Statistically Optimal Vibration Feature Selection for Fault Diagnosis in Wind Turbine Blade

Ahmed Ali Farhan Ogaili [*,**‡] iD, Alaa AbdulhadyJaber[*] iD Mohsin Noori Hamzah[*] iD

[*] Mechanical Engineering Department, University of Technology- Iraq/Baghdad

[**] Mechanical Engineering Department, University of Mustanisiryah -Iraq/Baghdad

(ahmed_ogaili@uomustansiriyah.edu.iq, alaa.a.jaber@uotechnology.edu.iq ,mohsin.n.hamzah@uotechnology.edu.iq,)

[‡]Corresponding author; Ahmed Ali Farhan Ogaili, Tel: +9647702833351

ahmed_ogaili@uomustansiriyah.edu.iq

Abstract- The present study identifies various faults in wind turbine blades from the acquired vibration signals. Various statistically obtained features were computed from time-domain vibration signatures, including kurtosis, skewness, standard deviation, variation, root mean square, and crest factor. To identify faults in wind turbines, a comparison was made with the performance of two classification models, Random Forest (RF) and Support Vector Machine (SVM), using the feature set obtained from the time-domain vibration signals. The results demonstrate these classifiers for fault diagnosis. The use of chi-square (χ^2) statistical feature selection techniques has been found to improve classification accuracy. To test the efficacy of this approach, we compared the proposed model with traditional models using several performance measurements. The findings confirmed that when chi-square (χ^2) is used in conjunction with RF, the proposed model achieved a significant improvement in precision, increasing from 75.3% to 83.315%. These results suggest that the chi-square (χ^2) can be valuable for optimizing feature selection and improving classification accuracy in machine learning models.

1. Introduction

Wind energy is a rapidly growing technology in the energy sector, known for its renewable, environmentally friendly and sustainable nature. The rotor blade of a wind turbine is a crucial component in wind energy production, and unpredictable failures can result in costly damage and repairs, as well as increased operating and maintenance expenses [1-5]. Therefore, early detection and diagnosis of wind turbine failures are crucial for effective condition monitoring. Vibration signals are often used as parameters for machinery conditions, with significant changes indicating developmental failures [6-7]. Wind turbine blades are particularly susceptible to damage due to environmental conditions and vibrations, which can affect them.

The cost of installations and accounts for up to 15- 20% of total costs [8-11]. Therefore, effective monitoring and diagnosis of wind turbine blades is essential for ensuring the sustainability and viability of wind energy as an energy source. The importance of monitoring wind turbine blades cannot be overstated, as it determines whether the monitoring equipment is operating optimally and producing the planned power output. Without sufficient information on the type of error, scheduling maintenance activities in advance becomes difficult. This lack of knowledge and understanding can fail to adequately prepare for potential errors. With greater knowledge and understanding, preparations could be made routinely and appropriately, ensuring maintenance activities are carried

out before any failures. Thus, ongoing research on wind turbine blade monitoring techniques is essential to enhance our understanding of these critical components and ensure the sustainability and effectiveness of wind energy as an energy source.

It is obvious that the blades of a wind turbine are constantly exposed to harsh environmental conditions and therefore susceptible to faults that can affect their performance and even cause catastrophic failure [12-14]. Changes in wind speed, foreign particle interaction, and environmental factors such as rain and snow can cause vibrations in the blades, leading to slow rotation and ultimately turbine failure. Therefore, monitoring the condition of the wind turbine blades is essential to ensure optimal performance and avoid costly shutdowns [14-17]. However, it can be challenging to accurately evaluate blade vibrations online due to the large size of wind turbines, their remote locations, and the various operating conditions they experience. Therefore, the development of effective fault detection systems is crucial to improve wind turbine blade condition monitoring and prevent costly and dangerous failures. With several studies conducted in this area, machine learning has emerged as a promising approach to diagnosing wind turbine defects. For example, Abdul Rahman and Al-Kindi [18] Experimental modal analysis was used to examine the detection of cracks in wind turbine blades. They used step beams to simulate wind turbine blades and applied experimental modal analysis techniques to identify cracks and their propagation. Although this study provided valuable information on the detection of cracks in wind turbine blades, more research is needed to improve effective wind turbine defect diagnosis systems. In a study by Tcherniak et al. [19], a structural health monitoring system based on vibrations of the wind turbine blades was demonstrated. The system detected structural flaws in the blades of the wind turbines, such as fractures, holes, and deformation. Using a semi-supervised learning algorithm, they simulated an artificially inserted edge opening in a blade, increasing the size from 15 cm to 45 cm over time. The system accurately classified the blade flaw. Sahoo et al. [20] incorporated the vibration of the wind turbine blade vibration with different machine learning methods, including decision trees, support vector machines (SVM), and k-nearest neighbor (KNN) to categorize the blade health status. They evaluated algorithms on healthy, bent, fractured, and eroded blades. They discovered that SVM was the most accurate in identifying blade defects (87%), followed by decision trees (82%) and KNN (80.0%). These studies demonstrate the potential of machine learning algorithms for detecting and monitoring faults in wind turbine blades, highlighting the importance of continued research to improve wind energy sustainability. Yang Tao [21-22] studied the effects of various types on the functioning of wind turbines by simulating impeller imbalance faults such as mass imbalance and asymmetric failure of the aerodynamic force. Specifically, mass imbalance faults were found to create a wave of electric power due to the impeller's output torque induced by the gravity of the imbalanced mass. However, asymmetric aerodynamic force faults caused the tower's vibration to produce a wave of the aerodynamic force on the impeller, resulting in a wave of electric power. This indicates that the aerodynamic force and tower vibration models are interconnected, highlighting the importance of considering both factors in diagnosing and monitoring.

Machine learning techniques have become increasingly popular for diagnosing various machine defects. To ensure accurate and effective fault classification, it is essential to identify the appropriate fault signatures for each type of machine element [23-24]. By leveraging machine learning algorithms and selecting the right fault signatures, it is possible to achieve a more precise and effective fault diagnosis, helping to prevent costly downtime and improve overall machine performance.

This investigation aims to detect faults in wind turbine blades using vibration signals as measurement parameters. The experimental work includes a healthy blade and three faulty blades. We developed the instrumentation equipment for the experiment and used it to obtain measured vibration signals. Statistical features were extracted from these signals and machine learning classifiers (including Random Forest and SVM) were trained on these features for fault categorization tasks. We use confusion matrices to evaluate accuracy to compare the classifier's performance. Ultimately, our goal is to identify whether a wind turbine blade operates normally accurately or has a defect and, if a defect is present, to classify the type of defect.

2. Machine Learning Model

2.1. Random Forest (RF)

RF, short for Random Forest, is an AI technique that uses machine learning to assess the health of machine components. It was originally developed by Breiman [25] as an extension of decision trees. In RF, decision trees are built using the CART (Classification and Regression Tree) method, without pruning, to their maximum size. The resulting assembly of trees produces a more accurate and robust model than any individual tree. For a visual representation of the RF construction, see Figure 1. To reduce variation and overfitting, Random Forest combines two techniques, bagging, and random feature selection. Each tree in the forest is constructed with a random subset of features and data samples. The data samples used to build a tree are known as "data in the bag," while the remaining samples are "out-of-bag observations" (OOBs) [26].
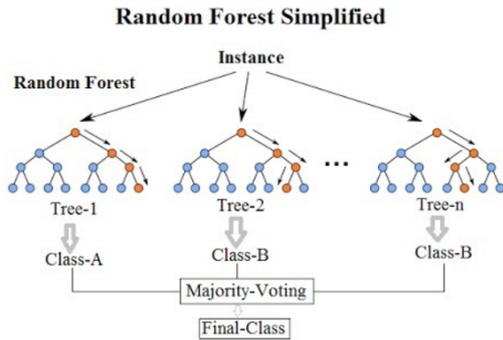
Fig. 1. Construction of a random forest [25].

To determine the class of a given sample, the forest takes a majority vote of all tree classifications. The OOBs estimate the model error rate without requiring a separate validation data set. For each tree t in the forest, the OOBt is the set of samples not used to build that tree. Classification error is typically measured using accuracy, precision, and recall metrics as shown in Eq. (1) [27].

$$errForesr = \frac{1}{n} Cart\{i \in \{1, \dots, n\}, y_i \neq \hat{y}i\} \qquad (1)$$

Where is y_i the most typical class trees suggest for which a sample i is in OOB$_t$.

2.2. Support Vector Machine (SVM)

Support Vector Machines (SVM) is a supervised machine learning technique that uses boundaries to separate and classify data points [28]. SVM works with both linearly separable and non-separable data, and the boundary selection depends on the structure and relationships. The support vectors represent the boundary, and the data points closest to the hyperplane axis. The SVM aims to identify the hyperplane with the maximum margin and the most considerable distance between the hyperplane and the support vectors. This approach results in a robust and accurate model. SVM is a fast and accurate classification algorithm that provides an excellent trade-off between model complexity and performance. For nonlinearly separable data, SVM uses kernel methods to identify similarities and relationships between data points, resulting in a hyperplane representation of the data. The SVM formulation is based on structural risk, which aims to minimize the model's generalization error. For binary classification problems, SVM maximizes the margin between two hyperplanes (H1), which can be used to categorize data into their respective classes. The H1 equation is expressed as follows:

$$x.w + b = 0 \qquad (2)$$

Here, x is the point on the separator plane (H1) and w is the vector on the plane. Normalization of the two class w parameters can be represented as

$$x_i.w + b \leq -1 + \xi_i \text{ for } y_i = -1 \qquad (3)$$

and

$$x_i.w + b \geq 1 + \xi_i \text{ for } y_i = +1 \qquad (4)$$

By combining Eqs. (3) and (4), we get the following.

$$y_i(x_i.w + b) \geq 1 - \xi_i \qquad (5)$$

Here ξ_i represents the slack parameter.

2.3. Validation and evaluation of performance

A confusion matrix is a useful tool to assess the efficacy of a classification method [29]. It enables visualization of a model's performance by comparing expected and actual values. The confusion matrix is typically used to analyze binary classification tasks and is organized as a table. For each class, the matrix displays the number of true positives (T_p), true negatives (T_n), false positives (F_p), and false negatives (F_n). By analyzing the confusion matrix, we can generate several performance indicators for the model, including accuracy, precision, recall, and the F1 score. Precision, calculated as the ratio of true positives to the total number of positive predictions Eq. (6), quantifies the model's ability to identify the positive class accurately. High accuracy suggests that the model finds positive cases effectively, whereas low precision shows several false positives. The confusion matrix is essential to evaluate the performance of a classification model and identify improvement areas.

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + F_n + T_n} \qquad (6)$$

It is crucial to note that this classification metric can lead to inaccurate results, as it measures the correlation between returns [30]. The true positive value is divided into true positives and false positives to assess precision. The recall is computed by dividing the number of true and false negative values by the total number of false positive and false negative values. The F1 score is a combined metric that considers both accuracy Eq. (7) and recalls Eq. (8). The equations used to calculate these metrics are as follows:

$$Recall = \frac{T_p}{T_p + F_n} \qquad (7)$$

$$precision = \frac{T_p}{T_p + F_p} \qquad (8)$$

The F1 score represents the harmonic mean of recall and precision Eq. (9), which offers a fair performance assessment [30]. It is advantageous when dealing with imbalanced class distribution datasets. The F1 score is calculated using the following equation:

$$F1 = \frac{2T_p}{2T_p + F_p + F_n} \qquad (9)$$

Receiver Operating Characteristics (ROC) curves are commonly used to evaluate and compare predictive models [31]. These curves show the correlations between the true and

False Positive Rates (TPR and FPR) to determine the optimal performance of the model. The TPR represents the percentage of positive cases that the model correctly recognized. In contrast, the FPR counts the percentage of negative cases that were mistakenly classified as positive. The uncertainty matrix is used to construct a ROC curve that shows the probabilities of the TPR and FPR values for different classification thresholds. Changing the threshold allows us to plot multiple points on the ROC curve. The area under the curve (AUC) can be calculated as a summary metric of the performance of the model. The equations used to calculate the TPR Eq. (10) and FPR Eq. (11) depend on the classification threshold and the number of true positives, false positives, and false negatives [31].

$$TPR = \frac{T_p}{T_p + F_n} \tag{10}$$

$$FPR = \frac{F_p}{F_p + T_n} \tag{11}$$

The ROC curve displays the true positive rate (TPR) against the false positive rate (FPR) at different classification thresholds, providing a visual representation of model performance [32]. The ROC curve can be adapted to accommodate multiple outcome variables, improving its ability to visualize complex data. The diagonal line of the ROC curve represents a classifier of random guessing. At the same time, the upper left triangle denotes high classification accuracy for both classes and the lower right triangle represents poor classification performance. The ROC curve is widely used in signal detection and in assessing prediction accuracy in various fields such as medicine, economics, climate prediction, and geoscience evaluation [32]. The ROC curve provides a valuable tool for evaluating and comparing machine learning models, allowing us to visualize the trade-off between sensitivity and specificity for different classification thresholds.

2.4. Evaluation Models

Supervised learning categorization was measured using industry standard measures of each method such as precision, recall, and F1 scores. Classification prediction error evaluation, confusion matrices, receiver operating characteristic curves, and precision recall curves were just a few of the evaluation approaches used to illustrate the classification results of the various machine learning models.

In this study, ten-fold cross-validation was used to evaluate the performance of the classification models. This method randomly divides the data into ten parts, each of which contains approximately the same proportion of each class as the original data set. Different parts are tested during each iteration while the model is trained on the remaining nine parts. The error rate is then calculated on the holdout set, and the process is repeated for each of the ten parts. The final error estimate is obtained by averaging the ten error estimates, providing an unbiased and reliable evaluation of the model's performance. All classification models developed in this study followed the ten-fold cross-validation method, ensuring a fair comparison of their performance on the given dataset.

3. Experimental Work

3.1. Experimental setup

The benchmark analysis used wind turbine blades, shown in Figure 2a and provided by Edibon Equipment's Computer-Controlled Wind Energy Unit (EEEC). The EEEC uses a laboratory-scaled aerogenerator to study how kinetic wind energy is converted into electrical energy and how variables affect it. The system consists of a stainless steel tube and computer-controlled variable-speed axial. The aerogenerator includes rotors and a generator that changes rotational speed with air velocity. The airflow from the fan rotates the wind energy, which is then converted into electrical energy by the generator. The angle of the blade can be adjusted and the configurable blades are detachable. EEEC is free-standing experimental test equipment with a 2000×550×550 mm stainless steel tunnel with two 1000×130 mm clear windows. The diameter of the aerogenerator is 510 mm and it can generate nearly 60 W with a charging current of 5 Am and 12 volts. The wind tunnel velocity ranges from 1.3 to 5.3 m/s and serves as a wind source for the start of the wind turbine. To simulate natural wind conditions, the wind speed was continuously altered. The experimental design is depicted in Figures 2(b) and 2(c).

To obtain vibration signals, a piezoelectric accelerometer was used as a transducer. It has a high frequency sensitivity for fault detection and is commonly used in condition monitoring. The PCB Piezotronics 352C65 uniaxial accelerometer was used, which has a sensitivity of (±10%) 100 mV/g (10.2 mV/(m/s²)), a measurement range of ±50 g pk (±491 m/s² pk), a broad resolution of 0.00016 g RMS (0.0015 m / s2) RMS and a frequency range of (±5%) 0.5 to 100,000 Hz. The accelerometer was mounted on the nacelle near the wind turbine hub using an adhesive mounting technique to record vibration data. A connection used to link it to the data acquisition device was an NI USB 4431 with five analog input channels, a sampling rate of 102.4 kilobits per second, and a resolution of 24 bits. The accelerometer is coupled to a signal conditioning instrument that contains an integrated charge amplifier and an analog-to-digital converter (ADC). ADC is used to acquire the vibration signal. The characteristic extraction technique was used to extract the characteristics of these vibration signals. Cables are connected to accelerometers and DAQ, and data capture devices are connected to Lenovo laptops incorporated with Core i7 CPUs, where LabVIEW software is used to collect data [33].
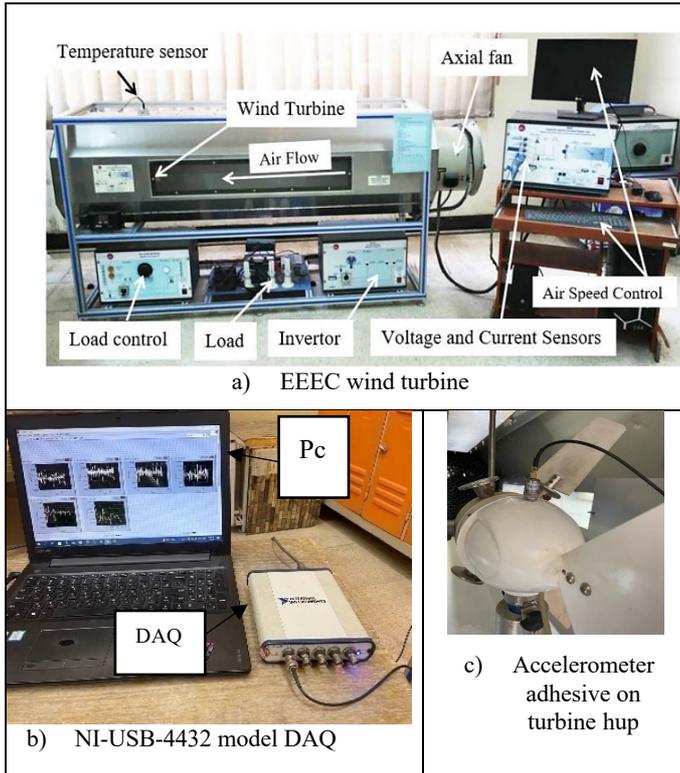
a) EEEC wind turbine

Pc

DAQ

c) Accelerometer adhesive on turbine hup

b) NI-USB-4432 model DAQ

Fig.2. Experimental Setup

3.2. Experimental Procedure

Signals were initially recorded when the wind turbine was healthy, without defects. The accelerometer was used to capture these signals based on the following requirements:

❖ In order to ensure consistency, the duration of the sample was determined after considering several factors. It is essential to have enough samples to make the statistical measurements meaningful. However, as the number of samples increases, so does the computation time. To strike a balance, the sampling frequency was set to meet the Nyquist sampling theorem, which requires the sampling frequency to be at least twice the maximum frequency. This study set the sampling rate at 1000 Hz to achieve this requirement.

❖ LabVIEW 2020 was used to record vibration signals, and at least 500 samples were captured for each state of the wind turbine blade.

❖ The manufacturer designed the wind turbine blades used in this study to closely approximate those of genuine commercial wind turbines. The blades were 300mm long, composed of fiber-reinforced polymer (FRP), and had a solid core. In addition, a variety of blades were identified. The following sections will discuss the models built using these wind turbine blades.

• When all other components are in excellent working condition, the following faults are simulated concurrently, and associated vibration signals are produced. The many failure scenarios modeled on the blade are shown in Figure 3.

• State1: Healthy blade without any defect

• State 2: Blade crack occurs due to damage from foreign objects on the blade while it is operating.

• State 3: Blade erosion. This fault is caused by high-velocity wind erosion of the outermost layer. The flawless surface was degraded using a sandpaper sheet to create an erosion effect.

• State 4: Due to the mass imbalance of the turbine blade used in the investigation, this study added 5 g to the blade at 18 cm from the root, as shown in Figure 3D.
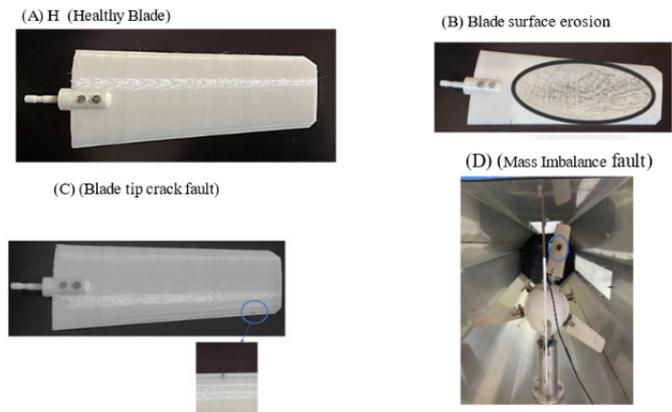

Fig.3. the simulated fault in this study.

3.3. Statistical feature extraction (calculation)

Appropriate signal processing techniques must be employed to extract useful information about the health or condition of components from the non-linear vibration signals obtained using the wind turbine. Time-domain analysis is one technique that allows for a direct examination of the signal pattern, making it easy to calculate characteristics. The characteristics are derived directly from the time waveform and this method has the advantage of simplicity in calculation.

Table 1. Time-domain features

Feature name	Formula
Kurtosis	$\dfrac{N \sum_{i=1}^{N}(x_i - \mu)^4}{[\sum_{i=1}^{N}(x_i - \mu)^2]^2}$
Root Mean Square (RMS).	$\sqrt{\dfrac{1}{N}\left[\sum_{i=1}^{N}(x_i)^2\right]}$
Variance (Var)	$\dfrac{1}{N}\left[\sum_{i=1}^{N}(x_i - \mu)^2\right]$
Standard Deviation (Std Dev)	$\sqrt{\dfrac{1}{N}\left[\sum_{i=1}^{N}(x_i - \mu)^2\right]}$

Vmax	$max\ (xi)$
Skewness	$\dfrac{N\ \sum_{i=1}^{N}(x_i - \mu)^3}{\sigma^3}$
Crest Factor	$\dfrac{Vmax}{RMS}$
Mean μ	$\dfrac{1}{N}\sum_{1}^{N} x_i$

Where *xi* is a signal for *i = 1, 2, N, N* is the number of data points.

3.4. Enhanced Models with Feature Selection

Feature selection necessitates selecting a tiny number of functions from the original subset of functions to reduce dimensionality without compromising the information contained. The characteristics to retain and the ones to eliminate are determined solely by the technique used. When redundant functions are employed, the classifier's efficacy is enhanced. To eradicate an irrelevant feature, it is preferable to employ feature selection criteria that quantify the importance of each feature in the feature set in terms of class labels. Feature classification techniques, such as chi-squared (χ^2), Fisher score, ReliefF, and Information Gain [34], are widely used in various problems. The method of ranking features aims to reduce the dimension, better separate features, and retain the necessary information. Before applying the RF models, this article selected the essential characteristics of the chi-square (χ^2) [35-36]. The Chi-square test (χ^2) selects characteristics based on correlations that determine the correlation between the characteristics and the expected class. Each nonnegative characteristic (Xi) calculates the chi-square statistic to determine which characteristic depends on the predicted attribute. The higher the chi-square score, the higher the chi-square (χ^2) score, as shown in Eq. (12). The feature refers to [37].

$$x^2 = \sum_{j=1}^{N} \frac{(Y_j - u_j)^2}{u_j} \qquad (12)$$

Y_j is the quantity of class observations j, and u_j is the expected value of Y_j. For $u_j = NP_j$, N is the number of observations, and P_j is the probability of occurrence.

4. Results and Discussion

4.1. Measurement of Vibration Signals

The vibration signals in different cases of wind turbine blades, shown in Figure 3, show that the characteristics of the vibration waveform differ considerably under the same working conditions. The state of the wind turbine blade includes a healthy and non-eroding surface, a crack at the tip, and an imbalance. The wind turbine blade rotated at 200 RPM when this condition was detected. Every 50 ms, the vibration signals are measured. Figure 3 shows the unfiltered and filtered vibration signals that the operating wind turbine blades produce. The vibration signals are shown to be healthy in

Figure 3a. In a typical engine room, there would be minimal vibrations. As a result, the measured signals showed that the wind turbine was operating normally.

At the same time, Figure. 3b shows the vibration signals when there is an imbalance fault in the wind turbine blade. Compared to signals in the healthy state, it could be found that vibration signals varied greatly from signals in the other cases. The highest vibration amplitude was approximately 0.248 g. Figure 3c shows the vibration signal on the surface of the damaged blade. It can be found that the vibration signal in this state is higher compared to the healthy state. The highest measured vibration amplitude measured was approximately 0.22 g. The final-state study of the blade tip crack in Figure (3-d) observed that the vibration signals raided in the healthy state and had a higher amplitude of approximately 0.021g.
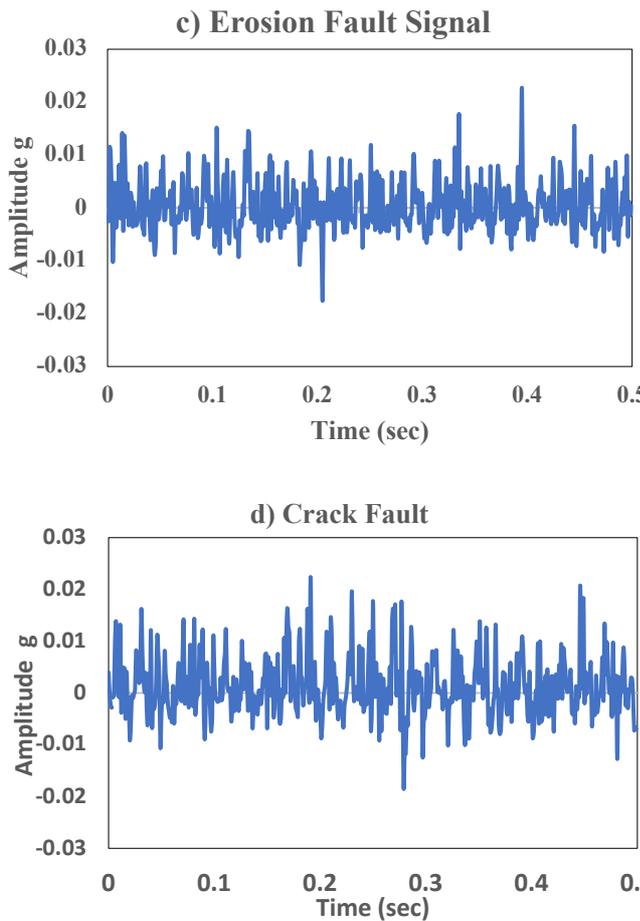

a) Healthy Signal

b) Imbalance Fault
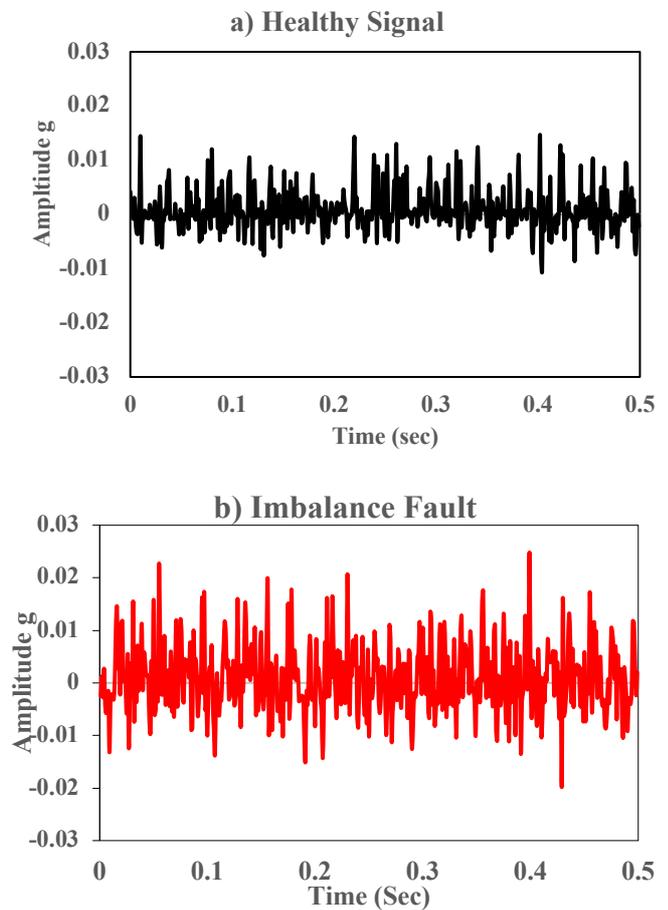
Fig.4. Vibration signals of a wind turbine in a) healthy state, b) imbalance state, c) erosion state, D) crack state.

4.2. Machine Learning Scoring and Evaluation

From the vibration signals, statistical features were extracted. The features were given as input to the RF algorithm and SVM. Classifier to determine the classification accuracy for fault detection in wind turbine blades. Not all features may be necessary for classification. Adding more irrelevant features may negatively impact the performance of the classification algorithm. They also increase the number of computational resources needed. Foreseeing which parameters will be helpful for classification using machine learning algorithms is impossible. Therefore, investigators must extract every descriptive statistical trait before choosing the best. The present investigation improved the model by applying the χ^2 statistical feature selection approach. The method of choosing features was used to choose the four most important ones. The essential characteristics that affect the models in Figure 4, which can be observed in kurtosis, std, dev, RMS, and skewness, are the best characteristics that influence the accuracy (CA) of the modal compared to another characteristic. Tables 2 and 3 illustrate the effectiveness of the FR and SVM models, respectively. Table 2 displays the results before the major feature selection approach.

In contrast, Table 2 displays the identical ones after the primary feature selection method. Tables (2&3) obtained the overall classification accuracy of both model classifiers with respect to the selected characteristic. The classification precision of the basic model was 75.2%, with eight characteristics.

However, the upgraded model that incorporates eight features demonstrated a classification accuracy of 63.6% when using the SVM and a higher accuracy when using the RF model. The maximum classification accuracy for fault detection and type reached 83%, comparable to the results obtained in the reference [16-20]. The computational training time for the RF model was 3.42 s, while the test time was 0.313 s. The improved accuracy also led to an increase in precision and recall. However, the influence of several assessment factors was not significant. Specifically, the specificity improved from 82.89% to 90.95% and the area under the AUC increased from 87% to 92.14%.
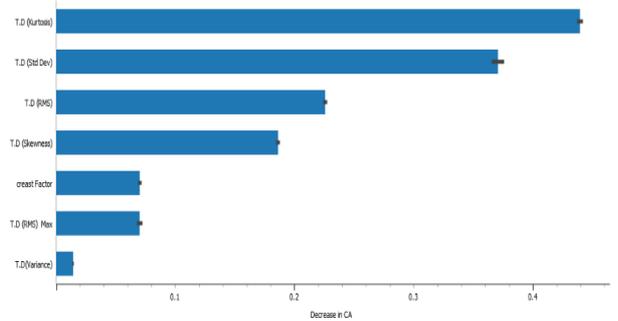
Fig.5. Important Features.

Table 2. Evaluation of Model Results Before Feature Ranking

Modal	RF	SVM
(Area under curve) AUC	0.931	0.878
Classification Accuracy (CA)	0.752	0.636
F1	0.753	0.612
Precision	0.7525	0.603
Recall	0.752	0.636
Specificity	0.917	0.89
Train time (sec)	3.248	1.63
Test time (sec)	0.298	0.225

Table 3. Evaluation after the ranking of features

Modal	RF	SVM
(Area UnderCurve) AUC	0.9623	0.9136
Classification Accuracy (CA)	0.8315	0.732
F_1	0.830788	0.753
Precision	0.8315	0.756
Recall	0.83154	0.75923
Specificity	0.94384	0.91984
Train time (sec)	3.424	1.983
Test time (sec)	0.313	0.254

Table 4. Classwise accuracy for RF and SVM with χ^2

For RF						
class	AUC	CA	F_1	Precision	Recall	Specificity
Healthy	0.9964	0.976	0.9526	0.948	0.957	0.9825
Erosion	0.922	0.855	0.716	0.7	0.733	0.89
Crack	0.991	0.961	0.923	0.9	0.936	0.969
Imbalance	0.941	0.871	0.73	0.766	0.7	0.929
For SVM						
Class	AUC	CA	F_1	Precision	Recall	Specificity
Healthy	0.998	0.984	0.9688	0.948	0.964	0.991
Erosion	0.862	0.777	0.487	0.572	0.424	0.894
Crack	0.993	0.96	0.922	0.9	0.946	0.965
Imbalance	0.89	0.798	0.63	0.578	0.705	0.829

Table 4 presents the class-wise accuracy in detail for RF and SVM with χ^2. The evaluation of classwise accuracy in this article involves the determination of different properties, such as the area under the curve (AUC), the accuracy of classification (CA), precision, recall, the F1 score, and specificity, as presented in Table 4. For the RF model, the AUC values range from 0.922 for erosion to 0.9964 for healthy, indicating that the model has adequate discriminatory power for all classes.

The CA ranges from 0.855 for erosion to 0.976 for healthy, indicating that the model correctly classifies most instances in each class. The F1 score ranges from 0.716 for erosion to 0.9526 for healthy, which is the harmonic mean of precision and recall. The precision ranges from 0.7 for erosion to 0.948 for healthy, which indicates the proportion of true positives among predicted positives. Recall ranges from 0.7 for erosion to 0.957 for healthy, signifying the proportion of true positives among actual positives. The specificity ranges from 0.89 for erosion to 0.9825 for healthy, demonstrating the proportion of true negatives among actual negatives.

For the SVM model, the AUC values range from 0.862 for erosion to 0.998 for healthy, indicating that the model has adequate discriminatory power for all classes except erosion. The CA ranges from 0.777 for erosion to 0.984 for healthy, indicating that the model correctly classifies most instances in each class. The F1 score ranges from 0.487 for erosion to 0.9688 for healthy. The precision ranges from 0.572 for erosion to 0.948 for healthy. The recall ranges from 0.424 for erosion to 0.964 for thriving. Specificity ranges from 0.829 for imbalance to 0.991 for healthy.

Overall, both models demonstrate good performance in the healthy and crack classes, with high AUC values, CA, F1-score, precision, recall, and specificity. Performance in erosion and imbalance classes is lower, particularly for the SVM model. The lower efficacy in these classes may be due to class imbalance or other factors specific to the data set. It would be beneficial to examine the confusion matrices or other performance metrics better to understand the model performance for each class.

Furthermore, the findings were strengthened by a confusion matrix (Figure 6) that evaluated each case individually. Figure 6a shows the confusion matrix for the RF model before the feature selection technique was applied. Illustrates the values for each of the four cases, where the total number of occurrences was divided by 4 to yield 2100 samples per case, with 8400 divided by 4. The results indicate that the healthy state was accurate in 846 cases. However, the erosion signals with erosion (state2) were sometimes misinterpreted as unbalanced states. On the contrary, more than half of the data for the unbalanced blade (state 4) were distributed among the three healthy and damaged states, making it challenging to determine the blade's actual health status. However,

incorporating the χ^2 with the RF model significantly improved the classification accuracy. The confusion matrix of the enhanced RF model is shown in Figure 6b. The model accurately predicted the condition of blades 1, 2, and 3. However, there was confusion between erosion and unbalance blades. Specifically, 712 erosion states were misclassified, resulting in an unbalanced state, and 738 states in the erosion state were classified as unbalanced. Another classifier, the SVM, produced the lowest accuracy in both cases before and after feature classification, as shown in Figure 7.
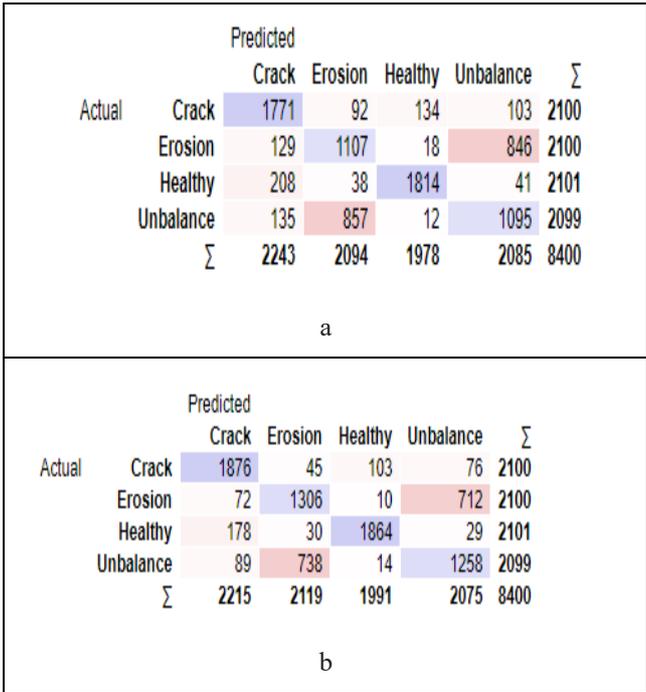
		Predicted				
		Crack	Erosion	Healthy	Unbalance	Σ
Actual	Crack	1771	92	134	103	2100
	Erosion	129	1107	18	846	2100
	Healthy	208	38	1814	41	2101
	Unbalance	135	857	12	1095	2099
	Σ	2243	2094	1978	2085	8400

a

		Predicted				
		Crack	Erosion	Healthy	Unbalance	Σ
Actual	Crack	1876	45	103	76	2100
	Erosion	72	1306	10	712	2100
	Healthy	178	30	1864	29	2101
	Unbalance	89	738	14	1258	2099
	Σ	2215	2119	1991	2075	8400

b

Fig.6. Confusion matrices for the RF: (a) before important feature selection, (b) after important feature selection.

		Predicted				
		Crack	Erosion	Healthy	Unbalance	Σ
Actual	Crack	1423	245	189	243	2100
	Erosion	90	333	5	1672	2100
	Healthy	1383	77	559	81	2100
	Unbalance	149	283	6	1661	2099
	Σ	3045	938	759	3657	8399

a

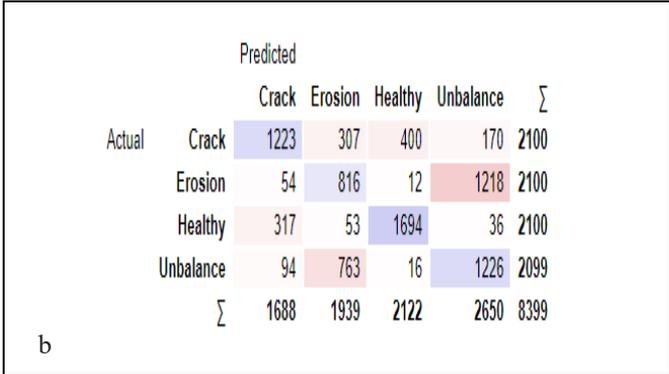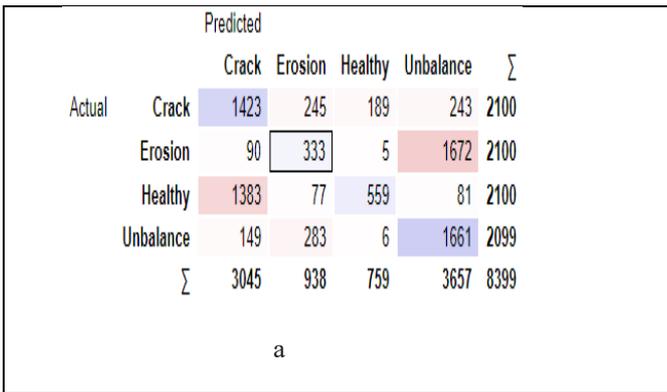|		Predicted				
		Crack	Erosion	Healthy	Unbalance	Σ
Actual	Crack	1223	307	400	170	2100
	Erosion	54	816	12	1218	2100
	Healthy	317	53	1694	36	2100
	Unbalance	94	763	16	1226	2099
	Σ	1688	1939	2122	2650	8399

b

Fig.7. Confusion matrices for SVM: (a) before the selection of important features, (b) after important feature selection.

The ROC curve helps to evaluate the performance of machine learning models. It helps to visualize the balance between sensitivity and specificity by plotting the TPR against the FPR. In Figure 8, we can see the ROC curves for a random forest model that was used to predict the fault state of the wind turbine blades. The results for each state are listed as follows: crack state = 0.977, healthy state = 0.985, erosion state = 0.882, and imbalance state = 0.887. This indicates that the random forest model is a suitable classifier for this task, particularly when increased by χ^2. Furthermore, Figure 9 shows the ROC curve for different states when using SVM as a classifier.
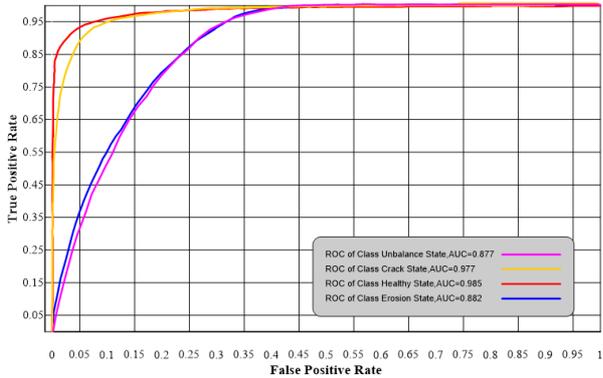

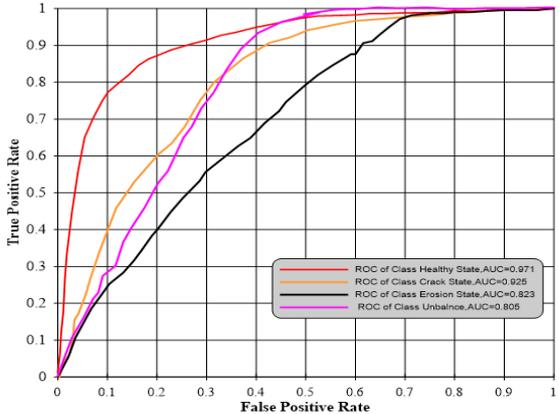**Fig.8.** ROC curve within the score of each class for Random Forest.

Fig.9. ROC curve within the score of each class for SVM.

Random forests are often the most effective when χ^2 and RF are used. However, they are immune to generalization problems. Breiman [25] proposed that the generalization errors of the random forest classification are related to the connection between two trees and the strength of individual trees. A tree with a low error rate was found to be an effective classifier. In contrast, the decline in correlation between classes decreases the generalization error. These factors may help to categorize the random forest classifier more accurately. Moreover, there is only one variable; the user must determine the number of trees. Therefore, the complexity of the classifier is reduced. The drawback of SVM is that it is fundamentally a binary classifier and can be used for multiclass classification with all algorithms. Therefore, the derived average result is computationally intensive and may contain bias. This may contribute to SVM's inferior performance compared to Random Forest.

5. Conclusions

The implementation of machine learning technology has revolutionized the process of identifying, tracking, and diagnosing faults in wind turbines, making them more resilient and easily accessible. Much of the success and proper functioning of artificial intelligence (AI) are based on the acquisition and classification of data. Therefore, the approach in this research involved the use of machine learning algorithms to detect the state of the wind turbine blade based on vibration data from a transducer. We selected three common types of blade failure for the analysis and built models for the customarily used and faulty states. Random Forest (RF) and SVM machine learning algorithms were used as fault classifiers. To improve fault diagnosis and prediction precision while reducing computational load, the χ^2 statistically optimal feature selection method was used to improve precision and efficiency. The results obtained were then used to evaluate the performance of the RF and SVM algorithms. It was discovered that the RF algorithm was more effective than SVM, with a performance increase from 75.3% to 83.315% when employing χ^2. It accurately predicted healthy, cracked, erosion, and imbalance states based on the high values of precision, recall, F_1 score, and analysis of the ROC curve found.

Overall, the findings suggest that machine learning techniques can effectively diagnose and predict wind turbine failures. Further research could focus on improving the accuracy of classifiers to detect erosion faults. Furthermore, more advanced signal processing techniques, such as wavelet transform, which deals with non-stationary vibration analysis, could be applied in future work to extract more indicative features.

Acknowledgments

References

[1] C. Amer and M. Sahin, "Structural analysis of a composite wind turbine blade," International Journal of Aerospace and Mechanical Engineering, vol. 8, no. 7, pp. 1264-1270, 2014.

[2] A. A. F. Ogaili, M. N. Hamzah and A. A. Jaber, "Integration of Machine Learning (ML) and Finite Element Analysis (FEA) for Predicting the Failure Modes of a Small Horizontal Composite Blade," International Journal of Renewable Energy Research (IJRER), vol. 12, no. 4, pp. 2168-2179, 2022.

[3] Y.Eldahab, H.Naggar H. Saad, and Abdalhalim Zekry, "Assessing Wind Energy Conversion Systems Based on Newly Developed Wind Turbine Emulator." International Journal of Smart Grid-ijSmartGrid ,Vol.4, pp 139-1270.4.

[4] M. Yesilbudak, S. Sagiroglu, and I. Colak, "A wind speed forecasting approach based on 2-dimensional input space," in 2012 International Conference on Renewable Energy Research and Applications (ICRERA), 2012: IEEE, pp. 1-5.

[5] M. Yesilbudak, S. Sagiroglu, and I. Colak, "A new approach to very short term wind speed prediction using k-nearest neighbor classification," Energy Conversion and Management, vol. 69, pp. 77-86, 2013.

[6] Ogaili, A.A.F., Hamzah, M.N., Jaber, AA "Free Vibration Analysis of a Wind Turbine Blade Made of Composite Materials" International Middle Eastern Simulation and Modeling Conference 2022, MESM 2022, 2022, pp. 203–209.

[7] Lu, B., Li, Y., Wu, X. and Yang, Z.: 'A review of recent advances in wind turbine condition monitoring and fault diagnosis, 2009 IEEE power electronics and Machines in wind applications, 2009, pp. 1-7

[8] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," Renewable and Sustainable Energy Reviews, vol. 13, no. 1, pp. 1-39, 2009.

[9] I. Colak, S. Sagiroglu, M. Yesilbudak, E. Kabalci, and H. I. Bulbul, "Multi-time series and-time scale modeling for wind speed and wind power forecasting part I: Statistical methods, very short-term and short-term applications," in 2015 International Conference on Renewable Energy Research and Applications (ICRERA), 2015: IEEE, pp. 209-214

[10] S. K. Abraham, V. Sugumaran, and M. Amarnath, "Acoustic signal based condition monitoring of gearbox using wavelets and decision tree classifier," Indian Journal of Science and Technology, vol. 9, no. 33, pp. 1-9, 2016.

[11] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, and M. Papaelias, "Condition monitoring of wind turbines:

Techniques and methods," Renewable energy, vol. 46, pp. 169-178, 2012.

[12] W. Y. Liu, B. P. Tang, J. G. Han, X. N. Lu, N. N. Hu, and Z. Z. He, "The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review," Renewable and Sustainable Energy Reviews, Volume 44, April 2015, Pages 466-472.

[13] A. Joshuva and V. Sugumaran, "A data-driven approach for condition monitoring of wind turbine blade using vibration signals through the best-first tree algorithm and functional trees algorithm: a comparative study," ISA (Instrum. Soc. Am.) Trans. 67 (2017) 160- 172.

[14] Yesilbudak, Mehmet, Seref Sagiroglu, and Ilhami Colak. "A novel implementation of kNN classifier based on multi-tupled meteorological input data for wind power prediction." Energy conversion and management 135 (2017): 434-444.

[15] O. Eyecioglu, B. Hangun, K. Kayisli, and M. Yesilbudak, "Performance comparison of different machine learning algorithms on the prediction of wind turbine power generation," in 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA), 2019: IEEE, pp. 922-926.

[16] A. Joshuva and V. Sugumaran, "Fault Diagnosis for Wind TurbineBlade through Vibration Signals Using Statistical Features and Random Forest Algorithm," IJPT, Vol. 9, April-2017.

[17] Joshuva, A., Kumar, R. S., Sivakumar, S., Deenadayalan, G. & Vishnuvardhan, R. (2020). An in-depth look at VMD to diagnose wind turbine blade faults using C4. 5 as feature selection and discriminating through multilayer perceptron. Alexandria Engineering Journal, 59(5), 3863-3879.

[18] K.F. Abdulraheem, G. Al-Kindi, A Simplified Wind Turbine Blade Crack Identification Using Experimental Modal Analysis (EMA), Int. J. Renew. Energy Research (IJRER) 7 (2) (2017)715–722.

[19] D. Tcherniak, L.L. Mlgaard, Active vibration-based structural health monitoring system for wind turbine blade: demonstration on an operating Vestas V27 wind turbine, Structural Health Monitoring. 16 (5) (2017) 536–550.

[20] Sahoo, S.; Kushwah, K.; Sunaniya, A.K. Health Monitoring of Wind Turbine Blades Through Vibration Signal Using Advanced Signal Processing Techniques. In Proceedings of the 2020 Advanced Communication Technologies and Signal Processing (ACTS), Silchar, India, 4–6 December 2020; pp. 1–6.

[21] Yang Tao, Ren Yong, Liu Xia et al. Research on modeling and simulation of wind turbine rotor imbalance fault [J]. Journal of Mechanical Engineering, 2012, 48(6):130-135.

[22] Ren Yong. Modeling and Simulation of the Unbalance of wind turbine blade[D].Wuhan: Huazhong University of Science and Technology,2012

[23] Gupta DL, Malviya AK, Singh S (2012) Performance analysis of classification tree learningalgorithms. Int J Comput Appl 55(6):0975–8887

[24] Amarnath M, Sugumaran V, Kumar H (2013) Exploiting sound signals for fault diagnosis of bearings using decision tree. Measurement 46:1250–1256.

[25] L. Breiman, Random forests, Machine Learning 45(1)(2001), 3-52.

[26] Yang, B. S., Di, X., & Han, T. (2008). Random forests classifier for machine fault diagnosis. Journal of Mechanical Science and technology, 22, 1716-1725.

[27] M. Cerrada, G. Zurita, D. Cabrera, RR Sanchez, M. Art' es' and C. Li, Fault diagnosis in spur gears based on genetic algorithm and random forest, Mechanical Systems and Signal Processin 70 (2016), 87–103

[28] Vapnik VN (1998) Statistical learning theory. Wiley, New York

[29] Maria Navin, JR, Pankaja, R., 2016. Performance Analysis of Text Classification Algorithms Using Confusion Matrix.

[30] Hossin, M., Sulaiman, M.N., 2015. Review of evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process 5 (2), 1.

[31] Tilaki-Hajian, K., 2013. Receiver Operating Characteristics (ROC) Curve Analysis Formal Diagnostic Test Evaluation. Caspian Journal of internal medicine 4 (2), 627.

[32] Antariksa, G., Muammar, R., & Lee, J. (2022). Performance evaluation of machine learning-based classification with rock physics analysis of geological lithofacies in the Tarakan Basin, Indonesia. Journal of Petroleum Science and Engineering, 208, 109250.

[33] Jaber, A.A.; Bicker, R. A simulation of nonstationary signal analysis using wavelet transform based on LabVIEW and Matlab. In Proceedings of the UKSim-AMSS 8th European Modeling Symposium on Computer Modelling and Simulation, EMS 2014,Pisa, Italy, 21–23 October 2014; pp. 138–144. https://doi.org/10.1109/EMS.2014.38.

[34] Vakharia, V., Gupta, V. K. & Kankar, P. K. (2017). Efficient Ball Bearing Failure Diagnosis using ReliefF and Random Forest Classifier. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 39(8), 2969-2982.

[35] Ali, L.; Zhu, C.; Zhou, M.; Liu, Y. Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample selection and features. Expert Syst. Appl. 2019, 137, 22–28.

[36] Liu, H.; Setiono, R. Chi2: Feature selection and discretization of numeric at tributes. In Proceedings of the 7th IEEE International Conference on Tools with Artificial Intelligence, Herndon, VA, USA, 5–8 November 1995; pp. 388–391.

[37] Al-Haddad, Luttfi A. and Alaa Abdulhady Jaber. 2023. "An Intelligent Fault Diagnosis Approach for Multirotor UAVs Based on Deep Neural Network of Multi-Reduction Transform Features" Drones 7, no. 2: 82. https://doi.org/10.3390/drones7020082.