Multi-Sensor Fault Diagnosis of Wind Turbine Gearbox Via Ensemble Filter Network

WEIXIONG JIANG, CHENGJIE WANG, JUN WU
and HAIPING ZHU

ABSTRACT

The wind turbine gearbox (WTG) is regarded as a pivotal component in modern power generation systems, primarily due to its capability to streamline structural complexity and optimize energy conversion efficiency. Nevertheless, the transmission systems are frequently subjected to premature degradation, which stems from prolonged exposure to stochastic loading patterns and extreme environmental stressors. To address this reliability challenge, a set of symptom parameters is systematically extracted from different aspects. Then, multiple symptom parameters are fed into the proposed ensemble filter network to achieve the precious and robust fault diagnosis of wind turbine gearbox. Experimental validation was conducted on the WTG fault simulation test platform of Beijing Jiaotong University. The results show that the proposed method is competitive with other existing methods in terms of diagnosis accuracy and performance stability.

INTRODUCTION

The operational reliability of wind turbine gearboxes (WTG) in distributed wind energy systems has been extensively documented due to their critical role in achieving high transmission efficiency and stabilized power output. However, the susceptibility to operational failures is significantly heightened by complex structural configurations and extreme environmental loading conditions. Empirical studies have identified predominant failure patterns including roller bearing abrasion, outer race spalling, gear tooth fracture, and so on. This necessitates the development of advanced diagnostic frameworks capable of achieving accurate and robust fault diagnosis to mitigate operational risks and optimize maintenance resource allocation.

The analysis of nonlinear and nonstationary vibration signatures constitutes a fundamental challenge in WTG fault diagnostics. Considerable research efforts have been dedicated to the development of advanced signal processing methodologies, encompassing high-fidelity signal decomposition [1], adaptive noise suppression [2], and automated feature engineering paradigms [3]. For example, independent component analysis (ICA) has been extensively employed for blind source separation of multivariate vibration data through non-Gaussian component extraction. Principal

Weixiong Jiang, Chengjie Wang, Jun Wu
Huazhong University of Science and Technology, School of Naval Architecture and Ocean Engineering, Luoyu Road, 1037, Wuhan, Hubei, China
Haiping Zhu
Huazhong University of Science and Technology, School of Mechanical Science and Engineering, Luoyu Road, 1037, Wuhan, Hubei, China

component analysis (PCA) and SAMMON mapping have been predominantly utilized for linear and nonlinear data projection over recent decades. Further algorithmic refinement was achieved through the implementation of curvilinear distance analysis (CDA), which incorporates nonlinear extensions of multidimensional scaling to enhance convergence characteristics. However, the persistent challenge underscores the necessity for developing robust feature extraction techniques to reflect the WTG operation condition overall.

On the other hand, multi-sensor fusion strategies have become indispensable for comprehensive WTG fault characterization [4]. The existing approaches require sophisticated integration of heterogeneous sensor data streams as diagnostic evidenced. For instance, a multi-sensor fusion framework was proposed for induction motor diagnostics under static operating conditions through binary classification schemes. The weighted ensemble learning architectures was proposed to enhance sensitivity for bearing defect identification in high-speed rail applications. A hybrid diagnostic model was proposed to significantly improve classification accuracy for rotating machinery in petrochemical facilities. However, the diagnostic reliability of these methods has been affected by the low signal-to-noise ratio (SNR) measurements from distributed sensor arrays. Therefore, the fault diagnosis is a troublesome challenge with multiple sensor data.

Fortunately, ensemble learning methods have exploited in recent years, and they are widely adopted for the multi-sensor fault diagnosis, such as random forest (RF), ensemble support vector machine (SVM), and stagewise additive modeling using a multi-class exponential loss function (SAMME). Inspired by the above, a novel WTG fault diagnosis method based on ensemble filter network is proposed to address the fault diagnosis issue under intricate sensor data.

In this article, we introduce the principle of ensemble filter network. Then, the wind turbine gearbox fault simulation test platform of Beijing Jiaotong University is established for the algorithm validation. Then, the diagnosis results are discussed. Finally, the conclusion is summarized.

ENSEMBLE FILTER NETWORK

The ensemble filter network is systematically divided into three phases. Initially, each Light gradient boosting machines (LightGBMs) are independently trained using multi-sensor data, with adaptive learning weights being iteratively recalibrated based on diagnosis accuracy. Subsequently, multi-sensor data with poor quality and sparse samples are eliminated through precision threshold filtering. Finally, an optimized ensemble of validated SAMME classifiers is synthesized to generate consolidated diagnostic outputs. The flowchart of the ensemble filter network (EFN) is comprehensively illustrated in Fig. 1, and the detailed implementation processes are outlined below.

(1) The raw vibration monitoring data are collected, and their multiple symptom parameters are extracted including the Root mean square, crest factor, waveform index and skewness.

(2) Multiple symptom parameters X_p (p=1,2,…,P) with corresponding fault category labels $L_p \in \{1,2,...,C\}$ are used to train LightGBM$_q$ (q=1,2,…,Q), and their diagnostic accuracies are reserved. Here, P denotes the total sample count, C

represents fault categories, and Q specifies the sensor quantity (Q=8 in current implementation).

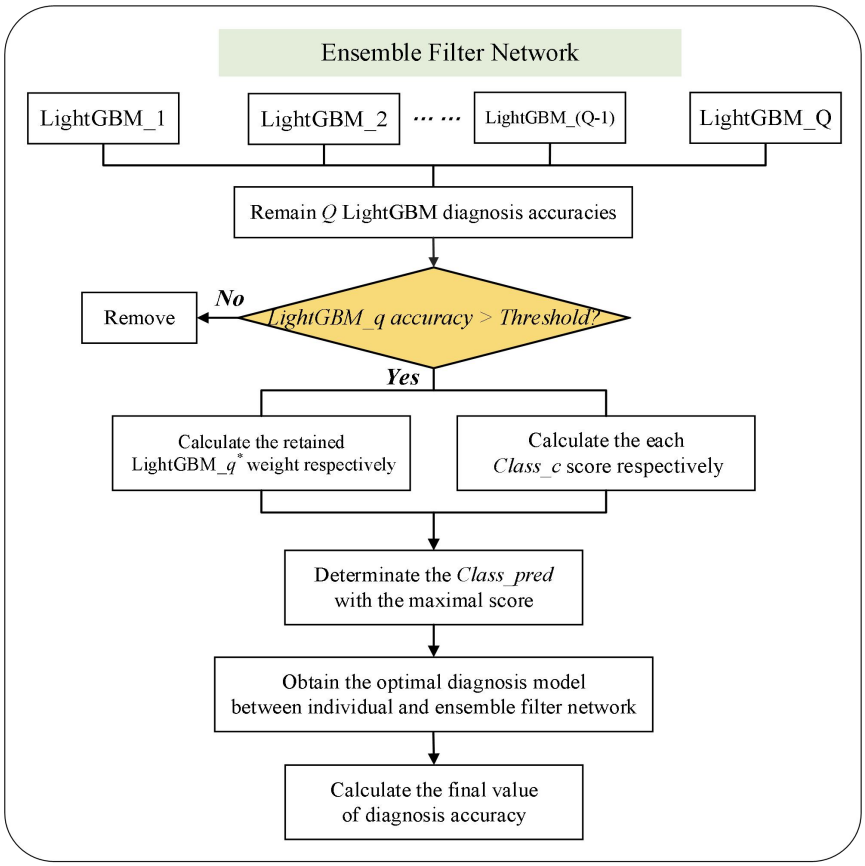

Figure 1. The flowchart of the ensemble filter network

(3) During LightGBM classifier optimization, the based learners are initially trained with uniform weighting coefficients $w_p = 1/P$. Classification error metrics for the *mth* weak learner are computed as:

$$err_m = \sum_{p=1}^{P} w_p \cdot \mathrm{II}(L_p \neq G_m(X_p)) / \sum_{p=1}^{P} w_p \tag{1}$$

where $\mathrm{II}(L_p \neq G_m(X_p)) = \begin{cases} -1 & \text{if } L_p = G_m(X_p) \\ 1 & \text{if } L_p \neq G_m(X_p) \end{cases}$, $G_m(\cdot)$ is the output of LightGBM model.

To enhance model discriminative capacity, adaptive learning weights are dynamically adjusted as:

$$w_p^{(m+1)} = \frac{w_p^{(m)} \cdot exp\left[\beta_m \cdot \mathbb{I}\left(y_p \neq \hat{y}_p^{(m)}\right)\right]}{\sum_{p=1}^{P} w_p^{(m)} \cdot exp\left[\beta_m \cdot \mathbb{I}\left(y_p \neq \hat{y}_p^{(m)}\right)\right]} \tag{2}$$

where $\beta_m = log\left[(1 - \ _m) / \ _m\right] + log(C-1)$, with the supplementary term $log(C-1)$ being incorporated to address multi-class imbalance challenges.

(4) Precision threshold τ is empirically established through cross-validation. LightGBM instances satisfying $Q^* \le Q$ are retained, forming a refined classifier subset $\left\{M_{q^*}\right\}_{q^*=1}^{Q^*}$ with corresponding accuracies $\left\{\alpha_{q^*}\right\}$, where $Q^* \le Q$.

(5) Diagnostic credibility weights γ_{q^*} are assigned proportionally to classifier performance:

$$\gamma_{q^*} = \frac{\alpha_{q^*}}{\sum_{q^*=1}^{Q^*} \alpha_{q^*}} \tag{3}$$

Soft voting integration is subsequently performed as:

$$S_c = \sum_{q^*=1}^{Q^*} \gamma_{q^*} \cdot \mathbb{I}\left(\hat{y}_p^{q^*} = c\right) \tag{4}$$

where $\hat{y}_p^{q^*}$ represents the categorical prediction from M_{q^*}. Final diagnostic determination is constrained by:

$$\hat{y}_p = \arg\max_c S_c \quad \text{s.t.} \quad S_c \ge \eta \cdot \max_{c'} S_{c'} \tag{5}$$

where $\eta \in (0,1)$ denotes confidence margin parameter.

(5) To ensure diagnostic superiority, EFN performance is rigorously validated against optimal constituent classifiers through:

$$\mathcal{A}_{EFS} = \max_{q^*} \alpha_{q^*} + \Delta_{\mathcal{A}} \tag{6}$$

where $\Delta_{\mathcal{A}}$ quantifies ensemble enhancement effects.

EXPERIMENTAL STUDY

The WTG fault simulation test platform of Beijing Jiaotong University is composed of an electric motor, a planetary gearbox, a fixed-axis gearbox, and a loading device, as shown in Figures 2 and 3. Four typical fault patterns are systematically simulated including gear tooth breakage, tooth surface wear, tooth root fracture, and missing teeth. Monitoring data under corresponding fault patterns are collected to validate the effectiveness of the proposed ensemble filter network.

The Sinocera CA-YD-1181 accelerometer is mounted on the base of the planetary gearbox to collect X-axis and Y-axis vibration signals from the sun gear. The rotational speed of the sun gear is captured using an encoder. A sampling frequency of 48 kHz is set for all channels, and monitoring data under various rotational speeds and operational states of the sun gear are acquired. Consequently, the total number of data points per operational state is calculated as 14,400,000 (48,000 Hz × 5 min × 60 s). Finally, the raw monitoring data are segmented into 450 samples per operational state, with each sample containing 2,048 data points. The dataset is allocated as follows: 350 samples are utilized for model training, 50 samples are reserved for validation, and 100 samples are designated for testing. The proposed algorithm was executed with MATLAB R2022b and Python v3.8, and conducted on a personal laptop with Intel Core i7 Processor 14900HK CPU, 32 GB memory, and Microsoft Windows 11 enterprise operation system.

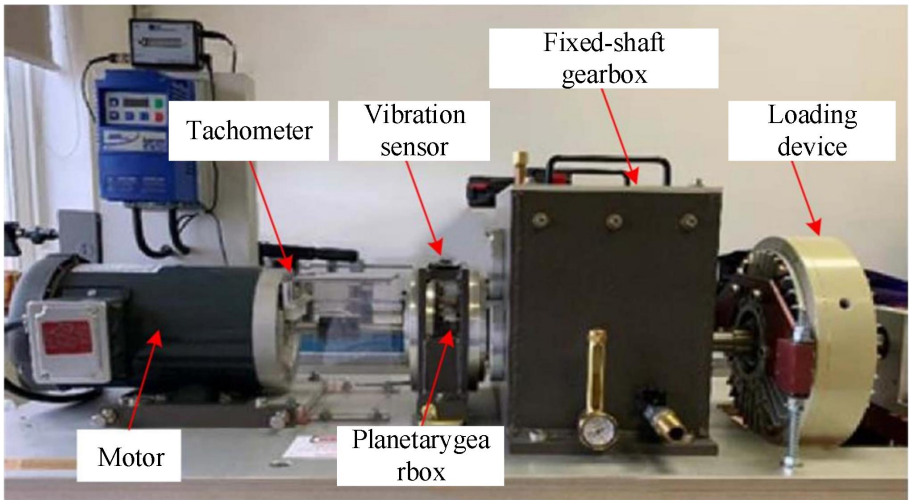

Figure 2. WTG fault simulation test platform

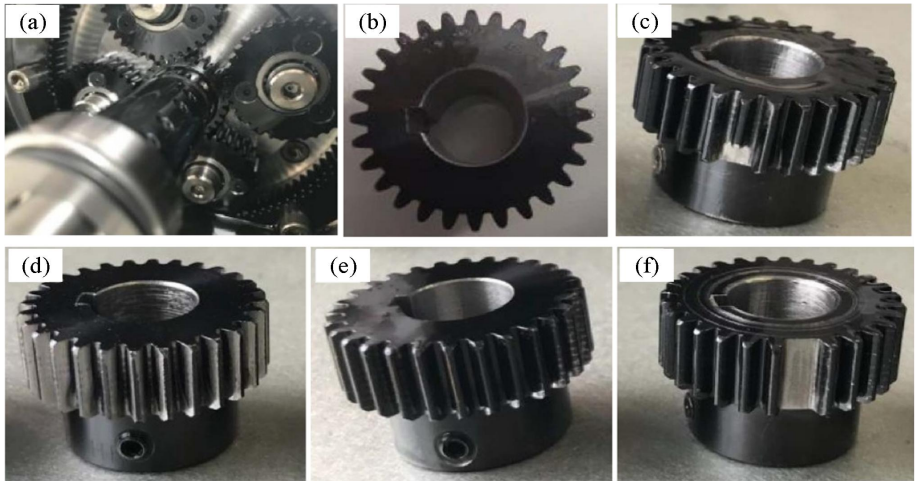

Figure 3. The fault patterns of planetary gearbox5
(a) Internal structure of the planetary gearbox (b) Gear health (c) Gear tooth breakage (d) Tooth surface wear (e) Tooth root fracture (f) Missing teeth

RESULTS AND DISCUSSION

TABLE I THE DIAGNOSIS ACCURACY AND SCORING WEIGHTS

Trial number	The scoring weights of each LightGBM under five repeated trials	
	LightGBM 1	LightGBM 2
1	0.491 (91.60%, 458/500)	0.509 (95.00%, 475/500)
2	0.491 (91.40%, 457/500)	0.509 (94.60%, 473/500)
3	0.494 (92.60%, 463/500)	0.506 (95.00%, 475/500)
4	0.493 (93.00%, 465/500)	0.507 (95.80%, 479/500)
5	0.502 (94.20%, 471/500)	0.498 (93.60%, 468/500)

TABLE II FAULT DIAGNOSIS ACCURACY OF EFN UNDER FIVE REPEATED TRIALS

No.	Trail-1	Trail-2	Trail-3	Trail-4	Trail-5
Accuracy	95.90% (959/1000)	96.30% (963/1000)	96.50% (965/1000)	96.70% (967/1000)	96.20% (962/1000)

The raw vibration signals under five fault patterns are illustrated in Figure 4. Symptom parameters are extracted from the raw vibration signals and inputted into the proposed Ensemble Fusion Network (EFN) for fault diagnosis. To eliminate randomness in diagnostic results, five repeated trials are conducted. The diagnostic accuracy and scoring weights of individual LightGBM models based on single-sensor data across the five trials are summarized in Table I.

As shown in Table I, the highest diagnostic accuracies achieved by the two LightGBM models across five repeated trials are reported as 94.20% and 95.80%, with average accuracies of 92.6% and 94.8%, respectively. The fault diagnosis accuracy of the proposed method over five repeated trials is tabulated in Table II. From the results, the proposed EFN is demonstrated to outperform any single-sensor diagnostic approach. This superiority is primarily attributed to the dynamic adjustment of scoring weights based on real-time sub-classifier accuracy, where higher weights are automatically assigned to sub-models with enhanced diagnostic performance due to superior data quality. Compared to manual optimal sensor selection, the automated sensor screening strategy is proven to be more objective and time-efficient, particularly when the optimal sensor for specific fault modes is not predetermined.

To further validate the advantages of the proposed method, comparative analyses are performed against conventional machine learning-, ensemble learning-, and deep learning-based fault diagnosis approaches, specifically: density-based spatial clustering of applications with noise (DBSCAN) [6], bagging [7], and deep residual shrinkage network (DRSN) [8]. To mitigate the impact of sample selection bias, all methods are trained ten times on identical datasets. The resulting diagnostic accuracies are compared in Table III.

From Table III3, the proposed diagnostic method achieves the highest accuracy (96.71%), followed by DRSN (85.43%), Bagging (84.78%), and DBSCAN (80.03%). In addition, the EFN only takes 38.51 s to implement the diagnosis task, followed by the Bagging (45.17 s), DBSCAN (57.21 s), and DRSN (79.14 s). The comparison analysis result indicates that the proposed EFN has advantage in terms of execution efficiency and diagnosis precise.

TABLE III THE FAULT DIAGNOSIS RESULTS OF COMPARATIVE METHODS

Diagnosis methods	Diagnosis strategies	Execution time (s)	Average accuracy ± standard deviation	
			Training set	Test set
Method-1 [6]	DBSCAN	57.21	82.51% ± 0.3181	80.03% ± 0.3805
Method-2 [7]	Bagging	45.17	86.51% ± 0.2354	84.78% ± 0.2951
Method-3 [8]	DRSN	**79.14**	90.81% ± 0.1513	85.43% ± 0.2841
Method-4 [Ours]	EFN	**38.51**	98.53% ± 0.0059	96.71% ± 0.0683

CONCLUSION

In this article, the multi-sensor fault diagnosis method named EFN is proposed. It can achieve the precious and robust fault diagnosis for the wind turbine gearbox. The experiment results indicate that the accuracy of the EFN is up to 96.71%, and the algorithm running only spend 38.51s. Compared with three popular diagnosis methods, the proposed method has competitive performance in terms of diagnosis efficiency and precision.

REFERENCES

1. Gim, H., Song, D., Hwang, J., Hwang, H., and Park, Y., 2015. Using engineering data comparison method for condition based maintenance system of Offshore Platform. Proceedings of the International Offshore and Polar Engineering Conference. 11(1), pp. 455–459.
2. He, K., Zhang, X., Ren, S., and Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 11(2), pp. 770–778.
3. Hwang, H., 2015. Introduction to a condition-based maintenance solution for offshore platforms. Proceedings of the International Offshore and Polar Engineering Conference. 12(1), pp. 460–463.
4. Jiang, W., Wu, J., Wang, C., Zhu, H., and Wang, X., 2024. Health assessment of wind turbine gearbox via parallel ensemble and fuzzy derivation collaboration approach. Advanced Engineering Informatics. 62(1), 102576.
5. Jiang, W., Wu, J., Zhu, H., Li, X., and Gao, L., 2023. Paired ensemble and group knowledge measurement for health evaluation of wind turbine gearbox under compound fault scenarios. Journal of Manufacturing Systems. 70(8), pp. 382–394.
6. Hahsler, M., Piekenbrock, M., Doran, D., 2019. Dbscan: Fast density-based clustering with R. Journal of Statistical Software, 91(1), pp. 1-13.
7. Cheng, G., Li ,Z., and Yao, X., 2017. Remote sensing image scene classification using bag of convolutional features. IEEE Geoscience and Remote Sensing Letters, 14(10) pp. 1-5.
8. Sun. Y., Pang, S., and Zhang, J., 2024. DRSN-GAF: deep residual shrinkage network (DRSN) for lithology classification through well logging data transformed by gram angle field. IEEE Geoscience and Remote Sensing Letters, 21(2), pp. 1-5.