses on performance, detected risks, predicted failures, and energy production levels. The developed model has been validated using 5-fold cross-validation based on *Accuracy, Precision, Recall, F1-Score,* and *ROC-AUC.* The model achieves approximately 89.68% *Accuracy,* 90.08% *F1-Score,* 95.65% *ROC-AUC,* and novel metric 65.13%, *Overall Performance*. The performance results demonstrate the promising potential of AI- and LLM-integrated systems for wind energy applications. Prototype data, labeled via the XGBoost model trained with SCADA data, was retrained using the LightGBM algorithm, achieving 98.37% *Accuracy*, 99.16% *F1-Score*, 98.95% *ROC-AUC* and 94.65% *Overall Performance*; the analysis proved that the newly added gas and sound sensors significantly improved the fault prediction performance of the system.

1 Introduction

1.1 Background

Maintaining the reliability of wind turbines is a significant engineering challenge due to their constant exposure to harsh environmental conditions. Since traditional reactive and periodic maintenance methods are costly and operationally inefficient, Artificial Intelligence (AI) and Machine Learning (ML) based predictive maintenance has emerged as a crucial solution. By analysing real-time sensor data, these systems can detect early failure patterns in critical components, such as gearboxes and electrical systems, preventing costly downtime.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Supporting this transition, research demonstrates that combining physical thermal models with ML significantly improves fault prediction and explainability (Corley, 2021; Khan, 2023). For instance, large-scale analyses using decision tree methods on bearing and lubricant temperatures have achieved approximately 92% accuracy (Hsu, 2020).

Furthermore, integrating Large Language Models (LLMs) has transformed traditional, static monitoring dashboards into interactive AI systems. Recent models, such as Generative Pre-trained Transformer (GPT)-4 and DeepSeek-V3, have demonstrated strong performance in generating structured classification workflows and synthetic datasets for fault diagnosis (Tan, 2025; Zhang, 2025). By allowing operators to query complex data in natural language, these multi-modal frameworks significantly enhance the scalability, predictive accuracy, and overall intelligence of modern energy monitoring systems.

1.2 Problem Statement

Wind turbines routinely face unexpected mechanical and electrical failures that disrupt production and drive maintenance costs upward. Existing Supervisory Control and Data Acquisition (SCADA)-based threshold methods offer limited diagnostic capability because they are unable to interpret complex correlations across heterogeneous sensor inputs. Consequently, modern predictive maintenance systems must incorporate AI-based analytics, Internet of Things (IoT) connectivity, and transparent explainability tools to achieve reliable and early fault prediction. Consequently, AI-driven, IoT-supported, and explainable predictive maintenance systems are required instead of reactive monitoring approaches.

In the remainder of this study, the related works and our contributions are presented in Sect. 2. The Materials and Methods, together with the detailed architecture of the Smart Wind Turbine Failure Prediction System (SWTFPS), are provided in Sect. 3. The LLM-based interactive user interface and system integration components are described in Sect. 4. The experimental results and model performance analyses are discussed in Sect. 5. Our Smart Wind Turbine Failure Prediction System (SWTFPS) is presented in Sect. 4. The results and model performance comparison are presented in Sect. 5. The discussion and limitations are in Sect. 6, followed by the Conclusion and Future Work in Sect. 7.

2 Related Works

Another study by Zhang et al. (2023) proposed a convolutional network, a multi-scale temporal convolutional network to balance the data, achieving 93.4% accuracy. Amin et al. (2022) combined cyclostationary analysis with a Convolutional Neural Network (CNN) to detect gearbox faults using generated data, achieving 89% accuracy even under variable wind conditions. In another study, a SCADA-based system was developed. Using neural networks for temperature anomaly detection in gearboxes and generators, failures were predicted up to 17 hours in advance with high accuracy (Alzawaideh, 2021). A prior study evaluated 16 supervised models on SCADA data, identifying Random Forest (RF) and Extra Trees (ET) as the best performers for detecting faults 24 hours in advance (de Lima Munguba, 2024). Inturi et al. (2021) applied multi-level data fusion of vibration, acoustic, and oil sensors with system, achieving 92% accuracy in gearbox fault classification. Meyer et al. (2022) developed a fully automated CNN approach using vibration spectrograms, achieving 100% accuracy on

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

laboratory data without manual feature engineering. Ogaili et al. (2024) used an eXtreme Gradient Boosting (XGBoost) classifier to detect blade faults, such as erosion and twisting, with 99.4% accuracy. Pham and Han (2025) designed a Bi-LSTM (Bidirectional Long Short-Term Memory) model with an Attention Mechanism, achieving 99.40% accuracy and providing 62-hour early warnings. Puruncajas et al. (2025) developed a normal behaviour model using SCADA temperature data, detecting gearbox anomalies 9 months before failure. Santiago et al. (2024) reviewed deep learning models for predictive maintenance, highlighting a research gap in their application to direct-drive and synchronous generators systems. Yu et al. (2021) proposed a convolutional network using wavelet packet coefficients, achieving over 99% accuracy in noisy environments. Zeng et al. (2023) introduced an estimation model to establish dynamic thresholds for anomaly detection, outperforming standard alarm systems. Zhang et al. (2023) combined CNN and (Long Short-Term Memory) LSTM to predict bearing temperatures from SCADA data, detecting anomalies earlier than single-model approaches. Finally, Zhang et al. (2014) used an Artificial Neural Network (ANN) with backpropagation on SCADA data to predict bearing temperatures, issuing warnings 3 months in advance without additional sensors.

2.1 The Research Gap and Our Contributions

The existing literature extensively utilises SCADA data and various deep learning architectures (CNNs, LSTMs, Transformers) for vibration analysis. However, significant gaps remain in integrating user interaction and failure prediction within a single system, because most studies rely on single-source data and present results through static dashboards rather than interactive interfaces. To address this research gap, we integrated the following state-of-the-art technologies as our contributions to wind turbine systems:

1. **Integrated LLM Interface:** Allows operators to conduct conversational monitoring and receive real-time, proactive explanations and recommendations for preventing faults.
2. **New & Traditional Data Integration:** A wide array of real-time sensor data is integrated with traditional SCADA data.
3. **Hybrid & Comparative AI:** A comprehensive comparison of multiple machine learning algorithms to identify the best model for wind turbine fault prediction.

3 Materials and Methods: System Architecture

The methodological framework of SWTFPS is built upon three phases: (1) the construction of **hardware** infrastructure equipped with multi-modal sensors for real-time data acquisition; (2) the development and evaluation of **machine learning models** to detect mechanical and electrical anomalies and (3) the integration of a **Large Language Model (LLM)** interface to facilitate interactive decision support and data visualisation.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

3.1 Materials: Custom-made Small Wind Turbine Prototype

A prototype has been designed and constructed to add newly available sensor data to traditional SCADA data. The prototype shown in Figure 1 consists of six main components:

1. The primary mechanical *Vertical Axis Wind Turbine (VAWT)* captures wind energy.
2. An *anemometer* measures real-time wind speed.
3. A *wind vane* monitors wind direction.
4. The *structural frame chassis* supports the turbine and electronic components.
5. A *DC generator* converts mechanical rotation into electrical energy.
6. The *control circuitry* houses the Espressif ESP32 microcontrollers and sensors. *Dual ESP32* units operate within a network environment with a minimum bandwidth of 10 Mbps and a latency below 50 ms.

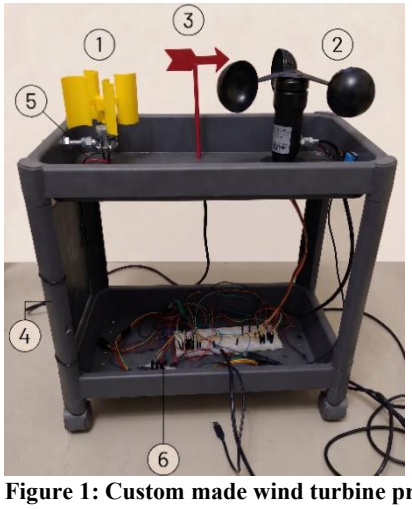


Figure 1: Custom made wind turbine prototype.

Two ESP32 microcontrollers are utilised to manage multiple real-time sensor inputs. All collected data is stored with timestamps. The circuit diagram for sensor integration is shown in Fig. 2. Three *DHT11* modules track temperature and humidity gradients across different sections of the system, while *gas and sound sensors monitor* atmospheric quality and acoustic emissions. A *piezoelectric sensor* is used to detect structural vibrations. A *line detect sensor* is used to measure the blade speed to ensure synchronisation with wind speed. An *INA219* power-monitoring module is integrated to capture precise current and voltage data from the turbine's output.

3.2 Materials: Data from the Prototype

By carefully mapping the physical components, a real-world industrial wind turbine was modelled at a laboratory scale. The prototype utilises an array of sensors to capture critical environmental and mechanical variables and preserves data structures for ensuring operational fidelity. Our physical model bridges the gap between theoretical SCADA datasets and live IoT data. Detailed descriptive statistics regarding the collected 2453 instance data from the prototype are presented in Table 1.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

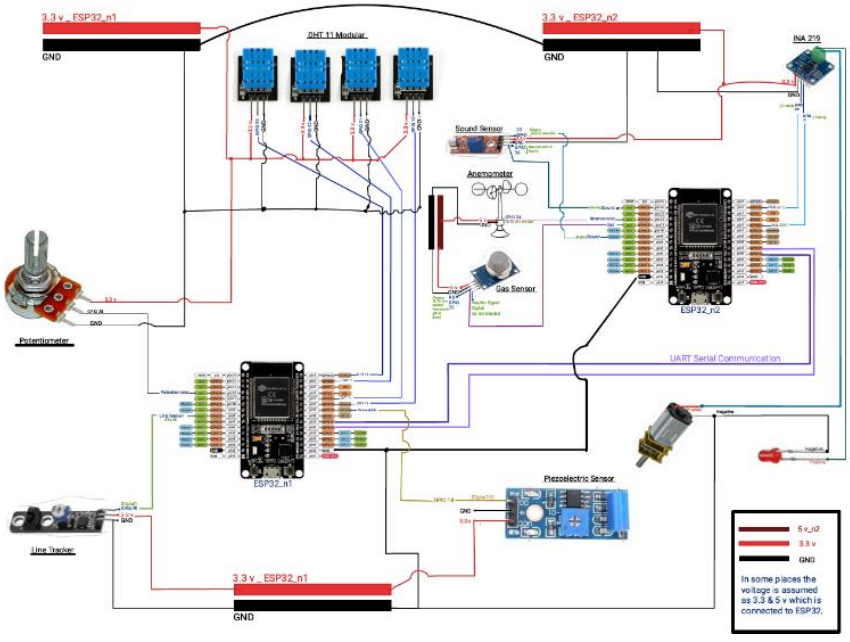


Figure 2: The circuit diagram of sensor integration on the prototype.

Table 1: Basic statistics of collected data from the prototype.

Attributes	Average	Standard Deviation	Coefficient of Variation	Min.	Max.	Range	IQR	MAPE
microphone	108.76	102.19	0.94	0	538	538	155	78.98
gas	368.79	57.15	0.15	113	545	432	65.75	12.38
wind_direction	167.95	140.4	0.84	21	360	339	309	78.33
vibration	525.49	612.26	1.17	0	2941	2941	974.75	99.85
wind_speed	24.58	23.18	0.94	0	250.99	250.99	38.84	75.04
ambient_temperature	28.84	6.42	0.22	0	50.9	50.9	6.9	16.63
gearbox_temperature	32.1	9.81	0.31	0	60.7	60.7	12.6	23.84
nacelle_temperature	33.6	7.19	0.21	0	58.4	58.4	9.4	16.66
voltage	1.61	0.88	0.55	0	3.17	3.17	1.73	51.18
current (mA)	0.29	0.12	0.41	0	0.89	0.89	0.14	30.89
power (mW)	0.46	0.33	0.72	0	2.47	2.47	0.46	58.6

The data collected from the prototype has been pre-processed. Principal Component Analysis (PCA) was used to visualise the data. To detect outliers, the 1.5×IQR (Interquartile Range) rule, rolling window analysis, K-Means clustering, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) were applied as in Fig. 3. A Min–Max scaler was used for normalization and feature scaling. Multi-class categorical features were mapped to numerical values. Non-predictive columns with zero standard deviation were removed. Features and target labels were prepared for machine learning. The dataset was initially divided into two subsets: 80% for model development and 20% as a hold-out test set for final evaluation. To reduce overfitting, the development subset was further evaluated using 5-fold cross-validation, where each fold consisted of 80% training data and 20% validation data. *Time Window Sliding* was employed to identify and filter

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

statistical outliers within localised temporal segments, removing extreme values while preserving the overall trend of the turbine's operational signals. The imbalanced dataset was handled through *Data-Level Methods.* To prevent data leakage, the dataset was *split chronologically.* A novel K-Means undersampling approach was then applied exclusively to the training set, balancing the classes by reducing "Normal" instances to representative centroids matching the "Failure" count.

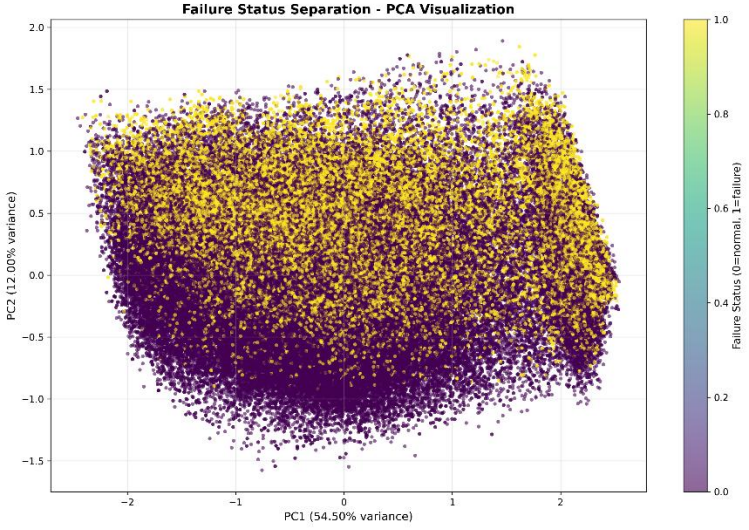


Figure 3: Final pre-processed dataset balanced via K-Means centroids.

The pre-processing pipeline ensured that the data were suitable for various machine learning algorithms, while mitigating issues such as skewness, imbalance, or feature misrepresentation.

3.3 SCADA Dataset Pre-processing

The specific dataset utilised is the large-scale **"Wind Turbine SCADA Data for Early Fault Detection"**. A detailed description of the data, engineering units, and sensor types is available in Kasimov (2024). The dataset covers 89 cumulative years of operational data derived from 36 industrial-scale wind turbines across three geographically distinct wind farms, as summarised in Table 2. The SCADA dataset initially had a significant imbalance, with 87.2% "Normal" operations and only 12.8% as "Failure" events. A balancing procedure was implemented.

Table 2: SCADA dataset summary.

Metric	Wind Farm A	Wind Farm B	Wind Farm C
Total Features	86	257	957
Resolution	10 Minutes	10 Minutes	10 Minutes
Training Duration	1 Year	1 Year	1 Year
Status Labels	Comprehensive	Comprehensive	Comprehensive

To simplify high-dimensional SCADA data, a feature-extraction process was implemented to capture the industrial three-phase alternating current (AC) power characteristics. The three separate voltage readings from the original dataset

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

(***sensor_32_avg***, ***sensor_33_avg***, and ***sensor_34_avg***) were combined into a single "Calculated Voltage" feature by averaging them. The final SCADA feature set is shown in Table 3.

Table 3: Reduced dataset obtained from original SCADA data.

Feature Name	Description
sensor_0_avg	Ambient temperature
sensor_1_avg	Wind absolute direction
sensor_43_avg	Nacelle temperature
sensor_11_avg	Gearbox temperature
wind_speed_3_avg	Windspeed
voltage	Calculated voltage (Mean of 3 phases)
sensor_50	Total active power

3.4 Methods: Our SWTFPS Architecture

The proposed system is a multi-layered architecture (Fig. 4) that integrates traditional data with real-time data stored on a cloud service. The system transmits operational metrics to the ThingSpeak IoT platform [20]. The integrated data are then analysed using machine learning algorithms and the Gemini API before being visualised in Oracle APEX.

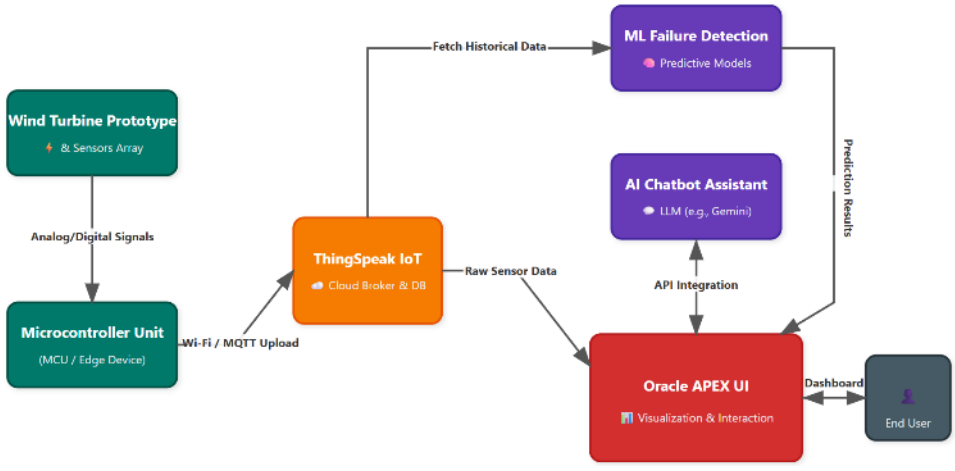


Figure 4: Overall system architecture.

For real-time operation, the system utilises the secure communication pipeline shown in Fig. 5, which employs both the Hypertext Transfer Protocol (HTTP) and Message Queuing Telemetry Transport (MQTT) protocols. The ESP32 units act as publishers sending data to distinct topics (*WindTurbine1*, *WindTurbine2*). The *Python AI server* acts as a subscriber, continuously retrieving raw data, performing failure predictions, and writing diagnostic feedback for the user.

3.5 Evaluated Models for Decision Support

The system analyses sensor data to detect potential faults in mechanical gearboxes and electrical systems at an early stage of operation. To achieve the highest precision, eleven different machine learning models (Table 7) were evaluated.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

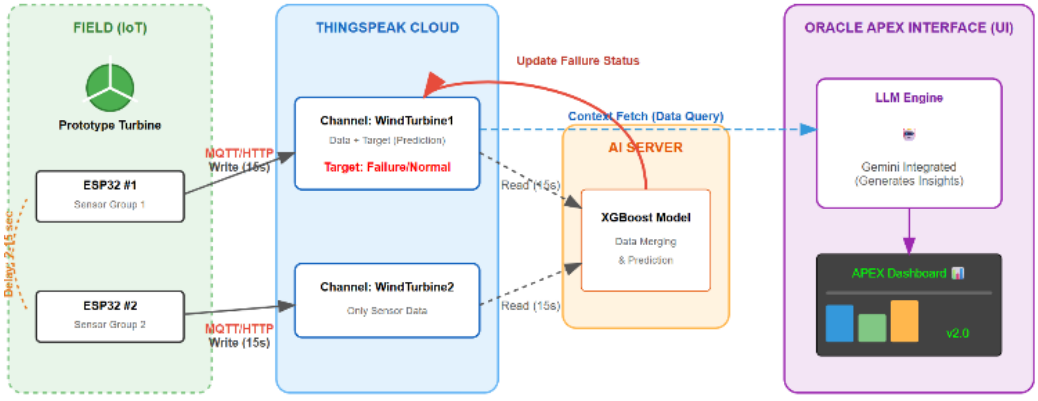


Figure 5: IoT and ML communication flow diagram.

Among the evaluated models, XGBoost emerged as the optimal choice due to its superior ability to handle non-linear sensor relationships and deliver the highest predictive performance. Ultimately, the system utilises this optimised model to provide actionable diagnostic outputs, including failure probability, risk levels, and the overall anomaly status of wind turbines. After the data pre-processing and model optimisation stages, the optimal XGBoost model was integrated into the physical system for real-time operational tests. Physical prototype tests have shown that the model successfully maintains a 15-second data collection cycle. Data transfer and analysis are completed within 5 seconds, panel updates within 3 seconds, and chat responses within 10 seconds. Synchronisation between ESP32 units was maintained, and the system returned to normal operation within 15 seconds without data loss.

3.6 Performance Metrics

The overall correctness and reliability of the models were evaluated using standard *supervised learning metrics: Accuracy, Recall, Precision, F1-Score,* and *ROC-AUC.* The mathematical formulations and detailed theoretical backgrounds of these standard machine learning evaluation metrics are further examined in (Naidu, 2023). The metrics collectively provide a comprehensive assessment of the model's ability to distinguish between faulty and healthy turbine states, especially in cases of unbalanced class distributions. In wind turbine failure prediction, the *F1-Score* is preferred because it better reflects the minority fault class in highly imbalanced datasets. The *ROC-AUC* is relegated to the background as they can be misleading in imbalanced data. To provide a more comprehensive and balanced assessment of classification performance under imbalanced operating conditions, an Overall Performance (OP) metric was additionally formulated by combining both success-based and error-based evaluation criteria. While conventional metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC individually describe specific aspects of model behaviour, they may not fully reflect the trade-off between predictive capability and classification reliability in fault diagnosis applications. In particular, high Accuracy or ROC-AUC values may still coincide with elevated false alarm rates or missed fault detections in highly imbalanced datasets. Therefore,

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

the proposed OP formulation integrates both positive performance indicators and error penalties into a single evaluation framework. The OP metric is defined as in Eq. (1):

$$Overall\ Performance = \frac{(0.25\times F1{-}Score)+(0.2\times ROC{-}AUC)+(0.15\times Accuracy)+(0.15\times Precision)+(0.15\times Recall\ )+(0.1\times(1-FPR))}{1+LogLoss\times Brier\ Error\times FNR}, \quad (1)$$

where Accuracy, Precision, Recall, F1-Score, and ROC-AUC represent success-oriented metrics, while LogLoss, Brier Error, False Positive Rate (FPR), and False Negative Rate (FNR) represent prediction uncertainty and classification error penalties. The constant value of 1 in the denominator is included to prevent numerical instability and disproportionate inflation of the metric when error terms approach zero. By jointly considering predictive strength and error behaviour, the OP metric provides a more interpretable and balanced comparison of machine learning models for real-time wind turbine fault prediction systems.

3.7 Feature Importance Analysis

Pearson correlation analysis was applied to optimise the high-dimensional SCADA dataset and reduce model complexity and noise (Fig. 6). Variables with strong and moderate correlations were retained as core components of the prediction model, whereas weakly correlated features were removed. The selective filtering created a more efficient diagnostic structure suitable for real-time processing, emphasising the most influential thermal, mechanical, and electrical relationships associated with fault conditions. The additional Scaled Feature Correlation Analysis based on Customised Thresholds ensured the removal of noise, allowing the AI to focus its learning on the most meaningful variables.

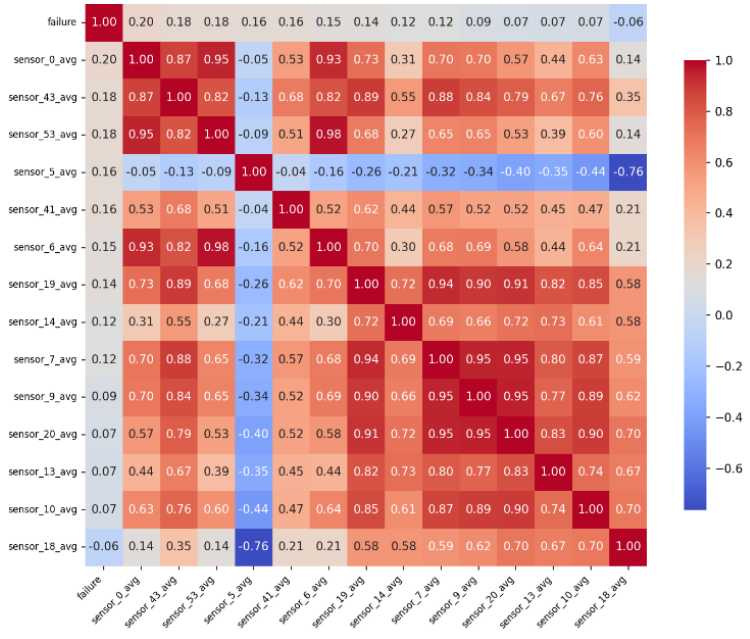


Figure 6: Heatmap of the most related features.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

The SHapley Additive exPlanations (SHAP) plot in Fig. 7 shows the model's most impactful features, from highest to lowest. These seven features form the basis of the appropriate final training phase. The features that impact real-time failure conditions most are as follows:

- ***sensor_0_avg*** (Ambient temperature)
- ***sensor_43_avg*** (Nacelle temperature)
- ***sensor_50*** (Total active power)
- ***sensor_11_avg*** (Gearbox temperature)
- ***wind_speed_3_avg*** (Windspeed)
- ***sensor_1_avg*** (Wind absolute direction)
- ***voltage*** (Mean of voltage phases)

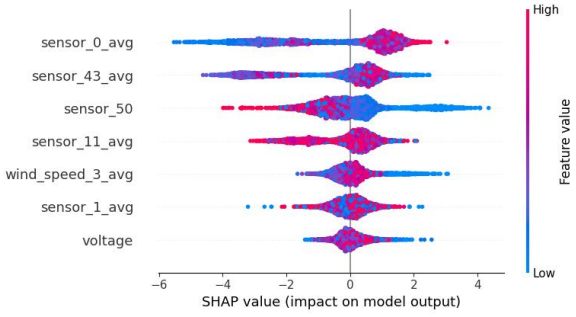


Figure 7: SHAP of XGBoost prototype model.

4 SWTFPS: LLM-Based Interactive User Interface

In the final phase of the SWTFPS project, a comprehensive Oracle APEX-based web interface was developed to make diagnostic data from the XGBoost model actionable for end users (Fig. 8). This platform acts as a bridge between complex AI logic and the operator.

The interface's technical architecture is built on modern web technologies (*HyperText Markup Language: HTML, Cascading Style Sheets: CSS, JavaScript: JS)* on the frontend and robust data management *(Structured Query Language: SQL, Procedural Language/Structured Query Language: PL/SQL)* on the backend. The system consists of the following core modules:

- Live IoT monitoring dashboard
- Instant AI failure predictions
- Interactive LLM maintenance assistant
- Transparent SHAP/Local Interpretable Model Agnostic Explanation (LIME) visual reports
- Natural Language system querying
- Fast understandable AI outputs

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

This integration transforms the prototype into a professional predictive maintenance platform. Supported by an LLM, it delivers real-time turbine health reports and optimisation suggestions on demand.

Figure 8: Oracle APEX real-time monitoring interface.

Figure 9 illustrates the typical page of a given response by the LLM interface to the user's question. The AI assistant can be updated quickly and easily by using its Retrieval Augmented Generation (RAG) capability. Examples of critical prompts and their AI assistant responses are given in Table 4.

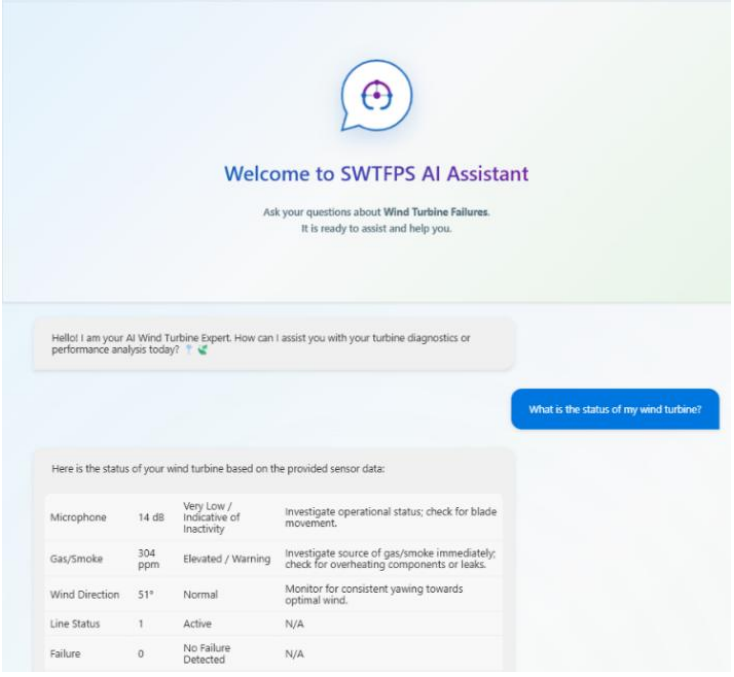


Figure 9: The SWTFPS AI assistant user interface.

Table 4 presents the most critical prompts and AI assistant responses Number 1-3, in a tabular format. When a status query is initiated, the backend fetches real-time sensor data, which the LLM then uses to interpret the system state.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

Table 4: Several complicated questions.

Number	Prompt	Response
1	What is the status of my turbine?	Based on the provided sensor data, your wind turbine is in a critical state and .... • **Microphone (dB): 110** -… • ... (other sensors) … **Summary and Suggestions: …**
2	What is the temperature of my turbine?	The ambient temperature around your wind turbine is **24.3°C**, the gearbox temperature is **22.2°C**, and the nacelle temperature is **23.4°C**. ….
3	Do you have suggestions to improve my wind turbine? Explain in detail.	**Comments and Suggestions based on Sensor Data:** The sensor data indicates a critical situation. … **Immediate Action Required:** …

5 Results and Comparison of Model Performances

5.1 Model Comparison and Final Choice

To maximise the predictive capability of the models and prevent overfitting, systematic hyperparameter tuning was conducted using 5-Fold Cross-Validation, following standard optimisation methodologies established in literature. Rather than relying on default parameters, optimal configurations for each algorithm (such as tree structures for ensemble methods, learning rates for boosting models, and specific distance metrics) were identified based on the highest mean *F1-Scores*.

The final performance metrics of the algorithms post-tuning are summarised in Table 5. The results indicate that CatBoost and LightGBM, achieved the highest accuracy levels. While XGBoost and Random Forest also demonstrated strong predictive capabilities, simpler models like Logistic Regression showed limited performance. This outcome highlights the need for tuned boosting models to capture the complex, non-linear relationships inherent in industrial sensor data. The architectural structures of the specific models and the technical definitions of parameters are consistent with the literature, and Islam et al. (2025) can be considered as a reference for comprehensive explanations of these parameters. The hyperparameter optimization process was carried out independently according to the characteristics of the existing dataset.

Table 5: Accuracy and optimal hyperparameters for all evaluated models.

Model	Accuracy	Key Optimal Parameters
CatBoost	89.66%	**learning_rate:** 0.1, **iterations:** 500
LightGBM	89.58%	**num_leaves:** 100, **n_estimators:** 200, **learning_rate:** 0.05
XGBoost	**88.98%**	**n_estimators:** 200, **max_depth:** 10, **learning_rate:** 0.05
RandomForest	88.02%	**n_estimators:** 100, **max_depth:** 20, **min_samples_split:** 2
DecisionTree	82.79%	**max_depth:** 20, **min_samples_split:** 10, **min_samples_leaf:** 4
MLP	80.20%	**hidden_layer_sizes:** (128, 64), **learning_rate_init:** 0.01, **alpha:** 0.001
KNN	79.87%	**n_neighbors:** 7, **weights:** 'distance'
Logistic	69.90%	**C:** 0.01, **penalty:** 'l2'

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

The performance of the k-Nearest Neighbors (KNN) and XGBoost models was compared using the raw dataset. The results clearly show the limitations of using unbalanced data. Although the KNN model has a high overall *Accuracy* (98.43%), its *Recall* (91.87%) and *F1-Score* (93.33%) for the "Failure" class indicate a strong bias toward the majority "Normal" class. On the other hand, the XGBoost model struggles significantly with raw high-dimensional data. Its overall accuracy drops to 66.09%, and its *Precision* for the Failure class falls sharply to 20.65%, indicating it generates many false alarms. While the initial baseline evaluation of the raw data showed that XGBoost performed worse than *KNN*, the model trained on the refined dataset emerged as the top performer.

According to the Confusion Matrix in Fig. 10, the XGBoost model can distinguish failure cases (Class 1) with high accuracy (1876 True Positives, **89.68%)**, and an F1-Score of **90.08%**. An Area Under the Curve (AUC) of **95.65%** is observed in the Receiver Operating Characteristic (ROC) curve in Fig. 10. The high predictive capability demonstrates the model's effectiveness in detecting faults and providing actionable insights.

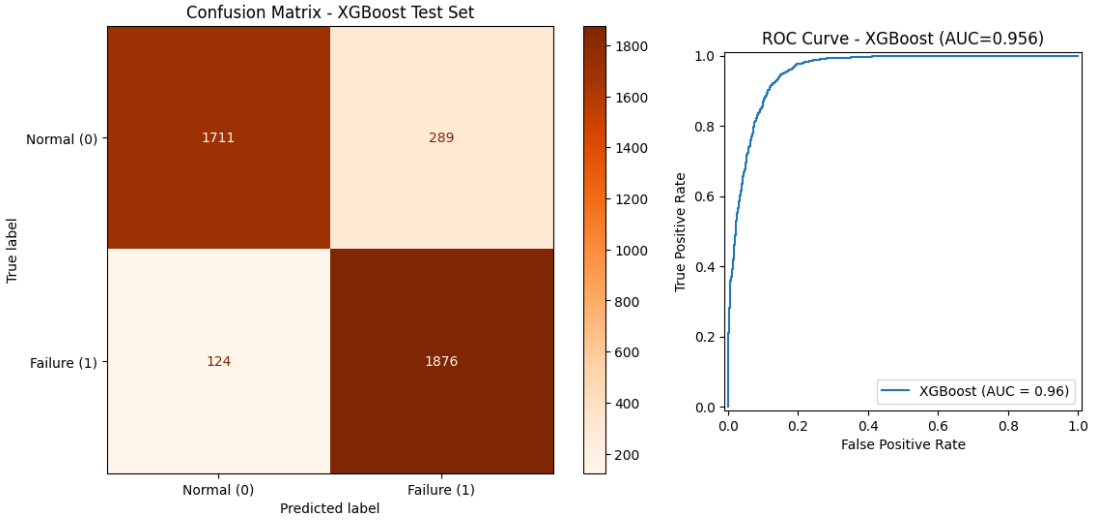


Figure 10: Confusion matrix and ROC curves of the XGBoost.

5.2 Explainability and Interpretation Capability

Following the initial assessment of the large-scale dataset, the SWTFPS was tested on a small-scale, refined subset of approximately 20,000 samples. The F1-Score was prioritised as the primary evaluation metric over ROC-AUC, as it better mitigates the imbalance in the wind turbine data. Again, the XGBoost model was selected for its superior performance. It achieved the highest *F1-Score* among the evaluated models, demonstrating a robust balance between Precision and Recall, in a small-scale experimental trial as well.

As detailed in Table 6, the *XGBoost* model provides a superior balance of high accuracy, fast inference speed, and architectural simplicity, making it the most reliable and cost-effective solution.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 6: Model performance comparison.

Model	Accuracy	F1-Score	ROC-AUC
XGBoost	**89.68%**	**90.08%**	**95.65%**
Stacking	89.58%	89.85%	96.39%
LightGBM	88.78%	89.32%	95.70%
CatBoost	88.23%	88.79%	95.61%
RandomForest	87.78%	88.22%	95.59%
MLP	80.18%	80.80%	88.37%
SVM	77.68%	77.89%	85.31%
KNN	77.83%	77.81%	85.39%
NaiveBayes	71.15%	71.11%	77.65%
DecisionTree	84.05%	84.28%	84.05%
Logistic	69.58%	69.64%	75.91%

Transforming the data into a noise-free, balanced dataset significantly improved the performance of all eleven evaluated models. However, the results show that all tree-based ensemble methods (Stacking Ensemble, Random Forest, LightGBM, CatBoost, and XGBoost) outperformed the simpler models (KNN, Logistic Regression, etc.).

6 Discussion and Limitations

6.1 Interpretation of Results

The XGBoost model optimised with 50% data balancing and the selection of seven key features demonstrated the highest performance **(89.68% Accuracy, 90.08% F1-Score, 95.65% ROC-AUC)**. This finding demonstrates that data quality is more critical than data quantity in fault prediction and that this compact approach is highly effective for real-time monitoring on hardware such as ESP32.

The global SHAP analysis revealed that ambient and nacelle temperatures are the primary drivers of predictions, which align well with turbine thermodynamics. To further complement the evaluation, local dynamics were examined using LIME (Fig. 11). The evaluated scenarios demonstrated the model's reliability in accurately detecting failures driven by thermal anomalies (*Case 1*). The model also recognises stable conditions (*Case 2*) and, critically, prioritises high-temperature warnings even when conflicting standard power metrics suggest normal operation (*Case 3*).

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

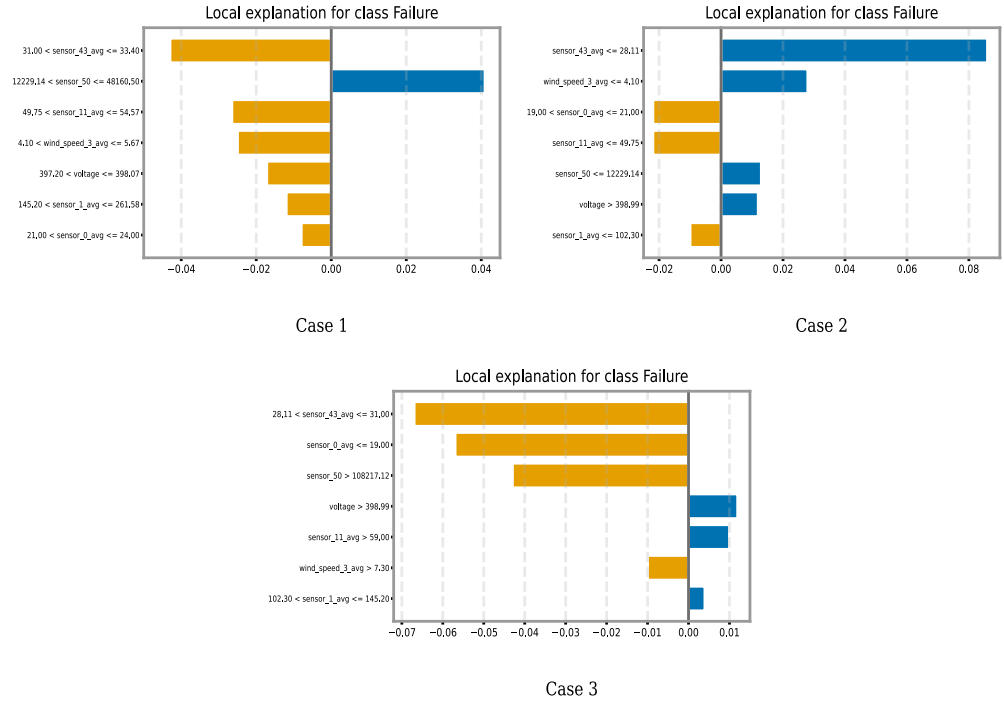


Figure 11: LIME cases of XGBoost.

To provide a more balanced evaluation of predictive capability and classification reliability under imbalanced conditions, the proposed Overall Performance (OP) metric introduced in Sect. 3.6 was additionally employed, and the result in Table 7 was obtained. As a result of the evaluation based on the overall performance metric, it is seen that the XGBoost model, whose performance was previously proven by the **F1-Score** ranking, stands out among the other models with its high **overall performance** of **65.13%.** The obtained result was compared with the results of similar studies in the literature.

Table 7: Model overall performance comparison.

Model	Accuracy	Precision	Recall	F1-Score	Log Loss	Brier Error	ROC-AUC	FPR	FNR	OP
XGBoost	89.68%	86.65%	93.80%	90.08%	25.36%	7.73%	95.65%	14.45%	6.20%	65.13%
Stacking	89.58%	87.57%	92.25%	89.85%	24.63%	7.48%	96.39%	13.10%	7.75%	64.94%
LightGBM	88.78%	85.17%	93.90%	89.32%	25.72%	7.94%	95.70%	16.35%	6.10%	64.40%
CatBoost	88.23%	84.70%	93.30%	88.79%	26.87%	8.19%	95.61%	16.85%	6.70%	63.18%
RandomForest	87.78%	85.12%	91.55%	88.22%	28.88%	8.78%	95.59%	16.00%	8.45%	61.07%
MLP	80.18%	78.32%	83.45%	80.80%	43.27%	13.87%	88.37%	23.10%	16.55%	47.12%
SVM	77.68%	77.15%	78.65%	77.89%	47.85%	15.52%	85.31%	23.30%	21.35%	42.88%
NaiveBayes	71.15%	71.21%	71.00%	71.11%	62.86%	20.15%	77.65%	28.70%	29.00%	34.16%
Logistic	69.58%	69.49%	69.80%	69.64%	58.95%	20.03%	75.91%	30.65%	30.20%	33.87%
KNN	77.83%	77.87%	77.75%	77.81%	117.78%	15.55%	85.39%	22.10%	22.25%	31.04%
DecisionTree	84.05%	83.09%	85.50%	84.28%	574.90%	15.95%	84.05%	17.40%	14.50%	11.91%

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

The results obtained in the articles in the literature are included in Table 8. However, the success and error metrics provided in the literature are insufficient to fill the table, and *Overall Performance (OP)* values could not be calculated. Similarly, Puruncajas et al.'s Artificial Neural Network (ANN) model and CPEM Study's Combined Probability Estimation model did not provide any scores and were therefore not directly included in the table.

Table 8: Model overall performance comparison among studies.

Article	Best Model	Accuracy	Precision	Recall	F1-Score	Log Loss	Brier Error	ROC - AUC	FPR	FNR	OP
Munguba et al.	Random Forest Classifier (RF)	99.94%	99.94%	99.94%	99.94%	-	-	100.0%	-	-	-
Zhang et al. (2023)	Adaptive Multivariate Time-Series CNN (AdaMTCN)	93.40%	-	-	99.60%	-	-	-	-	-	-
Zhang et al. (2025)	Hybrid CNN-LLM	91.50%	91.00%	90.00%	90.50%	-	-	96.00%	-	-	-
Ogaili et al.	Boost + ReliefF	99.40%	99.80%	99.80%	99.80%	0.005	-	99.90%	-	-	-
Pham & Han	Bi-LSTM + Attention Mechanism	99.40%	99.25%	99.52%	99.38%	-	-	99.80%	0.65%	-	-
Irfan et al.	Perturbed-Random Forest (P-RF)	99.88%	99.88%	99.88%	99.88%	-	-	100.0%	-	-	-
Tan & Carroll	DeepSeek-V3 Random Forest	-	97.00%	97.00%	97.00%	-	-	-	-	-	-
Inturi et al.	ANFIS (Adaptive Neuro-Fuzzy)	92.00%	-	-	-	-	-	-	-	-	-
Horodyvskyy et al.	LSTM Autoencoder / MLP	94.00%	-	94.00%	-	-	-	-	11.00 %	-	-
Hsu et al.	Random Forest	95.34%	-	-	-	-	-	-	-	-	-
Amin et al.	Cyclostationary-based CNN	89.00%	-	-	-	-	-	-	-	-	-

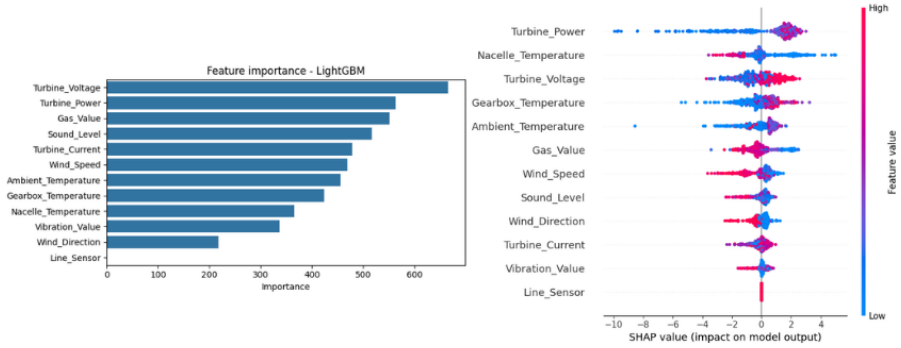
The data collected by the prototype was labeled by XGBoost which is the best model obtained from SCADA data. After the newly labeled dataset was retrained, LightGBM was selected as the best performing model with 98.37% *Accuracy*, 99.16% *F1-Score*, 98.95% *ROC-AUC*, and **94.65%** *OP* scores. Table 9 shows the comparison between the models trained with the data collected from the prototype. Figure 12 shows the (a) Feature Importance and (b) SHAP plots obtained from the LigthGBM model results. According to the feature importance result, the newly added sensors, **gas** and **sound** level sensors, were observed as *significant predictors* of failure output, while **vibration** was observed as a *moderate predictor.* The fact that the newly added sensors enabled the model to achieve a **98.86%** F1-Score (and **94.65%** OP) indicates that they have a high prediction performance above failure prediction.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 9: Data model overall performance comparison of labelled prototype.

Model	Accuracy	Precision	Recall	F1-Score	Log Loss	Brier Error	ROC-AUC	FPR	FNR	OP
LightGBM	98.37%	98.75%	99.58%	99.16%	10.27%	1.58%	98.95%	37.50%	0.42%	94.65%
CatBoost	98.17%	98.54%	99.58%	99.06%	5.30%	1.50%	98.66%	43.75%	0.42%	94.18%
XGBoost	97.96%	98.34%	99.58%	98.95%	5.81%	1.70%	98.87%	50.00%	0.42%	93.52%
Stacking	97.96%	98.34%	99.58%	98.95%	6.17%	1.53%	98.86%	50.00%	0.42%	93.51%
RandomForest	97.76%	98.33%	99.37%	98.85%	5.60%	1.55%	97.70%	50.00%	0.63%	93.21%
DecisionTree	98.37%	99.37%	98.95%	99.16%	58.73%	1.63%	90.10%	18.75%	1.05%	92.34%
MLP	97.76%	97.74%	100.00%	98.86%	8.56%	1.84%	91.76%	68.75%	0.00%	90.22%
KNN	96.95%	97.33%	99.58%	98.44%	21.52%	2.60%	89.68%	81.25%	0.42%	88.12%
SVM	96.74%	96.74%	100.00%	98.34%	10.55%	2.69%	93.86%	100.00%	0.00%	87.37%
Logistic	96.74%	96.74%	100.00%	98.34%	11.27%	2.84%	86.99%	100.00%	0.00%	86.00%
NaiveBayes	26.48%	100.00%	24.00%	38.71%	400.04%	68.63%	83.57%	0.00%	76.00%	50.51%



a) Feature Importance plot of LightGBM Model b) SHAP Plot of LightGBM Model

Figure 12: Feature importance and SHAP plots of LightGBM.

6.2 Limitations

Although SWTFPS is a successful prototype, it has some limitations for real-world use. First, there is a significant difference between our small prototype and massive industrial turbines; therefore, sensor patterns such as vibration and temperature may not perfectly match those of industrial systems. Second, the affordable sensors and ESP32 microcontrollers used are great for research but may not be sufficiently reliable or durable for the harsh environmental conditions of real wind farms.

7 Conclusion and Future Work

SWTFPS has successfully bridged the gap between industrial SCADA data and an IoT prototype. It demonstrates that strategic data enhancement and a balanced 7-feature model are highly effective in fault prediction. The key achievements are summarised as follows:

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

- **Integrated LLM:** Developed an LLM-centered interface that allows operators to communicate with the system that provides recommendations in natural language.
- **Real-Time Data Architecture:** Successfully established a live data collection that integrates traditional SCADA data with a wide array of new-generation sensor data, proving that systems can operate in full synchronization for real-time monitoring.
- **Optimised Performance via Hybrid AI:** Conducted a comprehensive comparison of multiple machine learning algorithms. By implementing Centroid-based K-Means balancing to eliminate data bias, the XGBoost model was optimized to achieve superior performance metrics: *Accuracy:* 89.68%, *F1-Score:* 90.08%, *ROC-AUC:* 95.65%, and *Overall Performance:* 65.13% while retrained model using the LightGBM algorithm, achieving 98.37% *Accuracy*, 99.16% *F1-Score*, 98.95% *ROC-AUC* and 94.65% *Overall Performance*.

To further develop the system, future work will focus on Advanced Diagnosis and Deep Learning for sub-categorization of specific mechanical faults; Edge Computing and Sensor Fusion for running models directly on offline hardware; and Digital Twin Integration for developing a 3D digital twin to provide a comprehensive, visual monitoring experience.

Code and data availability

The raw dataset analysed in this study was originally published by Kasimov (2024) and is publicly available at https://www.kaggle.com/datasets/azizkasimov/wind-turbine-scada-data-for-early-fault-detection. The source code used for data preprocessing, model training, and generating the figures presented in this manuscript is openly available in the GitHub repository at https://github.com/SaraMahyan/SWTFPS_Wind-Turbine-Failure-Prediction.

Author contributions

MHÖ and MSÜ contributed to the conceptualization and methodology of the study, supervised the research process, managed project administration, acquired funding, and reviewed and edited the manuscript. SM conducted the investigation, developed the software and hardware, provided resources, curated the data, and reviewed and edited the manuscript. İGS contributed to software development, conducted the investigation, curated the data, performed visualization and formal analysis, prepared the original draft of the manuscript, and reviewed and edited the manuscript. BA curated the data, performed formal analysis, contributed to software development and investigation, and reviewed and edited the manuscript.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Competing interests

The authors declare that they have no competing interests.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

The authors would like to express their sincere gratitude to Dokuz Eylül University, particularly the Department of Computer Engineering, for providing the necessary laboratory facilities, technical infrastructure, and academic support throughout this research. We are especially thankful to the faculty members and technical staff of the department for their valuable guidance, insightful suggestions, and continuous encouragement during the development of the prototype and experimental studies. We also extend our appreciation to Yaşar University for their academic contributions. This work would not have been possible without the collaborative research environment and resources provided by these institutions. The authors utilized VSCode Copilot (GPT-5.4 Mini architecture) to assist with code generation and adapting figures for color accessibility. Additionally, Gemini and Grammarly were used to refine the academic language and grammar of the manuscript. The authors thoroughly reviewed and take full responsibility for the final content.

Financial support

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Corley, B., Koukoura, S., Carroll, J., and McDonald, A.: Combination of thermal modelling and machine learning approaches for fault detection in wind turbine gearboxes, Energies, 14, 1375, https://doi.org/10.3390/en14051375, 2021.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

Khan, P. W., Yeun, C. Y., and Byun, Y. C.: Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning, Eng. Fail. Anal., 148, 107209, https://doi.org/10.1016/j.engfailanal.2023.107209, 2023.

Hsu, J. Y., Wang, Y. F., Lin, K. C., Chen, M. Y., and Hsu, J. H. Y.: Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning, IEEE Access, 8, 23427–23439, 2020.

Tan, Y. W. and Carroll, J.: LLMs in Wind Turbine Gearbox Failure Prediction, Energies, 18, 4659, https://doi.org/10.3390/en18174659, 2025.

Zhang, Y., Zhao, Z., Zhang, D., and Dai, J.: Optimisation of Large Language Models with Multi-modal Data Training for Fault Diagnosis in Wind Turbine Systems, in: Proceedings of the 2025 IEEE International Conference on Power Electronics and Energy (PEE), 172–176, https://doi.org/10.1109/peed63748.2025.00038, 2025.

Amin, A., Bibo, A., Panyam, M., and Tallapragada, P.: Vibration based fault diagnostics in a wind turbine planetary gearbox using machine learning, Wind Eng., 47, 175–189, https://doi.org/10.1177/0309524x221123968, 2022.

Alzawaideh, B., Baboli, P. T., Babazadeh, D., Horodyvskyy, S., Koprek, I., and Lehnhoff, S.: Wind turbine failure prediction model using scada-based condition monitoring system, in: 2021 IEEE Madrid PowerTech, 1–6, IEEE, 2021.

de Lima Munguba, C. F., Ochoa, A. A. V., Leite, G. D. N. P., da Costa, A. C. A., da Costa, J. Â. P., de Menezes, F. D., and de Souza, M. G. G.: Fault detection framework in wind turbine pitch systems using machine learning: Development, validation, and results, Eng. Appl. Artif. Intell., 138, 109307, 2024.

Inturi, V., Shreyas, N., Chetti, K., and Sabareesh, G.: Comprehensive fault diagnostics of wind turbine gearbox through adaptive condition monitoring scheme, Appl. Acoust., 174, 107738, https://doi.org/10.1016/j.apacoust.2020.107738, 2020.

Meyer, A.: Vibration fault diagnosis in wind turbines based on automated feature learning, Energies, 15, 1514, https://doi.org/10.3390/en15041514, 2022.

Ogaili, A. A. F., Hamzah, M. N., and Jaber, A. A.: Enhanced fault detection of wind turbine using extreme gradient boosting technique based on nonstationary vibration analysis, J. Fail. Anal. Prev., 24, 877–895, 2024.

Pham, D. A. and Han, S. H.: Advanced Machine Learning Model Based on Bi-LSTM and Attention Mechanism for Fault Detection in Wind Turbine Systems, J. Electr. Eng. Technol., 1–14, 2025.

Puruncajas, B., Castellani, F., Vidal, Y., and Tutivén, C.: Use of artificial neural networks and SCADA data for early detection of wind turbine gearbox failures, Machines, 13, 746, https://doi.org/10.3390/machines13080746, 2025.

Santiago, R. A. D. F., Barbosa, N. B., Mergulhão, H. G., Carvalho, T. F. D., Santos, A. A. B., Medrado, R. C., and Nascimento, E. G. S.: Data-driven models applied to predictive and prescriptive maintenance of wind turbine: A systematic review of approaches based on failure detection, diagnosis, and prognosis, Energies, 17, 1010, 2024.

Yu, X., Tang, B., and Zhang, K.: Fault diagnosis of wind turbine gearbox using a novel method of fast deep graph convolutional networks, IEEE Trans. Instrum. Meas., 70, 1–14, https://doi.org/10.1109/tim.2020.3048799, 2021.

Zeng, X., Yang, M., Feng, C., and Tang, Y.: A generalised wind turbine anomaly detection method based on combined probability estimation model, J. Mod. Power Syst. Clean Energy, 11, 1136–1148, 2022.

https://doi.org/10.5194/wes-2026-95
Preprint. Discussion started: 18 June 2026

Zhang, G., Li, Y., and Zhao, Y.: A novel fault diagnosis method for wind turbine based on adaptive multivariate time-series convolutional network using SCADA data, Adv. Eng. Inform., 57, 102031, 2023.

Zhang, Z. Y. and Wang, K. S.: Wind turbine fault detection based on SCADA data analysis using ANN, Adv. Manuf., 2, 70–78, 2014.

Kasimov, A.: Wind turbine SCADA data for early fault detection (Version 1) [data set], Kaggle, https://www.kaggle.com/datasets/azizkasimov/wind-turbine-scada-data-for-early-fault-detection, 2024.

MathWorks: ThingSpeak: An IoT analytics platform service, https://thingspeak.com, (accessed 29 May 2026).

Naidu, G., Zuva, T., and Sibanda, E. M.: A review of evaluation metrics in machine learning algorithms, in: Computer Science On-line Conference, 15–25, Springer, Cham, 2023.

Islam, M. M., Rifat, H. R., Shahid, M. S. B., Akhter, A., Uddin, M. A., and Uddin, K. M. M.: Explainable machine learning for efficient diabetes prediction using hyperparameter tuning, SHAP analysis, partial dependency, and LIME, Eng. Rep., 7, e13080, 2025.