Split-Point and Attribute-Reduced Classifier Approach for Fault Diagnosis of Wind Turbine Blade through Vibration Signals

A Joshuva *, M Arjun, B Sujith Adhithya, B Akash and S Abdul Wahaab

Centre for Automation & Robotics (ANRO), Hindustan Institute of Technology and Science, Chennai 603103, Tamilnadu, India.
*Email: joshuva1991@gmail.com

Abstract. This study proposes a data processing and analysis of wind turbine blade faults using split-point and attribute-reduced classifier (SPAARC) through statistical-machine learning approach. In this study, the fault like erosion, hub-blade loose connection, pitch angle twist, bend and crack faults have been simulated and the vibration data has been taken using a piezoelectric accelerometer. With the recorded data, statistical features where extracted and with the extracted features were used to classify the fault condition on the wind turbine blade through SPAARC. The classification accuracy was found to be 85.67% and validated through 10-fold-cross-validation.

Keywords: Turbine blade fault; SPAARC; J48; statistical; 10-fold cross classification

1. Introduction
The International Renewable Energy Agency (IREA) has estimated that the operation and maintenance costs of wind turbines compensate for a substantial 20-25% of the total electricity costs leveled. For on- and off-shore activities, servicing is directly responsible for up to 80% of these costs [1-3]. Corrective or unplanned servicing, which results in replacements if the pieces become broken, can be very costly, in particular when major components failure like wind turbine blade. The wind turbine is especially deficient in its blades due to various factors. Such corners are the central segment for the extraction of wind energy. The turbine must be shut down and the physical examination must be done to identify defects in wind turbine cuts. This will cause a massive reduction in the production of productivity and will require a high labor charge [4, 5].

Innovative iterative non-linear filter and morphological analysis, Liu and Zhang conducted research on naturally weakened wind turbine blade with fault detection [6]. The goal is to eliminate heavy noise from iterative nonlinear filters and to isolate poor vibrating vibrations. In order to diagnose a bearing defect in the frequency domain, a morphological transformations dependent envelope approach is then used. The findings of the diagnostics indicate that the method introduced can be an important diagnostic tool for very weak layers of blades and is preferable to conventional failure diagnostic procedures. Tu et al., [7] have been working to assess blade damages from portable devices by using wind turbine noise.
This article is intended to use the wind turbine noise for the blade inspection process and to conduct on-site wind farm inspections by utilizing built portable devices. The downside of this device is that during service it may check wind turbine blades.

Liu et al., [8] performed a study in an observational wavelet thresholding system on the vibration analysis of large wind turbine fault detection blades. Failure of blade coats will substantially reduce energy output, so the detection of defective blade costly maintenance and accidental loss is important to avoid. However, the key problems in diagnosing low speed blade are that several noise disruptions obscure the poor fault vibration signals and the effective vibrate data are very limited. Freeman et al., [9] performed the study of the fault signature of the marine hydrokinetic turbine blade with the help of continuous wavelet transform. This paper discusses the effects of hydrodynamic asymmetry problems on marine kinetic turbines and the difficulties faced by the non-intrusive identification and condition control approaches when trying to identify such defects.

Chen et al., [10] presented a study on the algorithm to find errors in deep learning on variable-speed wind turbines. To evaluate the signal characteristics of the fault a long-term memory model (LSTM) is developed. The focus system for improving efficiency is introduced in the LSTM. The results of the simulation show that the proposed approach can detect an imbalance error with more than 98 percent accuracy which demonstrates the effectiveness of the proposed approach on the imbalance detection of wind turbine sheets. Various works were done utilizing recreation investigation of issue and structure examination of wind turbine cutting edge and different parts of the wind turbine like gearbox; be that as it may, just a not many in the exploratory examination for wind turbine sharp edge was done [11-14]. Figure 1 shows the methodology of the work done.

Figure 1. Methodology

Figure 2. Experimental Setup

2. Experimental Studies
The trial arrangement, faults, and working technique are explained in detail in Joshuva and Sugumaran (2020) [15]. Figure 2 shows the experimental setup. The frequency of parameters used in the calculation was 12,000 Hz, with 10,000 information for each signal. The DAQ (NI-USB 4432) has been coupled with the Piezoelectric Accelerometer (DYTRAN 3055B1), which was used to store data. Figure 3 and Figure 4 shows the simulated fault on the blade and its vibration pattern.
3. Feature Extraction
In this study, descriptive statistical features have been extracted from the reordered signal and the procedure for feature extraction has been provided in [16]. The statistical features like mode, minimum, standard error, maximum sum, mean, median, range, skewness, kurtosis, standard deviation and sample variance.

4. Feature Selection
Feature Selection is the process where the user automatically or manually select those features which contribute most to the prediction variable or output [17]. Having irrelevant features in the data can decrease the accuracy of the models and make your model learn based on irrelevant features. In this study, J48 decision tree algorithm has been used as feature selection tool. The detailed procedure and the selection of technique has been provided in [18].

5. Feature Classification
SPAARC (Split-Point and Attribute-Reduction Classifier) [19] is a single tree classification algorithm based on the popular CART (Classification and Regression Tree) algorithm. It incorporates techniques for reducing the computational load, improving processing time whilst minimising effects on classification accuracy. These include ‘split-point sampling’ that reduces the number of split-points used when testing the suitability of an attribute at each node in the tree, plus ‘node-attribute sampling’, whereby a subset of attributes is tested at each alternate horizontal tree node level. Tests revealed the combined effect over a number of training datasets is the reduction in model build time of as much as 69%. This reduced run-time is an ideal feature well suited to constrained devices such as smartphones. The detailed information about SPAARC was provided in [20].

6. Result and Discussion
For good condition cutting edge and other defective wind turbine blade conditions [21], the vibration signals were reported. Maximum set of 600 items; 100 of these were good-fit blade tests. In the case of other failures, such as bending of the blade, corrosion, blade crack, hub-blade loose blade, pitch angle twist, 100 tests have been obtained for each case. Statistical features have been identified as attributes and are used for the classification The required performance of the algorithm will be the corresponding state of the classified data [22]. The sum of instances per leaf and the amount of data for the reduced error cutting was kept at 50 in order to choose 4 predominant features of the J48 decision tree algorithm. The batch size and debug ‘real’ default parameters for SPAARC have now been set as standard. Figure 5 shows the tree flow of the SPAARC classifier with the features. The values in the bracket represents the number of correctly classified instance to the incorrectly classified instances. Figure 6 shows the 10-fold-cross-validation result for SPAARC [23]. Out of 600 instances, 514 (85.67%) instances have been correctly classified and remaining 86 (14.33%) instances are misclassified. The other statistical measures like kappa statistic, mean absolute error, root mean squared error, relative absolute error, root relative squared error values are present (Figure 3).

Table 1 represents the confusion matrix for SPAARC. The first row of the confusion matrix, the primary component (area (1,1)) speaks to the quantity of accurately classified having a place with the equivalent which represents good condition (Class A). The subsequent component (area (1,2)) talks
about the quantity of good cases of bent fault erroneously assigned (bend) (Class B). The third component (area (1,3)) discusses the amount of good cases of crack loss (crack) that have been erroneously delegated (Class C). The fourth component (area (1,4)) speaks of the volume of positive events wrongly called loose fault situations of hub-blade (loose) (Class D). The fifth component (area (1,5)) speaks to the quantity of good occurrences that were inaccurately named pitch angle twist fault condition (PAT) (Class E). The sixth component (area (1,6)) speaks to the quantity of good examples that were mistakenly delegated erosion fault condition (erosion) (Class F) [24].

### Table 1. Confusion matrix for SPAARC

| Class | Class A | Class B | Class C | Class D | Class E | Class F |
|-------|---------|---------|---------|---------|---------|---------|
| Class A | 79 | 0 | 1 | 20 | 0 | 0 |
| Class B | 0 | 84 | 7 | 0 | 9 | |
| Class C | 0 | 8 | 89 | 3 | 0 | 0 |
| Class D | 18 | 0 | 7 | 75 | 0 | 0 |
| Class E | 0 | 0 | 0 | 0 | 98 | 2 |
| Class F | 0 | 4 | 0 | 0 | 7 | 89 |

### Table 2. Class-wise accuracy of SPAARC

| Blade Class | TP | FP | PRE | REC | F-M | MCC | ROC | PRC |
|-------------|----|----|-----|-----|-----|-----|-----|-----|
| Class A     | 0.79 | 0.03 | 0.81 | 0.79 | 0.80 | 0.76 | 0.94 | 0.79 |
| Class B     | 0.84 | 0.02 | 0.87 | 0.84 | 0.85 | 0.83 | 0.95 | 0.80 |
| Class C     | 0.89 | 0.03 | 0.85 | 0.89 | 0.87 | 0.84 | 0.96 | 0.80 |
| Class D     | 0.75 | 0.04 | 0.75 | 0.75 | 0.75 | 0.71 | 0.92 | 0.73 |
| Class E     | 0.98 | 0.01 | 0.93 | 0.98 | 0.95 | 0.94 | 0.98 | 0.89 |
| Class F     | 0.89 | 0.22 | 0.89 | 0.89 | 0.89 | 0.86 | 0.96 | 0.83 |

Thus, the subsequent column speaks to the second condition i.e bend condition. The third line speaks to the information focuses for the third condition, for example crack condition. The fourth line speaks to the information focuses for fourth condition, for example hub-blade loose fault condition. The fifth column speaks to the information focuses for the fifth condition, for example pitch angle twist fault condition. The sixth line speaks to the information focuses for the sixth condition, for example erosion condition.
fault condition. In the confusion matrix, the diagonal components speak to the accurately grouped examples and the others are misclassified instances [25]. Table 2 provides comprehensive reliability in class via the true positive rate (TP), false positive rate (FP), F-measures (F-M), Matthews coefficient of correlation (MCC), precision (PRE), area of the receiver's operating characteristics (ROC), recall (REC), region of precision recall curve (PRC). The true positive (TP) value should be almost 1 in a correct classification and the false positive (FP) should be less than 0. Figure 5 reveals that the TP average in most classes is near 1 and that the FP rate is close to 0. This supports that in Table 1, the outcome shown by the confusion matrix.

7. Conclusion
The wind turbine is extremely important in the day-to-day production of wind power. This study aims to provide the fault classification on wind turbine blade using split-point and attribute reduced classifier (SPAARC). The SPAARC has been validated through 10-fold-cross-validation and the correctly classified instance was found to be 85.67%. This classifier has the minimal mean absolute error (0.0676). SPAARC may therefore practically be used for wind turbine blade condition monitoring to identify conditions of the blade failure.

REFERENCES

[1] Fogaing MB, Gordon H, Lange CF, Wood DH and Fleck BA 2019 A review of wind energy resource assessment in the urban environment InAdvances in Sustainable Energy 2019 7-36.
[2] Liu Z and Zhang L 2020 A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings Measurement 149 107002.
[3] Ti Z, Deng XW and Yang H 2020 Wake modeling of wind turbines using machine learning Applied Energy 257 114025.
[4] Du Y, Zhou S, Jing X, Peng Y, Wu H and Kwok N 2020 Damage detection techniques for wind turbine blades: a review Mechanical Systems and Signal Processing 141 106445.
[5] Zuo H, Bi K and Hao H 2020 A state-of-the-art review on the vibration mitigation of wind turbines Renewable and Sustainable Energy Reviews 121 109710.
[6] Liu Z and Zhang L 2019 Naturally damaged wind turbine blade bearing fault detection using novel iterative nonlinear filter and morphological analysis IEEE Transactions on Industrial Electronics 30.
[7] Tu TH, Lo FC, Liao CC, Chung CF and Chen RC 2019 Using wind turbine noise to inspect blade damage through portable device InINTER-NOISE and NOISE-CON Congress and Conference Proceedings 2019 259 1787-91.
[8] Liu Z, Zhang L and Carrasco J 2020 Vibration analysis for large-scale wind turbine blade bearing fault detection with an empirical wavelet thresholding method Renewable Energy 146 99-110.
[9] Freeman B, Tang Y and VanZwieten J 2019 Marine hydrokinetic turbine blade fault signature analysis using continuous wavelet transform. In2019 IEEE Power & Energy Society General Meeting (PESGM) 2019 4 1-5.
[10] Chen J, Hu W, Cao D, Zhang B, Huang Q, Chen Z and Blaabjerg F 2019 An imbalance fault detection algorithm for variable-speed wind turbines: a deep learning approach Energies 12 2764.
[11] Joshuva A, Aslesh A and Sugumaran V 2019 State of the art of structural health monitoring of wind turbines International Journal of Mechanical and Production Engineering Research and Development 9 95-112.
[12] Jiménez AA, Zhang L, Muñoz CQ and Márquez FP 2020 Maintenance management based on machine learning and nonlinear features in wind turbines Renewable Energy 146 316-28.
[13] Vilar AÁ, Xydis G and Nanaki EA 2020 Small wind: a review of challenges and opportunities InSustaining Resources for Tomorrow 185-204.
[14] Movsessian A, Garcia D and Tcherniak D 2020 Adaptive feature selection for enhancing blade damage diagnosis on an operational wind turbine InProceedings of the 13th International Conference on Damage Assessment of Structures 594-605.

[15] Joshuva A and Sugumaran V 2020 A lazy learning approach for condition monitoring of wind turbine blade using vibration signals and histogram features Measurement 152 107295.

[16] Agrawal B, Dubey S and Dixit M 2019 Standard statistical feature analysis of image features for facial images using principal component analysis and its comparative study with independent component analysis InInternational Conference on Intelligent Computing and Smart Communication 1027-48.

[17] Nagra AA, Han F, Ling QH, Abubaker M, Ahmad F, Mehta S and Apasiba AT 2020 Hybrid self-inertia weight adaptive particle swarm optimisation with local search using C4.5 decision tree classifier for feature selection problems Connection Science 32 16-36.

[18] Sahu SK and Shrivas AK 2020 Comparative study of classification models with genetic search based feature selection technique InCognitive Analytics: Concepts, Methodologies, Tools, and Applications 773-83.

[19] Denison DG, Mallick BK and Smith AF 1998 A bayesian CART algorithm Biometrika 85 363-77.

[20] Yates D, Islam MZ and Gao J 2018 SPAARC: a fast decision tree algorithm. InAustralasian Conference on Data Mining 28 43-55.

[21] Solimine J, Niezrecki C and Inalpolat M 2020 An experimental investigation into passive acoustic damage detection for structural health monitoring of wind turbine blades Structural Health Monitoring 3 1475921719895588.

[22] Wu R, Zhang D, Yu Q, Jiang Y and Arola D 2019 Health monitoring of wind turbine blades in operation using three-dimensional digital image correlation Mechanical Systems and Signal Processing 130 470-83.

[23] Jung K, Bae DH, Um MJ, Kim S, Jeon S and Park D 2020 Evaluation of nitrate load estimations using neural networks and canonical correlation analysis with k-fold cross-validation Sustainability 12 400.

[24] Braun T, Spiliopoulos S, Veltman C, Hergesell V, Passow A, Tenderich G, Borggrefe M and Koerner MM 2020 Detection of myocardial ischemia due to clinically asymptomatic coronary artery stenosis at rest using supervised artificial intelligence-enabled vectorcardiography-a five-fold cross validation of accuracy Journal of Electrocardiology 8.

[25] Hasnain M, Pasha MF, Ghani I, Imran M, Alzahrani MY and Budiarto R 2020 Evaluating trust prediction and confusion matrix measures for web services ranking IEEE Access 8 90847-61.