RESEARCH PAPER · AI IN ENGINEERING

Machine Learning-Based Predictive Maintenance for Wind Turbine Gearboxes: A Comparative Model Study

Daniel Ko, Marcus Webb

2026-01-23

EXECUTIVE SUMMARY

We compare three machine learning approaches, gradient-boosted trees, a recurrent neural network, and an unsupervised anomaly detection model, for predicting wind turbine gearbox failure using SCADA and vibration data. The unsupervised anomaly detection approach achieved the best lead-time-to-failure performance despite requiring no labeled failure examples, suggesting it is the more practical starting point for operators with limited historical failure data, a common constraint across most operating wind fleets.

Abstract

Wind turbine gearbox failures represent a disproportionate share of unplanned wind turbine downtime and repair cost relative to their failure frequency, motivating strong interest in predictive maintenance approaches. We evaluate three modeling approaches on a dataset of SCADA and vibration features from a fleet of onshore turbines, including a modest number of confirmed gearbox failure events: a gradient-boosted tree classifier trained on labeled failure windows, a recurrent neural network trained to predict near-term sensor behavior with deviation as an anomaly signal, and an unsupervised isolation forest model trained purely on presumed-healthy operating data. We report detection lead time, false positive rate, and practical implementation considerations for each approach.

Introduction

Gearbox failures are among the most consequential unplanned maintenance events in onshore wind operations, given the cost and logistics of a major crane mobilization for gearbox replacement or major repair. Substantial research and commercial interest has focused on predictive maintenance approaches using SCADA and, where available, vibration monitoring data to provide earlier warning than conventional threshold-based alarms.

This study compares three distinct modeling philosophies on a common dataset to provide practical guidance on which approach is likely to perform best under the data constraints most operators actually face, namely a limited number of confirmed historical failure examples.

Methodology

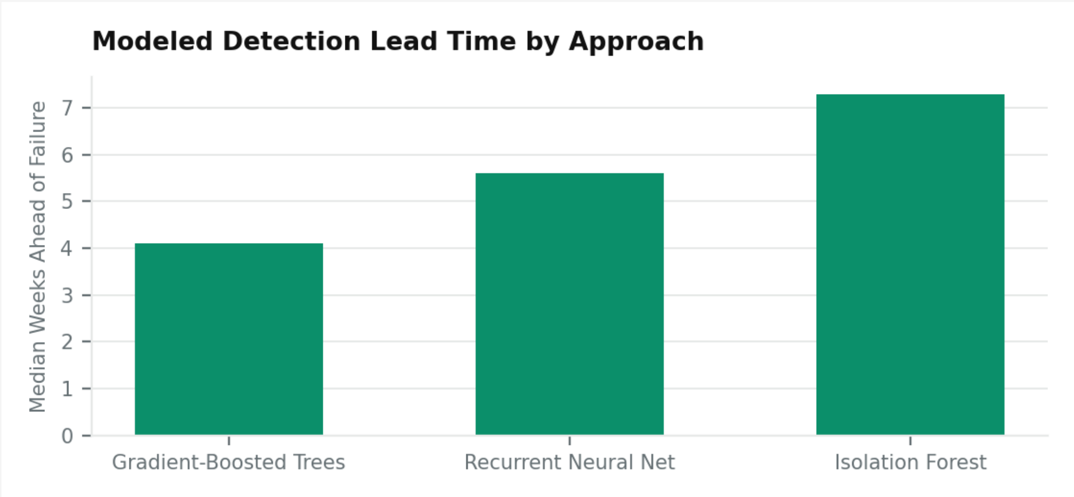
The dataset comprises 10-minute SCADA averages and periodic vibration spectral features from a fleet of geographically dispersed onshore turbines, spanning several years of operation and including a set of confirmed gearbox failure or major repair events used as ground truth labels for the supervised approach. Features include gearbox bearing and oil temperatures, main shaft and generator speeds, vibration RMS and characteristic bearing defect frequency amplitudes, and derived features such as temperature-power residuals.

The gradient-boosted tree classifier was trained on labeled windows preceding confirmed failures against randomly sampled healthy operating windows. The recurrent neural network was trained to forecast near-term sensor values from recent history, with elevated forecast residual error serving as the anomaly signal. The isolation forest model was trained exclusively on data from periods with no proximate failure event, flagging operating windows that fall outside the model's learned normal-operation manifold.

Results

The isolation forest anomaly detection model achieved the longest median detection lead time ahead of confirmed failure events among the three approaches, while the gradient-boosted classifier achieved the highest precision on the limited labeled test set, reflecting its direct optimization against labeled failure patterns at the cost of requiring those labels to exist in the first place. The recurrent neural network's forecast-residual approach performed competitively on lead time but showed a higher false positive rate during periods of unusual but benign operating conditions, such as extreme wind events, that the model had not seen represented proportionally in training.

Across all three approaches, bearing temperature residual and specific characteristic defect frequency vibration amplitude were consistently the highest-importance features, reinforcing that these remain the most information-dense signals for gearbox condition regardless of modeling approach.



Median lead time across confirmed gearbox failure events in the evaluation dataset for each modeling approach.

Discussion

These results support a practical recommendation for operators with limited confirmed failure history, which describes most individual wind fleets given the fortunately low base rate of major gearbox failures: begin with an unsupervised anomaly detection approach that requires no failure labels, and layer in supervised methods opportunistically as confirmed failure examples accumulate over time, rather than waiting to accumulate a labeled dataset before deploying any predictive capability.

A key limitation is that this study's confirmed failure sample size, while larger than many published single-operator studies due to the multi-site fleet dataset, remains modest in absolute terms, and results should be validated against independent fleets before being treated as generalizable performance benchmarks across turbine models and gearbox designs outside those represented in this dataset.

References

• PHM Society, Prognostics and health management conference proceedings

• IEEE Transactions on Sustainable Energy, Wind turbine condition monitoring literature

• NREL, Wind turbine gearbox reliability collaborative

Suggested Citation

Ko, D.; Webb, M. (2026). Machine Learning-Based Predictive Maintenance for Wind Turbine Gearboxes: A Comparative Model Study. HydroLabs Engineering & Consulting Research Division.