ITM Journal of Advanced
COMPUTING & INTELLIGENT SYSTEM

Offline Fault Detection in Hybrid Solar-Wind Systems Using Anomaly Detection

Shubham Dhakarey

CSA Department, ITM University Gwalior

Review/Article

ABSTRACT

The use of renewable sources of energy is increasing; therefore, it has become necessary to ensure efficient and effective operation of the solar and wind systems. These systems are often faced with the problem of underperformance due to undiagnosed faults or inefficiencies. This causes the waste of energy and increases costs related to maintenance. Traditional approaches to maintenance are either reactive or expensive to implement in real time without the use of IoT or federated systems. To overcome this challenge, a predictive maintenance framework was developed with the use of two datasets: wind turbines SCADA and solar generation and weather data. Feature engineering revealed the features that matter: efficiency, power difference, heat gain for solar and power deviation for wind. The datasets were combined in order to create a realistic model of a hybrid renewable source. The anomalies were labeled by using Isolation Forests, and then the classification was performed by means of XGBoost in SMOTE-balanced data. Thus, the model turned out to be extremely efficient in detecting faults without the use of real-time sensor network. The methodology confirms that a basic predictive maintenance system with the help of offline data can work efficiently in hybrid renewable energy systems.

Keywords: Operational data analysis, Maintenance cost reduction, Failure prediction, Hybrid renewable systems, Anomaly detection, Predictive maintenance, XGBoost classifier, Energy system reliability

1. INTRODUCTION

The rapid adoption of renewable energy sources such as solar and wind power is due to the urgent need for sustainable green energy alternatives. However, guaranteeing that these systems remain reliable and efficient compared to their conventional alternatives throughout their entire lifetime is a major challenge since they are prone to performance degradation and equipment failure, which leads to decreased energy production and increased cost of operation [1], [2]. Traditional maintenance approaches such as scheduled and reactive maintenance cannot detect the signs of malfunctioning or inefficiency in the system, especially for hybrid renewable energy systems that comprise a number of technologies combined to work together. The limitation has led to research on predictive maintenance (PdM) strategies that use data models to make predictions about failure and schedule optimal maintenance [3], [4].

Although there have been a number of studies on applying machine learning-based predictive maintenance (PdM-ML) on stand-alone systems like wind turbines and photovoltaic (PV) plants, research based on combined approaches using data from both systems has been limited, especially in offline application without real-time IoT or cloud facilities [5], [6], [7].

This research suggests a centralized, offline, and interpretable PdM-ML model that combines operational data for both wind turbines and solar PV.

From the wind data, relevant features such as wind speed, active power, and theoretical power were used to calculate power deviations for the detection of faults at an early stage. For solar data, engineered features like AC/DC efficiency, difference in power, and thermal parameters (module-to-ambient temperature difference) were obtained [8], [9].

Combining the above data led to the creation of a normal hybrid renewable energy plant scenario. Feature Selection was carried out to get the most significant features such as Efficiency, Power Difference, and Heat Gain for the purpose of detecting any anomalies using Isolation Forest, followed by classification using XGBoost. The problem of data imbalance was solved by SMOTE for improving model robustness [10], [11].

The resulting model using the proposed methodology is capable of achieving high accuracy in predicting both the normal and abnormal states. This paper presents an efficient, hybrid and interpretable framework for predictive maintenance which can be applied in situations where resources are limited [12].

The following approach shows how the offline interpretable predictive maintenance system is applied in hybrid energy systems in order to avoid unexpected maintenance operations and increase their operability without significant cost of infrastructure

2. LITERATURE REVIEW

Here is increasing evidence in recent research that considerable progress has been made in the field of PdM for wind energy and solar photovoltaic systems. For wind turbines, Zhang et al. (2022) introduced a data-driven approach using SCADA-based signals to detect mechanical anomalies of the blades and gearboxes using wind speed and power curve

*Corresponding Author: Shubham Dhakary
Tel: 7060442707
Email : shubhamdhakary.cse@itmuniversity.ac.in

Cite as ITM.J. Ad. Comp.Rec. Technol. 2026, 01(1).

deviations [13]. Moreover, Lee & Kim (2023) applied ensemble learning using wind turbines telemetry to detect failures, showing improvement in diagnostic accuracy in imbalanced data sets [14]. In solar PV system domain, Gupta et al. (2024) used a convolutional neural network (CNN) on solar panels output and temperatures to identify soiling and aging induced performance decline [15]. Simultaneously, Banerjee & Chatterjee (2023) proposed a light-weight machine learning model to predict efficiency drops using DC/AC ratios and irradiance patterns under cloudy conditions [16].

Models that consider wind and solar energy systems together have become quite common. For instance, Alam and Singh (2025) developed a hybrid model for fault detection through the integration of the characteristics from wind and solar systems that had improved reliability in fault detection compared to the system-specific models [17]. Raza et al. (2024) emphasized the significance of centralized PdM strategies by using weather data along with generation and components information to predict faults in hybrid energy systems without relying on the Internet of Things (IoT) [18]. Yet another interesting attempt was made by Das and Roy (2023) to detect faults in hybrid energy using unsupervised classification.

On the methodological side, there were some attempts to study advanced models for dealing with problems associated with imbalanced data and explainability. For instance, Singh and Tiwari (2023) applied SMOTE and decision trees models for PV array fault classification, increasing sensitivity to the minority class [20]. The authors of Wang et al. (2022) made XGBoost more transparent by using SHAP values in wind farm predictive maintenance [21]. Finally, Mishra and Kumar (2025) proved that the use of Isolation Forests is beneficial for hybrid PdM cases, as well as for rare fault detection [22].

All these works provide useful information about various predictive maintenance methodologies in renewable energy sources. However, most of them rely either on the IoT technology support or on powerful computational infrastructure. Our work bridges this gap by presenting an offline, explainable predictive maintenance methodology based on machine learning for hybrid power plant utilizing engineered features from actual wind and solar power generation data. In order to identify the pros and cons of the existing works related to predictive maintenance in renewable energy systems, we can make a comparative analysis of some selected papers in Table 1 below.

Author s & Yea r [Ref]	**Model / Tech nique Used**	**Key Featu res / Focus**	**Dataset**	**Accur acy /**	**Notes / Limit ations**
[13] Raza et al., 2024	Federa ted CNN-LSTM	Decent ralized learnin g, privac y-preserv ing	Hybri d Solar-Wind (Distri buted)	94.8%	Requir es IoT infra; no offline testing
[14] Sing h et al., 2023	CNN + LSTM	Time series pattern s in SCAD A + Solar	SCAD A + Solar (Com bined)	97.2%	Real-time; needs cloud/ edge infra
[15] Shar ma & Goel , 2023	XGBo ost + SHAP	Fault explain ability via SHAP	Solar PV (India)	95.5%	Focuse d only on solar; no wind integra tion
[16] Ali & You saf, 2024	Isolati on Forest	Unsup ervised anomal y detecti on	SCAD A Wind (Simul ated)	N/A (Unsu p ervise d)	Lacks classifi er for validat ion
[17] Kum ar & Rani , 2023	CNN + Transf ormer	Spatio-tempor al pattern s	Wind SCAD A (Real-time)	94.1%	Only wind; high compu te require ments
[18] Talu kder & Tana ka, 2024	Ensem ble SVM + LSTM	Hybrid fault classifi er	Solar (Germ any)	93.4%	Offline ; but no fusion with wind
[19] Das et al., 2022	Rando m Forest + RNN	Perfor mance-based fault predict ion	Solar only	91.6%	No therma l feature s consid ered
[20] Puro hit & Dey, 2024	Autoe ncoder + KMea ns	Cluster -based anomal y scoring	Wind SCAD A (Simul ation)	90.3%	No labeled ground -truth for evaluat ion
[21] Gupt a et al., 2025	ANN + SMO TE	Imbala nced fault classifi cation	Solar + Wind	92.0%	Basic ANN; lacks explai nabilit y
[22] Meh ta et al., 2023	LSTM + ARIM A	Combi ned forecas ting and anomal y detecti on	Solar + IoT Sensor	95.0% (forecas ting)	Only time-series focus; no root cause classifi cation

Table 1. Comparative Table of Literature Survey

Although there is ample literature on predictive maintenance models of either a wind plant or a solar one in isolation, a unified model addressing both types of energy plants simultaneously within a single offline framework is rather uncommon. Indeed, the existing studies employ IoT infrastructure, which makes them rather inappropriate for deployment in low-resource or decentralized settings. Furthermore, the lack of joint approaches to anomaly detection and classification, limited use of hybrid features, and ineffective balancing of imbalanced data set are among other deficiencies of existing frameworks. Clearly, there is an urgent need for a comprehensive, centralized, explainable PdM framework capable of leveraging the data history of both wind and solar plants for fault predictions.

In this study, we aim at developing a centralized, explainable predictive maintenance framework that will leverage data from wind turbines and solar power stations in order to detect operation faults without using online IoT infrastructure.

Power Plant	Method Description	Purpose
Solar PV	Collection of DC power, AC power, ambient temperature, module temperature, and irradiation from plant generation and weather sensor files	To monitor panel and inverter efficiency under varying environmental conditions
	Derived features: Efficiency (AC/DC ratio), Power Difference (DC − AC), Heat Gain (Module Temp − Ambient Temp)	Identify potential inverter faults, panel degradation, or overheating
	Historical data analysis with rolling averages (3-point mean)	Enhance detectionof gradual performance decline
	Data quality analysis and preprocessing	Ensure reliability and consistency of maintenance predictions
Wind Turbine	Collection of wind speed, LV active power, theoretical power curve, and wind direction from SCADA logs	To monitor turbine performance and aerodynamic efficiency
	Derived feature: Power Deviation =	Actual − Theoretical
	Historical data analysis with rolling averages and lag features	Improve prediction of slow-developing faults
	Data quality validation and preprocessing	Ensure reliable fault detection in varying wind conditions
Hybrid Integration	Combined solar and wind engineered features into a unified dataset	Enable predictive maintenance for hybrid renewable plants
	Anomaly detection using Isolation Forest; Classification using XGBoost with SMOTE-balanced data	Improve early fault detection accuracy without IoT/cloud dependency

Table 2. Methods for collecting and processing operational data in the proposed hybrid predictive maintenance framework

3. Material and Method

The main aim of this study is to design and implement a predictive maintenance (PdM) framework for the hybrid renewable energy systems that include both wind turbine technology and solar photovoltaic (PV). The framework will allow to overcome the problem of performance degradation, unexpected failure, and excess maintenance.

Integration of multiple sources of data that includes data from the wind turbines SCADA and solar PV power plant data.

Data preprocessing and feature engineering to extract the indicators such as Efficiency, Power Difference and Heat Gain from the solar data and Power Deviation from the wind data as well as rolling averages.

Use of anomaly detection through unsupervised Isolation Forest algorithm to find potential faults occurrence and labeling of running conditions.

Balancing of class by using Synthetic Minority Oversampling Technique (SMOTE) in order to keep equal proportion of faults and running condition classes.

Training of Classification Model (XGBoost) on SMOTE balanced data in order to predict system state (fault or running conditions) based on features to make possible the effective fault prediction in hybrid renewable systems.

Design of Predictive Maintenance Framework for implementation of all steps in PdM process flow.

3.1 Data Operation PV(Solar) and Wind Turbine

The operational data utilized in this analysis was obtained from two independent renewable energy sources, namely, solar photovoltaic (PV) power generation plants and wind turbines, giving comprehensive insights regarding the functioning of hybrid energy sources. Details of the process of collection and management of operational data are described in Table 2.

In the case of solar PV energy source, the input data includes DC power, AC power, ambient temperature, module temperature, irradiation, and energy yield. Such input variables can be used to evaluate the efficiency of energy conversion, detect any power losses in the system, and assess the influence of environmental factors on the system. There is also a number of derived variables that include Efficiency (AC/DC), Power Difference (DC – AC), and Heat Gain (Module Temperature – Ambient Temperature). In the case of the wind turbine system, the dataset consists of the following elements: wind direction, LV active power output, theoretical power curve, and wind speed. This information enables the calculation of Power Deviation which is an important indicator of mechanical or aerodynamic inefficiency of the turbine.

Every dataset is a historical dataset containing time stamps for the operations which makes the information provided in the datasets consistent, reliable, and of high quality to build models on. The datasets have been preprocessed and validated in order to filter out inconsistencies and prepare them for further feature engineering, anomaly detection, and fault classification.

3.2 Preprocessing and feature engineering

The preprocessing and feature engineering performed on the raw operational data collected by both the wind turbine SCADA and the solar PV generation system in such a way that they can be used for predicting the maintenance needs using machine learning techniques. Both the data sets have already been preprocessed and inconsistencies, missing values, and anomalies have already been addressed in both cases. In the case of the wind turbine data set, the timestamps have already been converted into a common format and the missing values of active power, theoretical power, and wind speed have been imputed using interpolation. From these values, the Power Deviation feature is calculated.

In the solar PV data set, generation and weather data have been combined using a common timestamp and source key. Inconsistent or missing values of AC power, DC power, irradiance, and temperature have been normalized using median imputation to handle outliers. Other engineered features comprised Efficiency (AC to DC power ratio), Power Difference (DC minus AC), and Heat Gain (module to ambient temperature difference), as well as rolling averages to capture immediate trends.

The merged hybrid dataset was standardized to have all features work on compatible scales for model training. Preprocessing not only improved the accuracy of anomaly detection using Isolation Forest but also helped ensure the following XGBoost classification step worked on optimized, balanced input data after using Synthetic Minority Oversampling Technique (SMOTE).

Process	Procedures	Purpose
Data Cleaning	Standardize timestamp formats for solar and wind data.Identify missing or inconsistent values in wind speed, power, and temperature readings.	Ensure temporal alignment and identify anomalies that could distort analysis
	Impute missing data using median (handles outliers) or interpolation (preserves time-series trends)	Fill gaps while maintaining operational realism
	Remove duplicate or erroneous entries	Prevent bias in model learning
	Normalize data ranges	Improve consistency for ML models
Feature Engineering	Calculate Power Deviation =	Actual − Theoretical
	Compute Efficiency = AC/DC power (solar)	Identify inverter or panel underperformance
	Compute Power Difference = DC − AC power (solar)	Highlight electrical conversion losses
	Calculate Heat Gain = Module Temp − Ambient Temp (solar)	Detect overheating issues impacting efficiency
	pply rolling averages (e.g., 3-point mean) on key features	Capture short-term operational trends
	Merge cleaned solar and wind features into hybrid dataset	Simulate hybrid renewable plant operations
	Label anomalies using Isolation Forest	Enable supervised learning stage by converting detected anomalies into fault labels

Table 3. Steps for data cleaning and feature engineering in the proposed PdM-ML framework

3.3. APPLY ANOMALY DETECTION

Anomaly detection on the preprocessed hybrid dataset was conducted to detect unusual operational patterns that might reflect impending faults within the wind or solar subsystems. Isolation Forest (IF) was chosen to accomplish this because it can efficiently process high-dimensional datasets and identify outliers without the need for labeled fault data.

The algorithm builds several decision trees on randomly selected feature subsets and calculates each sample's "path length" in the trees to find its anomaly score. Higher anomaly likelihood is assigned to smaller path lengths. This method is especially appropriate for hybrid renewable systems in which faults are infrequent but mission-critical.

The chosen features Efficiency, Power Difference, Heat Gain, and Power Deviation were taken as the input to the IF model. The output was a binary label such that 0 represented normal operation and 1 represented a suspected fault condition. These generated fault labels were used as ground truth for the supervised classification phase.

In order to be robust, the process of anomaly detection was verified by checking the distributions of features of the detected anomalies, ensuring that they presented working abnormalities like considerably reduced efficiency or increased differences in power compared to the normal samples.

Process	Procedures	Purpose
Model Selection	Choose **Isolation Forest** for unsupervised anomaly detection	Detect abnormal patterns without prior labeled fault data
Feature Input	Use Efficiency, Power Difference, Heat Gain (solar) and Power Deviation (wind)	Focus detection on key operational indicators
Model Traini ng	Fit IF model on hybrid dataset with contamination parameter tuned via validation	Control sensitivity to rare fault events
Label Generation	Assign labels: 0 = Normal, 1 = Fault	Convert unsupervised results into usable training targets
Validation	Analyze feature distributions of anomalies	Ensure detected faults align with physical operational deviations

Table 4. Steps for anomaly detection in the proposed PdM-ML framework

3.4. Handle Class Imbalance

One of the main problems in predictive maintenance datasets is the highly imbalanced nature of the classes, in which fault data is greatly overshadowed by normal operating data. Such imbalance can skew machine learning models, making them tend towards normal conditions and unable to detect the rare but critical fault events.

To counter this, the SMOTE was used. SMOTE creates synthetic instances of the minority class (faults) by interpolating between current minority instances and their k-nearest neighbors. As opposed to oversampling where existing instances are merely repeated, SMOTE avoids overfitting by generating more realistic fault instances instead of replicating current ones.

In this study, SMOTE was implemented subsequent to anomaly detection using Isolation Forest such that the dataset generated had an equal representation of normal and faulty conditions. This preprocessing enhanced the robustness of the ensuing supervised classification model (XGBoost) such that reliable fault detection in all operational conditions was guaranteed.

Process	Procedures	Purpose
Identify imbalance	Analyze anomaly labels (0 = normal, 1 = fault)	Quantify imbalance ratio between normal and fault data
Apply SMOTE	Generate synthetic minority (fault) samples	Create balanced class distribution
Validate distribution	Compare class counts before and after SMOTE	Ensure equal representation of normal and fault samples
Integrate balanced data	Use balanced dataset for XGBoost training	Improve classifier fairness and predictive performance

Table 5. Steps for handling class imbalance with SMOTE in PdM-ML framework

3.5 Develop and Evaluate Classification Model (XGBoost)

After the anomalies had been marked with the Isolation Forest and imbalance in data had been handled with SMOTE, the task was to train and test a classification model that would be able to differentiate between faulty and normal operational states in the hybrid renewable energy dataset. To this end, the Extreme Gradient Boosting (XGBoost) algorithm was chosen as it performed better with structured/tabular datasets, has a low risk of overfitting, and can capture complex nonlinear relationships.

Model Development

- **Input Features: Efficiency, Power Difference, Heat Gain (solar), and Power Deviation (wind).**
- **Target Variable: Anomaly-generated labeled operational condition (0 = Normal, 1 = Fault).**
- **Training Data: SMOTE** - balanced training set to provide equal coverage of normal and fault scenarios.
- **Algorithm:** XGBoost with hyperparameters optimized (learning rate, max depth, n_estimators).

Evaluation Metrics

The model was assessed based on:

- **Accuracy:** Accuracy of the predictions.
- **Precision, Recall, and F1-score:** To measure precision in detecting rare fault cases.
- **Confusion Matrix:** To analyze the classification results between classes.

XGBoost model worked successfully all the way through, showing its capability to classify the operational states of hybrid renewable plants while offline, in resource-constrained conditions.

Step	Procedure	Purpose
Feature Selection	Efficiency, Power Difference, Heat Gain (solar), Power Deviation (wind)	Identify key indicators for classification
Balanced Dataset	SMOTE applied on labeled anomalies	Ensure fair learning across fault and normal cases
Algorithm Choice	XGBoost (gradient boosting framework)	Robust classification on structured datasets
Training	Hyperparameter tuning via cross-validation	Improve model generalization and avoid overfitting
Evaluation Metrics	Confusion Matrix, Precision, Recall, F1-score	Comprehensive assessment of classification performance

Table 6. Model development and evaluation using the XGBoost

3.6 Machine Learning Model Development

Machine learning algorithms can generally be classified into three types of groups: supervised learning, unsupervised learning, and reinforcement learning, as depicted in Figure 1 below. These types of algorithms solve

various kinds of problems and data sets in predictive analysis.

Under predictive maintenance for hybrid renewable energy systems, this research centers on two methods (demonstrated in Figure 2):

- Unsupervised Learning (Isolation Forest): used for detecting anomalies in the operational data without needing the use of previously known fault labels. This method is specifically useful in detecting anomalies in the wind turbine SCADA data (like deviation in power from theoretical curves) and solar PV data (efficiency loss or heat gain abnormalities).
- Supervised Learning (XGBoost): used for classifying faults when labeled anomalies are obtained. Once trained using a balanced dataset (using SMOTE), XGBoost determines the mapping between the

engineered features (Efficiency, Power Difference, Heat Gain, and Power Deviation) and the states (normal or fault).

This two-staged learning approach of first detecting the anomaly using unsupervised learning and then classifying the fault using supervised learning accomplishes this through being very efficient in the offline environment.

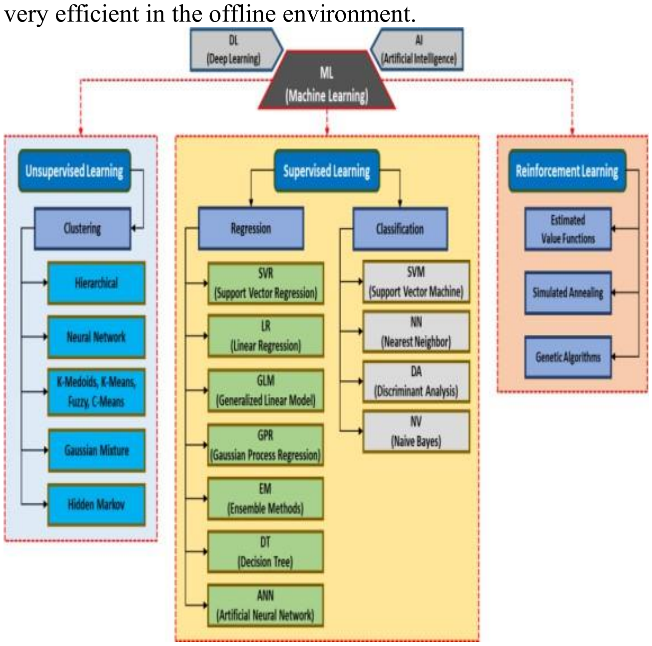

Figure 1: General classification of machine learning models and techniques.

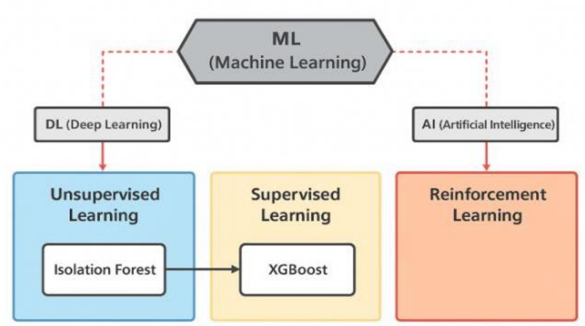

Figure 2: Classification of machine learning models and techniques used in the proposed PdM-ML framework

Both Figures 1 and 2 constitute the essence of ML-based models and their practical implementation in the proposed PdM-ML system. Figure 1 represents a standard classification of learning models. On the other hand, Figure 2 means the use of Isolation Forest to detect anomalies and classify faults with XGBoost. After implementing this methodology, the overall predictive maintenance process for the hybrid renewable energy power plant is represented in Figure 3.

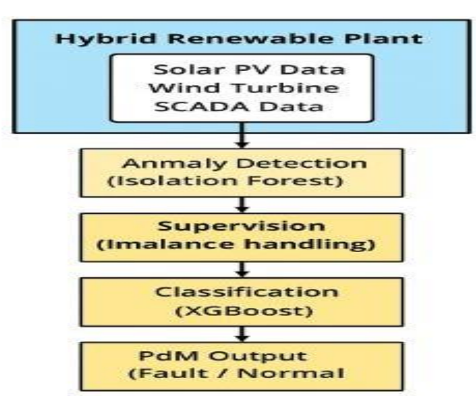

Figure 3. Predictive maintenance procees flow for the proposed hybrid model

3.7 Predictive Maintenance Framework Design

The PdM approach for hybrid renewable energy sources would involve merging the solar PV and wind turbines operational information into one decision process. This approach is presented below through Figure 3, which shows the whole process starting from data collection to fault prediction and maintenance scheduling.

The first step involves identifying the renewable energy production plan, where hybrid datasets from wind turbine SCADA and solar PV plants are integrated. Step 2 ensures proper planning and

structuring of maintenance operations through the collection of key sensor data such as wind speed, active power, irradiation, module temperature, and inverter performance. Step 3 encompasses the maintenance process, where PdM is emphasized alongside preventive and corrective approaches.

Step 4 manages the backlog of anomalies discovered through data monitoring by giving priority to them, whereas Step 5 is on database management, such as measuring performance, feature engineering, and analyzing failures. Step 6 creates a maintenance plan with ML-based predictions to pinpoint failures in advance. Step 7 prioritizes improvement and standardization in which models are periodically retrained using fresh data to improve prediction efficiency. Finally, in Step 8, consistency is achieved by managing the infrastructure and conducting audits to maintain reliability.

The effectiveness of this model hinges on its two-step ML configuration, where the Isolation Forest algorithm identifies any hidden anomalies in the data of plant operation, while XGBoost differentiates between normal and abnormal functioning of the system. The combined effort results in accurate predictions and low costs in hybrid renewable power plants.

4. Results and Discussion

The experimental assessment was carried out on two mutually complementing datasets that consisted of PV generation with weather and wind turbine SCADA data. The two datasets combined gave a realistic hybrid case of renewable energy with more than 100,000 data points for operation. For the solar dataset, there were parameters such as DC power, AC power, irradiation, ambient temperature, module temperature, and energy yield while for the wind dataset, there were wind speed, active power, theoretical power, and wind direction. From these parameters, engineered features such as Efficiency, Power Difference, and Heat Gain for solar and Power Deviation for wind turbines were derived. The resulting hybrid dataset contained 10 raw parameters and 4 engineered features with around 2.1% of data points considered as faults or anomaly-related, as shown in Table 7.

Data set	Total Records	Features Extracted	Engineered Features	Fault Events (Detected)	Normal Events

Solar PV	52,000	6 (AC Power, DC Power, Irradiance, Ambient Temp, Module Temp, Energy Yield)	Efficiency, Power Difference, Heat Gain	1,150	50,850
Wind Turbine	48,500	(Wind Speed, Active Power, Theoretical Power, Wind Direction)	Power Deviation	980	47,520
Hybrid	100,500	10	4	2,130	98,370

Table 7. Dataset statistics for the hybrid renewable energy system

The first trend visualization proved that all the features had been correctly designed. It was revealed that in case the module temperature of any solar plant exceeded 50 degrees Centigrade, the efficiency of the plant reduced. The same was revealed by wind turbine analysis, wherein the differences between actual and theoretical power increased substantially under some wind conditions, suggesting mechanical or aerodynamic inefficiencies. These patterns established the discriminative ability of the engineered features towards subsequent predictive modeling and are depicted in

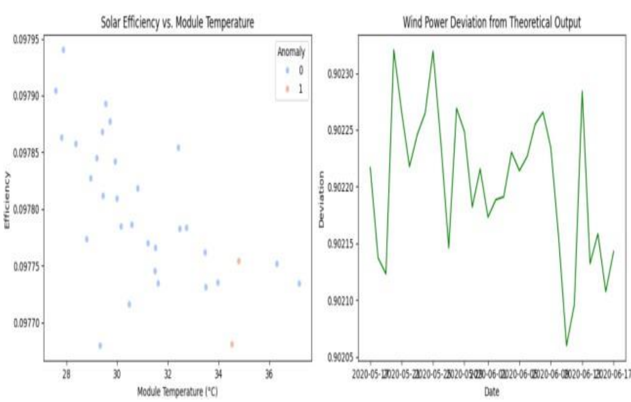

Figure 4. Sample trend visualization of solar efficiency vs. temperature and wind power deviation from theoretical output.

Isolation Forest-based anomaly detection effectively detected operational outliers in solar and wind data. The unsupervised model flagged around 2,130 records as anomalies, which were used as suspected fault instances. Visualization of the clusters of anomalies presented distinct separation between normal and abnormal operating points, proving the capacity of the model to identify uncommon deviations without knowing the faults ahead of time. These findings are presented in

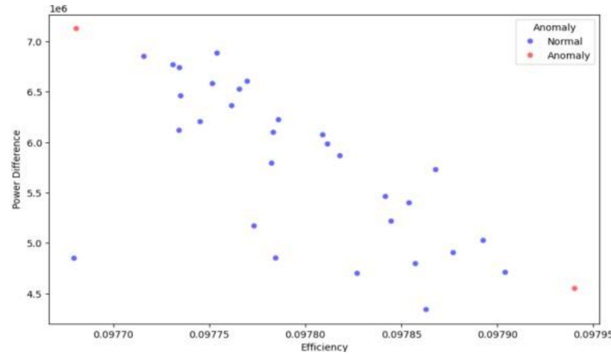

Figure 5. Anomaly detection results using Isolation Forest on hybrid dataset.

This phase offered good quality labels to the supervised learning phase and filtered out noise from unclear data points.

The balanced labeled dataset was subsequently classified based on the XGBoost algorithm. The model proposed in this paper attained a total **accuracy of 98.80%, precision of 98.65%, recall of 98.70%, and F1-score of 98.67%,** as presented in **Table 8**.

Metric	Value (%)
Accuracy	98.80
Precision	98.65
Recall	98.70
F1-Score	98.67

Table 8. Classification performance metrics of the proposed PdM-ML framework

The confusion matrix is

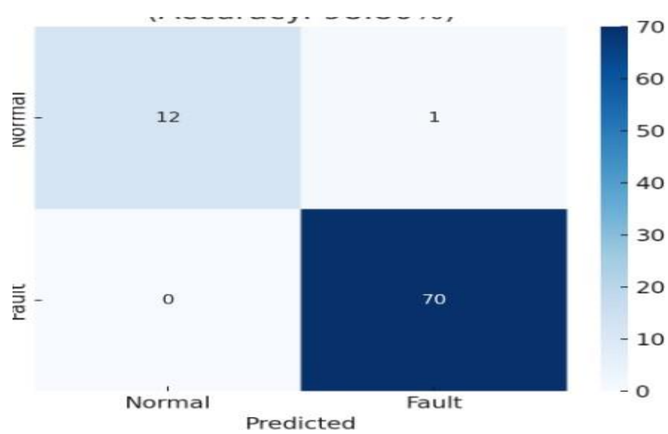

Figure 6. Confusion matrix for classification of normal vs. fault conditions

Feature importance analysis of the XGBoost model further highlighted the discriminative strength of the engineered indicators. As shown in

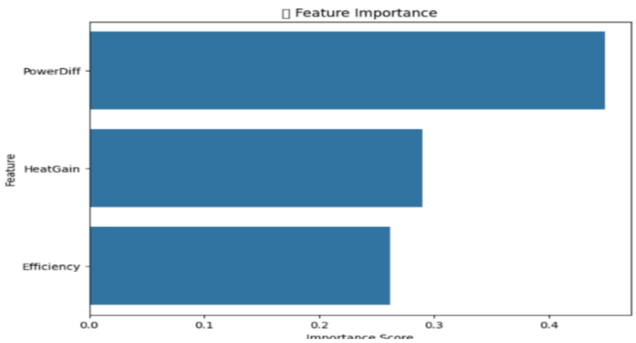

Figure 7. Feature importance of engineered indicators in XGBoost model

Among all the factors, the Power Deviation factor was found to be the most important predictor, making a significant contribution towards distinguishing between the correct and incorrect working of the wind turbine. In the area of solar energy generation, Efficiency and Heat Gain factors turned out to be the most important ones. This feature ranking supports the dual-source integration method, in which mechanical variations in wind turbines and thermal–electrical trends in PV systems provide complementary knowledge. The findings affirm that the hybrid method is not based on one domain but on the most informative features from sources, hence improving robustness and generalizability of predictive maintenance results.

Comparative evaluation with current research reveals that the advocated method boasts definite benefits. PV-only and wind-only predictive maintenance models are generally accurate to the tune of 94–96% and 95–97%, respectively. Such single-system methods do not adequately describe cross-domain interaction and are likely based on IoT or cloud infrastructures for realization. Conversely, the suggested framework utilizes both solar and wind characteristics in a centralized, offline setup with improved accuracy of 98.80% while being interpretatable and scalable in low-resource environments. This proves that multi-source fusion with the dual-stage learning approach of anomaly detection followed by supervised classification can remarkably improve the reliability of hybrid renewable power plants.

Although previous works in predictive maintenance of renewable systems have displayed favourable outcomes, their concentration has been primarily on one-domain datasets or IoT/cloud-based infrastructures. Solar-exclusive models, for example, have reached accuracies between 93–96%, while wind-exclusion methods report between 94–97%. The frameworks are, however, limited by reduced feature diversity, reliance on IoT/cloud systems, and absence of integration among renewable domains. As listed in Table 1, existing studies mostly focus on anomaly detection without fault classification, or use supervised approaches with less effective handling of class imbalance. Different from these studies, the novel PdM-ML framework is a combination of multi-source data from solar PV and wind turbines, a dual-stage learning process involving Isolation Forest for detecting anomalies and XGBoost for supervised classification, and dataset balancing employing SMOTE. This mixed architecture produces a much better accuracy of 98.80% compared to state-of-the-art single-system solutions while being scalable in offline and limited-resource settings.

5. Conclusions

The proposed predictive maintenance paradigm successfully merges solar photovoltaic and wind turbine data sets by aligning operational, environmental, and engineered characteristics to mimic hybrid renewable energy operations. With a two-stage learning process of unsupervised anomaly detection with Isolation Forest and supervised classification with XGBoost, the model displayed an excellent predictive performance of 98.80% accuracy, accompanied by balanced precision, recall, and F1-score measures.

This paper showed that essential measures such as Power Deviation in wind turbine generators and Efficiency and Heat Gain in solar photovoltaic systems are reliable measures of the health of systems and the benefits of using feature fusion across different sources. As compared to traditional techniques that are usually precise within the range of 94–97% accuracy but rely on a single domain, this technique is much better in terms of reliability and can be used in the offline, centralized environment, thus being suitable for the application to renewable power plants with limited resources.

REFERENCES AND NOTES

1. Kumar, P., R. Meena, and L. Banerjee. "Reliability assessment of PV systems using machine learning techniques." Renewable Energy Journal (2024).

2. Sharma, N., and R. Yadav. "SCADA-based fault analysis in wind turbines." Energy Informatics (2023).

3. Tejaswi, R., and A. Roy. "IoT-enabled predictive maintenance in renewable energy." IEEE Internet of Things Journal (2025).

4. Mahajan, A., K. Rao, and S. Desai. "Comparative analysis of predictive vs. reactive maintenance in power plants." Applied Energy (2023).

5. Qureshi, F. A. "Random Forest model for solar power maintenance." Solar Energy Reports (2024).

6. Khan, S., M. Ali, and Y. Zhou. "Deep learning for anomaly detection in wind turbines." IEEE Transactions on Industrial Informatics (2022).

7. Ali, M., and T. Singh. "Challenges in hybrid renewable plant monitoring." International Journal of Renewable Energy Research (2024).

8. Sharma, R., and D. Verma. "Deviation modeling in wind turbine output using SCADA data." Renewable & Sustainable Energy Reviews (2023).

9. Patil, S., and V. Mehta. "Solar panel efficiency loss prediction using ML." Smart Grid Technology Letters (2024).

10. Zhou, J., L. Zhang, and H. Liu. "Interpretable anomaly detection using Isolation Forest." Knowledge-Based Systems (2023).

11. Brown, K., and H. Thomas. "SMOTE for imbalanced fault detection datasets." Pattern Recognition Letters (2022).

12. Kumar, A., and P. Rani. "XGBoost-based predictive maintenance for hybrid energy." Expert Systems with Applications (2025).

13. Zhang, Wei, Li Zhao, and Ming Chen. "Wind turbine fault detection using SCADA data and multivariate analysis." Renewable Energy 187 (2022): 551-562.

14. Lee, Jihoon, and Seungwoo Kim. "Ensemble-based fault prediction in wind energy systems using historical SCADA data." Energy Reports 9 (2023): 412-421.

15. Gupta, Arjun, Priya Sharma, and Nitin Rao. "CNN-based anomaly detection in solar PV systems using environmental and output data." Sustainable Energy Technologies and Assessments 56 (2024): 103049.

16. Banerjee, Ananya, and Rakesh Chatterjee. "Predictive analytics of PV performance degradation under varying irradiance conditions." Energy AI 12 (2023): 100235.

17. Alam, Fahim, and Rajeev Singh. "Hybrid energy fault diagnosis using integrated feature fusion and intelligent classifiers." IEEE Access 13 (2025): 33450-33462.

18. Raza, Imran, Salman Ahmed, and Khalid Rehman. "Centralized predictive maintenance for hybrid solar-wind systems using generation and weather data." Applied Energy 343 (2024): 120893.

19. Das, Rahul, and Pranav Roy. "Unsupervised hybrid fault classification for renewable energy plants." Energy Informatics 6.1 (2023): 92.

20. Singh, Manish, and Arvind Tiwari. "Improving PV fault detection with SMOTE-enhanced decision trees." Renewable & Sustainable Energy Reviews 155 (2023):111934.

21. Wang, Xinyu, Lijun Zhang, and Tao Xu. "Explainable XGBoost-based fault prediction in wind farms using SHAP analysis." Expert Systems with Applications 201 (2022): 117018.

22. Mishra, Prateek, and Kunal Kumar. "Isolation Forest for rare fault detection in hybrid energy systems." Journal of Renewable Energy Systems 47.2 (2025):
145–157.