Rough Set Theory Based Blade Condition Classification on Wind Turbine through Statistical Features

A Joshuva*, K Rakesh Kumar, G S Sriram Gangadhar, S S Dhanush, M Arjun

Centre for Automation & Robotics (ANRO), Hindustan Institute of Technology and Science, Chennai 603103, Tamilnadu, India.

*Email: joshuva1991@gmail.com

Abstract. Wind energy has become a one of the alternative energy source due to fossil fuel crisis. These wind energies are being harvested from the wind through wind turbines. These wind turbines are subjected to various environmental factors and prone to severe vibration on blade. This vibration lead to the catastrophic calamities and cause severe capital loss and wind production loss. This study proposes a data processing and analysis of wind turbine blade faults using rough set theory based feature classification. The feature extraction (statistical features) and the feature selection (J48 decision tree algorithm) methods were used to identify the best features for fault classification. Using rough set theory, with five statistical features, 75.5% of classification accuracy have been obtained for the fault identification on wind turbine blade.

Keywords: Turbine blade fault; rough set theory; J48; statistical; 10-fold cross classification

1. Introduction

Due to total environmental pollution, wind energy has become the most important resource in the present period [1]. Wind power is an excellent renewable energy source and a normal generation alternative. The wind turbine is used for electricity transformation increasing wind energy. The capacity, efficiency, protection, durability, viability, and life-cycle of turbines should be improved to make wind power even more important [2]. The wind turns the rotor post and spins the shaft within the wind turbine. The sensitive, lengthy and versatile thin boundary blades are part of the most incredibly complex power and the friction on the straight edge blades will easily be carried on [3]. The strained and dangerous vibration induces rotor-edge pulverization and harm [4]. The blade is one of the most important components that is often damaged. The vibration of the blade when operating is hard to figure due to its massive development and working condition [5]. The study uses vibration source and rough set theory to diagnose damage on wind turbine blades.

Hsu et al., [6] carried out a research on diagnostics and automated management of wind turbines by way of computational process control and machine learning. This research has studied sensor data by combining elements in the maintenance test list in data mining systems with practitioners' insight. The findings show a high degree of precision: 92.68% for the software Decision Tree and 91.98% for the Random Forest Software. The study shows that wind turbine faults can be observed and components replacement requirements can be forecast by data mining and simulation. Rizk et al., [7] also performed hyperspectral imagery used to identify damage to wind turbine blades and icing. In this analysis it is important to use the hyperspectral imaging device for detecting harm to the wind turbine blade and for icing. Within this study, the harm types, triggers and methods used for identification are mentioned.
specifically. In addition, current issues and successful attempts to examine real harm identification to turbine blades are addressed. The results showed that surface and subsoil defects and icing incidents could be observed in their early occurring phases through hyperspectral imaging.

Ghanbarpour et al., [8] conducted dependable energy extraction in wind turbines using a concept of predictive failure resistance. The point of this article is to achieve control objectives in the event of partial load area disturbances, instability, sensor and actuator faults for wind turbines. The goal is to map the best generator speed by harvesting maximal energy in wind turbine control systems in the partial load region. Jiang et al., [9] proposed to operate on the main fault position in wind turbine systems for graphically semi-supervised temporal deep learning. A different approach to identifying the central failure chain in wind turbine systems utilizing unlabeled data is presented in this paper. The first approach is to train and derive high-level functions from the data is a deep learning algorithm, a stacked sparse auto-encoder. A semi-monitored time study algorithm is then used to generate a pseudo-labeled data set with an unlabeled data set.

![Figure 1. Methodology](image1)

![Figure 2. Experimental Setup](image2)

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 [10]. For the discovery of faults in cutting edges of wind turbine, the machine learning method was considered; however, use was limited on drawing. The analysis was performed under a very limited system of imperfections [11]. This is particularly valid on account of fault finding of the wind turbine cutting edge. Consequently, there is a solid need to plan an issue determination framework that can deal with numerous flaws in wind turbine edges utilizing rough set theory draws near. This study aims to detect distinctive cutting-edge problems with a rough set theory and statistical analysis. The methodology of the work carried out is shown in Figure 1.

2. Experimental Studies
The trial arrangement, faults, and working technique are explained in detail in Joshuva and Sugumaran (2020) [12]. Figure 2 shows the experimental setup. The sampling frequency used in the assessment was 12,000 Hz and each signal has a length of 10,000 data. Piezoelectric accelerometer (DYTRAN 3055B1) was used along with DAQ (NI-USB 4432) for recording the data. Figure 3 and Figure 4 shows the simulated fault on the blade and its vibration pattern.
3. Statistical based 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 [13]. The statistical features like mode, minimum, standard error, maximum sum, mean, median, range, skewness, kurtosis, standard deviation and sample variance.

4. J48 Decision Tree based 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. 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 [14].

5. Rough Set Theory based Feature Classification

The notion of rough set theory (RST) was introduced by Pawlak [15]. It is a principle that emerges from basic research into the conceptual features of information systems. Rough-set principle in comparison repositories was the technique for the retrieval or exploration of information. It is closely related to fuzzy theory. One can use rough set approach to discover structural relationship within imprecise and noisy data [16]. Rough sets and fuzzy sets are complementary generalizations of classical sets. The approximation spaces of rough set theory are sets with multiple memberships, while fuzzy sets are concerned with partial memberships. Soft Computing includes along with rough sets, at least fuzzy logic, neural networks, probabilistic reasoning, belief networks, machine learning, evolutionary computing, and chaos theory. The working procedure of rough set theory is given in [17].

6. Result and Discussion

For good condition blade and other wind turbine blade faults the vibrating signals were reported. A minimum of 600 samples, including 100 from a good condition weapon, have been obtained. In fact,
100 samples of every state were tested for other faults including blade bending, corrosion, blade crack, hub blade loose links, twisting of the pitch. The statistical attributes were derived as features and are used for classification purposes [18]. The relevant performance of the algorithm is the corresponding state of the classified data. For choosing four main features of the J48 decision tree algorithm, the total number of instances and the amount of data used on reduced error pruning were 50. The parameters for RST such as batch size (100) and “real” debug choice were configured as usual. The 10-fold cross-validation result for RST [19] is shown in figure 5.

For 600 instances, 453 (75.5%) were correctly categorized and 123 (20.5%) remaining instances became incorrectly classified with 24 (4%) unclassified instances. 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 5) [20]. Table 1 represents the confusion matrix for rough set theory. The primary component (area (1,1)) speaks to the quantity of accurately classified having a place with the equivalent (Class A). The following component (area (1.2)) refers to the number of good instances of bend failure (bend) (Class B). The third component (area (1,3)) refers to the number of good instances in which crack (crack) was erroneously delegated (Class C). The fourth component (area (1,4)) refers to the quantity of good instances which were wrongly named loose (loose) hub blade defects (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) [21].

| Time taken to build model: 0.64 seconds |
|----------------------------------------|
| === Stratified cross-validation ===     |
| === Summary ===                        |
| Correctly Classified Instances         | 453     | 75.5% |
| Incorrectly Classified Instances       | 123     | 20.5% |
| Kappa statistic                        | 0.7436  |
| Mean absolute error                    | 0.0712  |
| Root mean squared error                | 0.2668  |
| Relative absolute error                | 26.4927 |
| Root relative squared error            | 73.0653 |
| InClassified Instances                 | 24      |       |
| Total Number of Instances              | 600     |       |

Figure 5. 10-fold-cross-validation result for rough set theory

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 [22]. The sixth line speaks to the information focuses for the sixth condition, for example erosion fault condition. In the confusion matrix, the diagonal components speak to the accurately grouped examples and the others are misclassified instances. Table 2 [23] demonstrates the thorough precision of the section. The class-specific consistency refers to real positive (TP) performance, false positives (FP), precision (PRE), recall (REC), F-Measure (F-M), Matthews coefficient of correlation (MCC), receiver area (ROC) area, PRC area [24]. For a better categorizer, the true positive (TP) value
should be almost 1 in a correct classification and the false positive (FP) should be less than 0. Table 2 reveals that the rate of TP in most groups is similar to 1 and the rate of FP almost 0. It means that the finding shown in Table 1 is a confusion matrix [25].

Table 1. Confusion matrix for rough set theory

| Blade Class | Class A | Class B | Class C | Class D | Class E | Class F |
|-------------|---------|---------|---------|---------|---------|---------|
| Class A     | 76      | 1       | 0       | 20      | 0       | 0       |
| Class B     | 2       | 78      | 8       | 0       | 0       | 9       |
| Class C     | 4       | 7       | 61      | 11      | 7       | 2       |
| Class D     | 22      | 1       | 6       | 62      | 2       | 0       |
| Class E     | 1       | 0       | 2       | 1       | 91      | 5       |
| Class F     | 0       | 2       | 2       | 0       | 8       | 85      |

Table 2. Class-wise accuracy of rough set theory

| Blade Class | TP  | FP  | PRE | REC | F-M | MCC | ROC | PRC |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Class A     | 0.78| 0.06| 0.72| 0.78| 0.75| 0.70| 0.85| 0.59|
| Class B     | 0.80| 0.02| 0.87| 0.80| 0.83| 0.80| 0.87| 0.72|
| Class C     | 0.66| 0.03| 0.77| 0.66| 0.71| 0.66| 0.78| 0.53|
| Class D     | 0.66| 0.66| 0.66| 0.66| 0.66| 0.59| 0.77| 0.47|
| Class E     | 0.91| 0.03| 0.84| 0.91| 0.87| 0.84| 0.93| 0.78|
| Class F     | 0.87| 0.03| 0.84| 0.87| 0.85| 0.85| 0.82| 0.90|

7. Conclusion

This study aims to provide the fault classification on wind turbine blade using rough set theory (RST). The RST has been validated through 10-fold-cross-validation and the correctly classified instance was found to be 75.5%. This classifier has the minimal mean absolute error (0.0712). Hence rough set theory for condition monitoring of the wind turbine blade can be employed practically to identify the conditions in blade.

REFERENCES

[1] Wu X, Hu Y, Li Y, Yang J, Duan L, Wang T, Adcock T, Jiang Z, Gao Z, Lin Z and Borthwick A 2019 Foundations of offshore wind turbines: a review Renewable and Sustainable Energy Reviews 104 379-93.
[2] Ti Z, Deng XW and Yang H 2020 Wake modeling of wind turbines using machine learning Applied Energy 257 114025.
[3] 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 7-36.
[4] Madi E, Pope K, Huang W and Iqbal T 2019 A review of integrating ice detection and mitigation for wind turbine blades Renewable and Sustainable Energy Reviews 103 269-81.
[5] Igwemezie V, Mehananparast A and Kolios A 2019 Current trend in offshore wind energy sector and material requirements for fatigue resistance improvement in large wind turbine support structures—a review Renewable and Sustainable Energy Reviews 101 181-96.
[6] Hsu JY, Wang YF, Lin KC, Chen MY and Hsu JH 2020 Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning IEEE Access 8 23427-39.
[7] Rizk P, Al Saleh N, Younes R, Ilinca A and Khoder J 2020 Hyperspectral imaging applied for the detection of wind turbine blade damage and icing Remote Sensing Applications: Society and Environment 5 100291.

[8] Ghanbarpour K, Bayat F and Jalilvand A 2020 Dependable power extraction in wind turbines using model predictive fault tolerant control International Journal of Electrical Power & Energy Systems 118 105802.

[9] Jiang N, Hu X and Li N 2020 Graphical temporal semi-supervised deep learning–based principal fault localization in wind turbine systems Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 12 0959651819901034.

[10] 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.

[11] Xu D, Liu PF, Chen ZP, Leng JX and Jiao L 2020 Achieving robust damage mode identification of adhesive composite joints for wind turbine blade using acoustic emission and machine learning Composite Structures 236 111840.

[12] 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.

[13] Anh NX, Nataraja RM and Chauhan S 2020 Towards near real-time assessment of surgical skills: A comparison of feature extraction techniques Computer Methods and Programs in Biomedicine 187 105234.

[14] Singh J, Bagga S and Kaur R 2020 Software-based prediction of liver disease with feature selection and classification techniques Procedia Computer Science 167 1970-80.

[15] Pawlak Z 1998 Rough set theory and its applications to data analysis Cybernetics & Systems 29 661-88.

[16] Tay FE and Shen L 2003 Fault diagnosis based on rough set theory Engineering applications of artificial intelligence 16 39-43.

[17] Suraj Z 2004 An introduction to rough set theory and its applications ICENCO 3 80.

[18] Lin Z and Liu X 2020 Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network Energy 24 117693.

[19] Huo Y, Xin L, Kang C, Wang M, Ma Q and Yu B 2020 SGL-SVM: a novel method for tumor classification via support vector machine with sparse group Lasso Journal of Theoretical Biology 486 110098.

[20] Xu J, Zhang Y and Miao D 2020 Three-way confusion matrix for classification: a measure driven view Information Sciences 507 772-94.

[21] Meehan AJ, Latham RM, Arseneault L, Stahl D, Fisher HL and Danese A 2020 Developing an individualized risk calculator for psychopathology among young people victimized during childhood: a population-representative cohort study Journal of affective disorders 262 90-8.

[22] Chicco D and Jurman G 2020 The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation BMC genomics 21 6.

[23] Düntsch I and Gediga G 2020 Indices for rough set approximation and the application to confusion matrices International Journal of Approximate Reasoning 118 155-72.

[24] Leong S, Zhao Y, Ribeiro-Rodrigues R, Jones-López EC, Acuña-Villaorduña C, Rodrigues PM, Palaci M, Alland D, Dietze R, Ellner JJ and Johnson WE 2020 Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB Tuberculosis 7 101898.

[25] Zeng G 2020 On the confusion matrix in credit scoring and its analytical properties Communications in Statistics-Theory and Methods 49 2080-93.