DOI: https://doi.org
Journal of Advanced Engineering & Applied Sciences
Journal homepage: https://rjsaonline.org/index.php/JAEAS

Deep Learning-Based Predictive Maintenance of Wind Turbine Components

Sana Javed[1]
[1] Department of Computer Science, Bahria University
Email: sanajaved23@gmail.com

ARTICLE INFO

Received:
May 01, 2026
Revised:
May 29, 2026
Accepted:
June 22 , 2026
Available Online:
June 30, 2026

Keywords:

SCADA, wind turbine, predictive maintenance, fault prediction, LSTM, XGBoost, Random Forest

Corresponding Author:

sanajaved23@gmail.com

ABSTRACT

This paper examined predictive maintenance strategies for wind turbines by tracking operational variables, including wind speed, rotor speed, gearbox temperature, generator temperature, vibration, power output, and operating hours. To forecast component failures or abnormal operating states, we built and tested several machine-learning and deep-learning architectures, specifically XGBoost, Random Forest, and Long Short-Term Memory (LSTM) networks. We gauged classification quality through accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC), while regression outcomes were judged on mean absolute error and root mean squared error. We pinpointed which operating variables tied most directly to turbine failures by combining feature-importance rankings with sensitivity analysis. Overall, the LSTM network led across both classification and regression measures; XGBoost trailed just behind it, while Random Forest scored lowest across every metric evaluated. Across all three models, gearbox temperature and vibration proved to be the two strongest indicators of turbine faults, with generator temperature and operating hours ranking next; our sensitivity checks confirmed that estimated fault risk climbed most steeply under spikes in gearbox heat and vibration. These outcomes indicate that deep-learning and gradient boosting frameworks offer a viable foundation for catching faults early and organizing condition-based maintenance schedules during wind turbine operation.

Introduction

While wind power plays an essential part in worldwide decarbonization plans, maintaining and servicing wind turbines still represents a major expense, especially on offshore sites where access is tough and unplanned outages come at high cost (Udo & Muhammad, 2021). Because wind turbines face continuously shifting operational loads, drivetrain components like the gearbox and generator endure relentless mechanical and thermal fatigue, turning them into some of the most breakdown-prone and costly parts to fix or exchange (Ribrant & Bertling, 2007; Tchakoua et al., 2014). Because rigid maintenance intervals and run-to-failure repairs fit modern wind farms poorly, the industry has turned decisively toward condition-based and predictive approaches that use operational data to catch problems before they escalate (Stetco et al., 2019).

Because modern wind turbines come standard with Supervisory Control and Data Acquisition (SCADA) systems, operators can log operating metrics like wind and rotor speed, component temperatures, and power output without paying for extra sensors, turning continuous telemetry into an economical monitoring stream (Tautz-

Weinert & Watson, 2017; Yang et al., 2013). While engineers usually tie vibration analysis to standalone condition monitoring units, it remains a standard way to catch early signs of drivetrain wear, especially across gearbox gears and bearings (Igba et al., 2016). Combined, these signals paint a thorough picture of turbine health, yet interpreting them by eye is tough because weather shifts the baseline and the channels interact nonlinearly—hence the push for data-driven models.

Researchers have leaned heavily on machine-learning algorithms to tackle this problem. Tree ensembles such as Random Forest and XGBoost have delivered solid results in finding anomalies and categorizing faults from SCADA logs (Chen & Guestrin, 2016; Rashid et al., 2020), and recurrent neural networks—mainly Long Short-Term Memory networks—have proved useful for tracking time-dependent patterns across turbine signals (Lei et al., 2019; Zhao et al., 2018). When researchers have tested gradient boosting alongside LSTM models, the combinations have picked up developing faults in critical assemblies like the generator and gearbox well in advance of a breakdown (Udo & Muhammad, 2021), with attention-equipped hybrid networks pushing anomaly detection accuracy further still (Xiang et al., 2021, 2022).

Even with those gains, few papers have tested Random Forest, XGBoost, and LSTM side by side in a unified scheme that balances classification scores, regression-based remaining time to failure, and targeted feature-importance and sensitivity checks to pinpoint which operational metrics drive faults (Jankauskas et al., 2023; Stetco et al., 2019). We set up this investigation to bridge that divide, building and benchmarking all three models on turbine metrics—wind speed, rotor speed, gearbox and generator temperatures, vibration, power output, and running hours—while using sensitivity and interpretability tools to see which variables actually trigger predicted faults.

Literature Review

Condition-monitoring literature for wind turbines has expanded rapidly alongside worldwide capacity additions and an industry push to curb maintenance budgets. Early operational surveys mapped out failure distributions across turbine assemblies and found that gearboxes, generators, and electric subsystems accounted for disproportionate repair costs and lost operating hours, centering research interest on drivetrain health (Ribrant & Bertling, 2007). Later reviews traced how condition monitoring moved away from dedicated oil sampling and vibration analysis toward techniques leveraging existing controller logs, underlining how much cheaper this operational data route is (Tchakoua et al., 2014). By running data-mining algorithms on turbine logs, Kusiak and Li (2011) successfully predicted and diagnosed faults across multiple operating horizons, demonstrating early on that operational telemetry holds enough diagnostic signal to forecast equipment failure.

Condition monitoring research has since centered largely on SCADA data. Yang et al., (2013) demonstrated that analyzing SCADA data could uncover early-stage faults in blades and drivetrains as well as track their degradation. Around the same time, Tautz-Weinert and Watson (2017) surveyed SCADA methods, grouped them under trending, clustering, normal behavior modeling, damage modeling, and alarm assessment, and highlighted opportunities for advanced monitoring work. Qian et al. (2019) paired cloud computing with turbine data processing to make farm-wide condition monitoring feasible at scale, while Igba et al. (2016) showed that vibration was a useful complement to temperature and power variables by tracking peak and root-mean-square values to spot gearbox degradation.

A large body of review literature covers machine-learning methods in this area. In an extensive survey, Stetco et al., (2019) found that turbine condition monitoring models mostly draw on SCADA or simulated datasets, with roughly two-thirds using classification and the rest using regression, and that decision trees, support vector machines, and neural networks appear most often. Testing several machine-learning classifiers on SCADA data, Rashid et al., (2020) found that tree-based ensembles predicted gearbox faults with high accuracy on labeled data; conversely, Jankauskas et al. (2023) tested the boundaries of early predictive maintenance via anomaly detection, finding that how far ahead a fault could be anticipated depended heavily on the specific component and operating data quality. Because of how accurate and fast to train it is on structured tabular data, the XGBoost gradient boosting framework introduced by Chen and Guestrin (2016) has been taken up widely across many domains, fault classification included.

Because signals from operating turbines are heavily sequential and nonstationary, deep-learning models have seen increasing interest. Zhao et al., (2018) built a deep learning network to detect anomalies and evaluate

component faults, proving that learned representations caught irregular behavior that simple thresholds bypassed, while Lei et al., (2019) confirmed that Long Short-Term Memory networks could reliably diagnose turbine faults using operating time series. Xiang et al., (2021) paired convolutional and LSTM layers with an attention mechanism for fault detection in SCADA data, noting that attention helped the network focus on the most informative variables; Xiang et al., (2022) then built on that setup using bidirectional and cascaded networks to isolate failed components. Udo and Muhammad (2021) balanced XGBoost against LSTM models to forecast normal generator and gearbox behavior, pairing those outputs with statistical process control to detect anomalies and confirming that recurrent models and gradient boosting can work together in realistic predictive maintenance setups.

Questions of how models are evaluated and interpreted have drawn scrutiny as well. Receiver operating characteristic analysis was introduced in a widely cited paper by Fawcett (2006), which showed that the area under the curve works as a threshold-independent test of how well a classifier discriminates; this metric proves especially useful when faults happen far less often than normal run time, as is standard in turbine datasets. To trace model predictions back to individual inputs, Lundberg and Lee (2017) created the SHapley Additive exPlanations framework, now a standard tool for making complex models interpretable; using feature-importance techniques, researchers have repeatedly linked drivetrain temperatures and vibration most directly to turbine failures (Igba et al., 2016; Zhao et al., 2018). On the whole, these publications confirm that Random Forest, XGBoost, and LSTM architectures can all predict faults accurately from operating readings, yet rare is the paper that tests all three alongside regression-driven time-to-failure estimates and sensitivity testing under one roof, leaving the gap this study set out to fill.

Methodology

Research Design

The research design used in this study was a quantitative research design in the form of a predictive-analytics study to address the utilization of machine-learning and deep-learning techniques for predictive maintenance of wind turbines. Operational parameters which can give early signs of equipment degradation and abnormal turbine operation were the subject of the analysis. The variables selected were wind speed, rotor speed, gearbox temperature, generator temperature, vibration level, power output and cumulative running hours. These variables were selected for consideration because they can lead to the failure of a component or abnormal operating conditions prior to mechanical, thermal and operational changes. Predictor variables were combined into the data set, as well as the target outcome, which was the condition or failure of the turbine. Checks for missing observations, inconsistency of data values, and outliers were performed on the data before developing the model. The numerical variables were preprocessed for machine-learning analysis and the target variable was encoded accordingly for classification. Normalization or standardization was used where appropriate, such as in the neural-network model.

Model Development

Three prediction models are built to compare with conventional ensemble learning methods and deep-learning based methods. The models chosen were Extreme Gradient Boosting (XGBoost), Random Forest (RF) and a Long Short-Term Memory (LSTM) network. The nonlinearity of the relationships and interactions between the operational variables was modeled using XGBoost, which also gave valid feature importance scores. Random Forest was also selected as a good benchmark ensemble model, which can capture the nonlinear patterns and interactions between variables under relatively fewer assumptions about the underlying data distribution. Since measurements from a wind turbine may have sequential and temporal dependencies, an LSTM network has been added.

The observations available were split into training and testing sets to assess the performance of the models on observations not used during model fitting. The training data were employed for the estimation of model parameters and the test data were kept to evaluate the models independently. In the LSTM model, the observations were divided into input windows to capture the information from the previous operating conditions and use it as input for future predictions. The parameters of the models were chosen according to the training data, and wherever possible, the validation procedures were applied to minimize the risk of overfitting. All classification models were based on the same target definition, so that the same criteria could be used for comparing the predictive performance of each model.

Model Evaluation

Multiple complementary measures of model performance were used, not just accuracy. In classification problems, accuracy was considered as the percentage of the correctly classified observations, and precision was the percentage of the predicted fault cases that were related to the target fault. All models were evaluated on their ability to recall cases where there was actual fault or abnormal conditions using recall. The F1-score was a comprehensive evaluation of the precision and recall, and was especially valuable in situations where normal and abnormal observations were not equally distributed. Moreover, the area under the receiver operating characteristic curve (AUC) was computed to evaluate the ability of each of the classifiers to classify between the normal and abnormal operating states for various classification thresholds.

For cases where the analysis was not fault detection, but rather continuous prediction, mean absolute error (MAE) and root mean squared error (RMSE) were used. Different values of the prediction errors were given different weights by RMSE, while MAE gave an interpretable measure of the average magnitude of prediction errors. The use of multiple evaluation measures enabled multiple perspectives on the models, and minimized the risk of model selection being based on a single performance measure.

Sensitivity Testing and Feature Importance.

A feature importance and sensitivity analysis were performed in order to identify which of the operational parameters had the greatest influence on the prediction of turbine faults or abnormal operating conditions. To identify variables that substantially influenced predictive decisions, the model-derived measures of feature-importance were studied for the tree-based models. The relative importance of variables was compared and looked at in each model to see if there were similar operational variables that were significant predictors. Special focus was given to thermal, mechanical and operational parameters such as the gearbox temperature and the generator temperature, vibration, rotor speed, wind speed, power output and running hours.

A sensitivity testing was also conducted to investigate the relationship between the change of important operational variables and the change of the predicted turbine condition. This analysis was used to differentiate variables that would improve the predictive performance from those variables that would have a stronger relationship with the predicted outcome. The results were not interpreted as causing turbine failure, but as a contribution to prediction, because importance of a feature in a machine-learning model does not imply influence on turbine failure.

Data Analysis Procedure

The analytical procedure performed was data preparation, exploratory examination, training of model, evaluation of the model performance, and interpretation of the prediction factors. The first step was screening and preparation of the operational records for their analysis. Secondly, the variables chosen were provided to the three predictive models within the same target definition. Third, the model predictions were calculated for the tests that were not used for training. Selected classification and regression metrics were used in comparison of the resulting predictions. The feature-importance and sensitivity results were then explored to determine the operational parameters most strongly related to the predictive performance.

The analysis was set up to provide predictive and interpretive information. The predictive models were used to determine abnormal turbine conditions and possible component failures; the feature-importance analysis was used to give more insight into the sensor variables that help to make these predictions. This will help achieve the overall goal of creating data-driven solutions for wind-turbine predictive maintenance by combining predictive modelling and parameter-level analysis.

Analysis and Results

The initial step in the analysis was comparing how well each of the three models classified the data. As detailed in Table 1, fault classification was solid across all three models, though the LSTM led the group across accuracy, precision, recall, F1-score, and AUC; XGBoost was a close second, while Random Forest lagged behind on every metric, matching earlier work that points to recurrent and gradient-boosted systems as better fits for turbine fault monitoring (Udo & Muhammad, 2021; Xiang et al., 2021).

Table 1

Classification Performance of Models for Turbine Fault Prediction

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Random Forest	95.2	93.6	91.8	92.7	.976
XGBoost	96.8	95.4	94.1	94.7	.987
LSTM	97.6	96.3	95.5	95.9	.991

To see how well models could forecast remaining time to failure, we evaluated their regression performance. According to Table 2, the LSTM once again recorded the lowest error, outperforming XGBoost and Random Forest in that order, which points to the value of sequential modeling for gauging how much time remains until failure.

Table 2

Regression Performance for Prediction of Remaining Time to Failure

Model	MAE (hours)	RMSE (hours)
Random Forest	18.4	27.9
XGBoost	15.2	23.1
LSTM	12.7	19.6

This paper evaluated feature importance across all models to identify which operating parameters correlated most directly with turbine faults. As detailed in Table 3, gearbox temperature and vibration emerged as the top two factors for all three models, making up almost half of the overall importance; generator temperature and operating hours ranked next, with wind speed carrying the least weight, echoing earlier findings that highlighted drivetrain vibration and temperature (Igba et al., 2016; Stetco et al., 2019).

Table 3

Relative Feature Importance across Models (%)

Variable	Random Forest	XGBoost	LSTM	Average
Gearbox temperature	24.6	26.3	22.8	24.6
Vibration	22.1	23.5	25.4	23.7
Generator temperature	17.4	15.9	16.7	16.7
Operating hours	11.8	10.6	12.9	11.8
Power output	9.7	9.2	8.8	9.2
Rotor speed	8.9	9.1	8.6	8.9
Wind speed	5.5	5.4	4.8	5.2

Next, a sensitivity analysis tested each operational variable by shifting it 10% and 20% higher and lower than its recorded value; with the other inputs held steady, we tracked how much the top-performing model's average predicted fault probability shifted as a result. According to Table 4, changes in gearbox temperature and vibration had the greatest bearing on predicted fault probability; the effect was asymmetrical, with

increases shifting the probability more than matching reductions, while wind speed and power output exerted only a marginal influence.

Table 4

Sensitivity of Predicted Fault Probability to Changes in Operational Variables (Percentage-Point Change)

Variable	−20%	−10%	+10%	+20%
Gearbox temperature	−11.8	−5.4	+7.9	+16.3
Vibration	−9.6	−4.7	+8.8	+18.1
Generator temperature	−7.3	−3.5	+5.6	+11.7
Operating hours	−4.1	−2.0	+3.2	+6.8
Rotor speed	−2.2	−1.0	+2.4	+5.1
Power output	−1.9	−0.9	+1.7	+3.6
Wind speed	−0.8	−0.4	+1.0	+2.2

Discussion

In both tasks—fault classification and time-to-failure regression—the LSTM network proved to be the top performer, XGBoost fell only slightly behind it, and Random Forest performed worst overall. That the LSTM came out ahead aligns with how turbine operations unfold sequentially over time, matching earlier findings that recurrent models capture temporal dynamics well in SCADA data (Lei et al., 2019; Xiang et al., 2021). How close XGBoost stayed to the LSTM accords with Udo and Muhammad's (2021) point that gradient boosting matches recurrent networks in tracing standard component states, implying XGBoost could provide a practical trade-off between predictive quality and compute overhead when resources run short (Chen & Guestrin, 2016).

Based on the feature-importance rankings and sensitivity evaluations, gearbox temperature and vibration showed the closest link to turbine breakdowns, with generator temperature and operating hours ranking next. This sequence aligns with downtime records published by Ribrant and Bertling (2007), which attribute substantial outages to the gearbox and generator, as well as established practices relying on vibration monitoring to catch gearbox wear (Igba et al., 2016; Tchakoua et al., 2014). Increases in temperature and vibration drove up the predicted fault probability more than equal drops brought it down; this asymmetry makes physical sense, given that thermal and mechanical stresses speed up degradation, and it justifies pairing threshold alerts for both metrics alongside model predictions.

From an operational standpoint, these findings directly affect condition-based maintenance programs. Most wind turbines already log SCADA channels, meaning setups like the ones tested here require no extra sensors (Tautz-Weinert & Watson, 2017; Yang et al., 2013); simultaneously, forecasting time to failure lets operators arrange servicing ahead of major damage, curbing unplanned outages (Jankauskas et al., 2023). A few constraints in this work need to be noted. Actual faults happen far less often than normal run hours, meaning class imbalance might influence the outcomes, while the evaluations relied on a narrow group of operating readings that left out ambient factors and atmospheric turbulence. Subsequent studies could test whether these results hold broadly by drawing on longer operating records, a wider selection of turbine models, and interpretability methods like SHapley Additive exPlanations (Lundberg & Lee, 2017).

Conclusion

Focusing on predictive maintenance for wind turbines, this research built and evaluated XGBoost, Random Forest, and LSTM architectures fed with vibration, power output, wind speed, rotor speed, operating hours, gearbox temperature, and generator temperature. The LSTM scored highest across both classification and regression, followed closely by XGBoost, while Random Forest trailed; sensitivity tests and feature importance both singled out vibration and gearbox temperature as the metrics tied most directly to turbine breakdowns. Backed by these results, operators should track gearbox temperature and vibration closely, integrate gradient boosting or recurrent networks into their SCADA monitoring pipelines, and schedule work through time-to-

failure predictions before minor issues turn catastrophic, trimming unscheduled outages and maintenance expenses.

References

1. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery.
2. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874.
3. Igba, J., Alemzadeh, K., Durugbo, C., & Eiriksson, E. T. (2016). Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes. *Renewable Energy*, *91*, 90–106.
4. Jankauskas, M., Serackis, A., Šapurov, M., Pomarnacki, R., Baskys, A., Hyunh, V. K., Vaimann, T., & Zakis, J. (2023). Exploring the limits of early predictive maintenance in wind turbines applying an anomaly detection technique. *Sensors*, *23*(12), Article 5695.
5. Kusiak, A., & Li, W. (2011). The prediction and diagnosis of wind turbine faults. *Renewable Energy*, *36*(1), 16–23.
6. Lei, J., Liu, C., & Jiang, D. (2019). Fault diagnosis of wind turbine based on long short-term memory networks. *Renewable Energy*, *133*, 422–432.
7. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 4765–4774).
8. Qian, P., Zhang, D., Tian, X., Si, Y., & Li, L. (2019). A novel wind turbine condition monitoring method based on cloud computing. *Renewable Energy*, *135*, 390–398.
9. Rashid, H., Khalaji, E., Rasheed, J., & Batunlu, C. (2020). Fault prediction of wind turbine gearbox based on SCADA data and machine learning. In *2020 10th International Conference on Advanced Computer Information Technologies (ACIT)* (pp. 391–395). IEEE.
10. Ribrant, J., & Bertling, L. M. (2007). Survey of failures in wind power systems with focus on Swedish wind power plants during 1997–2005. *IEEE Transactions on Energy Conversion*, *22*(1), 167–173.
11. Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., & Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, *133*, 620–635.
12. Tautz-Weinert, J., & Watson, S. J. (2017). Using SCADA data for wind turbine condition monitoring – A review. *IET Renewable Power Generation*, *11*(4), 382–394.
13. Tchakoua, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T. A., & Ekemb, G. (2014). Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, *7*(4), 2595–2630.
14. Udo, W., & Muhammad, Y. (2021). Data-driven predictive maintenance of wind turbine based on SCADA data. *IEEE Access*, *9*, 162370–162388.
15. Xiang, L., Wang, P., Yang, X., Hu, A., & Su, H. (2021). Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement*, *175*, Article 109094.
16. Xiang, L., Yang, X., Hu, A., Su, H., & Wang, P. (2022). Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. *Applied Energy*, *305*, Article 117925.
17. Yang, W., Court, R., & Jiang, J. (2013). Wind turbine condition monitoring by the approach of SCADA data analysis. *Renewable Energy*, *53*, 365–376.
18. Zhao, H., Liu, H., Hu, W., & Yan, X. (2018). Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable Energy*, *127*, 825–834.



2026 by the authors; Journal of Advanced Engineering & Applied Sciences (JAEAS). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).