Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37907

Modeling Hidden Degradation States and Predicting Failure Risk in Wind Turbine Gearboxes

Boyu Hu

Wuhan-Britain China School, Wuhan, China
boyuhu828@gmail.com

Abstract. Wind turbine gear boxes have to work in a fluctuating-load and harsh-environment setting. Over time, wear and performance degradation gradually accumulate; it becomes difficult to know when they are out of service and when to carry out maintenance. Hidden states are taken as the starting point in this paper to explore how hidden markov model (HMM) can be used to model gearbox degradation and assess changes in risk over time. First, organize the sensor monitoring data and extract state-related features from signals such as vibration and temperature in the study. An HMM is then built to model the unobservable health status of the equipment as several hidden states. Observation probabilities and state transition relationships are used to show how the equipment changes from a normal operating state to various states of degradation gradually. Based on the above changes in time, the study determines whether the risk is rising. To further clarify this way, the three related methods SCTCN, NTNNet and DT-SegRNN are also studied. These faults are relatively small in scale, such as missing data, and they also differ from those in the simulation. Based on the above analysis, the primary strength of HMM is not fault classification. Crucially, it links the observations of sensors with the deterioration status of equipment that is difficult to monitor directly and supports research on prolonged health changes and risk trends.

Keywords: Wind turbine gearbox, hidden markov model (HMM), gearbox degradation, unobservable health status

1. Introduction

With the development of industry in recent years, energy consumption has become increasingly severe. Therefore, wind power has been used to reduce energy consumption in recent years [1]. Over a long period of operation, the gears and bearings in a wind turbine gearbox have been exposed to various changes in load and temperature, vibration, etc. With the accumulation of the above effects over time, wear and failure will eventually occur [2]. If maintenance is only organized after a failure occurs, both the cost of repair and losses due to downtime will be higher. Therefore, early detection of condition changes in operating data and tracking the development of risk are feasible problems in gearbox health management [3].

Therefore, the previous studies have used all sorts of methods. Kang et al. processed the operational data with BP neural networks, K-means clustering and SVM, and used the results for

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37907

fault early warning [4]. Meng et al. used SCADA data to enhance data processing and spatiotemporal feature fusion, thus reducing the response time of fault alarms [5]. Wang et al. proposed a hybrid model for gearbox temperature prediction and fault diagnosis [6]. Cheng et al. attempted to integrate digital twins and transfer learning to aid fault prediction using prior knowledge [7]. These studies have their own specific focuses, but most of them still focus on fault identification or RUL results.

A gearbox generally shows a gradual deterioration process before an actual failure; thus, it is difficult to identify the type of fault directly, and this state is often not easily observed by any single sensor. Sensors collect external indicators, such as vibration and temperature; therefore, it is necessary to deduce whether the equipment is operating normally, showing slight degradation, or has reached a serious level based on this data. Therefore, this paper takes the hidden states as the key entry point and uses hidden markov model (HMM) to connect the observed data with the degradation states. First, organize the monitoring signals and extract state features; then, estimate the hidden states and state transition relationships; finally, observe the risk trend based on changes in the sequence of states. Previous studies have used HMM for fault prediction in wind turbine gearboxes, this paper will focus more on applying it to long-term degradation state description.

2. Related research in wind turbine gearboxes

This section analyzes three representative works, and their main contents are shown in Figure 1.

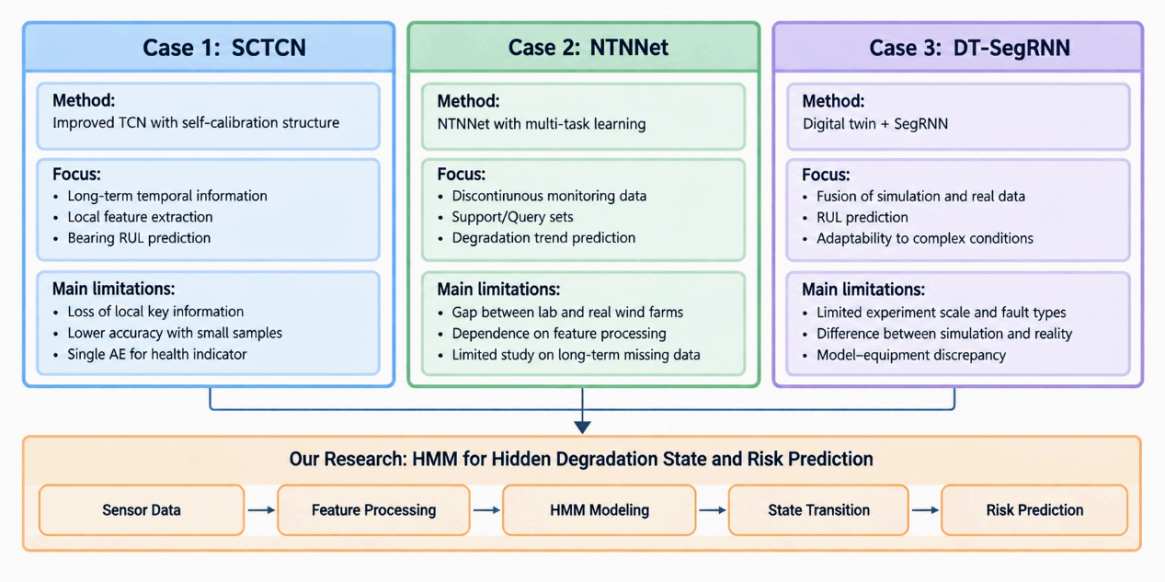

Figure 1. Analysis and summary of representative works (picture credit: original)

2.1. Case 1: improved TCN for RUL prediction of wind turbine gearbox bearings

TCN can leverage long-range temporal information through dilated causal convolutions, and is therefore commonly used for RUL prediction. However, expanding the receptive field of convolutions does not mean that all local variations can be fully preserved; some short-term changes during bearing degradation may be weakened. He et al. proposed SCTCN, which adds a self-calibration structure to TCN, enabling the network to re-focus on local features while also leveraging long-term information [8]. Relevant experimental results show that this method can improve the RUL prediction performance for wind turbine gearbox bearings.

DOI: 10.54254/2755-2721/2026.37907

2.2. Case 2: NTNNet for degradation trend prediction of wind turbine gearbox components

Yu et al. proposed NTNNet to predict the degradation trend of key gearbox components, addressing the practical issue that wind turbine monitoring data are often intermittent [9]. Rather than requiring long, complete, and continuous time series, this method segments the available data into shorter samples, constructs multiple support sets and query sets, and carries out multi-task learning. Through repeated parameter updates, the model enhances its capability to represent degradation characteristics of key components. Compared with existing methods, this approach maintains good prediction performance even when monitoring data are incomplete, and it provides some reference for actual health assessment.

2.3. Case 3: DT-SegRNN for improving RUL prediction performance

In RUL prediction, relying solely on measured data is often constrained by factors such as insufficient fault samples, while using simulated data alone may fail to match the actual equipment. Lu et al. proposed DT-SegRNN, which combines digital twins with SegRNN [10]. In their study, they first used digital twins to simulate the gearbox life cycle and obtain theoretical RUL values, then combined these with actual experimental data for prediction. However, there are still discrepancies between simulated data and real equipment, and this is a key issue that requires careful consideration in subsequent work.

3. Analysis of limitations and solutions

3.1. Analysis of limitations

3.1.1. Limitations of case 1

SCTCN improves TCN's handling of both long-term and local information, but this improvement is still built on convolutional feature extraction. When training samples are limited, local degradation features may still be weakened, especially fluctuations or abrupt changes that occur over a short period. Besides, if feature fusion relies on fixed thresholds, it won't adapt well to different operating conditions; and using AE alone to construct health indicators doesn't make full use of the complementary information that different sensors might provide. Existing comparisons show that TCN underperforms LSTM for small overlap sizes, suggesting that local information preservation remains a challenge.

3.1.2. Limitations of case 2

NTNNet is a good way to handle the problem of irregular monitoring data, but its validation conditions are still far from those in a real wind farm. Temperature, load and operating conditions are relatively easy to control in the test bench; however, real wind turbines experience more fluctuations in operating conditions over a long period of time. Therefore, it is unknown whether the experimental results can be directly applied to wind farms at present. In addition, the model input still needs feature extraction, selection and construction of degradation indicators; thus, the quality of the front-end processing directly affects the prediction results. Most of the previous research has focused on short-term abnormalities and data deficiencies, but has not adequately addressed extended data gaps or simultaneous sensor failures.

DOI: 10.54254/2755-2721/2026.37907

3.1.3. Limitations of case 3

The first reason for the problem with DT-SegRNN is not a lack of data fusion, but rather that the simulated data does not match real-world equipment. On the one hand, the current experimental scale and types of faults are relatively limited; therefore, their generalization ability under complex operating conditions has not been fully verified. On the other hand, the load, lubrication and vibration conditions in the experiments are relatively standardized and do not fully reflect the long-term fluctuations of actual wind farms. The theoretical RUL obtained by the digital twin is still based on a dynamic model. If the model parameters are far from the actual equipment, adding virtual data will not improve the prediction result.

Although the above studies analyzed the performance of cleaning robots from different perspectives, there are still certain limitations, as shown in Table 1.

Table 1. Limitations of three cases

Cases	Limitations
Case 1	Loss of key local information; reduced prediction accuracy with small samples; strong reliance of the health indicator on a single AE feature
Case 2	Discrepancies between laboratory data and actual wind farms; reliance on front-end feature extraction and degradation indicator construction; lack of research on long-term data loss and simultaneous failure of multiple sensors.
Case 3	Limited experimental scale and fault types; differences between standardized operating conditions and real industrial environments; possible discrepancies between the digital twin model and actual equipment.

3.2. Solutions and recommendations

3.2.1. Solutions and recommendations of case 1

Future work can continue to improve in two aspects: local feature extraction and health indicator construction. In terms of network architecture, the self-calibration concept of SCTCN can be retained, and a dedicated branch can be set up specifically for extracting short-term variations, then fused with long-term features to reduce the smoothing-out of local information. For the pooling part, an adjustable convolution method can also be tried, enabling the sampling range to adjust according to data changes. A temporal convolutional autoencoder can be employed for health indicator analysis to avoid using fixed thresholds and, instead, verify the monotonicity and trend consistency of the indicators.

3.2.2. Solutions and recommendations of case 2

The following verifications are to create a simulated environment for a wind farm. Collect long-term data on all kinds of wind speeds, loads, temperatures and operating conditions, and carry out continuous tests of real turbine units. At the same time, data gaps of different extents can be

DOI: 10.54254/2755-2721/2026.37907

artificially created by adding missing values or using recorded fault events to determine the range of applicability of the model; that is, extended interruptions and simultaneous sensor failures. Front-end feature extraction can also be simplified to learn degradation features more effectively from data closer to the raw vibration signal. The effect of different sample sizes on the results for the Support Set and Query Set will be compared. Finally, in the online test, record the prediction time, computational resources and false alarms at the same time to conduct a full-coverage evaluation of its engineering value.

3.2.3. Solutions and recommendations of case 3

The three levels for improvement are data, digital twin modelling and practical application. Increase the number of samples at different rotational speeds, loads, temperatures and lubrication conditions, add faults such as wear, cracks, pitting and tooth breakage, and perform cross-validation with public datasets for the data. A model can be built based on real sensor data to continuously update material parameters, contact parameters and degradation parameters in a digital twin, thus bringing the simulation results closer and closer to those obtained in practice. Attention mechanisms, transfer learning and uncertainty estimation can be used to address the stability problem caused by noise and changes in operating conditions in the data-driven part. Finally, an extended-duration online test will be conducted on a real wind turbine, and at this time, the accuracy of prediction, response speed and computational cost will all be evaluated.

In response to the above shortcomings, improvements can be made in the following aspects, as shown in Table 2.

Table 2. Proposed improvement strategies for the three cases

Cases	Improvement Suggestions
Case 1	Adopt a dual-branch self-calibration structure to enhance local feature extraction; introduce a learnable sampling strategy; use a temporal convolutional autoencoder to improve the health indicator.
Case 2	Add long-term data from actual wind farms; expand discontinuous-data scenarios; reduce reliance on manually constructed degradation indicators; investigate the impact of Support/Query Set size, and conduct online tests.
Case 3	Expand experiments under multiple operating conditions and fault types; validate with public datasets; update digital twin parameters using real-time sensor data; conduct long-term online validation.

4. Conclusion

This paper investigates the long-term degradation status of wind turbine gearboxes and explores why HMMs are suitable for this problem. The three ways to solve the problems of local temporal information, intermittent monitoring data and the use of simulated data are SCTCN, NTNNet and DT-SegRNN, respectively. However, changes in operating conditions and data quality of real wind farms are also reasons for reduced model performance. HMM is a direct prediction method; it does not have the state layer that HMM adds. Sensor data, such as vibration and temperature, are used as observations for the model, and the health status of the equipment is represented by hidden states. Determine whether the degradation is slowing down, remaining constant or increasing in the future

DOI: 10.54254/2755-2721/2026.37907

based on the model. Although this formula cannot directly conduct fault diagnosis or remaining useful life prediction, it is more intuitive in displaying changes in equipment health. The following work will still need to integrate long-term data from real wind farms to practically verify the state partitioning, transition probabilities and risk thresholds.

References

[1] Carroll, J., Koukoura, S., McDonald, A., et al. (2019). Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques. *Wind Energy, 22*(3), 360-375.

[2] Zeng, C., Yao, X., Tang, D., et al. (2016). Fault prediction of wind turbine gearbox based on improved HMM. *Acta Energiae Solaris Sinica, 37*(4), 1017-1023.

[3] Liu, C. (2016). Fault prediction of wind turbine gearbox based on hidden semi-Markov model. Master's thesis, North China Electric Power University, Beijing, China.

[4] Kang, J. (2023). Fault warning analysis of wind turbine gearbox based on machine learning. *Electrical Technology and Economy, 9*, 156-158.

[5] Meng, K. (2023). Research on fault warning method of wind turbine gearbox system based on graph convolutional neural network. Doctoral dissertation, North China Electric Power University, Beijing, China.

[6] Wang, H., Zhao, X., & Wang, W. (2023). Fault diagnosis and prediction of wind turbine gearbox based on a new hybrid model. *Environmental Science and Pollution Research, 30*(9), 24506–24520.

[7] Lu, Q., Cheng, L., Zhu, D., et al. (2026). Fault prediction method towards gearbox based on digital twin and deep transfer learning. *Engineering Research Express, 8*, Art. no. 085506.

[8] He, K., Su, Z., Tian, X., et al. (2022). RUL prediction of wind turbine gearbox bearings based on self-calibration temporal convolutional network. *IEEE Transactions on Instrumentation and Measurement, 71*, 1–12.

[9] Yu, X., Zhang, Z., Tang, B., et al. (2025). Non-temporal neural networks for predicting degradation trends of key wind-turbine gearbox components. *Renewable Energy, 243*, Art. no. 122438.

[10] Lu, Q., Li, M., & Huang, X. (2025). Digital twin and data-driven remaining useful life prediction of gearbox. *IEEE Access, 13*, 111614–111627.