Aerodynamic Noise-Based Fault Detection for Wind Turbine Blades: An Online Unsupervised Approach

Genwu Chen
Anhui Yufeng Energy Technology Co., Ltd.
Shanghai, China
genwuchen@163.com

Yiwei Wei
Anhui Yufeng Energy Technology Co., Ltd.
Shanghai, China
yiweiwei@proton.me

Chuanqing Pu*
College of Smart Energy
Shanghai Jiao Tong University
Shanghai, China
sashabanks@sjtu.edu.cn
Corresponding author

Feilong Fan
College of Smart Energy
Shanghai Jiao Tong University
Shanghai, China
feilongfan@sjtu.edu.cn

Nengling Tai
College of Smart Energy
Shanghai Jiao Tong University
Shanghai, China
nltai@sjtu.edu.cn

Wentao Huang
Department of Electrical Engineering
Shanghai Jiao Tong University
Shanghai, China
hwt8989@sjtu.edu.cn

***Abstract*—Wind turbine blades are vulnerable to environmental damage, which can compromise the energy conversion efficiency of wind turbines and precipitate safety hazards. This paper proposes an online unsupervised method for detecting faults in wind turbine blades based on aerodynamic noise. The proposed method uses the Empirical Mode Decomposition (EMD) algorithm to process the aerodynamic noise signals and extract feature vectors. A domain of normal operation is delineated using these feature vectors derived from standard aerodynamic noise. During the real-time operation, the latest aerodynamic noise sample is integrated into this normative domain and analyzed using the DBSCAN clustering algorithm to pinpoint anomalies. Case studies conducted on the MIMII dataset indicate that the proposed method is more effective in detecting anomaly status of wind turbine blades from the aerodynamic noise. The proposed unsupervised approach is designed to address the challenge of limited anomaly samples in real-world datasets, demonstrating significant robustness. It is anticipated to facilitate rapid deployment within wind farm SCADA systems.**

***Index Terms*—Wind turbine blades, Aerodynamic noise, Fault detection, Unsupervised learning, Empirical Mode Decomposition, DBSCAN**

I. Introduction

Wind power generation has become one of the significant sources of electricity in power systems. As of 2023, the cumulative installed capacity of wind power in China has exceeded 400 million kilowatts [1]. The blades of wind turbines are among the key components of wind power generators. Natural environmental factors such as windblown sand, salt fog, and lightning can cause varying degrees of damage to the blades [2]–[5], thereby affecting the energy conversion efficiency of wind turbines and even leading to major safety incidents. Therefore, it is crucial to monitor the operational status of wind turbine blades.

Correct and rapid detection can facilitate timely responses to blade damage. The high cost of manual inspections necessitates the development of automated detection algorithms. Common technologies employed for assessing blade damage include infrared thermography [6], X-Ray [7], acoustic emission [8]–[10], and ultrasonography [11]. However, most of these methods are invasive, requiring not only costly hardware but also necessitating turbine shutdown for sensor installation and replacement.

During the operation of wind turbines, the rotation of the blades will stir the surrounding air and generate various states of airflow, resulting in aerodynamic noise [12]. The aerodynamic noise produced by the blades under normal and abnormal operating status exhibits different characteristics [13]. The difference can serve as a basis for fault detection. Acoustic analysis methods offer advantages such as simple and flexible equipment installation, low cost, and non-contact detection. The collection and analysis algorithms of aerodynamic noise can be integrated into the existing Supervisory Control and Data Acquisition (SCADA) system of wind farms as a complement to current state monitoring methods.

Identifying the operational status of wind turbine blades (WTBs) through aerodynamic noise can be framed as a pattern recognition problem. Existing research employs supervised learning techniques to accomplish this task. Supervised learning methods involve learning the relationship between aerodynamic noise generated by WTBs and the operational status of the WTB based on the samples that have been labeled with different categories. Current detection models in this task include Support Vector Machines (SVM) [14], Logistic Regression [15], and Convolutional Neural Networks (CNNs) [16]. However, there are significant limitations to the supervised learning approach. Firstly, during the operation of wind turbines, there is a wide variety of blade faults. It is impractical to intentionally create or collect comprehensive patterns of anomaly aerodynamic noise. Secondly, labeling a large volume of aerodynamic noise data incurs high labor costs. Thirdly, the probability of blade faults in wind turbines is relatively low, often leading to an imbalance between normal and anomaly samples in the dataset, which results in a low recall rate for the detection model.

In summary, the primary challenge in detecting faults in WTBs through aerodynamic noise lies in learning from datasets that contain only a few anomaly samples, or even datasets without any anomaly samples, to build a detection model. To address these challenges, this paper contributes the following:

1) A non-invasive method for detecting WTB faults is proposed, which has the potential to be integrated into online monitoring systems. The proposed method utilizes the aerodynamic noise collected during the operation of wind turbines as the basis for detection, requiring no downtime for monitoring and allowing for real-time adaptive threshold adjustments.

2) The proposed method for detecting faults in WTBs is unsupervised. In the training phase, the method uses the collected aerodynamic noise during the normal operation of wind turbines to construct a domain of normal operation. In the real-time operation phase, it identifies anomaly aerodynamic noise through the DBSCAN clustering method. The proposed method only requires a sufficient number of normal samples for learning, without the need for expensive manual labeling.

The rest of this paper is organized as follows. Section II introduces the implementation framework of the proposed method. Section III presents the fault detection algorithm for aerodynamic noises that can be deployed in SCADA systems, and Section IV validates the effectiveness of the framework through case studies. Section VI concludes the paper.

II. Proposed Framework

The implementation framework of the proposed method is illustrated in Figure 1. This framework is mainly divided into a preparation phase and an online monitoring phase. During the preparation phase, aerodynamic noise generated from wind turbines operating under normal conditions is sampled. The raw aerodynamic noise samples are then processed through a signal decomposition and feature extraction module to obtain a set of representative feature vectors as compressed information for model recognition. These normal feature vectors are used to construct the normal operational domain of these wind turbines to be detected. In the online monitoring phase, the latest sampled signals, after undergoing the same signal decomposition and feature extraction, are merged with the original normal operational domain. The determination of whether the latest sampled signals are anomalies is made through an unsupervised pattern recognition method. The anomaly samples are then sent to the SCADA system for further analysis and response, and the normal operational domain is updated accordingly.

III. Fault Detection Algorithm

A. Signal Decomposition and Feature Extraction

For each wind turbine, the aerodynamic noise $X(t)$ transmitted from the microphone sensor unit will be processed through the Empirical Mode Decomposition (EMD) algorithm

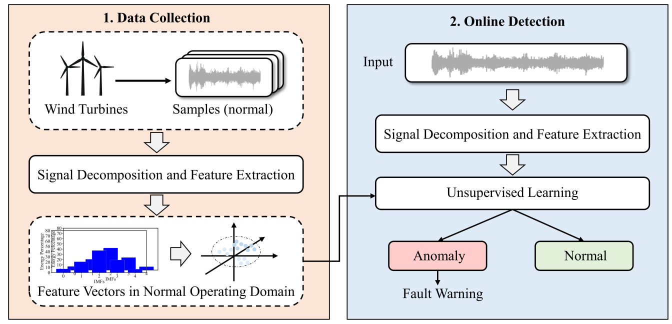

Fig. 1. The implementation framework of the proposed method.

[17] to obtain different Intrinsic Mode Functions (IMFs). The process of EMD is as follows:

1) Identify all the local maxima and minima of the signal $X(t)$.

2) Connect all the local maxima and minima with cubic spline lines to form the upper and lower envelopes $u(t)$ and $v(t)$ of the signal.

3) Subtract the mean envelope from the original signal, as follows:

$$h(t) = X(t) - \frac{u(t) + v(t)}{2} \tag{1}$$

4) Juge whether $h(t)$ is an IMF. If $h(t)$ satisfies the (2) and the two envelopes of $h(t)$ are symmetrical about the abscissa, it is an IMF:

$$|N_{extr}(t) - N_{zero}(t)| \leq 1 \tag{2}$$

where $N_{extr}(t)$ and $N_{zero}(t)$ are the number of extrema and zero crossings of $h(t)$, respectively. If $h(t)$ is an IMF, the process is terminated; otherwise, $h(t)$ is treated as the original signal $X(t)$ and the process is repeated.

5) Subtract $c(t)$ from the original signal $X(t)$ to obtain the residual signal $r(t)$, as follows:

$$r(t) = X(t) - c(t) \tag{3}$$

6) If $r(t)$ satisfies the termination condition, the process is terminated; otherwise, $r(t)$ is treated as the original signal and the process is repeated.

The original signal $X(t)$ is decomposed into a series of IMFs and a residual signal $r(t)$, as shown in (4):

$$X(t) = \sum_{i=1}^{N} c_i(t) + r(t) \tag{4}$$

where $c_i(t)$ is the i-th IMF, and N is the number of IMFs.

The original signal $X(t)$ is decomposed into a series of IMFs through EMD, with the first N IMFs being taken as representative components of the original signal. The energy spectra of these filtered IMFs are used as feature vectors. By the discrete signal energy formula and Parseval's theorem, the energy of each IMF is:

$$E_i = \sum_{t=1}^{T} c_i^2(t) \tag{5}$$

where $c_i(t)$ is the i-th IMF, and T is the length of the signal. The energy percentage of each IMF is expressed as:

$$P_i = \frac{E_i}{\sum_{i=1}^{N} E_i} \tag{6}$$

The feature vectors are constructed from the energy percentages of these IMFs, as expressed in (7). The feature vectors represent the composition of modal components with different characteristics in the WTB aerodynamic noise signals, thereby enabling effective differentiation of distinct aerodynamic noise features of WTBs.

$$\mathbf{F_j} = [P_1, P_2, \cdots, P_N] \tag{7}$$

where index j represents the j-th sampled data/ The feature vectors of the normal operational domain are obtained through the same process. The normal operational domain can be represented as a set of feature vectors $\mathcal{F} = \{\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_M\}$, where M is the number of normal samples.

B. Unsupervised Pattern Recognition

During the online deployment and operation phase of the proposed method, pattern recognition will be conducted on the latest sampled aerodynamic noise data, based on the established normal operational domain. The latest sampled data are processed through the aforementioned signal decomposition and feature extraction module, resulting in a new set of energy spectrum vectors $\mathbf{F}_{new}$. The latest energy spectrum vectors are merged with the constructed normal operational domain to form clusters of energy spectrum vectors: $\mathcal{F}_{new} = \{\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_M, \mathbf{F}_{new}\}$.

The DBSCAN clustering algorithm [18] is then applied to the new feature vectors to identify the anomaly samples. The DBSCAN algorithm is a density-based clustering algorithm that groups together points that are closely packed together, marking as outliers points that lie alone in low-density regions. The DBSCAN algorithm is defined by two parameters: ε and $MinPts$. ε is the maximum distance between two samples for one to be considered as in the neighborhood of the other, and $MinPts$ is the minimum number of samples in a neighborhood for a point to be considered as a core point. The DBSCAN algorithm is described as follows:

1) For each point $\mathbf{F}_i$ in the dataset $\mathcal{F}_{new}$, the ε-neighborhood of $\mathbf{F}_i$ is defined as:

$$N_\varepsilon(\mathbf{F}_i) = \{\mathbf{F}_j \in \mathcal{F}_{new} | \|\mathbf{F}_i - \mathbf{F}_j\| \leq \varepsilon\} \tag{8}$$

2) A point $\mathbf{F}_i$ is a core point if the number of points in its ε-neighborhood is greater than or equal to $MinPts$.

3) For each core point $\mathbf{F}_i$, the points in its ε-neighborhood are grouped together as a cluster. If two clusters have overlapping points, they are merged into a single cluster.

4) Points that are not core points and do not belong to any cluster are marked as outliers.

The DBSCAN algorithm is capable of identifying clusters of arbitrary shapes and size, and is robust to noise. The algorithm is also capable of identifying outliers, which are often the anomaly samples. The proposed method uses the DBSCAN algorithm to identify the anomaly samples in the latest sampled aerodynamic noise data. If the abnormal point identified by DBSCAN is $\mathcal{F}_{new}$, then the current sampling point corresponds to the abnormal working condition of WTB.

IV. Case Studies

We illustrate a set of aerodynamic noise signals that were correctly classified by the proposed method, including the original signal, frequency spectrum, and feature vectors obtained after EMD processing, as shown in Figure 2. For the aerodynamic noise of WTBs under normal and abnormal conditions, both the frequency spectrum and energy spectrum vectors exhibit significant differences, providing a recognizable basis for the detected algorithm.

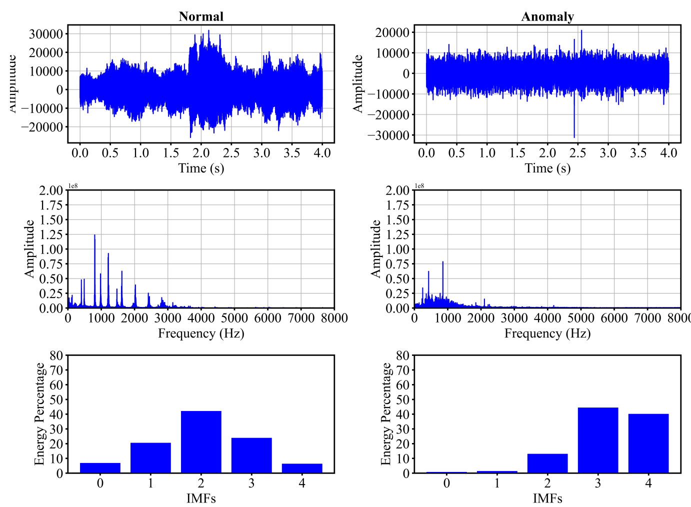

Fig. 2. Aerodynamic noise signals and their feature vectors.

The performance of the proposed method was further evaluated on the MIMII dataset [19]. The MIMII dataset is a sound dataset for the investigation and inspection of faulty industrial machines, composed of sounds from real machines collected by eight microphones. The anomalous sounds in these datasets were collected by intentionally damaging the target machines. The dataset for each machine contains about 1,000 normal sound samples for training, as well as 100-200 samples each of normal and abnormal sounds for testing. The distribution of normal and anomaly samples of the data set used is shown in Table 1.

TABLE I
Distribution of Normal and Anomaly Sample

Type	Normal	Anomaly	Total	Length(s)
train	911	2	913	10
test	100	407	507	10

The dataset related to fan blades was used to validate the proposed detection method. The normal operating domain will be constructed using the provided training set. We comprehensively compared the performance of the proposed model with other supervised learning methods. The results are shown in Table 2. The precision, recall, and F1 score are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F1Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{11}$$

where TP is the number of true positive samples, FP is the number of false positive samples, and FN is the number of false negative samples.

The compared supervised learning models include Support Vector Classification (SVC), lightGBM, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GaussianNB), and Multi-Layer Perceptron (MLP). The compared supervised learning model aims to minimize the binary cross-entropy of classification, which is defined as:

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^{N} y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \tag{12}$$

where N is the number of samples, y_i is the true label, and $\hat{y}_i$ is the predicted label.

TABLE II
PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH SUPERVISED LEARNING METHODS

Model	Precision	Recall	F1 Score	Size(MB)
Proposed	0.75	0.9	0.82	/
SVC	0.197	0	0	/
lightGBM	0.2	0.0049	0.0096	0.0023
Decision Tree	0.2	0.0049	0.0096	9
Random Forest	0.2	0.0049	0.0096	9
KNN	0.197	0	0	/
GaussianNB	0.22	0.047	0.077	
MLP	0.197	0	0	/

The proposed method achieved a higher recall rate and F1 score than the supervised learning methods. Due to the imbalance of the dataset, the SVM, KNN, and MLP models achieved almost 0 recall rate and F1 score. The Decision Tree, Random Forest, and lightGBM models also failed to achieve good metrics. The proposed method not only achieved a classification accuracy of 75%, but also achieved a recall rate of 0.9 and an F1 score of 0.82, indicating that the proposed method is more effective in detecting the aerodynamic noise of wind turbine blades. In addition, the proposed method also has a smaller model size, making it more suitable for deployment in SCADA systems.

The confusion matrix of the proposed method and the supervised learning methods on the test set is shown in Figure 3. It can be seen from the confusion matrix that the predicted results of the proposed method on the test set are basically consistent with the true results, and the proportion of misclassified normal samples and missed anomaly samples is relatively small. Due to the lack of anomaly samples in the training set, the supervised learning methods tend to classify the anomaly samples as normal samples in the training set.

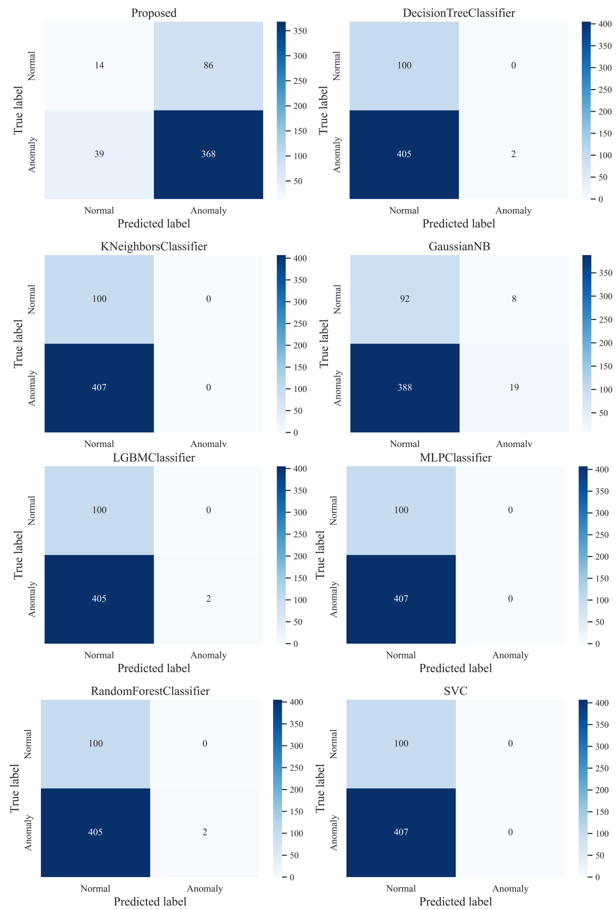

Fig. 3. Confusion matrix of the proposed method on the test set

To demonstrate that the evaluation of the proposed method is not heavily dependent on the number of anomaly samples in the training set, we adjusted the number of anomaly samples in the training set. The number of anomaly samples in the training set was increased from 2 to 260, after which both the proposed method and other methods were evaluated. As the number of anomaly samples in the training set increased, the changes in accuracy, recall, and F1 score for different models on the test set are shown in Figure 4. It is evident that the proposed method achieves good performance under varying numbers of anomaly samples. This indicates that the proposed method has good robustness. Conversely, the performance of the supervised learning methods is heavily dependent on the number of anomaly samples in the training set. The supervised learning methods exhibit poor performance when the number of anomaly samples is low. The performance of the supervised learning methods is significantly improved as the number of anomaly samples in the training set increases. However, these methods still do not match the performance of the proposed method. This indicates that the supervised learning methods are not suitable for detecting the aerodynamic noise of wind

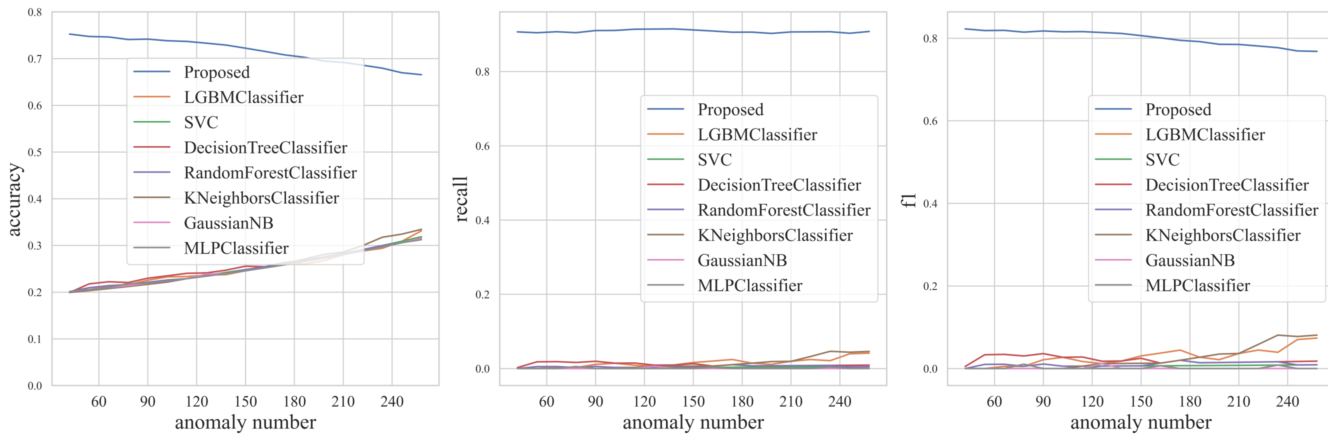

Fig. 4. Performance of different models with varying numbers of anomaly samples in the training set

turbine blades, as the number of anomaly samples in the training set is often limited.

V. Conclusion

This paper proposed an online unsupervised method for detecting faults in wind turbine blades based on aerodynamic noise. The proposed method was validated on the MIMII dataset, and the results show that the proposed method achieved a higher recall rate and F1 score than supervised learning methods, indicating that the proposed method is more effective in detecting anomaly samples in the aerodynamic noise of WTB. The proposed method has the potential for rapid deployment in WTB detection systems and can adapt to datasets without labeled anomaly samples. Even in the absence of labeled samples, online monitoring can preliminarily be realized by collecting aerodynamic noise samples from normally operating wind turbines. It is noted that while the performance of the proposed method on the MIMII dataset surpasses some supervised learning algorithms, it still lacks in accuracy. It is necessary to optimize the hyperparameters of the proposed method, such as the selection of the number of IMFs and the initial parameter settings of DBSCAN. Future work is awaited to deploy the method presented in this paper onto actual wind farm SCADA systems for validation, to further improve the effectiveness of the proposed method through real aerodynamic noise sampling.

Acknowledgment

This work was supported in part by Shanghai Pujiang Program under Grant 22PJ1405000, in part by the National Key R&D Program of China under Grant 2023YFA1011301, and in part by the National Natural Science Foundationof China under Grant 52337006.

References

[1] National Energy Administration, ”National power industry statistics in 2023”, National Energy Administration, 2023. [Online]. Available: https://www.nea.gov.cn/2024-01/26/c_1310762246.htm. [Accessed: 20-Feb-2024].
[2] B.Chen, et al. ”Acoustical damage detection of wind turbine blade using the improved incremental support vector data description,” in Renewable Energy, vol. 156, pp. 548-557, 2020.
[3] Tang, J., et al. ”An experimental study of acoustic emission methodology for in-service condition monitoring of wind turbine blades,” in Renewable Energy, vol. 99, pp. 170–179, 2016.
[4] Liu, W., et al. ”The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review,” in Renewable and Sustainable Energy Reviews, vol. 44, pp. 466–472, 2015.
[5] Du, Y., et al. ”Damage detection techniques for wind turbine blades: A review,” in Mechanical Systems and Signal Processing, vol. 141, pp. 106445, 2020.
[6] Yang, B., et al, ”Non-destructive testing of wind turbine blades using an infrared thermography: A review,” in 2013 International Conference on Materials for Renewable Energy and Environment, 2013, pp. 407–410.
[7] Mishnaevsky Jr, L., et al. ”Micromechanisms of leading edge erosion of wind turbine blades: X-ray tomography analysis and computational studies,” in Wind Energy, vol. 23, no. 3, pp. 547–562, 2020.
[8] Tang, J., et al. ”An experimental study of acoustic emission methodology for in service condition monitoring of wind turbine blades,” in Renewable Energy, vol. 99, pp. 170–179, 2016.
[9] Z. Liu, X. Wang, L. Zhang. ”Fault diagnosis of industrial wind turbine blade bearing using acoustic emission analysis,” in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 9, pp. 6630–6639, 2020.
[10] Tang, J., et al. ”A pattern recognition approach to acoustic emission data originating from fatigue of wind turbine blades,” in Sensors, vol. 17, no. 11, pp. 2507, 2017.
[11] C. Xuefeng. ”Research and application of condition monitoring and fault diagnosis technology in wind turbines,” in Journal of mechanical engineering, vol. 47, no. 09, pp. 45, 2011.
[12] W. Liu. ”A review on wind turbine noise mechanism and de-noising techniques,” in Renewable Energy, vol. 108, pp. 311–320, 2017.
[13] Zhu, W., et al, ”Wind turbine blade fault detection by acoustic analysis: Preliminary results,” in 2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2021, pp. 1–5.
[14] Q. Cai, et al. ”Wind turbine blade defect identification based on acoustic feature extraction and support vector machine,” (in Chinese) in Software Engineering and Applications, vol. 10, pp. 454, 2021.
[15] B. Liu, et al. ”Fault diagnosis of wind turbine blades based on MFCC sound feature signal extraction,” in Equipment Management and Maintenance, no. 12, pp. 148–149, 2022.
[16] Xiang, L., et al. ”Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism,” in Measurement, vol. 175, pp. 109094, 2021.
[17] Rilling, G., et al, ”On empirical mode decomposition and its algorithms,” in IEEE-EURASIP workshop on nonlinear signal and image processing, 2003, pp. 8–11.
[18] Schubert, E., et al. ”DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN,” in ACM Trans. Database Syst., vol. 42, no. 3, 2017.
[19] Harsh Purohit, undefined., et al, ”MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection,” 2019.