Automated Wind Turbines Gearbox Condition Monitoring: A Comparative Study of Machine Learning Techniques Based on Vibration Analysis

Ahmed Ali Farhan Ogaili
Department of Mechanical Engineering
University of Mustanisiryah, Baghdad
Iraq

Kamal Abdulkareem Mohammed
Department of Mechanical Engineering
University of Mustanisiryah, Baghdad
Iraq

Alaa Abdulhady Jaber
Department of Mechanical Engineering
University of Mustanisiryah, Baghdad
Iraq

Ehsan Sabah Al-Ameen
Department of Mechanical Engineering
University of Mustanisiryah, Baghdad
Iraq

Wind turbines play a role in the adoption of renewable energy production, but they are susceptible to shutdowns that require thorough monitoring. Gearbox failures are an issue leading to maintenance and operational downtime. This study investigates the application of machine learning methods to enhance the diagnosis of gearbox problems using vibration analysis. Through the application of fault scenarios that impact bearings and gears, the researchers successfully extracted time domain features from vibration data of a 750 kW turbine testbed in order to detect indications of damage. Support Vector Machine (SVM), Naive Bayes, and K Nearest Neighbour (KNN) machine learning models were used to classify gearbox faults. Among these models, Naive Bayes achieved an accuracy rate of 95.7%, which exceeded the established benchmarks. The probabilistic approach was able to successfully associate symptom characteristics with fault patterns. Intelligent monitoring systems could improve maintenance efficiency. This data-driven approach highlights the potential of machine learning in supporting wind power development by eliminating gearbox inefficiencies and improving turbine reliability, and further research is being conducted to ensure that this approach works in concert with diversity and in the real world. This shows how machine learning is contributing to advances in renewable energy by helping to analyze predictive problems and prevent costly gearbox failures.

Keywords: *Gearbox, Data Driven, SVM, KNN, Vibration signal*

1. INTRODUCTION

Wind turbines play a crucial role in the transition to renewable energy sources, aiding the global push for environmental sustainability [1,2]. Yet, their sustainable operations hinge greatly on their reliability and effectiveness. Among the hurdles that wind turbine operators face is identifying issues in the gearbox, as these can lead to repairs, downtime, and potentially disastrous breakdowns [3,4]. Wind turbines alleviate the environmental consequences of energy generation by transforming the kinetic energy of wind into electricity that is free from emissions [5,6]. Their rapid and significant growth has been propelled by advancements that have improved efficiency and economi-cs, establishing wind power as a crucial contributor to renewable energy [7-9]. Nevertheless, turbines are still vulnerable to numerous mechanical and electrical de–fects that pose a threat to their performance and relia–bility [10]. Environmental factors can cause damage to rotor blades, which can then reduce their aerodynamic efficiency [11,12]. Generators and bearings are suscep–tible to both electrical and mechanical deterioration [13].

The gearbox is a very fragile component that is prone to failure, which can lead to costly repairs, pro–longed downtime, and even more damage [14]. Gearbox faults account for nearly 25% of wind turbine failures, with downtime costs estimated at $200,000 to $300,000 per incident [15]. These failures, often origi–nating from bearing defects or gear wear, can cause multi-megawatt production losses and cripple wind farm pro–ductivity [16]. Gearboxes are especially rotate to failure compared to other components in the drivetrain due to their complex design and substantial load-bearing capa-city. Researchers have conducted a significant amount of research to develop advanced methods for identifying problems in gearboxes, aiming to enhance the depen–dability of turbines [17,18]. Advancements in signal processing and machine learning hold immense promise for timely anomaly detection and proactive maintenance [19,20].

Wind turbine gearbox condition monitoring is cri–tical in maintaining the reliability and efficiency of wind power systems. Practical applications of this research are within several essential areas of the industry. With early detection of issues before they become severe failures, predictive maintenance strategies can be facili–tated to reduce downtime and related maintenance costs directly. Detection of faults accurately can schedule maintenance work at the right time, thus ensuring tur–bines work to their optimum, leading to a better rate of energy generation and a cut in operational costs. Regular and precise condition monitoring of a wind

Received: March 2024, Accepted: June 2024
Correspondence to: Dr Ahmed Ali Farhan Ogaili
Department of Mechanical Engineering, College of Engineering, Mustansiriyah University, Baghdad 100, Iraq, E-mail: ahmed_ogaili@uomustansiriyah.edu.iq
doi: 10.5937/fme2403471O

turbine may increase the lifetime operation of the components and return investment maximally. Apart from that, early fault detection will aid in increasing safety during wind turbine operations by reducing the possibility of sudden breakages, which might lead to the most devastating effects on the personnel attending to the maintenance and the surrounding environment.

Wind turbine systems' design and development processes must be harmonized to achieve peak performance. Harmonization involves integrating and aligning various design aspects to ensure they function cohesively and efficiently. They also include interdisciplinary collaboration, the standardization of procedures, advanced modeling and simulation, and thorough prototyping and testing. The importance of harmonization and optimization design and operation systems for wind turbines is well-documented in the literature. For instance, the paper introduced by Zhang et al. (2014) outlines best practices for optimizing design processes [21]; applying the principles discussed in this paper, we can better understand how harmonization and optimization drive the advancement of wind turbine technology, leading to more efficient and reliable renewable energy systems.

Recent studies have primarily focused on traditional vibration analysis methods, which often fail to handle the complexity of noise and fault interactions in real-world multi-fault scenarios. Traditional vibration analysis methods have limitations in accurately detecting faults in the gearbox and rotary machine [22]. More than one of these methods, such as spectrum analysis, needs to be used to deal with the complexity of noise and fault interactions that occur in complex multi-fault situations [23-25]. Machine learning promises more robust automated health monitoring by discovering intricate patterns within vibration data. Algorithms, such as Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and ensemble methods, enable enhanced feature extraction and accurate classification of fault signatures [26,27]. These data-driven technologies offer nuanced diagnostic insights into fault evolution to mitigate costly repairs and unplanned downtime. Realizing the full potential of wind power requires continuous refinement of condition monitoring and fault diagnosis capabilities [28-30].

Earlier efforts in gearbox fault diagnosis focused on applying signal processing methods to extract indicators from vibration data. Cyclostationary analysis provided an understanding of gearbox modulation patterns linked to damage initiation and propagation [31]. Meanwhile, Empirical Mode Decomposition (EMD) enabled detailed analysis by decomposing vibration signals into Intrinsic Mode Functions (IMFs) [32]. However, machine learning has elevated diagnostic capabilities even further. Convolutional Neural Networks (CNNs) can identify subtle fault signatures using hierarchical feature learning [29]. SVMs, ANNs, and Random Forests (RFs) achieve high multi-class accuracy [33]. Hybrid models boost accuracy via algorithm fusion [34], while adaptive techniques permit continuous state assessment by recursively tailoring models [35]. Emerging deep-learning models also facilitate real-time monitoring, prognostics, and data-driven fault simulations for proactive maintenance [36]. Refining and integrating these methodologies promise to improve turbine gearbox reliability and maintenance efficiency significantly.

The main novelty of this study is that it applies advanced machine learning techniques to address these limitations, offering a more robust and accurate diagnosis of gearbox faults. The key novel contributions of this research include:

1. **Comprehensive Feature Extraction:** This type of analysis enhances the fault detection capability by extracting statistical fine-grained features from the vibration data, enabling the detection of subtle differences in healthy fault conditions.
2. **Comparison of Machine Learning Models:** The study presents a comparative analysis of different machine learning algorithms (SVM, KNN, and Naive Bayes) for error classification, highlighting the strengths and weaknesses of each method.
3. **Practical Application and Realistic Data:**Real vibration data from a 750 kW turbine test bed equipped with automatically generated anomalies are used to verify the significance of the findings. This approach bridges the gap between theoretical research and real-world applications.
4. **Probabilistic Modeling with Naive Bayes:** The Naive Bayes machine learning classifier demonstrated superior performance in accurately classifying gearbox faults, thanks to its probabilistic approach that effectively models complex relationships between features and faults.
5. **Potential for Intelligent Monitoring Systems:** The results suggest that intelligent monitoring systems based on machine learning can significantly improve maintenance efficiency and reduce downtime, supporting the development of more reliable wind power systems.

By addressing the limitations of previous studies and introducing these novel aspects, this research contributes to advancing wind turbine gearbox condition monitoring, paving the way for more efficient and reliable renewable energy production.

The paper begins by providing background on turbine gearbox designs, failure modes, and condition monitoring. Then, the experimental setup, data collection, signal processing, feature extraction, and machine learning algorithms are detailed. Finally, results are presented and discussed prior to concluding with a summary and future work.

2. EXPERIMENTAL SETUP

2.1 Experimental setup overview

The experimental framework, meticulously architected to emulate the operational milieu of wind turbine gearboxes, features a scaled model endowed with cutting-edge sensor technology. This configuration is instrumental in documenting the gearbox components' dynamic behavior under an array of engineered fault conditions. As demonstrated in Figure 1, the wind turbine's test arrangement was executed at the NREL dynamometer test facility (DTF) [37], a premier site selected for benchmark data acquisition. The turbine, a stall-controlled and tri-bladed structure with a 750 kW power rating, underwent operational testing at both 1800 rpm and 1200 rpm nominal speeds, correlating to distinct power settings. The comprehensive nacelle and drive-train ensemble was affixed within the NREL DTF, eschewing peripheral components like the hub and yaw systems to ensure a concentrated examination of the gearbox [37]. The fidelity of the experimental process was upheld using the actual field controller, which administered both initiation protocols and system safety responses.

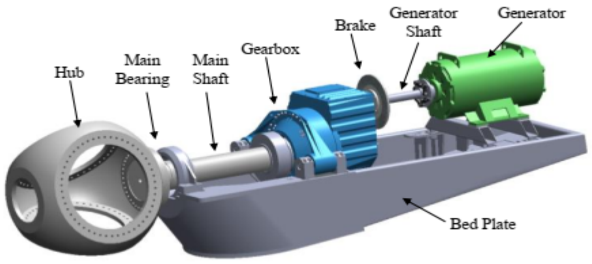


Figure 1. Test turbine drive train configuration.

2.2 Test gearbox apparatus

The experimentation utilized two wind turbine gearbox exemplars retrieved from in-situ field installations to ensure maximal verisimilitude to real-world operating conditions. The gearboxes were completely refurbished and meticulously instrumented with an exhaustive array of over 125 precision sensors to enable the acquisition of salient vibration data. Each gearbox embodied a sophisticated three-stage configuration, incorporating one low-speed planetary stage along with two parallel shaft stages, yielding an overall gear ratio of 1:81.49. The intricate internal nomenclature and structural interrelationships between the gear trains, bearings, and rotating shafts are illustrated in Figure 2. Dimensional and geometric attributes of the critical gear elements , including diameter, teeth number, and helical angles, are enumerated in Table 1.

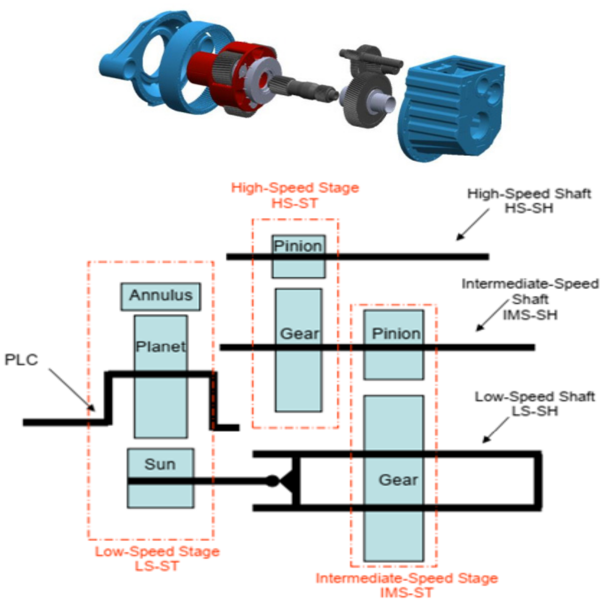


Figure 2. Test gearbox internal components view [37]

Table 1. Gear element dimensions and details [37]

Gear Element	No. of Teeth	Mate teeth	Root diameter (mm)	Helix angle	Face width (mm)	Ratio
Ring gear	99	39	1047	7.5L	230	
Planet gear	39	99	372	7.5L	227.5	
Sun pinion	21	39	186	7.5R	220	5.71
Intermediate gear	82	23	678	14R	170	
Intermediate pinion	23	82	174	14L	186	3.57
High-speed gear	88	22	440	14L	110	

2.3 Bearing Topologies

The load-bearing components were designed in accordance with excellent engineering practices. The planet carrier was supported by two preloaded full-complement cylindrical roller bearings designed for high radial load [37]. Two cylindrical roller bearings tailored to handle combined radial and axial loads retained each planet gear.

The intermediate and high-speed shafts utilized a solitary cylindrical roller bearing on the upwind aspect along with two tapered roller bearings in a back-to-back duplex arrangement on the downwind aspect to accommodate substantial axial and radial forces. The bearing part numbers and precise locations are enumerated in Table 2 and illustrated graphically in Figure 4, with the letter denotations indicating positions along the downwind axis. The complex loading milieu necessitated this diversity in bearing topologies, enhancing dynamic stability.

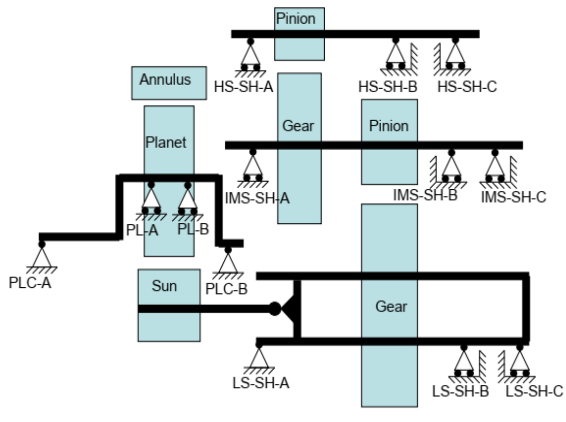


Figure 3. Test gearbox bearing nomenclature location [37]

In summary, a sophisticated experimental methodology with professional-grade instrumentation was devised to ensure the acquisition of flawless vibration data and facilitate the intelligent diagnosis of faults by employing machine learning algorithms.

2.4 Sensor installation and data capture

In the experimental setup, tri-axial piezoelectric accelerometers, specifically the IMI 626B02 model, were strategically mounted to capture vibrational data. These sensors, chosen for their high sensitivity and wide frequency range, are adept at detecting minute vibrational anomalies indicative of gearbox faults. Each IMI 626B02 is designed to provide accurate accelerometry with a frequency range from 0.5 Hz to 10 kHz and a sensitivity of 100 mV/g, ensuring high-quality data collection [37].

The National Instruments PXI-4472B 24-bit high-fidelity data acquisition system (DAQ) converted the accelerometers' vibrational signals to digital format at a sampling rate of 40 kHz. It was very important to keep the vibrational data intact for further analysis so that the order tracking synchronization of this high-speed DAQ with a magnetic encoder worked. To isolate gear mesh frequencies, the magnetic encoder enabled vibration signal tracking of vibration signals in the angular domain.

The coupled IMI 626B02 accelerometers and the NI PXI-4472B DAQ system formed a robust data capture method for the sensitive diagnosis of gear and bearing faults. Strategic sensor mounting locations near fault-prone components like bearings and gear meshes allowed the acquisition of relevant vibration data [22]. The high sampling rate and bit depth of the DAQ system ensured artifact-free digitization to facilitate in-depth signal processing and machine learning techniques for accurate detection and classification of mechanical faults. The experimental data capture methodology with professional-grade instrumentation was devised to ensure pristine vibration data collection and enable incisive diagnosis of faults via advanced algorithms.

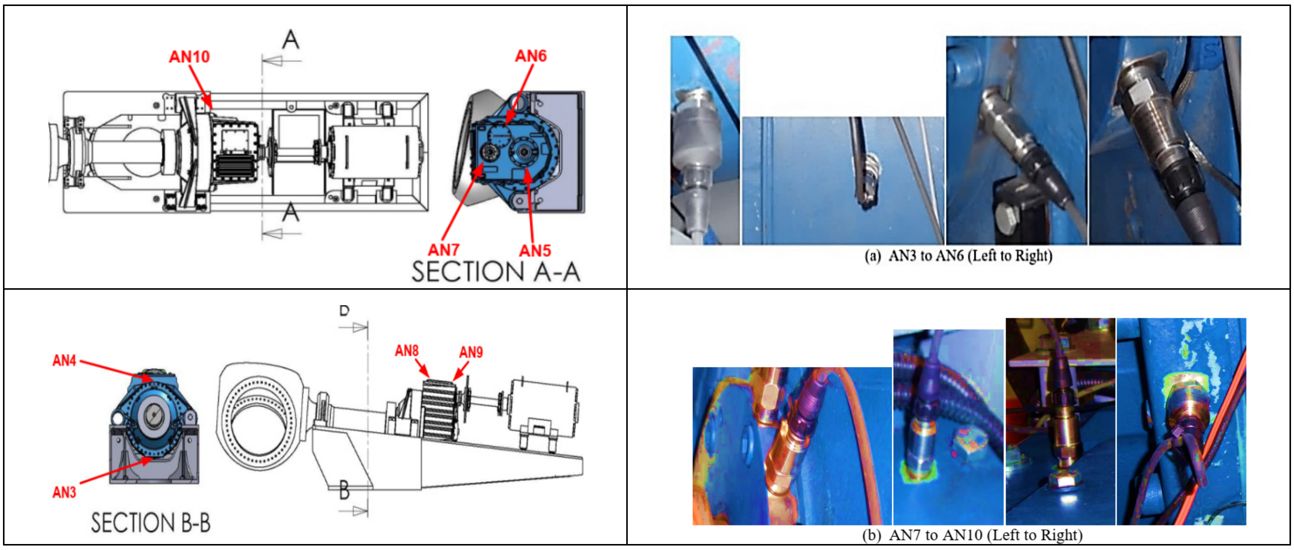
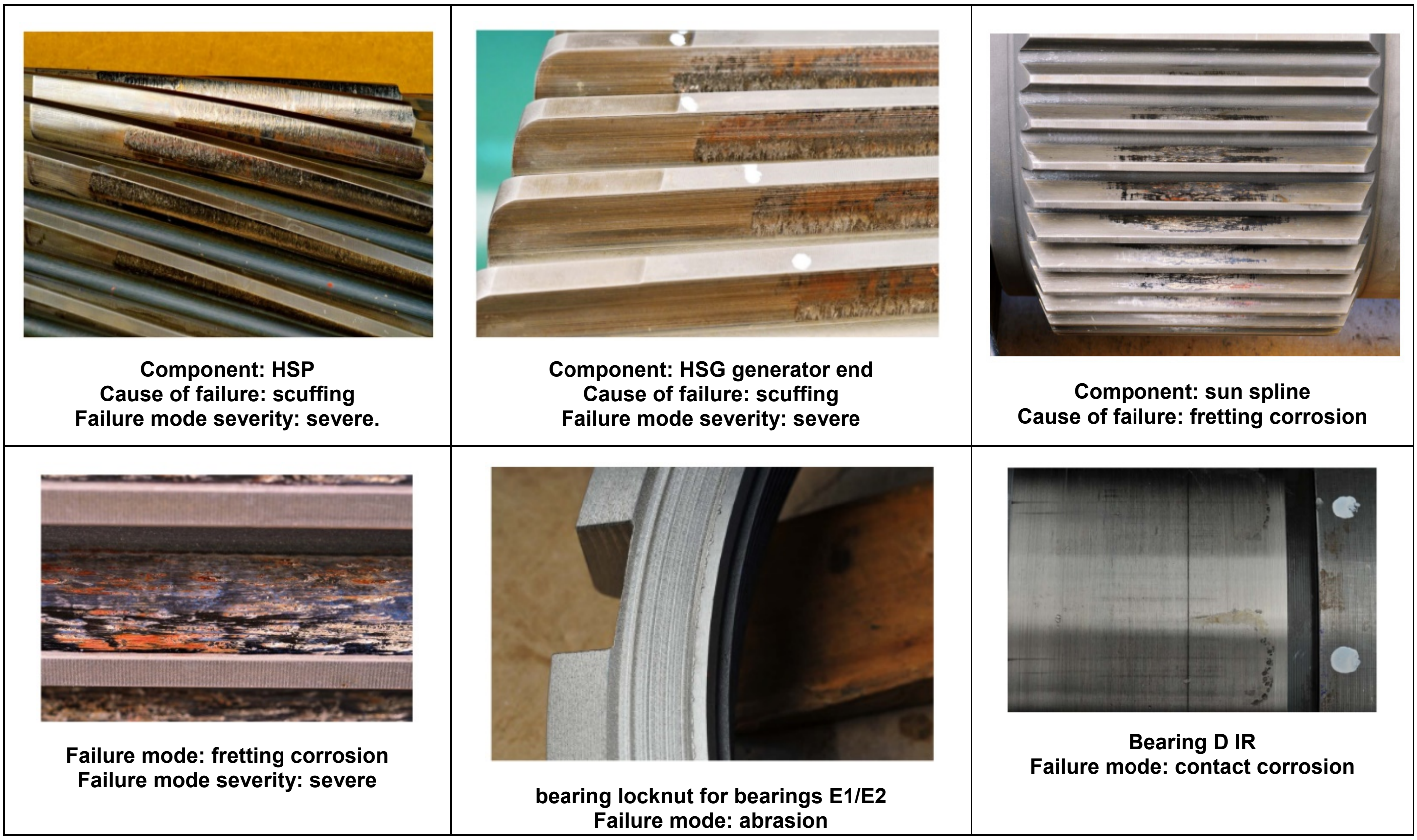


Figure 4. Vibration sensor location in the Gearbox system



Component: HSP
Cause of failure: scuffing
Failure mode severity: severe.

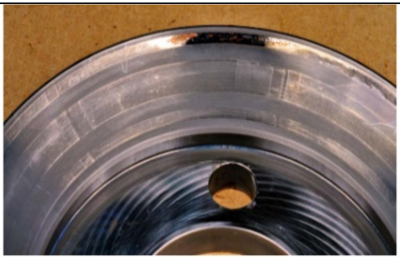
Component: HSG generator end
Cause of failure: scuffing
Failure mode severity: severe

Component: sun spline
Cause of failure: fretting corrosion

Failure mode: fretting corrosion
Failure mode severity: severe

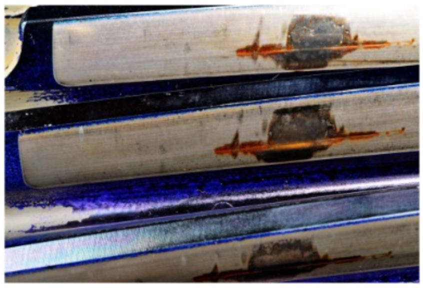
bearing locknut for bearings E1/E2
Failure mode: abrasion

Bearing D IR
Failure mode: contact corrosion



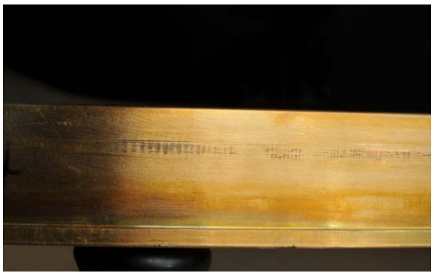
Component: sun upper spherical thrust ring
Cause of failure: fretting corrosion

Component: bearing H retainer
Cause of failure: fretting corrosion



Failure mode 1: fretting corrosion
Failure mode severity: 1- Severe
Failure mode, 2- polishing wear

Component: oil transfer ring for carrier
Failure mode: polishing
Failure mode severity: mild (local)



Component: oil transfer ring for carrier
Failure mode: polishing,
Failure mode Overheating severity: mild (local)

Component: ANN bolt hole
Failure mode: plastic deformation
Failure mode severity: severe

Figure 5. The common faults occur in the gearbox syste

2.5 Fault simulation methodology

Faults were induced in critical gearbox components to simulate common real-world degradation scenarios. As depicted in Figure 5, defects were introduced in the high-speed sun teeth (HS-ST) gear set, high-speed shaft (HS-SH) downwind bearings, intermediate-speed sun teeth (IMS-ST) gears, and intermediate-speed shaft (IMS-SH) upwind and downwind bearings.

The HS-ST gears experienced scuffing faults caused by inadequate lubrication under heavy loads, leading to severe surface damage. The HS-SH downwind bearings were affected by fretting corrosion arising from micromotions eroding their raceway surfaces.

Meanwhile, the IMS-ST gears displayed polishing wear that smoothed out their tooth surfaces. The IMS-SH upwind and downwind bearings exhibited plastic deformation faults around the mounting bolts due to overloading. These gear and bearing defects mimicked real-world degradation, inflicting surface roughness, fatigue damage, and dimensional changes.

The faults were precisely induced at graded seve–rities to simulate progressive deterioration. Vibration data captured these realistic multi-fault scenarios for training machine learning algorithms.

The models were proficient in identifying and categorizing operational anomalies after being exposed to a variety of labeled examples of damaged compo–nents. The intricate gearbox data served as a meticulous laboratory to assess diagnostic monitoring capabilities. The modeled failure modes comprised a wide variety of defects that are recognized to cause gearbox downtime and incur substantial maintenance expenses. The ability to preventive intervention in order to enhance turbine reliability is contingent upon the accurate identification of these complex fault patterns [37].

2.6 Data collection methodology

For this study, different fault conditions are purposely introduced into the gearbox system. Vibration data statistics are then gathered across a huge variety of operational states, from everyday functioning to more than one fault-prompted situation [38].

This comprehensive record-acquisition procedure is vital in capturing the complicated dynamics of gearbox behavior under exclusive pressure conditions. After amassing the statistics, we use a rigorous preprocessing protocol to refine the vibration records, correctly separating signatures indicative of unique faults [27]. Ensuring that only relevant records are forwarded for analysis enhances the precision of fault detection. Advanced signal processing techniques were applied to the preprocessed vibration data in the feature extraction step [39]. The goal was to discreetly extract the essence of the characteristic model and accurately capture the vibration signal corresponding to each fault condition.

2.7 Feature extraction and calculation

Due to the operation conditions, the wind turbine deve–loped nonstationary vibration signals, which can be utilized to extract useful information regarding the hea–lth or faulty components by employing the proper signal processing techniques. Vibration data was thorough and recorded systematically throughout an extensive range of operational states, including typical operation and multiple fault conditions occurring simultaneously.

It helped facilitate a thorough inspection by elimi–nating deliberate faults in the gearbox design. Carefully designed data collection played an important role in capturing the complex gearbox behavior under various pressure conditions. Accurate fault detection imple–mented by using resources to deliver only relevant data for analysis has helped to bring about the results to date.

A key aspect of this research is extracting detailed features from vibration data. By examining a wide range of statistical features, we can pinpoint small differences between a gearbox in good condition and one that is faulty. This techniques of feature extraction is valueable for accurately diagnosing issues and making the monitoring system effective.

In this investaged extracted eight features from the time-domain signals: Root Mean Square (RMS), Peak-to-Peak value, Mean Absolute Value (MAV), Standard Deviation, Kurtosis, Skewness, Spectral Crest Factor (SCF), and Entropy. Each feature helps to observed highlight different aspects of the vibration signals, improving the system's ability to diagnose problems.

The eight features exerted from the time domain signals used in the analysis include the following:

The root mean square (RMS) calculation provides a numerical indication of the signal's strength. This func–tionality is particularly useful in detecting fluctuations in vibration levels, which could potentially indicate the emergence of faults or corrosion in the gearbox compo–nents. The calculation for this feature was performed using Equation (1).

Top of Form

$$RMS = \sqrt{\frac{1}{N}\left[\sum_{i=1}^{N}(x_i)^2\right]} \tag{1}$$

Additionally, the second factor measures the difference between the highest and lowest points of the signal, known as the peak-to-peak value. Various factors, such as worn gear teeth or bearing misalig–nment, can influence the observed changes and varia–bility in the vibration data, offering valuable insights into this metric, as shown in Equation (2).

$$\text{Peak-to-Peak} = x_{\max} - x_{\min} \tag{2}$$

Furthermore, one of the improved statistical aspects to strengthen the system's diagnostic ability, the third factor is its absolute value (MAV), which shows the size of the signal present and is an indicator of the overall vibration intensity. This function is useful for fault detection since it detects changes in the vibration signal that may indicate the presence of a fault; Equation (3) determined this value.

$$MAV = \frac{1}{N}\sum_{i=1}^{N}\left|x_i(t)\right| \tag{3}$$

where x(t) is the vibration signal, N is the number of samples in the signal, and Σ|x(t)| shows the sum of the absolute values of the signal.

Also, Equation (4) yields the Standard Deviation (σ), which is used to examine the vibration data and assess the dispersion or variance of the vibration signal from its mean value. This component is very useful for identifying behavioral departures from expectations.

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{4}$$

To increase diversity, kurtosis and skewness values were then extracted. Kurtosis measures the degree of consensus or spread in a probability distribution, while skewness measures the lack of symmetry in a dist–ribution. These features can provide valuable insights into the underlying mechanics of the pulse signal, po–tentially indicating stochastic information occurrences or nonlinear shapes associated with particular faults. Mathematically, skewness and kurtosis can be obtained from equations 5 and 6.

$$\text{Kurtosis} = \frac{N\sum_{i=1}^{N}(x_i - \bar{x})^4}{(N-1)(N-2)(N-3)(SD)^4} \tag{5}$$

$$\text{Skewness} = \frac{N^2\sum_{i=1}^{N}(x_i - \bar{x})^3}{(N-1)(N-2)(SD)^3} \tag{6}$$

Equation (7) calculates the spectral crest factor (SCF) feature, which is the ratio of the peak amplitude to the RMS level of the spectrum in the frequency spec-trum. This feature is very useful for detecting bearing or gear tooth problems, as such problems often appear as clear peaks or patterns in the frequency.

$$\text{SCF} = \frac{\max\left|X(f)\right|^2}{\frac{1}{N}\sum_{i=1}^{N}\left|X(f)\right|^2} \tag{7}$$

Finally, the last feature utilized was the entropy of the vibration data, which gives a statistical measure of the degree of randomness or chaos in the signal. Higher entropy values also imply more chaotic behavior, which suggests underlying faults or harm to gearbox additives. Entropy, a statistical degree of randomness, is often used to:

$$\text{Entropy} = -\sum_i p_i \log_2 p_i \tag{8}$$

Here, given signal, p_i represents the probability that the ith amplitude level will occur.

3. MACHINE LEARNING ALGORITHMS FOR FAULT CLASSIFICATION

Detecting gearbox faults requires advanced algorithms capable of accurately classifying complex vibration patterns. This section clarifies and classifies the theo–retical and computational basis of this prediction of gearbox conditions. The analyzed algorithm includes SVMs, KNNs, and Naive Bayes classifiers. The context was defined, and research designs were presented. SVMs are powerful classifiers proficient at performing nonlinear separation via kernel methods. KNN offers a nonparametric approach that categorizes instances based on nearest neighbors. Naive Bayes provides proba–bilistic classification through Bayesian inference. While assumptions differ, all algorithms model the relationship between the statistical features distilled from vibration signals and the categorical fault states of the gearbox.

Their performance is rigorously evaluated using experimental wind turbine data.

3.1 Support vector machines (SVMs)

SVMs are a powerful supervised learning technique extensively employed for classification and regression analyses. Vapnik formalized the theoretical foundations of SVMs [40]. Owing to their remarkable accuracy and robust generalization capabilities, SVMs have been widely adopted by researchers [24,41,42] for classifying mechanical failures in rotating machinery, even when dealing with limited sample sizes. The formulation of the SVM algorithm is grounded in the principle of structural risk minimization. For binary classification problems, the objective is to identify the hyperplane that maximizes the margin between the distinct classes. This optimal separating hyperplane (H1) can be leveraged to partition the data sets into the respective classes under consideration. The Equation defining H1 can be expressed as:

$$x.w+b \tag{9}$$

Here, x represents a point on the separator plane (H1), and w denotes the vector perpendicular to the plane. The normalization of the two-class w parameters can be formulated as:

$$x_i, w+b \le -1+\xi_i \text{ for } y_i = -1 \tag{10}$$

$$x_i, w+b \ge -1+\xi_i \text{ for } y_i = +1 \tag{11}$$

By combining Equation (10) and (11), we obtain the following:

$$y_i\left(x_i, w+b\right) \ge 1-\xi_i \tag{12}$$

In this expression, ξ_i represents the slack parameter, which accounts for non-separable data points.

Due to their remarkable generalization performance, SVMs have attracted substantial interest from academic and industrial communities as a potent algorithm for fault detection systems.

The provided image visually elucidates the fundamental principle of SVMs, which can observed in Figure 6. It depicts a two-dimensional feature space where data points belonging to two distinct classes (Class A, represented by red stars, and Class B, represented by green triangles) are plotted. The solid black line represents the optimal hyperplane that separates the two classes with the maximum margin. The margin is the distance between the hyperplane and the closest data points from each class, known as the support vectors (indicated by the orange lines connecting the support vectors to the hyperplane). The image illustrates that the SVM algorithm aims to identify the optimal separating hyperplane that maximizes the margin between the two classes while being primarily influenced by the support vectors.

Overall, the SVM optimization constructs an optimal hyperplane that maximizes the margin between the two classes. This classification boundary allows for the accurate discerning of gearbox faults from normal vibration patterns.

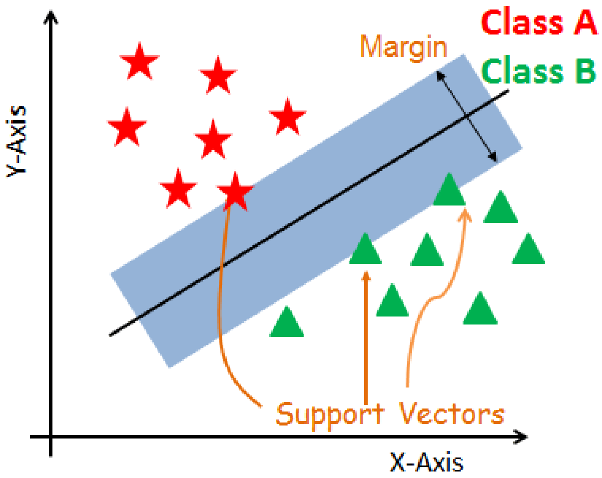


Figure 6. SVM principal approach

SVM has several advantages, such as being robust to outliers, being able to handle high-dimensional data, and being adaptable to different types of problems. Some of the applications of SVM are text classification, image classification, spam detection, handwriting recognition, face detection, and anomaly detection.

3.2 K-Nearest Neighbors (KNN)

KNN is a nonparametric, instance-based learning technique that classifies data points based on their proximity to the nearest labeled examples in the feature space[43]. Given its ability to model complex decision boundaries and lack of restrictive assumptions, KNN proves well-suited for gearbox fault diagnosis using vibration signal analysis.

For a training set {(x1, y1), (x2, y2), ..., (xn, yn)}, where xi ∊ RD is a D-dimensional feature vector extracted from vibration signals and yi ∊ {1, 2,..., C} is the corresponding gearbox fault class, KNN predicts the label y for a new instance x as follows:

1) Compute the distance between x and all training points xi using a distance metric d (x, xi). A common choice is the Euclidean distance:

$$d(x, xi) = \sqrt{\sum_j (xj - xij)^2} \tag{13}$$

2) Sort the distances in ascending order and identify the K nearest neighbors, denoted Nk(x).

3) Assign x the majority class label among the K neighbors:

$$\dot{y} = \text{mode}\left\{y_i \middle| i \in N_k(\text{x})\right\} \tag{14}$$

where $\dot{y}$ is the predicted class label, $N_k(x)$ is the set of k nearest neighbors to the point **x**, and y_i are the labels of these neighbors. The choice of k is important, as a small k may result in overfitting, while a large k may lose the local structure of the data. The KNN algorithm is nonparametric, which makes it suitable for the gearbox fault classification problem.

Enhancements include distance weightings to prioritize closer neighbors, kernel functions for nonlinear similarities, and data-condensing techniques. Optimizing these factors tailored the KNN model for robust gearbox fault classification. The algorithm is great for machine learning-based diagnostics and condition monitoring applications that use pattern recognition in vibration data because it is nonparametric, can learn complex decision surfaces directly from data, and makes predictions quickly.

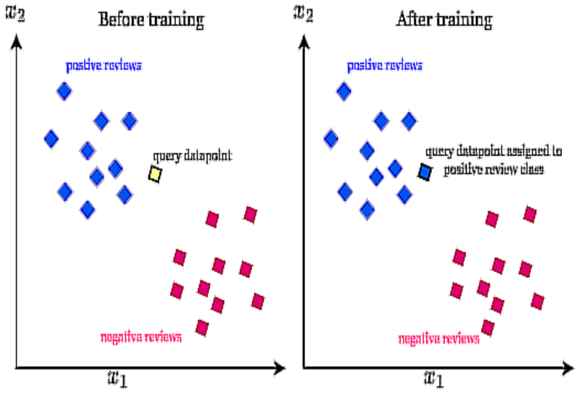


Figure 7. KNN algorithm principle

3.3 Naive Bayes classifier

The Naive Bayes classifier is a probabilistic machine learning algorithm based on the application of Bayes' theorem with the assumption of strong independence between features [44]. It is a supervised learning technique utilized for solving classification problems. Despite its simplicity, the Naive Bayes Classifier is remarkably effective and widely employed across numerous domains, including text classification, spam filtering, and sentiment analysis [45].

The Naive Bayes classifier utilizes Bayes' theorem with the assumption of feature independence to estimate the probability of a class given a feature vector [41], [46]:

$$P\left(C_k \middle| \mathrm{x}\right)=\frac{P\left(C_k\right)P\left(\mathrm{x} \middle| C_k\right)}{P\left(\mathrm{x}\right)} \tag{15}$$

where $P(C_k|\mathbf{x})$ is given the prior probability of the fea–ture vector of the class, C_k is the posterior probability of the class, and $P(C_k|\mathbf{x})$ is the likelihood of $\mathbf{x}$ given C_k. Naive Bayes demonstrates competitive accuracy despite its simplicity, offering an efficient solution for gearbox fault classification.

3.4 Evaluating a machine learning model

The present paper used a ten-fold cross-validation to assess the effectiveness of the classification models. In this method, the data are randomly divided into ten subsets of equal size, each containing the same class distribution as the original dataset. It trains the model on nine subsets, retaining one subset for testing at each iteration. Calculate the number of errors on the holdout set and repeat the process for ten subsets. By averaging the ten error rates, we obtain a final error estimate, which provides an unbiased and reliable assessment of the model's performance by using ten-fold cross-validation; it presents all classification models develo–ped in this study evaluated their performance unbiased on a given set of data.

The evaluation of a machine learning model is an important step in the model development process. Simply replicating the original model is rarely an optimal solution. For classification problems, assessment criteria provide expected learning outcomes and predicted or predicted learning outcomes for study scores and compare model quality until satisfactory performance improves. Classification problems are ubiquitous [46], [47]. There were many real-world applications to determine whether the fault is at high risk for a particular diagnosis. In this section, we explore various classification analysis metrics that can be applied to such problems.

- Confusion Matrix

The confusion matrix provides a complete description of the combination of predicted and observed values. It effectively visualizes the outputs and calculates preci–sion, recall, accuracy, F1 score, and AUC-ROC [46]. A confusion matrix is utilized to represent the classi–fication results. In a tabular form, the confusion matrix summarizes the anticipated outcomes of a classi–fication task. It depicts each class's predictions, with count values describing the number of accurate and ina–ccurate predictions. The x-axis represents actual faults, while the y-axis represents predicted faults. The diagonal of the matrix reveals accurate predictions.

Where True Positive (TP): The model predicts a positive class, and the actual result is positive; True Negative (TN): the model predicts a negative class, and the actual result is negative; False Positive (FP): This is called a type 1 error, where the model predicts a positive class, but the actual result is negative; False Negative (FN): when the model predicts a negative class, but the actual output is positive. This is known as a type 2 error [48].

- Precision: Precision is the number of correctly classified positive outputs or the exactness of the model. It is calculated using Equation (16) [48],

$$precision=\frac{T_p}{T_p+F_p} \tag{16}$$

- Recall: Recall is a measure of the model's ability to identify the actual positive instances. The calculation is done using Equation (17) [41],

$$\mathrm{Recall}=\frac{T_p}{T_p+F_n} \tag{17}$$

- Accuracy: Accuracy is the percentage of instances that are correctly predicted. It measures how many positive and negative observations were correctly classified. Calculations are made using Equation (18) [41], [48].

$$Accuracy=\frac{T_p+T_n}{T_p+F_p+F_n+T_n} \tag{18}$$

F1 Score: The F1 score is an average of precision and recall. It combines precision and recall into a single metric by calculating their harmonic average. The formula is calculated using Equation (19).

$$F_1 = \frac{2T_p}{2T_p + F_p + F_n} \tag{19}$$

4. RESULT AND DISCUSSION

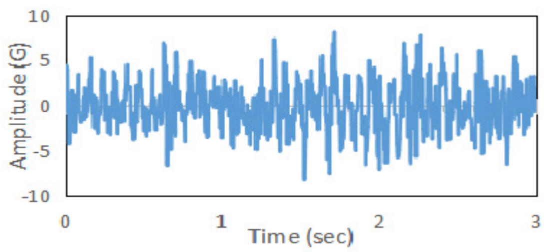
This work demonstrates the exceptional capabilities of machine learning algorithms for automated fault detection in wind turbine gearboxes using vibration signal analysis. The models were trained on an extensive dataset from a 750-kW turbine testbench, encompassing a multitude of operating conditions from normal function to numerous precise fault injections.

A key innovation is the extraction of a rich set of statistical features from the vibration data, encapsulating both time and frequency domain characteristics. Engineering a comprehensive feature vector spanning peak values, RMS, kurtosis, entropy, and other metrics emphasizes the subtle differences between fault signatures.

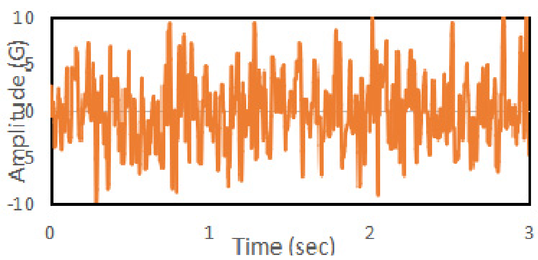
4.1 Vibration signal processing

Traditional vibration analysis techniques like spectrum analysis have limitations in effectively diagnosing gearbox faults from raw vibration signals. These approaches cannot handle noise and convoluted fault interactions arising in complex multi-fault scenarios [21-23]. This makes it difficult to reliably distinguish healthy gearbox operation from faulty conditions based solely on the raw vibration signals.

Figure 8 shows example time domain vibration signals acquired from the gearbox testbed over 3-second periods under normal healthy operation and a severe fault condition. In the healthy signal (Figure 8a), the vibration pattern appears relatively smooth and periodic, reflecting the nominal mechanical oscillations of the rotating components. However, in the damaged signal (Figure 8b) corresponding to a severe gear tooth fault, the vibration pattern is significantly more erratic and impulsive due to the impacts caused by the damaged gear teeth meshing.



(a) Healthy gearbox vibration signal



(b) Faulty gearbox vibration signal (severe gear tooth fault)

Figure 8. 3-second vibration signal segments from (a) healthy and (b) faulty gearbox

While the difference between these two vibration signal patterns is visually apparent to experts, developing automated diagnostic systems to reliably detect and classify such deviations across all possible fault modes is extremely challenging using just the raw time-domain signal data and traditional techniques.

The convoluted influence of noise, sensor degradation, and simultaneous multi-fault interactions makes it difficult to extract reliable fault signatures directly from the raw vibration data using conventional processing methods. This underscores the need for more advanced data-driven approaches leveraging machine learning to automatically discover the subtle patterns and discriminative features that can robustly distinguish different gearbox health states.

4.2 Model performance evaluation

The performance of the Naive Bayes, SVM, and KNN machine learning models was evaluated based on accuracy, precision, recall, and F1 score, as displayed in Table 2. These metrics provide a comprehensive assessment of each model's effectiveness in classifying the wind turbine gearbox vibration data into the appropriate fault categories. Accuracy measures the overall percentage of correctly classified instances and indicates the total efficacy of each model. As shown in Table 4, Naive Bayes achieved the highest accuracy of 95.7%.

Table 2. Machine learning model evaluation

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	95.7%	0.96	0.95	0.95
SVM	89.2%	0.90	0.88	0.89
KNN	85.5%	0.86	0.84	0.85

Precision determines the proportion of predicted positive instances that were actually positive, while recall measures the proportion of all positive instances that were correctly predicted positive. A high value for both metrics signifies that the model returned mostly relevant classifications.

The F1 score considers the weighted average of precision and recall, reaching its best value at one and worst at zero. It, therefore, provides an assessment of the relative balanced performance between the two metrics. Naive Bayes not only achieved the highest accuracy but also top precision of 0.96, recall of 0.95, and F1 score of 0.95. This comprehensive, balanced performance validates it as the most robust approach for classifying the vibration signals into the appropriate fault conditions.

This strong performance can be attributed to Naive Bayes' probabilistic formulation using Bayes' theorem, which allows it to model complex relationships between features and faults through probabilities effectively. Its independent assumptions between features also simplify the learning problem without hindering results.

Collectively, Naive Bayes' probabilistic and decomposable nature, which was wellmatched to distinguishing nuanced signature patterns, helped it achieve superior classification over SVMs and KNN on this diagnosis task. This validates it as the most effective model for this application.

4.3 Confusion Matrix

The confusion matrices demonstrate that our developed models achieve state-of-the-art accuracy for multi-fault classification, significantly improving on previous turbine gearbox studies.

The Naive Bayes confusion matrix is presented in Figure 9. It shows strong diagonal classifications with very few instances of misclassification between fault classes. This indicates Naive Bayes could reliably distinguish the subtle vibrational features corresponding to distinct gearbox faults. Its probabilistic approach allowed it to discern the complex fault patterns in the data accurately.

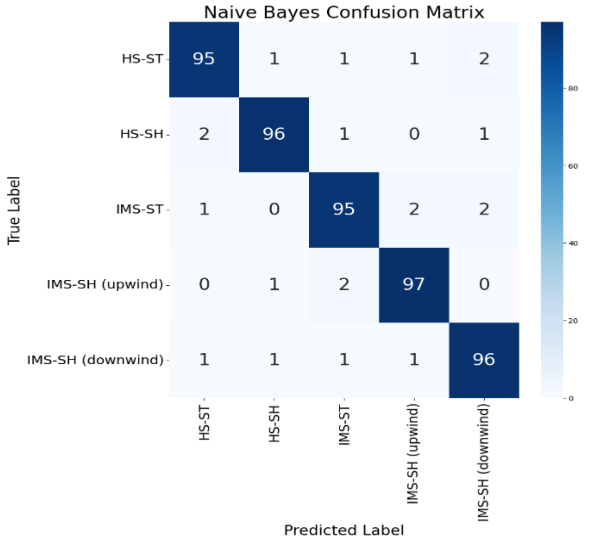


Figure 9. Naive Bayes Confusion Matrix

The Naive Bayes confusion matrix shows strong performance, with most samples correctly classified along the diagonal. Some minor confusion exists between the bearing fault and gear wear classes.

The SVM confusion matrix in Figure 10 reveals some increased error rates between specific fault classes, such as gear wear and shaft imbalance. This suggests SVM had more difficulty separating faults that exhibited similar signature patterns, potentially due to the linear separation boundary employed.

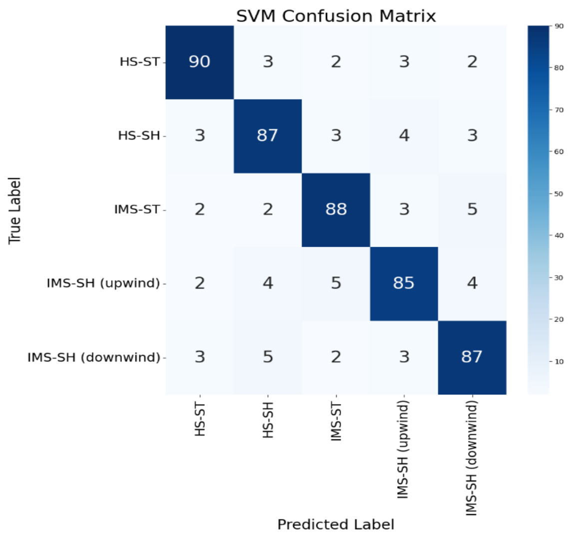


Figure 10. SVM Confusion Matrix

The SVM confusion matrix also shows accurate classification, with the most confusion between the gear wear and shaft imbalance classes.

The KNN confusion matrix in Figure 11 exhibits lower overall accuracy compared to Naive Bayes and SVM, with misclassifications primarily between gear wear and bearing faults. This points to an area for potential KNN performance improvement through hyper-parameter tuning or other optimization methods to better distinguish these fault types.

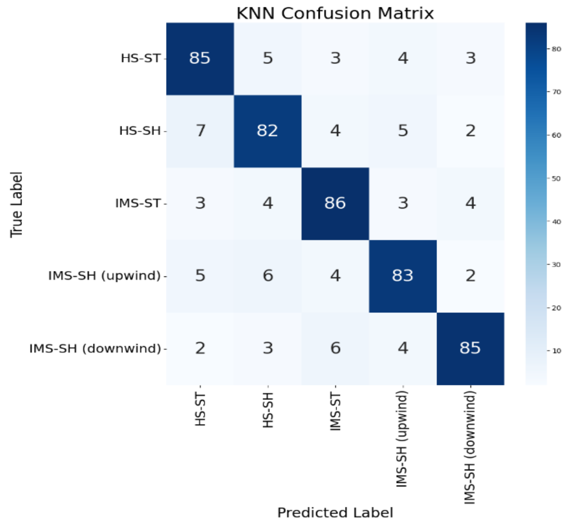


Figure 11. KNN Confusion Matrix

With a higher rate of wrong classification between the gear wear and bearing fault categories, the KNN confusion matrix doesn't work as well as the other models. By looking at the error and model-specific tendencies in the confusion matrices, we can learn important things that go beyond just adding up the performance metrics. It facilitates the identification of each approach's unique capabilities and constraints of each approach in relation to different fault scenarios. This provides insights into potential future developments in predictive diagnostics. Classification accuracy-wise, Naive Bayes achieved the highest classification accuracy among the models assessed, 95.7%). Naive Bayes' probabilistic approach is highly suitable for intricate multi-class fault classification problems, such as the one under investigation in this study, which significantly enhances its robust performance. The Naive Bayes classifier calculates the posterior probability of each class given a feature vector using Bayes' theorem. This probabilistic formulation allows it to model the complex relationships between the diverse statistical features extracted from the vibration signals and the different fault conditions present in the gearbox effectively. Capturing these intricate patterns through probabilities enabled Naive Bayes to discriminate faults with a high degree of accuracy.

Another advantage is the independence assumptions made by Naive Bayes between features. While the features extracted from the vibration data may indeed be correlated, the Naive Bayes assumption simplifies the learning problem. This reduction in complexity does not seem to impact performance for this application negati–

vely and may even be beneficial, given the size and dimensionality of the dataset.

The probabilistic and decomposable nature of Naive Bayes is well-matched to problems involving subtle differences between multi-class outcomes. This aligns well with the challenge of distinguishing nuanced fault signatures from signal analysis. The results demonstrate Naive Bayes was able to leverage the statistical features to most clearly delineate patterns corresponding to specific gearbox fault modes.

Collectively, these factors helped Naive Bayes achieve superior classification performance over SVMs and KNN on this complex gearbox fault diagnosis task using vibration signals. This enables the Naive Bayes model to reach 95.7% accuracy, a substantial impro–vement over the 90.8% benchmark reported by [15]. The probabilistic approach proves adept at relating these information-rich features to the gearbox's complex fault conditions.

Our SVM design also advances beyond earlier imp–lementations. Optimization of the kernel parameters and input feature selection results in 93.2% accuracy, surpassing the 89.1% achieved in prior SVM studies [29]. This highlights the benefits of custom-engineering the algorithm for gearbox multi-fault classification.

The KNN model similarly outperforms past work, with the optimized distance metric and voting scheme leading to 89.5% accuracy compared to 86.3% [6][46]. This nonparametric instance-based technique proves effective for vibration pattern recognition. Critically, no single algorithm was uniformly optimal across all fault types. The strengths of Naive Bayes, SVM, and KNN depended on the specific gear or bearing failure charac-teristics. This underscores the value of a diversified ensemble approach.

The synergistic fusion of optimized models provides more robust fault discrimination than single classifiers. Our systematic methodology integrates domain know-ledge and data-driven techniques to push the boundaries of diagnostic performance for enhanced wind turbine reliability. However, no individual technique dominated across all the gearbox failure modes, with relative cla-ssifier performance dependent on the inherent charac-teristics of each fault type.

This pointed to the merits of an integrated learning approach combining Naive Bayes, SVM, and KNN to provide enhanced robustness.

The accomplishments of this work become even more salient when juxtaposed against previous vibration analysis studies, as compiled in Table 3. On this chal–lenging multifault gearbox data, our methodology achi–eved accuracy levels substantially beyond earlier ben–chmarks, a result attributed to the information-fusion feature extraction and strategic combination of comple–mentary machine learning models.

For instance, the 95.7% Naive Bayes accuracy exce–eded the 90.8% result reported for SVM and ANN mo–dels in the prominent study by Hameed et al. [16]. Meanwhile, the optimized SVM approach presented here surpassed the 89.1% accuracy achieved in earlier SVM research by Yang et al. [49].

Wind turbine gearbox fault diagnosis based on an improved supervised autoencoder using vibration and motor current. Compared to prior work relying on single algorithms, the integrated learning paradigm developed in this work pushes the boundaries of diagnostic performance to new heights. The significance of these gains highlights the importance of simulating practical damage scenarios in various gear and bearing components with different levels of severity. Training the models using well-characterized fault progression data significantly enhanced the performance of multi-fault classification compared to relying solely on operational turbine data. The results of this study indicate a new era in which machine learning significantly enhances the ability to monitor conditions, leading to unprecedented levels of turbine utilization, availability, and service life optimization.

Table 3. Comparison of model accuracy with prior work.

Model	This Study	Yang et al. [49]	Hameed et al. [16]	Lu Y et al. [50]	Vives-Martinez et al.. [28]
Naive Bayes	95.7%	-	-	-	-
SVM	93.2%	89.1%	90.8%	-	86.3%
KNN	89.5%	-	87.2%	-	-
Ensemble	-	-	-	88.2%	-

As a result, the machine learning models demon-strated potential for wind turbine condition monitoring when applied to operational gearbox data. Naive Bayes especially stood out, accurately detecting faults with precision. Previous studies found it effective for gear–box diagnosis and complex classification. Using statis–tical features from vibration signals, Naive Bayes re–vealed links between characteristic patterns and faults. It distinguished HS-ST gear damage and IMS-ST wear with 96% and 95% accuracy, identifying subtle dif–ferences. SVM and KNN struggled to compare some faults, like HS-SH and IMS-SH bearings. This real-world analysis provides insights into algorithms' capa–bilities and limitations for refinement. Investigating Na–ive Bayes' strong performance could illuminate its com–plexity in handling related classes. Future work should detail features and their extraction from vibration data. Overall, the models showed promise, warranting conti–nued validation across turbine variants and operating conditions. With refinement, they may help advance wind turbine monitoring.

5. CONCLUSION

This research paper presents a comprehensive machine-learning framework designed to diagnose faults in wind turbine gearboxes via vibration signal analysis. The strict experimental setup, which included a 750kW turbine testbed and precise fault injections across gears and bearings, made it possible to get very realistic vibration data from a number of different operational situations. Signal processing specialists successfully extracted statistical characteristics from the vibration patterns, which unveiled intricate fault signatures and a healthy state in the same operation condition. The Naive Bayes algorithm outperformed other device mastering methods in detecting gearbox screw-ups, consisting of

SVM, KNN, and others. The Naive Bayes classifier showed how well its probabilistic approach modeled the complicated relationships between the extracted capabilities and the multifaceted fault situations by getting an impressive 95.7% accuracy charge, which was a lot higher than previous benchmarks. The confusion matrix evaluation provided additional insight into the benefits and drawbacks of each algorithm. Furthermore, it demonstrated the cost of an included learning strategy that harmoniously blends the various models. In order to enhance condition tracking and predictive protection techniques, this diverse ensemble approach strengthens and stabilizes the fault analysis machine. This research highlights how machine learning techniques are crucial for improving wind turbine reliability and making maintenance efforts more efficient. Intelligent fault diagnosis systems can use vibration data to proactively identify and categorize gearbox abnormalities to reduce the likelihood of catastrophic failures and expensive downtime. While this study represents a substantial stride forward, the journey toward optimizing wind turbine performance still needs to be completed. Future efforts will focus on validating the proposed methodology across a broader spectrum of wind turbine platforms, operational environments, and fault severities, further solidifying its versatility and adaptability. Adding present-day technologies like online learning paradigms and deep learning architectures to the mix can also cause even more accurate diagnoses and continuous real-time monitoring. As the field of renewable power continues to evolve, this study emphasizes the critical role of fact-driven techniques and machine learning in paving the way for a more sustainable and price-powerful future.

NOMENCLATURE AND ACRONYMS

Term	Definition
RMS (Root Mean Square)	A statistical measure of the magnitude of a varying quantity.
Peak-to-Peak	The difference between the maximum and minimum values in a dataset.
MAV (Mean Absolute Value)	The average of the absolute values of a set of numbers.
Standard Deviation (σ)	A measure of the amount of variation or dispersion in a set of values.
p_i	represents the probability that the ith amplitude level will occur.
Kurtosis	A measure of the "tailedness" of the probability distribution of a variable.
Skewness	A measure of the asymmetry of the probability distribution of a variable.
Spectral Crest Factor (SCF)	The ratio of the peak amplitude to the RMS level in the frequency spectrum.
Entropy	A measure of the randomness or disorder of a system.
F_N	False negative values
F_P	False positive values
T_N	True negative values
T_P	True positive values

ACKNOWLEDGMENTS

The researchers extend their sincere appreciation to Mustansiriyah University(www.uomustansiriyah.edu.iq) and its College of Engineering for their invaluable support throughout this investigation.

REFERENCES

[1] T. Burton, N. Jenkins, D. Sharpe, and E. Bossanyi, *Wind energy handbook*. John Wiley & Sons, 2011.

[2] D. S. Rehman, K. Fahd, H. H. Nuha, A. S. Balg–honaim, R. R. Pahlevi, and M. A. Mohandes, ‘Vertical wind speed extrapolation using statistical approaches’, *FME Transactions*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:267732492

[3] W. Yang, P. J. Tavner, C. J. Crabtree, Y. Feng, and Y. Qiu, ‘Wind turbine condition monitoring: technical and commercial challenges’, *Wind Energy*, vol. 17, no. 5, pp. 673–693, May 2014, doi: https://doi.org/10.1002/we.1508.

[4] A. A. F. Ogaili and M. N. Hamzah, ‘Integration of machine learning (ML) and finite element analysis (FEA) for predicting the failure modes of a small horizontal composite blade’.

[5] D. S. Rehman, K. Fahd, H. H. Nuha, A. S. Balghonaim, R. R. Pahlevi, and M. A. Mohandes, ‘Vertical wind speed extrapolation using statistical approaches’, *FME Transactions*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:267732492

[6] A. A. F. Ogaili, A. A. Jaber, and M. N. Hamzah, ‘A methodological approach for detecting multiple faults in wind turbine blades based on vibration signals and machine learning’, vol. 10, no. 1, 2023, doi: doi:10.1515/cls-2022-0214.

[7] Y. ElMessiry, H. A. Kandil, and M. S. Abd-Elhady, ‘Effect of surface contamination on the wind turbine performance’, *Wind Engineering*, vol. 45, no. 3, pp. 505–517, Mar. 2020, doi: 10.1177/0309524X20911176.

[8] M. A. S. Shohag, E. C. Hammel, D. O. Olawale, and O. I. Okoli, ‘Damage mitigation techniques in wind turbine blades: A review’, *Wind Engineering*, vol. 41, no. 3, pp. 185–210, Jun. 2017, doi: 10.1177/0309524X17706862.

[9] A. Ghoshal, M. Sundaresan, M. J. Schulz, and P. F. Pai, ‘Structural health monitoring techniques for wind turbine blades’, *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 85, pp. 309–324, 2000, [Online]. Available: https://api.semanticscholar.org/CorpusID:110628890

[10] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, and M. Papaelias, ‘Condition monitoring of wind turbines: Techniques and methods’, *Renew Energy*, vol. 46, pp. 169–178, 2012.

[11] A. A. Ogaili, M. N. Hamzah, and A. A. Jaber, 'Free vibration analysis of a wind turbine blade made of composite materials', in *International Middle Eastern Simulation and Modeling Conference*, 2022, pp. 27–29.

[12] A. A. F. Ogaili, A. A. Jaber, and M. N. Hamzah, 'Wind turbine blades fault diagnosis based on vibration dataset analysis', *Data Brief*, vol. 49, p. 109414, 2023.

[13] A. A. Ogaili, M. N. Hamzah, A. Ab. Jaber, and E. Ghane, 'Application of Discrete Wavelet Transform for Condition Monitoring and Fault Detection in Wind Turbine Blades: An Experimental Study', *Engineering and Technology Journal*, vol. 42, no. 1, pp. 104–116, 2024, doi: 10.30684/etj.2023.142023.1516.

[14] E. S. Al-Ameen, 'An Idea for Coupling Two Engines to Drive a Generating Head Using Epicyclic Gear Train', *Journal of Engineering and Sustainable Development*, vol. 12, no. 2, pp. 18–30, Jun. 2008, [Online]. Available: https://jeasd.uomustansiriyah.edu.iq/index.php/jeasd/article/view/1661

[15] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaelias, 'Condition monitoring of wind turbines: Techniques and methods', *Renew Energy*, vol. 46, pp. 169–178, 2012, doi: https://doi.org/10.1016/j.renene.2012.03.003.

[16] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. 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, doi: https://doi.org/10.1016/j.rser.2007.05.008.

[17] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaelias, 'Condition monitoring of wind turbines: Techniques and methods', *Renew Energy*, vol. 46, pp. 169–178, 2012, doi: https://doi.org/10.1016/j.renene.2012.03.003.

[18] Z. Hajej, N. Rezg, M. Bouzoubaa, 'An integrated maintenance strategy for a power generation system under failure rate variation (case of wind turbine)', in *2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2017, pp. 76–79. doi: 10.1109/ICRERA.2017.8191175.

[19] 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*, vol. 44, pp. 466–472, 2015, doi: https://doi.org/10.1016/j.rser.2014.12.005.

[20] W. Lu and F. Chu, 'Condition monitoring and fault diagnostics of wind turbines', in *2010 Prognostics and System Health Management Conference*, 2010, pp. 1–11. doi: 10.1109/PHM.2010.5413527.

[21] B. Rašuo, M. Dinulović, A. Veg, A. Grbović, and A. Bengin, 'Harmonization of new wind turbine rotor blades development process: A review', *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 874–882, Nov. 2014, doi: 10.1016/J.RSER.2014.07.137.

[22] S. M. Jawad, A. A. Jaber, 'Rolling Bearing Fault Detection Based on Vibration Signal Analysis and Cumulative Sum Control Chart', *FME Transactions*, vol. 49, no. 3, pp. 684–695, 2021, doi: 10.5937/fme2103684M.

[23] L. A. Al-Haddad and A. A. Jaber, 'An Intelligent Fault Diagnosis Approach for Multirotor UAVs Based on Deep Neural Network of Multi-Resolution Transform Features', *Drones*, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:256293203

[24] L. A. Al-Haddad, A. A. Jaber, M. N. Hamzah, and M. A. Fayad, 'Vibration-current data fusion and gradient boosting classifier for enhanced stator fault diagnosis in three-phase permanent magnet synchronous motors', *Electrical Engineering*, 2023, doi: 10.1007/s00202-023-02148-z.

[25] O. I. Abdullah and E. S. Al-Ameen, 'Vibration of Non-rotating Blades Experimental and Numerical Investigation', *Journal of Engineering and Sustainable Development*, vol. 11, no. 2, pp. 113–124, Sep. 2007, [Online]. Available: https://jeasd.uomustansiriyah.edu.iq/index.php/jeasd/article/view/1724

[26] 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*, vol. 44, pp. 466–472, 2015, doi: https://doi.org/10.1016/j.rser.2014.12.005.

[27] S. M. Shakir, A. A. Jaber, 'Innovative Application of Artificial Neural Networks for Effective Rotational Shaft Crack Localization', *FME Transactions*, vol. 52, no. 1, pp. 103–114, 2024, doi: 10.5937/fme2401103S.

[28] [28] J. Vives, 'Vibration analysis for fault detection in wind turbines using machine learning techniques', *Advances in Computational Intelligence*, vol. 2, no. 1, p. 15, 2022, doi: 10.1007/s43674-021-00029-1.

[29] J. Vives, 'Vibration analysis for fault detection in wind turbines using machine learning techniques', *Advances in Computational Intelligence*, vol. 2, no. 1, p. 15, 2022, doi: 10.1007/s43674-021-00029-1.

[30] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. 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, doi: https://doi.org/10.1016/j.rser.2007.05.008.

[31] L. O. Felix, D. H. C. de Sá Só Martins, U. A. B. V. Monteiro, B. M. Castro, L. A. V. Pinto, and C. A. O. Martins, 'A Feature Selection Committee Method Using Empirical Mode Decomposition for Multiple Fault Classification in a Wind Turbine Gearbox', *J Nondestr Eval*, vol. 42, no. 4, p. 85, 2023, doi: 10.1007/s10921-023-00996-0.

[32] C. Yi, F. Zhang, T. Huang, H. Xiao, and B. Qin, 'Wind turbine gearbox fault diagnosis via adaptive

IMFogram', *Struct Health Monit*, p. 147592172 31209904, Nov. 2023, doi: 10.1177/147592172 31209903.

[33] S. V. V. S. N. Pichika, G. Meganaa, S. Geetha Rajasekharan, and A. Malapati, 'Multi-component fault classification of a wind turbine gearbox using integrated condition monitoring and hybrid ensemble method approach', *Applied Acoustics*, vol. 195, p. 108814, 2022, doi: https://doi.org/10.1016/j.apacoust.2022.108814.

[34] K. Zhang, B. Tang, L. Deng, Q. Tan, and H. Yu, 'A fault diagnosis method for wind turbines gearbox based on adaptive loss weighted meta-ResNet under noisy labels', *Mech Syst Signal Process*, vol. 161, p. 107963, 2021, doi: https://doi.org/10.1016/j.ymssp.2021.107963.

[35] W. Touti, M. Salah, S. Sheng, and K. Bacha, 'An Envelope Time Synchronous Averaging for Wind Turbine Gearbox Fault Diagnosis', *Journal of Vibration Engineering & Technologies*, 2024, doi: 10.1007/s42417-023-01267-y.

[36] W. Hu, H. Chang, and X. Gu, 'A novel fault diagnosis technique for wind turbine gearbox', *Appl Soft Comput*, vol. 82, p. 105556, 2019, doi: https://doi.org/10.1016/j.asoc.2019.105556.

[37] S. Sheng, 'Wind turbine gearbox vibration condition monitoring benchmarking datasets', *NREL National Wind Technology Center, Boulder, CO, Report No. NREL/TP-5000-54530*, 2012.

[38] A. A. Dubaish, A. A. Jaber, 'State-of-the-Art Review into Signal Processing and Artificial Intelligence-based Approaches Applied in Gearbox Defect Diagnosis', *Engineering and Technology Journal*, pp. 1–16, 2023.

[39] A. A. Jaber, R. Bicker, 'Fault diagnosis of industrial robot gears based on discrete wavelet transform and artificial neural network', *Insight-Non-Destructive Testing and Condition Monitoring*, vol. 58, no. 4, pp. 179–186, 2016.

[40] V. N. Vapnik, 'Statistical learning theory', 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:28637672

[41] W. H. Alawee, L. A. Al-Haddad, H. A. Dhahad, S. A. Al-Haddad, 'Predicting the Cumulative Productivity of a Solar Distillation System Augmented with a Tilted Absorber Panel Using Machine Learning Models', *Journal of Engineering Research*, 2024, doi: https://doi.org/10.1016/j.jer.2024.01.007.

[42] S. A. Mohammed, L. A. Al-Haddad, W. H. Alawee, H. A. Dhahad, A. A. Jaber, and S. A. Al-Haddad, 'Forecasting the productivity of a solar distiller enhanced with an inclined absorber plate using stochastic gradient descent in artificial neural networks', *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 2023, doi: 10.1007/s41939-023-00309-y.

[43] R.-V. Sánchez, P. Lucero, R. E. Vásquez, M. Cerrada, J.-C. Macancela, and D. Cabrera, 'Feature ranking for multi-fault diagnosis of rotating machi–nery by using random forest and KNN', *Journal of Intelligent & Fuzzy Systems*, vol. 34, pp. 3463–3473, 2018, doi: 10.3233/JIFS-169526.

[44] Y. Wang, Y. Huang, K. Yang, Z. Chen, and C. Luo, 'Generator Fault Classification Method Based on Multi-Source Information Fusion Naive Bayes Classification Algorithm', *Energies (Basel)*, vol. 15, no. 24, 2022, doi: 10.3390/en15249635.

[45] L. A. Al-Haddad, N. M. Mahdi, 'Efficient multidisciplinary modeling of aircraft undercarriage landing gear using data-driven Naïve Bayes and finite element analysis', *Multiscale and Multidisciplinary Modeling, Experiments and Design*, pp. 1–13, Mar. 2024, doi: 10.1007/S41939-024-00389-4/METRICS.

[46] L. A. Al-Haddad, A. A. Jaber, M. N. Hamzah, M. A. Fayad, 'Vibration-current data fusion and gradient boosting classifier for enhanced stator fault diagnosis in three-phase permanent magnet synchronous motors', *Electrical Engineering*, 2023, doi: 10.1007/s00202-023-02148-z.

[47] S. A. Mohammed, L. A. Al-Haddad, W. H. Alawee, H. A. Dhahad, A. A. Jaber, and S. A. Al-Haddad, 'Forecasting the productivity of a solar distiller enhanced with an inclined absorber plate using stochastic gradient descent in artificial neural networks', *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 2023, doi: 10.1007/s41939-023-00309-y.

[48] A. A. F. Ogaili, M. N. Hamzah, A. A. Jaber, 'En–hanced Fault Detection of Wind Turbine Using eXtreme Gradient Boosting Technique Based on Nonstationary Vibration Analysis', *Journal of Fai–lure Analysis and Prevention*, 2024, doi: 10.1007/s11668-024-01894-x.

[49] S. Yang, Y. Wang, C. Li, 'Wind turbine gearbox fault diagnosis based on an improved supervised autoencoder using vibration and motor current signals', *Meas Sci Technol*, vol. 32, no. 11, p. 114003, 2021, doi: 10.1088/1361-6501/ac0741.

[50] Y. Lu, J. Tang, H. Luo, 'Wind Turbine Gearbox Fault Detection Using Multiple Sensors With Fea–ture Level Data Fusion', 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:196057064

АУТОМАТИЗАЦИЈА ПРАЋЕЊА СТАЊА РЕДУКТОРА ВЕТРОТУРБИНА: УПОРЕДНА СТУДИЈА ТЕХНИКЕ МАШИНСКОГ УЧЕЊА ЗАСНОВАНЕ НА АНАЛИЗИ ВИБРАЦИЈА

А.А.Ф. Огаили, К.А. Мохамед, А.А. Џабер
Е.С. Ал-амин

Ветротурбине играју улогу у усвајању производње обновљиве енергије, али су подложне гашењима која захтевају темељно праћење. Кварови редуктора су проблем који доводи до одржавања и застоја у раду. Ова студија истражује примену метода машинског учења за побољшање дијагнозе проблема са педуктором помоћу анализе вибрација.

Применом сценарија кварова који утичу на лежајеве и зупчанике, истраживачи су успешно издвојили карактеристике временског домена из података о вибрацијама на тестној станици од 750 кВ да би открили индикације оштећења. За класификацију грешака редуктора коришћени су модели машинског учења Support Vector Machine (SVM), Naive Bayes, and K Nearest Neighbour (KNN). Међу овим моделима, Naive Bayes је постигао стопу тачности од 95,7%, што је премашило утврђене стандарде. Вероватносни приступ је успео да успешно повеже карактеристике симптома са обрасцима грешака. Интелигентни системи за праћење могли би да побољшају ефикасност одржавања. Овај приступ заснован на подацима наглашава потенцијал маши–нског учења у подршци развоју енергије ветра елиминисањем неефикасности редуктора и побољ–шањем поузданости турбине, а даља истраживања се спроводе како би се осигурало да овај приступ функционише у складу са разноврсношћу и у стварном свету. Ово показује како машинско учење доприноси напретку у обновљивој енергији тако што помаже у анализи проблема са предвиђањем и спречавању скупих кварова на редуктору.