Journal of Global Research in Electronics and Communication
Volume 2, No. 06, June 2026
Available Online at: www.jgrec.info
RESEARCH PAPER

Machine Learning- Based Condition Monitoring of Wind Turbines Using SCADA Dataset for Early Fault Detection

Aravindh Balan
Freelance Post-doctoral scholar
Project Manager
Inline hydraulics GmbH, Germany
draravindhbalan@gmail.com

***Abstract*—There has been a steady increase in the amount of wind power installed worldwide as a result of global efforts to replace fossil fuels and lower the average global temperature. The LCO of wind energy includes, among other things, the costs of operating and maintaining wind farms. Using machine learning and SCADA data, this research proposes a framework for early defect identification in wind turbines. The approach is designed for use with wind turbines. The proposed methodology contains data preprocessing methods like data cleaning, data normalization, feature selection, label encoding and SMOTE for the class imbalance issue. Processed data is then classified into fault and non-fault conditions using XGBoost and LightGBM (LGBM) models to effectively classify the data. The experimental outcomes showed that the proposed models have high fault detection performance, with accuracy of 95.2% for XGBoost, and the highest accuracy of 95.6% for LGBM, which has better values of precision, recall, and F1 score. The performance of the proposed LGBM is related to other machine learning models like NB, KNN and SVM by obtaining better classification performance with low misclassification rates. This framework could greatly aid predictive maintenance, keep turbines up and running, cut down on maintenance expenses, and improve the operational stability of wind power systems.**

***Keywords*—SCADA Dataset, Wind Turbine, Machine Learning, Early Fault Detection, Predictive Maintenance, Fault Classification, Renewable Energy Systems.**

I. INTRODUCTION

Wind power is most promising renewable energy sources for producing large quantities of electricity, has attracted global interest due to the growing depletion of petrochemical energy. Nevertheless, most wind turbine installations are situated in remote areas like mountains and deserts. These places are notorious for their harsh weather conditions and complicated geological environment, which can lead to frequent malfunctions and unforeseen shutdowns [1][2]. The wind farms are typically situated in areas of difficult weather conditions, resulting in complex operating conditions and a higher risk of unanticipated faults and downtime [3]. Over past ten years, researchers have examined a variety of condition monitoring techniques, with a large number of them focusing on early problem detection through the use of SCADA data.

One of the most frequently used techniques for detecting component failure is used to classify outliers or other abnormal patterns in data [4]. The SCADA system's data, vibration data, failure logs, and infrequently status and maintenance logs [5][6]. The main goal of this study is to provide a system for early defect identification with validated SCADA data and supplementary failure facts [7].

First, SCADA data can be utilized to track the turbine's health and performance in a similar fashion as other tracking techniques [8]. However, this data can be analyzed in a way that identifies problems that might not be picked up by any one of the individual sensors [9][10]. Furthermore, the SCADA system can provide up-to-the-minute information on various turbine parameters, such as generator output, ambient temperature, rotor speed and blade pitch angle. This enables the operator to promptly identify and address any possible issues with the turbine's performance, ultimately leading to optimal operation and maximum efficiency [11][12]. Finally, SCADA data can save time and money by decreasing the requirement for manual inspection and processing of sensor data. To enhance performance of fault identify contributing variables, SCADA data can be processed using a variety of approaches, such as control charts and data mining.

An efficient quality monitoring system must include wind turbine condition monitoring as one of its primary functions. It is possible to automate it using ML approaches. In order to identify distinct wind turbine problems, researchers have used a variety of ML methods, including deep learning networks, convolutional neural networks, decision tree algorithms, and random forests [13]. The multiple linear regression model (MLRM) is a useful for predicting generator temperatures and wind turbine temperatures in general by examining the association between the temperatures observed at each time step.

A. *Motivation and contribution*

Wind turbines operate in complex environments, which may cause unexpected faults and lower output. Maintenance practices are expensive, and have historically not been able to detect problems early enough. Machine learning techniques can be applied well to intelligent fault detection and predictive maintenance, as SCADA systems continuously produce huge volume of operational data. The goal of this effort is to improve operational reliability, decrease downtime, and increase wind turbine efficiency by developing a trustworthy framework for condition monitoring. Detailed below are the main contributions.

- Proposed a wind turbine condition monitoring and early fault identification system with intelligent machine learning approach on the wind turbine SCADA dataset.

© JGREC 2026, All Rights Reserved

- Implements various pre-processing methods e.g. data cleaning, data normalization, feature selection, label encoding and class balancing using SMOTE.
- Trained and tested XGBoost and LightGBM (LGBM) models with high accuracy for fault and non-fault classification.
- Performed comprehensive evaluation using confusion matrix, ROC curve, and accuracy, metrics to validate the reliability of the proposed framework.
- Showed potential of the suggested framework for predictive maintenance, decreasing downtime and increasing the operational reliability of wind turbines.

The literature review is given in Section II, planned methodology is defined in Section III, experimental outcomes and comparative analysis is discussed in Section IV and Section V summarizes paper and suggests directions for future work.

II. Literature Review

Machine learning and deep learning methods are effective for identifying wind turbine faults using SCADA data, according to a recent study.

A. Benabdesselam et al. (2026) suggested to aggregate IeI values over time, we define Cumulative Event Intensity, an integral of the event magnitude. This approach is tested on 207 wind turbines installed in 20 wind farms, and belonging to three different manufacturers, component types, and attaining 87.25% of the temperature threshold exceedance events 15 days ahead of time with a false positive rate of less than 21%[13]. In 2025, D. Nayak et al. presented a novel ML framework for spatiotemporal Graph Neural Network (GNN) to model the topological and temporal dependencies in wind energy systems. It combines the graph attention mechanisms with gated recurrent units to improve the detection of faults, Remaining Useful Life (RUL) forecasting and identification of anomalies in a multi-task learning manner, as opposed to the conventional models. The methodology is validated with the CARE dataset consisting of 89 years of SCADA data from 36 turbines, where the fault detection accuracy is 94.2%, the MAPE for RUL is 6.4% and the maximum lead time predictions for fault detection is 8.4 weeks [14].

L. Natrayan (2025) proposed a hybrid model combining the spatio-temporal fault analysis using Comvest networks with an Explainable AI (XAI) technique using SHAP for the identification of the root cause of the fault. The model was tested against a wind turbine dataset from Kaggle based on the SCADA, which gave the model a 94.18% accuracy, 92% F1-score and AUC of 0.972. More importantly, the interpretability of this SHAP-based approach is shown to have a coverage of 93.4%, which allowed the most important constraints (rotor speed and rotor pitch angle) to be easily identified as the most critical parameters to be used as the fault indicators [15].

J. Maldonado-Correa et al. (2024) created two models with Deep Learning (DL), inspired by the Transformer, to forecast IGBT module failures in an onshore wind farm located in Ecuador. With a high degree of accuracy and an approximation of 4.25 months before to failure occurrence, data demonstrate that two suggested models function admirably [16]. F. Shaheen and M. M. M. Al-Khalidy (2024) suggested an ML model that uses SCADA data to detect wind turbine problems early. Tested models consisting Logistic Regression, SVM, NB, Random Forest, Decision Trees, MLP, and LightGBM. LightGBM performed best with an average accuracy of 78.59% in complicated and imbalanced datasets [17].

Similarly, S. R. Mohapatro et al. (2024) employed early fault detection as a vital element to reduce downtime for achieving this goal. Using ML algorithms, SCADA data may be utilized for condition monitoring and issue identification. A framework for defect prediction with fewer features is developed using support vector machines (SVMs) and hyperparameter optimization; resulting F1-scores range from 62% to 94% and accuracy is among 73% and 98%. Furthermore, costs are further reduced by reduced sensor counts and soft sensing methods that estimate accuracy in the range of 59% to 84% and f1scores between 54% and 80%. Then, Artificial Neural Networking (ANN) is introduced, which further achieves the accuracy point of 91% and the f1-scores of 88% [18].

III. Methodology

Figure 1 shows the steps involved in the suggested methodology for defect identification in wind turbines using the SCADA dataset. These steps include collecting data, preparing it, training the model, and assessing its performance. Initially, the SCADA data are pre-processed using data cleaning, normalization, feature selection, label encoding, and SMOTE-based class balancing to improve data quality and reduce imbalance. The next step is to train XGBoost and LightGBM (LGBM) ML models to distinguish between fault and non-fault states. Lastly, proposed models are assessed for their effectiveness in early failure detection and wind turbine condition monitoring using the following metrics: recall, accuracy, ROC curve, and confusion matrix.

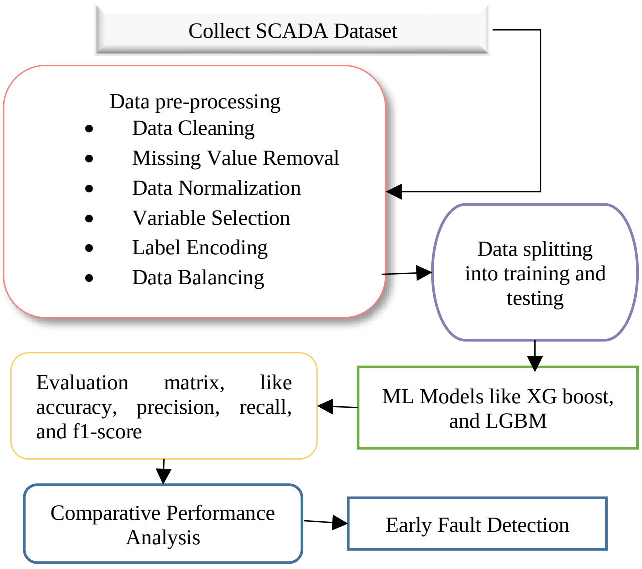

Fig. 1. Proposed Methodology for Wind Turbine Fault Detection

The steps involved in the proposed methodology are briefly explained as follows.

A. Data Collection

The SCADA data collected from a wind farm in Inner Mongolia, China, is applied in this study. The wind farm is equipped with 1.5 MW rated power variable speed constant

© JGREC 2026, All Rights Reserved

frequency turbines. The SCADA data is sampled every 30 seconds. There are twenty-five separate pieces of data included in each record, including the current condition of the turbine, the time stamp, the yaw state, and so on. Additionally, 49 continuous parameters. You can access the SCADA data for most of the turbines between July 1st and September 23rd, 2014. From July 1st, 2014, to August 31st, 2014, this article examines the SCADA data from thirteen accessible turbines.

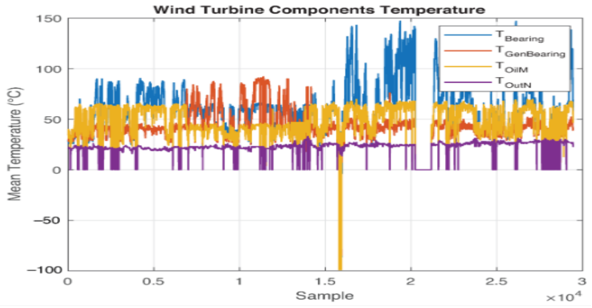

Fig. 2. Wind Turbine Components Temperature Analysis

Figure 2 depicts the temperature difference of various wind turbine components from the SCADA system such as the bearing, generator bearing, oil inlet and outdoor temperature. The graph displays temperature changes of the components in various samples, and any sudden changes or peaks that do not adhere to the normal pattern could suggest operational problems or fault conditions at an early stage. These temperature variations are crucial to detect anomalies, enhance condition monitoring, and to facilitate predictive maintenance in wind turbine systems.

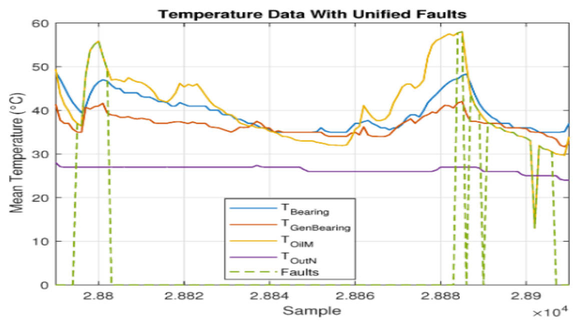

Fig. 3. Temperature Data with Unified Faults

The unified fault occurrences and wind turbine component temperatures are plotted against each other in Figure 3, based on SCADA data. The illustrates temperature trend of the bearing, generator bearing, oil inlet and outdoor temperature during the period of time when the fault was detected as indicated by dashed lines. It is observed in the vicinity of the fault regions that there are large temperature changes in the components and sudden rises in temperature, indicative of abnormal operation. The analysis proves that temperature monitoring is a valid approach for early fault pattern identification and improving predictive maintenance in wind turbine systems.

The correlation matrix for the important wind turbine parameters derived from the SCADA data is shown in Figure 4, which involves generator bearing temperature, generator phase temperature, generator RPM, humidity,wind speed and nacelle temperature. The heatmap displays the correlation strength and direction between features: greater positive correlation means that there is a strong dependence between the operational parameters.

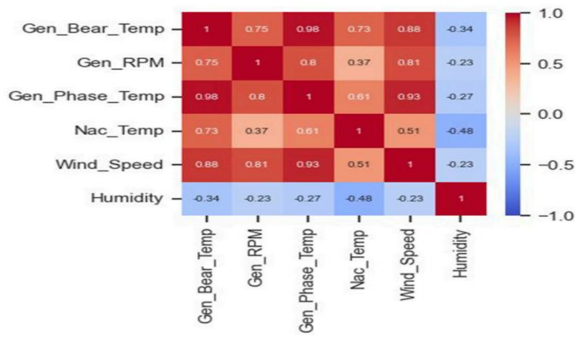

Fig. 4. Correlation Matrix of Wind Turbine SCADA Features

B. Data Preprocessing

Any machine learning model must begin with data pre-processing. This is due to the fact that a combination of human and business processes is usually involved in the creation, processing, and storage of raw data, which frequently leads to errors such as inaccurate, inconsistent, illogical, duplicate, or missing values. Appropriate operation of the algorithms depends on fixing these flaws. Finding and dealing with (or removing) outliers is, hence, a crucial part of preprocessing. By isolating the training and evaluation sets, we can ensure that the models are taught and assessed on data that accurately represents normal turbine operating. Therefore, the models are better able to spot outliers in the experimental data.

- **Data Cleaning:** Due to the dynamic nature of wind turbines and their intricate electromechanical design, the SCADA system gathers and stores enormous volumes of highly dimensional working data, which includes normal and unusual data from defects, turbulence, shutdowns, and device failures such as acquisition, communication, and storage devices. Therefore, acquiring healthy records for training models requires cleaning the data on raw datasets before simulating.
- **Data Normalization:** Different features typically have distinct dimensions. So, to make model training easier by removing dimensional effects, need to normalize these input dimensions so that the range of values is [0, 1] according to Equation (1).

$$X' = \frac{X - min\,(X)}{max\,(X) - min\,(X)} \tag{1}$$

- **here max (X) and min (X):** denote its max and min values and X is the raw data.
- **Variable Selection:** Wind turbines often have hundreds of operational state characteristics together and kept by SCADA system. These include both continuous data. It is recommended to choose status parameters (such as main bearing temperature) that have a strong correlation with the goal output when considering the model's complexity and processing efficiency.
- **Label Encoding:** Converting numerical representations of categorical attributes.
- **Data Balancing (Synthetic Minority Over-sampling Technique):** To fix the ML class imbalance, use SMOTE. It creates minority class synthetic samples using surrounding instances and interpolation. Rebalancing improves model performance, reduces bias, and prevents overfitting.

© JGREC 2026, All Rights Reserved

The class distribution of the SCADA dataset after applying Synthetic Minority Oversampling Technique (SMOTE) is shown in Figure 5. The balancing process resulted in equal number of fault and non-fault samples, alleviating class unevenness problem in the data. This is because of the balanced distribution which contributes to the learning capability of the ML models, thus the classification of faults during wind turbine condition monitoring is more accurate and unbiased.

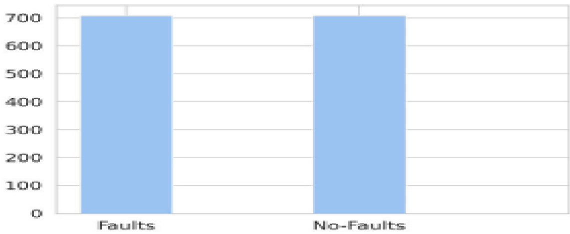

Fig. 5. Class Distribution After Balancing using SMOTE

- **Train Test split:** The dataset is split into 80% training and 20% testing.

C. Classification with ML models

The processed SCADA dataset is classified using machine learning models such as XGBoost and LightGBM (LGBM) for accurate identification of fault and non-fault conditions in wind turbine systems.

1) XG Boost Model

XGBoost is scalable end-to-end gradient tree boosting. To minimize an objective function in a tree regressor model, select the appropriate parameters θ. An objective function can be defined as Equation (2):

$$Obj = \sum_{i=1}^{n} L(y_{xi}\hat{Y}_{xi}) + \sum_{m=1}^{M} \Omega\,(f_m) \qquad (2)$$

The second-order derivable loss function L, with n as the sample size, estimates the variance among the predicted target variable and the actual target variable (yxi) .

$$\Omega(f_m) = \Upsilon T + \frac{1}{2}\beta \sum_{j=1}^{T} \omega_j^2 \qquad (3)$$

In Equation (3) where γ and β are hyperparameters to control difficulty of tree, wj is score of each leaf node, T is the number of leaf nodes in tree, and . γ and β signified by alpha and lambda in XGBoost provided by SKLearn.

2) Light GBM Model

LightGBM is a decision tree algorithm-based, distributed, high-performance gradient-boosting system that sees extensive use in a variety of ML tasks, including ranking, classification, and regression. Combining numerous weak machine learning models into one strong one is the goal of this Boosting algorithm type. So that incorrectly categorized data gets more emphasis in the subsequent training cycle, boosting techniques raise weights of data while lowering weights of well-categorized data. Equation (4) may be used to illustrate the main idea:

$$\mathrm{f(x)} = \sum_{\mathrm{q}=1}^{\mathrm{Q}} \alpha_q\, T(x, \theta_q) \qquad (4)$$

where Q is no. of base learners; x is training sample, $f(x)$ is target value corresponding to training sample, $T(x, \theta_q)$ is the q^{th} base learner α_q is weight coefficient of q^{th} base learner and θ_qThe parameter for learner's classification involved in training.

D. Evaluation Metrics

Classifying the identified faults using the derived dataset containing deep-learning features has been satisfactorily achieved. In the analysis, evaluate the method's performance using the following metrics: accuracy (A), recall(R), F1-score (F1) and precision (P). The suggested method's classification metrics are represented by the following equations. Equations (5) through (8) demonstrate how to calculate the performance measurements.

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+Fp+TN+FN}} \qquad (5)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad (6)$$

$$Recall = \frac{\mathrm{TP}}{TP+FN} \qquad (7)$$

$$F1 - Score = \frac{2(Precision*Recall)}{\mathrm{Precision+Recall}} \qquad (8)$$

Model ACC is number of correct predictions divided by the total observations. PRE is ratio of correctly anticipated positive observations to positive class labels. F1 score estimates the harmonic mean of PRE and REC, whereas recall is the percentage of positive observations correctly detected.

FP stands for total number of false positives, while TP represents total number of positive observations. Similarly, FN for total number of real positive observations that were first categorized as negative and TN stands for the total number of negative observations.

IV. Result Analysis and Discussion

In this study, machine learning algorithms were created using the "Python" programming language and associated libraries, including "scikit_learn." Windows 10 was installed, Jupyter Notebook was utilized for development, and the hardware configuration consisted of 64 GB RAM, "Intel Core" i7 CPU. When it comes to defect detection in wind turbines using the SCADA dataset, Table I compares performance of XGBoost and LightGBM (LGBM) machine learning models. In terms of accuracy, the LGBM demonstrated the highest accuracy (95.6%), with a better value for recall and F1-score, thus showing higher performance in correctly finding turbine faults among the two models. The both models achieved the same precision of 96.5%, indicating high reliability for fault classification.

TABLE I. Proposed Result Performance for Wind Turbines using SCADA Dataset

Performance Measures	XGBoost	LGBM
Accuracy	0.952	0.956
Precision	0.965	0.965
Recall	0.938	0.945
F1-score	0.951	0.955

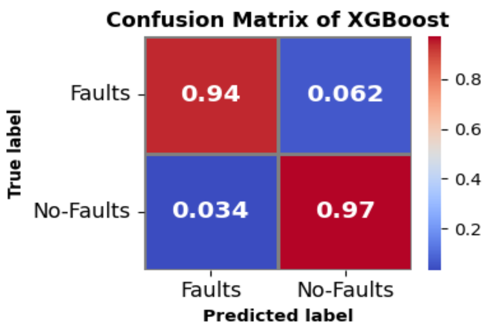

Fig. 6. Confusion Matrix of XG Boost Model

Figure 6 shows the XGBoost model's confusion matrix. The model is correctly classify 94% of fault conditions and 97% of non-fault conditions, thus showing a good classification ability. The misclassification of faults as non-faults is 6.2% while that of non-faults as faults is 3.4% of the total samples.

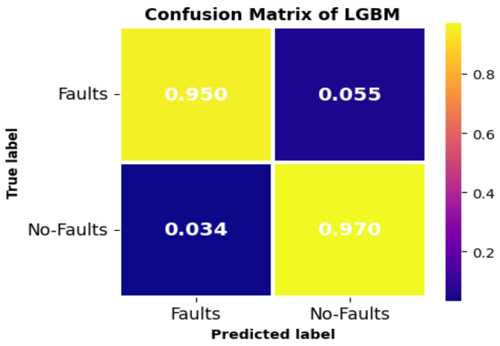

Fig. 7. Confusion Matrix of LGBM Model

The classification performance of the model is very good with 95% of the fault conditions and 97% of the non-fault conditions being correctly identified, as shown in Figure 7. It is only 5.5% of fault samples and 3.4% of non-fault samples that are misclassified. Results reveal that LGBM model is effective and dependable for early wind turbine defect identification and condition monitoring.

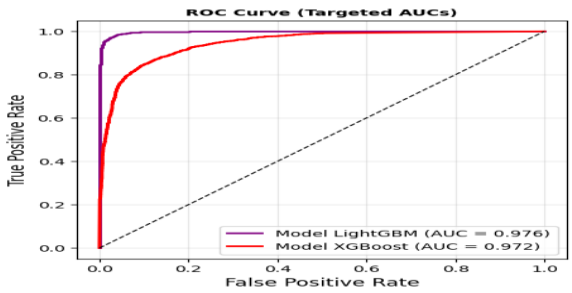

Fig. 8. ROC Curve of Proposed Models

The ROC curve for the proposed XGBoost and LightGBM (LGBM) SCADA models for wind turbine failure detection is shown in Figure 8. ROC curves demonstrate TPR-FPR trade-offs for both models. The LGBM model has the best Area Under the Curve (AUC) of 0.976, followed by XGBoost at 0.972. The obtained AUC values are almost equal to one, which proves the excellent classification ability and good discriminative power for early fault identification in wind turbine systems.

A. *Comparative Analysis and Discussion*

Table II shows that the proposed XGBoost and LightGBM (LGBM) models outperform existing ML-based fault identification models for wind turbines using the SCADA dataset. The traditional classifiers like Naïve Bayes (NB), KNN and SVM had comparatively low accuracy and F1-score results, suggesting their inability to cope with the complex fault patterns. The existing XGBoost model had a good accuracy of 92.2% but the proposed XGBoost and LGBM models outperform the existing one with an accuracy of 95.2% and 95.6%, respectively. The LGBM model had the highest PRE, REC, and F1 for overall performance, demonstrating its capacity to reliably and early diagnose wind turbine faults for condition monitoring system.

TABLE II. COMPARATIVE PERFORMANCE ANALYSIS OF EXISTING AND PROPOSED MODELS FOR WIND TURBINE FAULT DETECTION

Models	ACC	PRE	REC	F1-score
NB[19]	0.850	0.858	0.850	0.883
XGBoost[19]	0.922	0.930	0.922	0.927
KNN[20]	0.7462	0.9411	0.50	0.6530
SVM[20]	0.7014	0. 80	0.50	0.6153
XGBoost	0.952	0.965	0.938	0.951
LGBM	0.956	0.965	0.945	0.955

The SCADA data is suggested to be accomplished using the XGBoost and LightGBM (LGBM) algorithms. Reliable predictive maintenance is made possible by the system's ability to accurately differentiate between fault and non-fault states and conduct thorough analyses of operational parameters. The proposed model has several merits over conventional machine learning models, including better fault detection performance, better performance for high dimensionality SCADA data, faster processing speed and lower misclassification rates. The benefits include reduced maintenance costs, reduced downtime of the turbines, enhanced reliability and efficiency, and improved overall turbine performance and sustainability.

B. *Justification and Novelty*

The study is justified due to the increasing demand for accurate and reliable fault detection in wind turbines using SCADA data to minimize maintenance expenses and operational downtimes. The novelty of this work is based on an efficient machine learning-based framework that uses XGBoost and LightGBM models together with the data pre-processing strategy, normalization, feature selection and balancing using SMOTE technique for more accurate fault detection with lesser computational complexity and reduced misclassification rate.

C. *Limitations*

Due to the limited scope of historical SCADA data, the suggested framework cannot account for all probable uncertainties and unusual fault conditions in wind turbines. The quality of the data, missing data and sensor noise can also impact the performance of the models. Furthermore, the study does not apply complex architectures of deep learning models to temporal fault analysis, such as those for time series. Additionally, the study integrates only the machine learning models, namely XGBoost and LGBM models, without using advanced temporal fault analysis deep learning architectures. The system proposed is tested with a limited data set, which may result in different performance in other environmental and operational conditions.

D. *Existing Work Gap*

While several existing studies have shown promising performance with wind turbine fault detection from SCADA data, many of these approaches employ complex deep learning architectures and are not easily implementable in real time and have a high computational cost. Additionally, some models suffer from problems of class imbalance, false positive prediction and decreased scalability at varying operating conditions. In this context, an efficient and reliable machine learning-based solution that can deliver accurate early fault indication while simultaneously achieving reduced computation complexity and enhanced predictive maintenance capabilities in wind turbine systems is required.

© JGREC 2026, All Rights Reserved

V. CONCLUSION AND FUTURE WORK

Many areas of study rely heavily on artificial intelligence, which is a rapidly expanding area of study. It has great potential as a tool for optimizing maintenance in the WT sector, where it can assist in predicting failures early on and guarantee energy output all year round. In order to improve the fault identification process with a dataset that is extremely imbalanced, this study tests three data pre-treatment procedures. This research proposes a methodology for early defect identification in wind turbines using SCADA datasets that is machine learning-based and uses the XGBoost and LightGBM (LGBM) models. To enhance the data quality and minimize class imbalance, data pre-processing methods like data cleaning, normalization, feature selection, label encoding, and SMOTE-based data balancing are applied in the methodology. In experiments, the XGBoost model had the best accuracy of 95.2% while the LGBM model had 95.6% accuracy and superior precision, recall, and F1-score. In comparison, the proposed models outperformed NB, KNN, and SVM classifiers. The proposed framework can significantly contribute towards reduced maintenance expenses, reduced time out of service, and enhanced operational reliability. Deep learning techniques like CNN and LSTM can be used to improve temporal defect identification, while IoT technology can provide real-time monitoring and explainable AI methods for intelligent predictive maintenance in wind turbine systems.

REFERENCES

[1] C. Gück, C. M. A. Roelofs, and S. Faulstich, "CARE to Compare: A Real-World Benchmark Dataset for Early Fault Detection in Wind Turbine Data," *Data*, vol. 9, no. 12, p. 138, Nov. 2024, doi: 10.3390/data9120138.

[2] R. K. Pandit and D. Infield, "SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes," *IET Renew. Power Gener.*, vol. 12, no. 11, pp. 1249–1255, Aug. 2018, doi: 10.1049/iet-rpg.2018.0156.

[3] M. Kari, "Deep Learning-Based Fault Prediction Models for Enhanced Network Security Monitoring," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 3, p. 492, Jun. 2023, doi: 10.48175/IJARSCT-11600I.

[4] E. Y.-K. Ng and J. T. Lim, "Machine Learning on Fault Diagnosis in Wind Turbines," *Fluids*, vol. 7, no. 12, p. 371, Dec. 2022, doi: 10.3390/fluids7120371.

[5] O. T. Bindingsbø, M. Singh, K. Øvsthus, and A. Keprate, "Fault detection of a wind turbine generator bearing using interpretable machine learning," *Front. Energy Res.*, vol. 11, Dec. 2023, doi: 10.3389/fenrg. 2023.1284676.

[6] R. K. Pandit, J. Wang, and L. Tinsley, "SCADA-based early detection of wind turbine yaw motor current failures using Gaussian process surrogate models and Fisher's combined probabilistic alarm thresholding," *Wind Energy Eng. Res.*, vol. 5, p. 100031, Jun. 2026, doi: 10.1016/j.weer.2026.100031.

[7] S. K. Raju, M. Periyasamy, A. A. Alhussan, S. Kannan, S. Raghavendran, and E.-S. M. El-kenawy, "Machine learning boosts wind turbine efficiency with smart failure detection and strategic placement," *Sci. Rep.*, vol. 15, no. 1, p. 1485, Jan. 2025, doi: 10.1038/s41598-025-85563-5.

[8] S. Murumkar and C. Tayal, "Semantic Web Platform for Predicting Patient Health Awareness Levels Using Data Mining and Ontology," in *2026 IEEE Madhya Pradesh Section Conference (MPCON)*, Gwalior, India: IEEE, Mar. 2026, pp. 1649–1655. doi: 10.1109/MPCON69668.2026.11508203.

[9] A. K. Padhy, T. Pravinbhai Patel, V. Soni, S. Shivam, G. B. Thokala, and B. Vulugundam, "Machine Learning-Based Fault Prediction in Large- Scale Distributed Systems," in *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)*, IEEE, Feb. 2026, pp. 1–6. doi: 10.1109/ICAIC67076.2026.11395778.

[10] A. Santolamazza, D. Dadi, and V. Introna, "A Data-Mining Approach for Wind Turbine Fault Detection Based on SCADA Data Analysis Using Artificial Neural Networks," *Energies*, vol. 14, no. 7, p. 1845, Mar. 2021, doi: 10.3390/en14071845.

[11] A. Stetco *et al.*, "Machine learning methods for wind turbine condition monitoring: A review," *Renew. Energy*, vol. 133, pp. 620–635, Apr. 2019, doi: 10.1016/j.renene.2018.10.047.

[12] N. Jia *et al.*, "AI-Driven Fault Detection and O&M for Wind Turbine Drivetrains: A Review of SCADA, CMS and Digital Twin Integration," *Energies*, vol. 19, no. 5, p. 1370, Mar. 2026, doi: 10.3390/en19051370.

[13] P. W. Khan, C. Y. Yeun, and Y. C. Byun, "Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning," *Eng. Fail. Anal.*, vol. 148, p. 107209, Jun. 2023, doi: 10.1016/j.engfailanal.2023.107209.

[14] D. Nayak, S. Singh, N. Sharma, and K. Palani, "Spatiotemporal Graph Neural Network for Predictive Maintenance in Wind Energy Systems Using Multi-Task Learning," in *2025 International Conference on Computing and Communications (COMPUTINGCON)*, IEEE, Sep. 2025, pp. 1–6. doi: 10.1109/COMPUTINGCON64838.2025.11377265.

[15] L. Natrayan, "XAI-Driven ConvLSTM Framework for Root Cause Analysis in Wind Turbine Systems," in *2025 3rd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIHEI)*, IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/IDICAIHEI65991.2025.11379369.

[16] J. Maldonado-Correa, J. Torres-Cabrera, S. Martín-Martínez, E. Artigao, and E. Gómez-Lázaro, "Wind turbine fault detection based on the transformer model using SCADA data," *Eng. Fail. Anal.*, vol. 162, p. 108354, Aug. 2024, doi: 10.1016/j.engfailanal.2024.108354.

[17] F. Shaheen and M. M. M. Al-Khalidy, "Wind turbine fault detection and diagnosis using machine learning techniques," *IET Conf. Proc.*, vol. 2024, no. 37, pp. 74–80, Mar. 2025, doi: 10.1049/icp. 2025.0824.

[18] S. R. Mohapatro, M. Mulchandani, N. Mohanty, M. R. Sethi, and S. Sahoo, "Fault Detection in Wind Turbine using Soft Sensing Approach," in *2024 3rd International Conference for Advancement in Technology (ICONAT)*, IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/ICONAT61936.2024.10775020.

[19] C. Lee and E. D. C. Maceren, "Wind energy system fault classification and detection using deep convolutional neural network and particle swarm optimization-extreme gradient boosting," *IET Energy Syst. Integr.*, vol. 6, no. 4, pp. 479–497, Dec. 2024, doi: 10.1049/esi2.12144.

[20] C. Velandia-Cardenas, Y. Vidal, and F. Pozo, "Wind Turbine Fault Detection Using Highly Imbalanced Real SCADA Data," *Energies*, vol. 14, no. 6, p. 1728, Mar. 2021, doi: 10.3390/en14061728.

© JGREC 2026, All Rights Reserved