



Bearing faults classification using a new approach of signal processing combined with machine learning algorithms

F. Gougam¹ · A. Afa² · A. Soualhi³ · W. Touzout¹ · C. Rahmoune¹ · D. Benazzouz¹

Received: 25 September 2023 / Accepted: 11 December 2023 / Published online: 10 January 2024
© The Author(s), under exclusive licence to The Brazilian Society of Mechanical Sciences and Engineering 2024

Abstract

Vibration analysis plays a crucial role in fault and abnormality diagnosis in various mechanical systems. However, efficient vibration signal processing is required for valuable diagnosis and hidden patterns' detection and identification. Hence, the present paper explores the application of a robust signal processing method called maximal overlap discrete wavelet packet transform (MODWPT) that supports multiresolution analysis, allowing for the examination of signal details at different scales. This capability is valuable for identifying faults that may manifest at different frequency ranges. MODWPT is combined with covariance and eigenvalues to signal reconstruction. After that, health indicators are specifically applied on the reconstructed vibration signal for feature extraction. The proposed approach was carried out on an experimental test rig where the obtained results demonstrate its effectiveness through confusion matrix analysis of machine learning tools. The ensemble tree model gives more accurate results (accuracy and stability) of bearing faults classification and efficiently identify potential failures and anomalies in mechanical equipment.

Keywords Vibration signal · Fault diagnosis · Health monitoring · Signal processing · Machine learning

1 Introduction

Bearings are significantly recognized as the predominant mechanical components in rotating machinery [1]. Their primary function is to limit frictions among the various moving parts within rotary devices such as alternators, compressors, turbines, drives, and more. Despite their crucial role, bearings are susceptible to deterioration and are regarded as the weakest links in industrial equipment. In fact, studies indicate that approximately 40% of electric motor failures can be attributed to bearing defects [2]. Consequently, it becomes

imperative to detect and diagnose rolling element defects at early stages for unplanned failures' prevention, production downtime prediction, and efficiency degradation estimation [3].

Indeed, vibration analysis has become a key tool in the field of diagnostic engineering, providing valuable information regarding mechanical equipment health and performance. By capturing and analyzing vibrational signals, engineers and researchers can detect faults, identify abnormal behavior, and predict potential failures [4]. Basically, the effective processing and interpretation of vibrational signals require significant challenges due to their complex nature and noise effects; however, vibration signals in rotating machines exhibit nonlinear and non-stationary behavior due to the variation harmonics in the working conditions mainly load and speed. This provides a challenge in fault diagnosis as fault signatures are often obscured within a complex and overwhelming signal content, particularly during the early stages.

In general, mechanical defects generate non-stationary signals associated with background noise during acquisition of experimental signals. However, conventional methods such as fast Fourier transform (FFT) encounter complexity when analyzing such signals [5, 6], the primary limitation of

Technical Editor: Marcelo Areias Trindade.

✉ F. Gougam
f.gougam@univ-boumerdes.dz

¹ Solid Mechanics and Systems Laboratory (LMSS), Department of Mechanical Engineering, University M'Hamed Bougara Boumerdes, Boumerdes, Algeria

² Department of Mechanical and Process Engineering, Faculty of Mechanical Engineering and Process Engineering, Houari Boumediene University of Science and Technology, Algiers, Algeria

³ Laboratoire LASPI, University Lyon, UJM-Saint Etienne, 42100 Saint-Etienne, France

FFT lies in its inability to identify the dominant frequencies present in the raw signal in addition to its uselessness in the time domain. This basically implies that, although the presence of a fault at a certain signal frequency can be identified, the exact occurrence timing remains unknown. Moreover, this information is crucial for identifying resonance frequencies, detecting harmonic components, and identifying specific fault signatures in the frequency domain [7]. Therefore, to overcome this issue, researchers have explored alternative techniques such as time–frequency analysis, which offers the potential to detect abnormalities due to the occurred faults. Furthermore, various methods have been investigated for fault diagnosis, including short-time Fourier transform (STFT) [8], wavelet transform (WT) [9], wavelet packet transform (WPT) [10], empirical mode decomposition (EMD) [11], local mean decomposition [12], and empirical wavelet transform (EWT) [13]. Those techniques are designed to improve fault detection using temporal and frequency information simultaneously.

As previously mentioned, considerable effort has been devoted in recent decades to develop fault detection methods, specifically for rotating machines, with particular focus on bearings. Shanbr et al. [14] have explored the application of kurtogram analysis on sets of vibration signals to assess its ability to extract fault frequency characteristics in wind turbine gearboxes. Rahmoune et al. [15] have used kurtogram analysis to monitor gearbox condition using electrical signatures. Gougam et al. [13] introduced EWT as a denoising technique for raw vibration signals, preserving impulsive EWT modes during the reconstruction process. Additionally, maximal overlap discrete wavelet packet transform (MODWPT) offers a powerful tool for time–frequency analysis in vibration analysis [16]. By decomposing the signal into time–frequency coefficients, MODWPT provides localized information about changes in vibration characteristics over time. The application of MODWPT to fault diagnosis, transient analysis and frequency modulation effects in vibration signals is further discussed in Sect. 4.

In machine learning, the main purpose is to find and extract the most relevant features that significantly contribute to determining, influencing, and validating the predicted results. Consequently, feature extraction represents a challenging procedure which implies the adoption of advanced algorithms based on robust signal processing techniques [17, 18]. The initial step in the feature extraction process involves applying an efficient signal processing technique that filters out and eliminates the noise impact, which can ultimately lead to erroneous diagnostic results [19–21]. Besides, several sources can contribute to model errors. This includes bias, variance, and noise. Bias refers to the simplifying assumptions made by a model, which may cause it consistently or systematically miss significant data patterns [22]. Variance, meanwhile, refers

to the model's sensitivity to fluctuations in the training dataset, potentially leading to overfitting. Noise refers to the presence of random and irrelevant information in the data, which can also have an impact of the model performance. Through ensemble learning, the machine learning algorithms stability and accuracy are improved, resulting more reliable and robust predictions. The combination of several models overcomes the limits of individual models and improves overall generalization and predictive power [23–26].

In this paper, a novel approach combining MODWPT and autocovariance is investigated to process the original vibration data in order to eliminate any remaining noise on the raw vibration signals. Subsequently, several time-domain features (TDFs) are computed from the reconstructed signal to build feature matrices containing the relevant information. Afterward, a feature selection method is employed to identify and select relevant features associated with distinct fault signatures. This study involves four types of bearing faults under varying conditions, as well as several health states. The proposed approach is evaluated using real-time datasets obtained from bearings operating under various realistic conditions. Additionally, several machine learning classifiers are used to improve the automatic fault classification process.

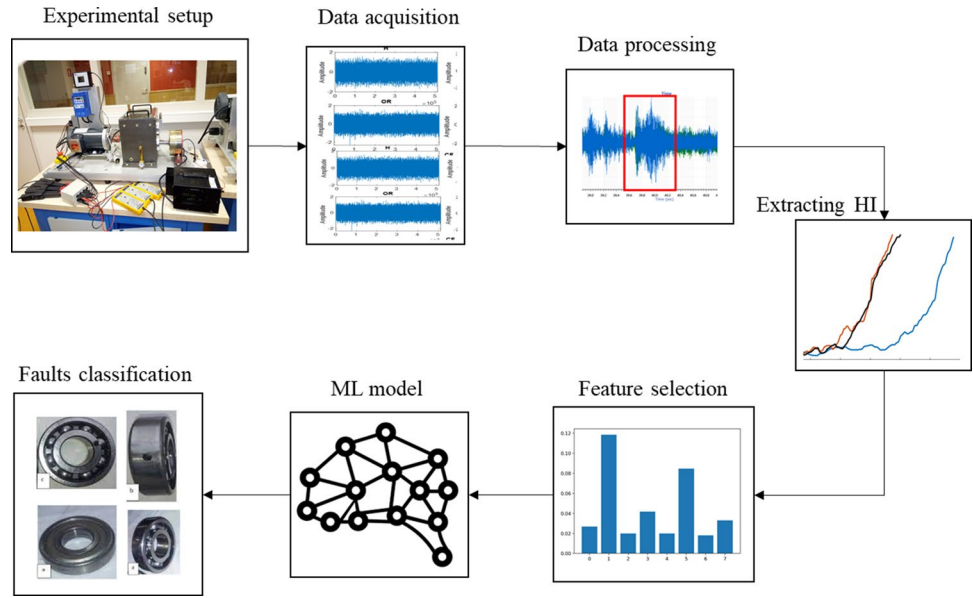
2 Methodology of the proposed approach

- MODWPT decomposition and rearrangement of MODWPT coefficients in a 2D array.
- Subtract the MODWPT coefficients from their mean signal.
- Calculate the covariance matrix of the resulting signal after subtraction.
- Extract singular values and sort eigenvectors in descending order of eigenvalues.
- Project the MODWPT coefficients to the new principal axes.
- Reconstruct the signal using the new MODWPT coefficients.

3 Flowchart

From Fig. 1, the vibration signals are acquired using accelerometers in the experimental setup. Subsequently, we apply the proposed approach which is based on MODWPT (maximal overlap discrete wavelet packet transform) decomposition, for robust signal extraction. A set of health indicators (HIs) is computed from the reconstructed signal for feature extraction. After that, a feature selection

Fig. 1 Steps of faults diagnosis



algorithm (Relief) is integrated to select the most relevant features for fault classification. Finally, we use machine learning (ML) tools for automated fault diagnosis, employing various models.

4 Mathematical background

4.1 Maximal overlap discrete wavelet packet transform (MODWPT)

MODWPT represents a significant advancement in the field of wavelet transforms, taking the concepts of DWT (discrete wavelet transform) and MODWT (maximal overlap wavelet transform) as outlined by Afia et al. [16]. Similarly, to Mallat's algorithm, MODWPT uses quadrature mirror filters. In this context, g and h are denoted as the low-pass and high-pass filters, respectively, with L representing their length (where L is an even number). The design of these filters follows the approach detailed by Afia et al. [16].

$$\begin{cases} \sum_{i=0}^{L-1} \tilde{h}_i^2 = \frac{1}{2} \\ \sum_{i=0}^{L-1-2k} \tilde{h}_i \tilde{h}_{i+2k} = 0, k = 1, 2, \dots, \frac{1}{2}(L-2) \\ \tilde{g}_l = (-1)^{l+1} \tilde{h}_{L-l-1}, l = 0, 1, \dots, L-1 \end{cases} \quad (1)$$

In contrast to Mallat's algorithm, which employs a two-base decimation operation, MODWPT utilizes interpolation. Specifically, at each level of the MODWPT, $2^{j-1} - 1$ zeros are inserted between consecutive adjacent

coefficients of \hat{g}_1 and \hat{h}_1 . This ensures that the resulting wavelet coefficients (WTs) for each wavelet sub-band maintain the same length as the original signal [27–29].

For a discrete time, sequence $\{x(t), t = 0, 1, \dots, N - 1\}$, WT's $\{W_{j,n,t}\}$ of the n th sub-band at level j are computed as follows:

$$W_{j,n,t} = \sum_{i=0}^{L-1} \tilde{f}_{n,l} \tilde{W}_{j-1, [n/2], (t-2^{j-1}) \bmod N, t=0, 1, \dots, N-1} \quad (2)$$

where $n = 0, 1, \dots, 2^{j-1}, W_{0,0,t} = x(t)$

$$\tilde{f}_{n,l} = \begin{cases} \tilde{g}_l, & \text{if } n \bmod 4 = 0 \text{ or } 3 \\ \tilde{h}_l, & \text{if } n \bmod 4 = 1 \text{ or } 2 \end{cases} \quad (3)$$

4.2 Covariance

Covariance is a fundamental component in various statistical analyses and models. It serves as the basis for calculating correlation coefficients, which further quantify the strength and direction of the linear relationship between variables. The covariance is defined as the measure of the relationship between two vectors of variables, A and B [30].

$$\text{cov}(A, B) = \frac{1}{N-1} \sum_{i=1}^N (A_i - \mu_A) * (B_i - \mu_B) \quad (4)$$

The covariance matrix of two random variables represents a matrix containing pairwise covariance calculations between each variable.

$$C = \begin{pmatrix} \text{cov}(A, A) & \text{cov}(A, B) \\ \text{cov}(B, A) & \text{cov}(B, B) \end{pmatrix} \tag{5}$$

For a matrix A , in which each column represents a random variable composed of observations, the covariance matrix is calculated by computing the pairwise covariances between each combination of columns.

$$C(i, j) = \text{cov}(A(:, i), A(:, j)) \tag{6}$$

Measures Relationship Covariance provides a measure of the relationship between two variables. A positive covariance indicates a direct relationship, where both variables tend to increase or decrease together [31].

Scale Independence Covariance is scale-independent, meaning it is not affected by changes in the units or scales of the variables. This makes it a useful measure for comparing relationships between variables with different units or scales [32].

Provides Insights Covariance can provide insights into the direction and strength of the linear relationship between variables. It can be used to identify patterns or dependencies in the data and to assess the extent to which changes in one variable are associated with changes in another [33].

4.3 Signal reconstruction

The reconstructed signal is based on transformed coefficients with covariance and the singular values of centered coefficient:

$$C_C = C - M_C \tag{7}$$

where C_C is centered coefficients and C the MODWPT coefficient, while M_C is the mean of MODWPT coefficients.

Table 1 Equations of statistical features

Dimensional features	Symbol	Formula
Kurtosis	X_{kur}	$\frac{\sum_{i=1}^n [x_i - \bar{x}]^4}{\sum_{i=1}^n [x_i - \bar{x}]^2} \frac{\sum_{i=1}^n [x_i - \bar{x}]^2}{(N-1)X_{sd}^2}$
Standard deviation	X_{sd}	$\sqrt{\frac{1}{N-1} \sum_{i=1}^N [x_i - \bar{x}]^2}$
Maximal	X_{max}	$\max(x_i)$
Root mean square	X_{rms}	$\sqrt{\frac{1}{N} \sum_{i=1}^N [x_i]^2}$
Peak to peak	X_{p2p}	$\max(x_i) - \min(x_i)$
Peak to rms	X_{p2rms}	$\frac{X_{max}}{X_{rms}}$

$$CV_C = \text{Cov}(C_C) = \text{Cov}(C - M_C) \tag{8}$$

$$v = \text{eigen}(CV_C) \tag{9}$$

where v is eigenvectors of covariance of centered coefficients.

$$X_{Rec} = \sum C_{Rec} = \sum (C_{TC} * v' + M_C)' \tag{10}$$

in which, C_{TC} is transformed coefficients and C_{Rec} reconstructed coefficients, while M_C and X_{Rec} are the mean coefficients and the reconstructed signal, respectively.

4.4 Scalar features extraction

Extracting scalar features from vibration signals involves analyzing the signal to obtain numerical parameters that describe various vibration aspects. These features can then be used for other analyses, such as fault detection, condition monitoring, and machine learning algorithms. The following are some commonly used scalar features in vibration signal analysis:

Mean The average value of the vibration signal.

Standard deviation A measure of the dispersion of the signal values around the mean.

Skewness Indicates the asymmetry of the signal distribution.

Kurtosis Describes the peakedness or flatness of the signal distribution.

Formulas for the most frequently used features are given in Table 1.

Within the context of bearing condition monitoring, the average spectral kurtosis (SK) is a robust feature used for fault detection. SK is highly effective in identifying impulsive boundary signatures, as important information may be concealed in the raw signal and revealed through spectral analysis. The SK of a given vibration signal refers to the kurtosis of its spectral components (Eq. 11) and is expressed as the normalized fourth-order spectral moment [26].

$$X_{msk} = \frac{\langle |X^4(t, f)| \rangle}{\langle |X^4(t, f)| \rangle^2} - 2 \tag{11}$$

where $\langle \cdot \rangle$ is the average time–frequency operator and $X^4(t, f)$ is the 4th order cumulant-filtered bandpass of the input signal X . Figure 2 illustrates the method procedure.

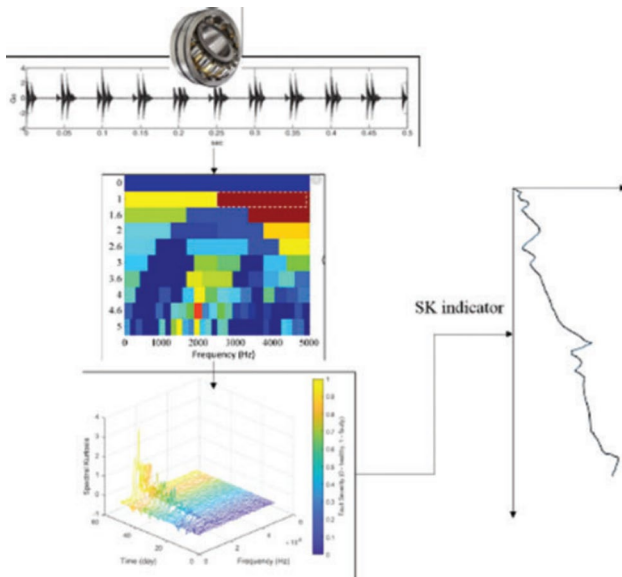


Fig. 2 Spectral kurtosis indicator

5 Automatic feature selection

Relief is a feature selection algorithm used for classification tasks. It evaluates the relevance of features by considering their ability to distinguish instances of different classes. The algorithm assigns weights to features based on observed differences in attribute values for nearest instances of the same class and nearest instances of different classes [26] (Fig. 3).

The main steps involved in the Relief algorithm are as follows [26]:

- Initialize feature weights to zero.
- For each instance, find the nearest instance of the same class (nearest hit) and the nearest instance of a different class (nearest miss).

- Update the feature weights based on the attribute value differences between the current instance and the nearest hit and miss instances.

$$W_i = W_i - (x_i - \text{nearHit}_i)^2 + (x_i - \text{nearMiss}_i)^2 \quad (12)$$

- Repeat the above steps for a specified number of iterations or until convergence.
- Rank the features based on their final weights, with higher weights indicating greater relevance.

By applying the Relief algorithm, the most relevant features for classification are selecting, which can improve the performance and interpretability of ML models.

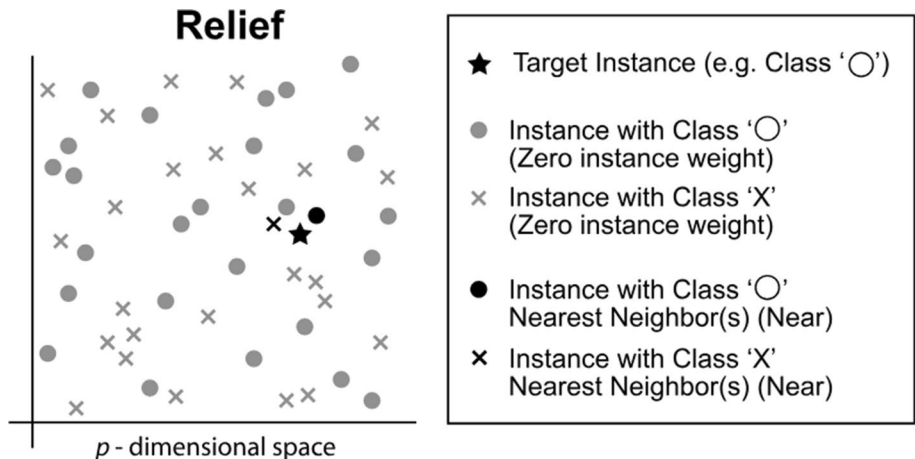
6 Ensemble learning

6.1 Definition

Ensemble learning is a technique used to enhance machine learning outcomes by combining multiple models [34]. This approach offers an improved classification performance compared to using a single model. The fundamental idea is to train a group of classifiers, known as experts, and allow them to collectively make predictions. Two common types of ensemble learning methods are bagging and boosting. Both approaches aim to reduce the variance of individual estimates by aggregating predictions from different models. Consequently, the resulting ensemble model tends to be more stable. The two terms are briefly discussed in the following [35]:

Bagging: bagging (bootstrap aggregating) is an intuitive and straightforward ensemble learning technique that exhibits good performance. As the name suggests, it combines two concepts: "bootstrap" and "aggregation." Bootstrap is another sampling method in which subsets of observations

Fig. 3 Bagging process flow



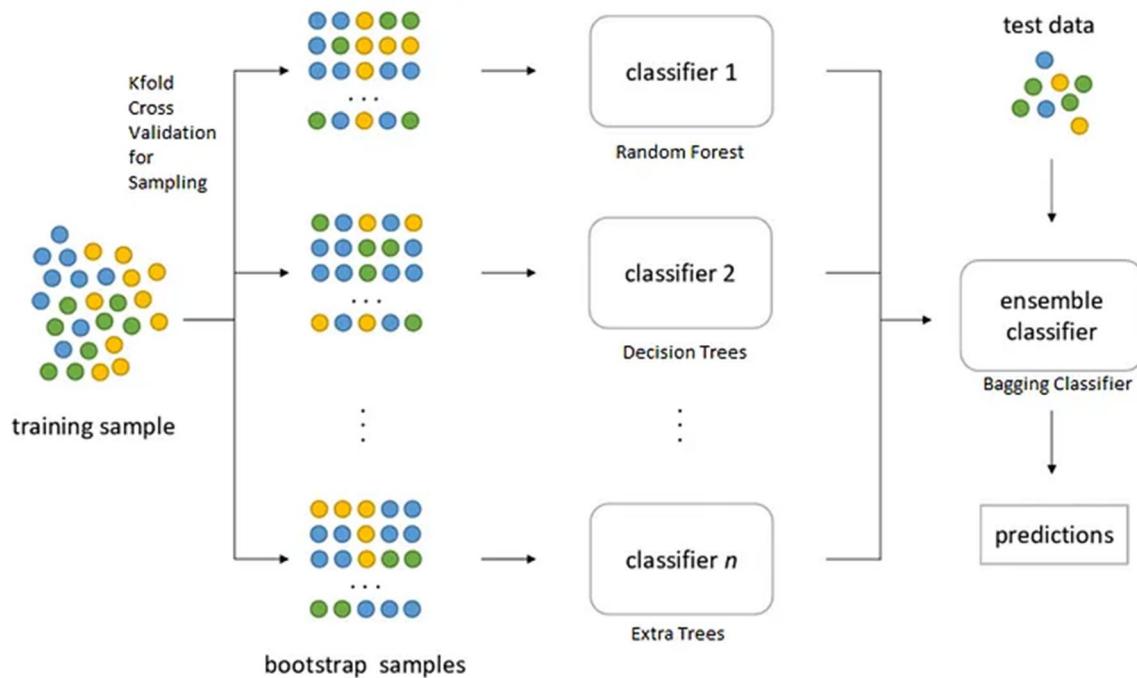


Fig. 4 Bagging process flow

are created by randomly selecting from the original dataset with replacement. This method involves training multiple homogeneous weak learners independently and in parallel. Their predictions are then combined to determine the model's average output.

Boosting: Similarly, boosting utilizes homogeneous weak learners, but it differs from bagging in its approach. In this method, learners are trained sequentially and adaptively, with each subsequent learner aiming to improve the model's predictions based on the previous learners' performance (Fig. 4).

6.2 Implementation steps of bagging

The ensemble learning process can be summarized as follows [36]:

- **Creation of Multiple Subsets** The original dataset is divided into multiple subsets, each containing an equal number of tuples. This is achieved by randomly selecting observations from the dataset, allowing for replacement (i.e., the same observation can be selected more than once in a subset).
- **Base Model Creation** A base model, typically a weak learner, is created for each of these subsets. Each base model is trained on a specific subset of the data.
- **Parallel Training** The base models are trained in parallel and independently of each other. Each model learns from

its respective training subset, using the chosen learning algorithm.

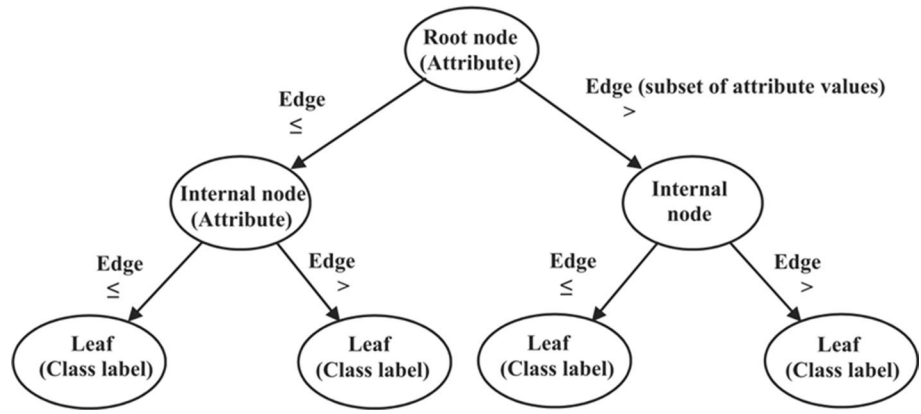
- **Combining Predictions** The final predictions are determined by combining the predictions made by all the base models. This can be done through various methods such as majority voting (in which the most common prediction is selected) or averaging (in which the predictions are averaged).

By following these steps, ensemble learning leverages the diversity of multiple models to improve overall prediction accuracy and robustness.

7 Decision tree (DT)

Decision tree (DT) is used to provide a prediction of target variable according to input variable by learning decision rules. Essentially, the fundamental concept of this algorithm is to identify the optimal attribute that carries representative information within the dataset. DT is a classification tree that is designed to classify or predict classification algorithms in machine learning. Classification tree models produce a discrete number of values for the target variables [37]. To construct a classification attribute model according to the other attributes, DT takes as input an object defined by a range of properties. The decision is made at every tree stage depending on the

Fig. 5 Decision tree form



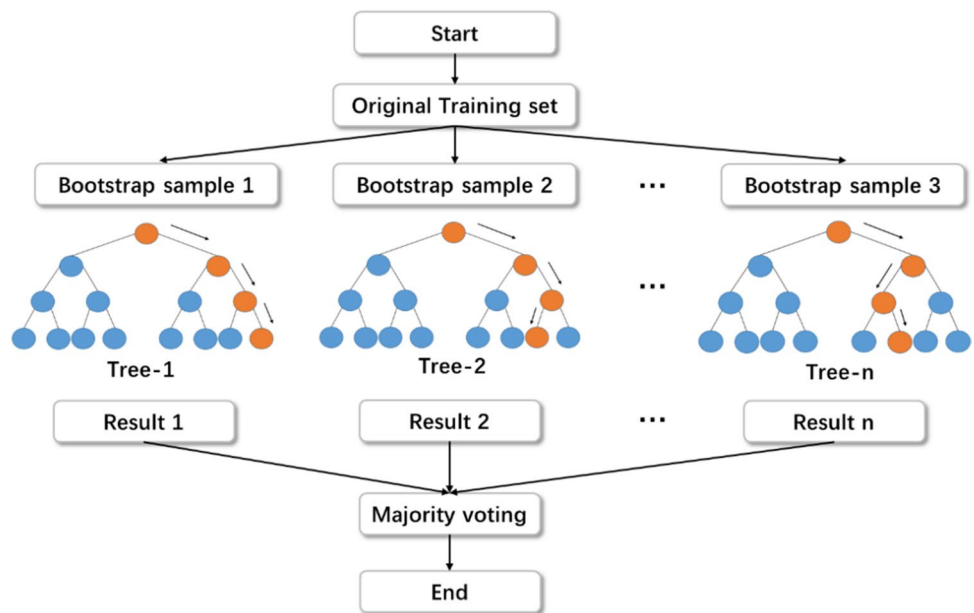
previous branching operations. DT is comprised of nodes which form a rooted tree, which is a directed tree with a node known as "root" with no incoming edges. The tree starts from the root node, then tests along the edge, and the test is repeated until the end node (leaf) is reached. After reaching the leaf node, the tree will predict its associated outcome, i.e., the class label. Iterative Dichotomiser 3 (ID3) and its successor C4.5 are the reference tools for DTs. DT is a sequential design that combines a series of basic tests in an efficient and consistent way, in which, at each test, a numerical feature is compared to a threshold value [38]. Conceptual rules are considerably easier to create than numerical weights in a neural network. Figure 5 gives an overview of the DT form [39].

8 Random forest (RF)

The random forest (RF) is a comprehensive tree-based classifier initially proposed by Leo Breiman [40]. It demonstrates remarkable adaptability to datasets, making it suitable for both extensive and limited regression and classification problems. RF is constructed from numerous decision trees (DTs), each contributing a classification for the input data. Subsequently, RF consolidates these classifications to identify the most heavily voted prediction as the final output.

RF relies on the construction of N decision trees using bagging ("bootstrap aggregating"). In this process, each tree draws a random sample from the data, and every node in the tree is split based on the optimal variable within the subset of input features obtained through the Gini index as a measure of attribute selection [41]. This index is computed after evaluating the impurity of attributes

Fig. 6 Typical structure of random forests



concerning the classes. Figure 6 provides a schematic representation of the RF algorithm.

The procedure for constructing a random forest is outlined as follows:

- Randomly choose n sub-training sets along with their corresponding test sets from the original dataset.
- During the split process of each node in the decision tree, randomly select m attributes. From these m attributes, choose one based on the information gain method as the splitting attribute.
- Iterate through step 2 until further splitting is not feasible.
- Repeat steps 1–3 to generate a substantial number of decision trees, constituting the random forest.

9 Multi-support vector machine (MSVM)

The support vector machine (SVM) is a machine learning technique adept at handling classification tasks efficiently, even with limited available data [42, 43]. SVMs achieve this by identifying a hyperplane in an n -dimensional space that can effectively separate data points belonging to different classes [44]. The hyperplane is positioned to maximize the distance from the data points [44]. Support vectors, which are data points closest to the hyperplane, exert the most influence on determining the hyperplane's position [45]. If these support vectors are removed, the hyperplane shifts to a new location [45]. Figure 7 provides a graphical representation of SVMs, showcasing support vectors in tree form [45].

To facilitate multi-class classification, SVM is extended into multi-class SVM (MSVM), which decomposes the

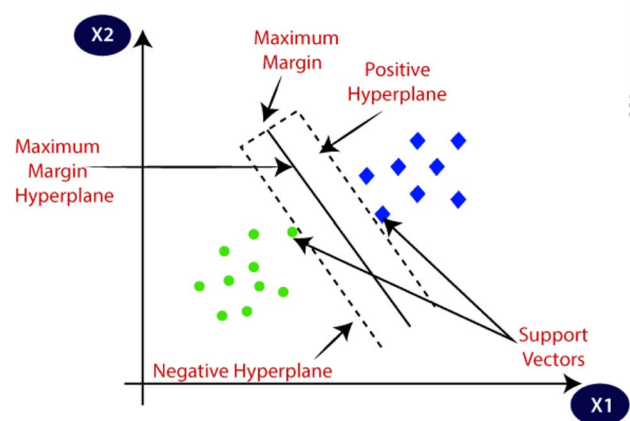


Fig. 7 Representation of binary classification using support vectors

Table 2 Characteristics of the LASPI test bench motor

Characteristic parameters	Value
Protection class (IP)	43
Normalized operating temperature	40 °C
Nominal voltage between phases	380 V
Power supply frequency	50 Hz
Nominal speed	2850 rpm
Nominal power output	1.5 kW
Power factor	0.80
Nominal current	11.4 A

problem into smaller sub-problems, each treated as a binary classification task [46]. Various approaches can be employed to address multi-class classification tasks, such as one versus one, one versus all, and directed acyclic graph. [42].

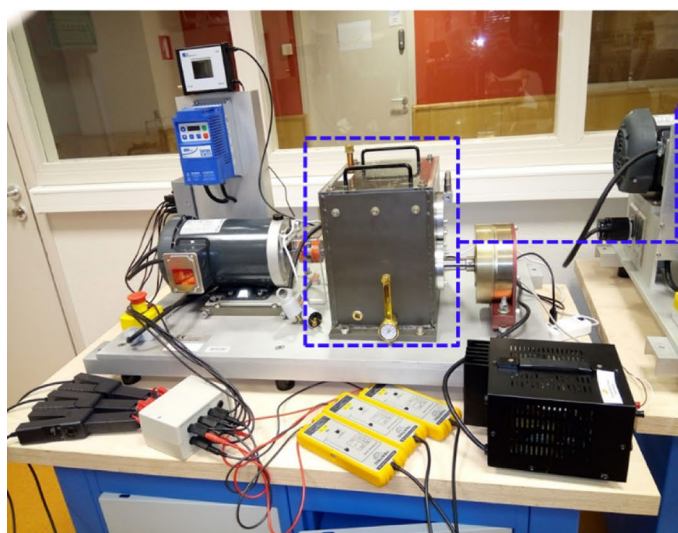
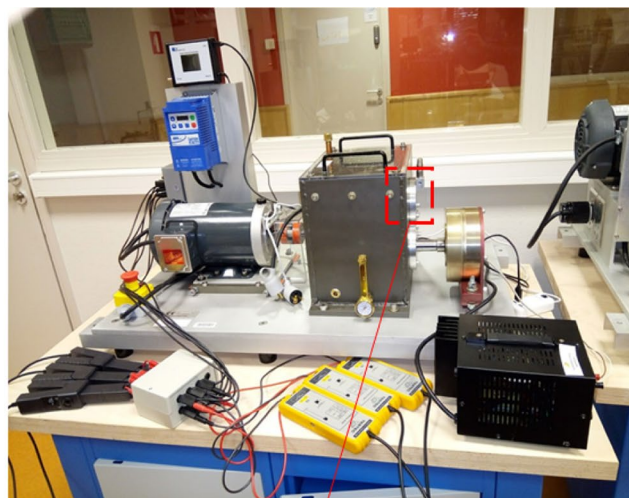
10 Experimental study

The authors have explored a test rig from Industrial Signals and Processes Analysis Laboratory (*Laboratoire d'Analyse des Signaux et Processus Industriels-LASPI*-) for bearing fault monitoring in a gearbox system [47]. It consists of a three-phase inverter controlling a 1.5 kW induction motor driving the gearbox. An electromagnetic brake connected to the gearbox to simulate the motor load. The motor characteristics are given in Table 2, and an overview of the test bench is shown in Fig. 8. The gearbox consists of three shafts: the input and intermediate and output shafts, where the bearings studied are located on the intermediate shaft. The input shaft, connected directly to the motor, has a gear with 29 teeth and two bearings with nine balls each. The bearings have a diameter of 0.3125 inch, a pitch diameter of 1.5157 inch, and a 0-contact angle.

Measurements are made in continuous mode for 10 s, with a sampling frequency of 51.6 kHz. Vibration signals are acquired with an accelerometer sensor sensitivity of 100 mV/g. Using the specified test rig instrumentation, a total of five distinct health states were examined, such as encompassing healthy bearing states, inner bearing ring defects, outer bearing ring defects, and combined bearing defects. In addition, each of these conditions was tested at two different speeds: 35 Hz and 45 Hz. In addition, each speed condition was tested at two load levels: 35% and 45%.

The gearbox contains three stages where the tested bearing is located on the intermediate shaft. Figure 9 shows the details of the experimental bench [37].

Fig. 8 Experimental setup



Three-phase asynchronous motor

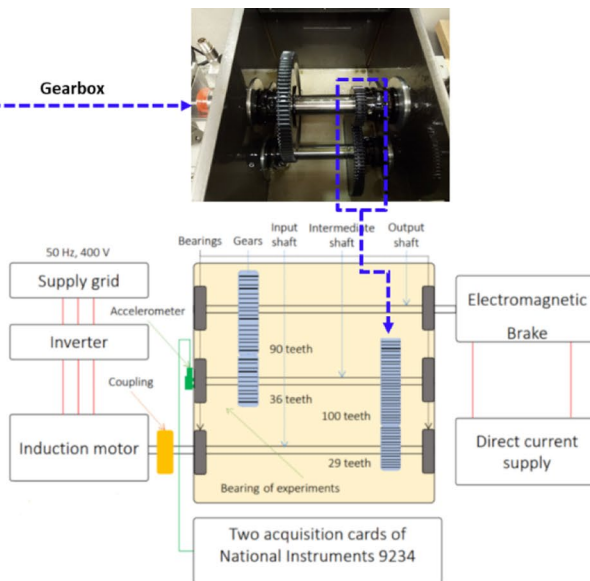


Fig. 9 Experimental schematic detail of setup

11 Results and discussion

Vibration signal collected in this study consists of data obtained from a benchmark scenario, encompassing both the healthy and four different bearing defects: inner race fault (IRF), outer race fault (ORF), ball fault (BF), and cage

fault (CF). Figure 6 shows the vibration data associated with different fault types plus the healthy state (Fig. 10).

In the proposed approach, raw signals are processed with a signal decomposition technique named MODWPT. MODWPT decomposes the signal by five levels producing a total of 32 coefficients. These coefficients are used for post-process extraction. Figure 11 shows a sample of the coefficients produced by MODWPT.

Fig. 10 Vibration signals of different bearing health state

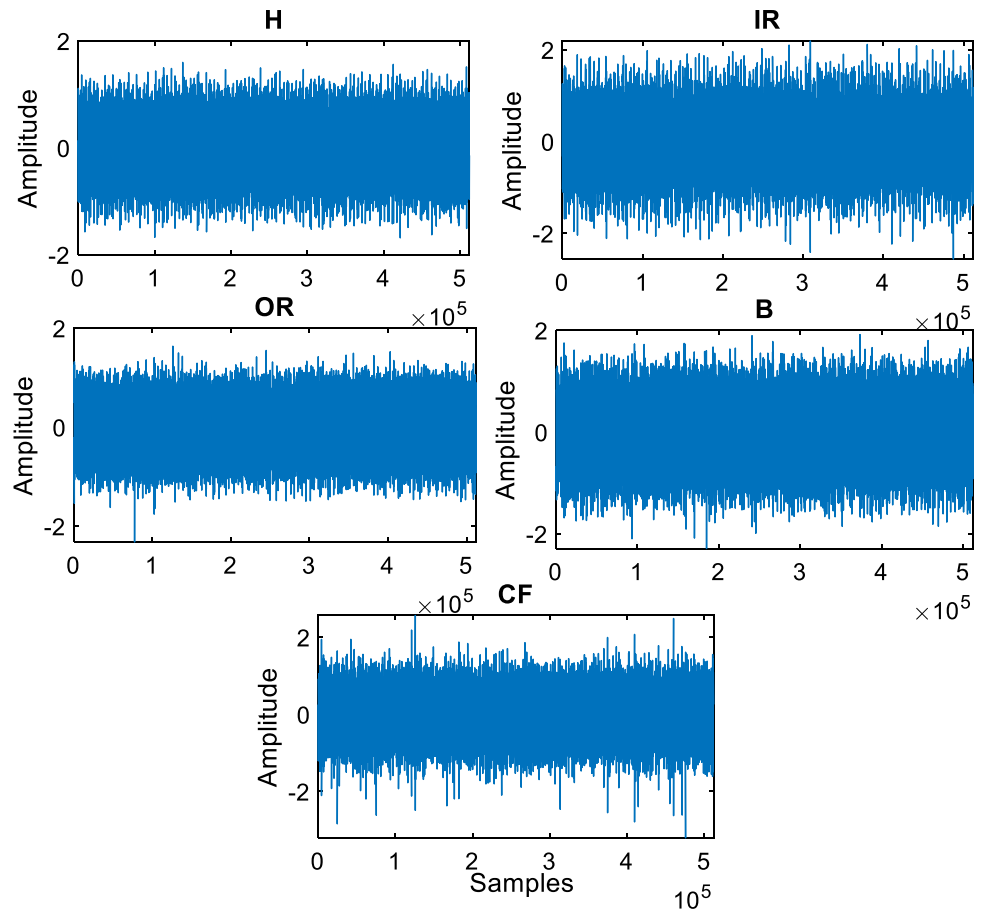


Fig. 11 MODWPT wavelet coefficients

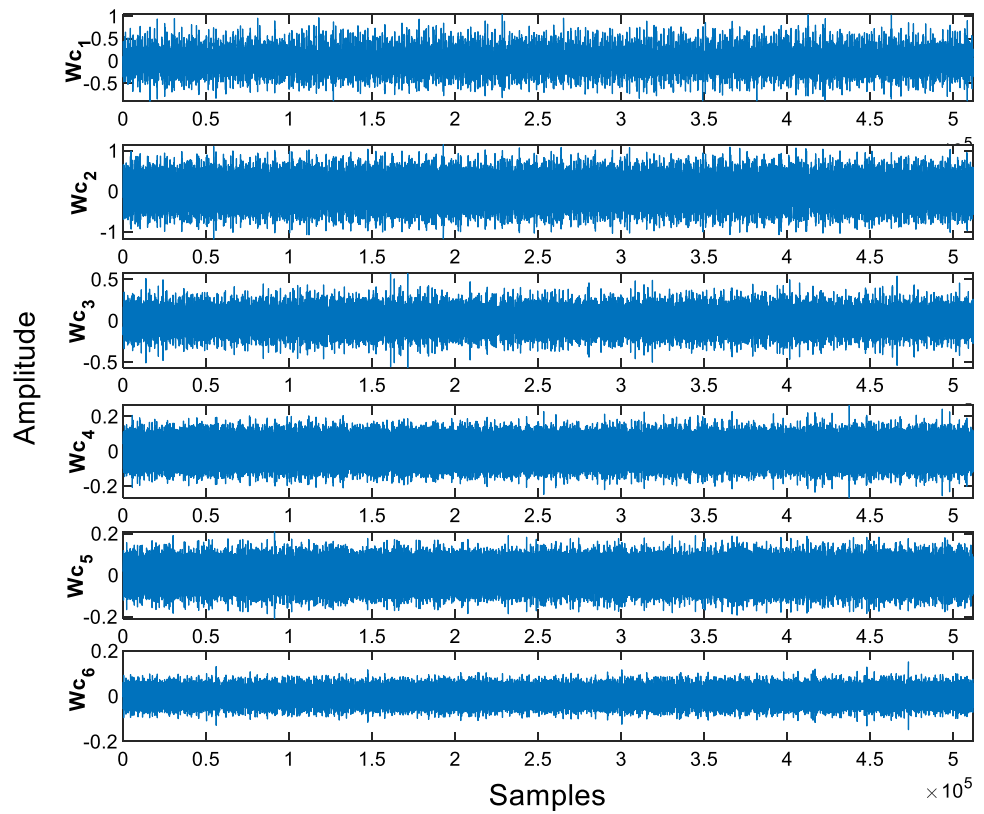


Table 3 Covariance H 35 Hz 35prc

–	Wc_1	Wc_2	Wc_{32}
Wc_1	0.0152	0.0061	$4.7494e^{-04}$	$7.0328e^{-05}$	$-1.1358e^{-04}$.	.	$-4.4964e^{-07}$
Wc_2	0.0061	0.0351	0.0181	0.0017	$7.5836e^{-04}$.	.	$9.1122e^{-07}$
.
.
Wc_{32}	$-4.4964e^{-07}$	$9.1122e^{-07}$	$6.6957e^{-07}$	$1.8427e^{-07}$	$1.4713e^{-06}$.	.	$7.40321e^{-07}$

In fault diagnosis, subtraction of the MODWPT coefficients from their mean signal allows analysis and identification of deviations or anomalies in vibration data. This process isolates variations from the normal behavior, emphasizing unique fault-related patterns. By enhancing fault-related information, efficient fault diagnosis and condition monitoring of various systems and machines are achieved. Covariance plays a very important part in statistical analyses and models, serving as the fundamental component for correlation coefficient calculations. These coefficients allow us to estimate the magnitude and direction of vectors. More precisely, in our study, covariance quantifies the relationship between the rows of obtained coefficients or signals from the subtraction of MODWPT coefficients and their mean.

In statistics, the covariance matrix is a square matrix that summarizes the variances and covariances between variables in a dataset. The diagonal elements of the covariance matrix represent the variances of individual variables. Table 3 gives covariance matrix of healthy bearing and IRF.

Extracting singular values and sorting eigenvectors in descending order of eigenvalues allows to identify and prioritize the most significant patterns of variation in a dataset. By analyzing the singular values, which represent the extent of variation, and ranking the corresponding eigenvectors according to their eigenvalues, dominant patterns or structures in the data can be determined. This process reduces dimensionality and allows the focus to be placed on the key factors contributing to the observed variations, thus providing a better insight into the signal's behavior (Table 4).

The projection of MODWPT coefficients onto the new principal axes expresses the data in terms of the most significant modes of variation. By transforming the coefficients into a reduced dimensional space defined by the eigenvectors, this projection simplifies data analysis and highlights

essential features. Defect detection and classification are facilitated by capturing important information while minimizing less significant variations.

To assess the disparity between the signal obtained by the proposed approach and the original vibration data, a comparative graph is generated to presents the dissimilarities and variances between the two signals in Fig. 12.

Statistical features are then extracted from the reconstructed signal. These features encompass a series of parameters that capture various aspects of the signal's statistical properties, such as its mean, standard deviation, skewness, and kurtosis. These statistical features are valuable indicators for assessing the presence of different fault types. By analyzing the extracted features from Fig. 13, patterns and anomalies associated with specific defect types can be identified by reducing confusion sample states. This information helps in developing effective diagnostic techniques and machine learning algorithms for fault detection and classification.

Health indicators (HI) extracted from statistical features are used as inputs for the classification of a prediction model. Different machine learning models are used to evaluate the proposed approach effectiveness. By training these models on labeled data and using HIs as features, the models learn to classify vibration signals into specific types of defects. Model performance is assessed using the confusion matrix and by calculating accuracy, precision, and recall. This evaluation validates the proposed approach and helps to determine the most appropriate model or combination of models for accurate bearing fault detection and classification.

The occurrence of a fault in a bearing has discernible effects on machinery performance. This is evident through distinct fault characteristics that manifest in the spectral domain, where signals from sensors are analyzed in terms of frequency content. Bearings, crucial for machinery

Table 4 covariance IR 35 Hz 35prc

–	Wc_1	Wc_2	Wc_{32}
Wc_1	0.0669	-0.0050	$4.5010e^{-04}$	$2.1891e^{-04}$	$-1.3811e^{-04}$.	.	$-6.9285e^{-07}$
Wc_2	-0.0050	0.0454	-0.0201	0.0017	$8.7386e^{-04}$.	.	$1.0640e^{-07}$
.
.
Wc_{32}	$-6.9285e^{-07}$	$1.0640e^{-06}$	$7.4863e^{-07}$	$1.8427e^{-07}$	$2.0476e^{-06}$.	.	$1.4109e^{-06}$

Fig. 12 Comparison of vibration signals

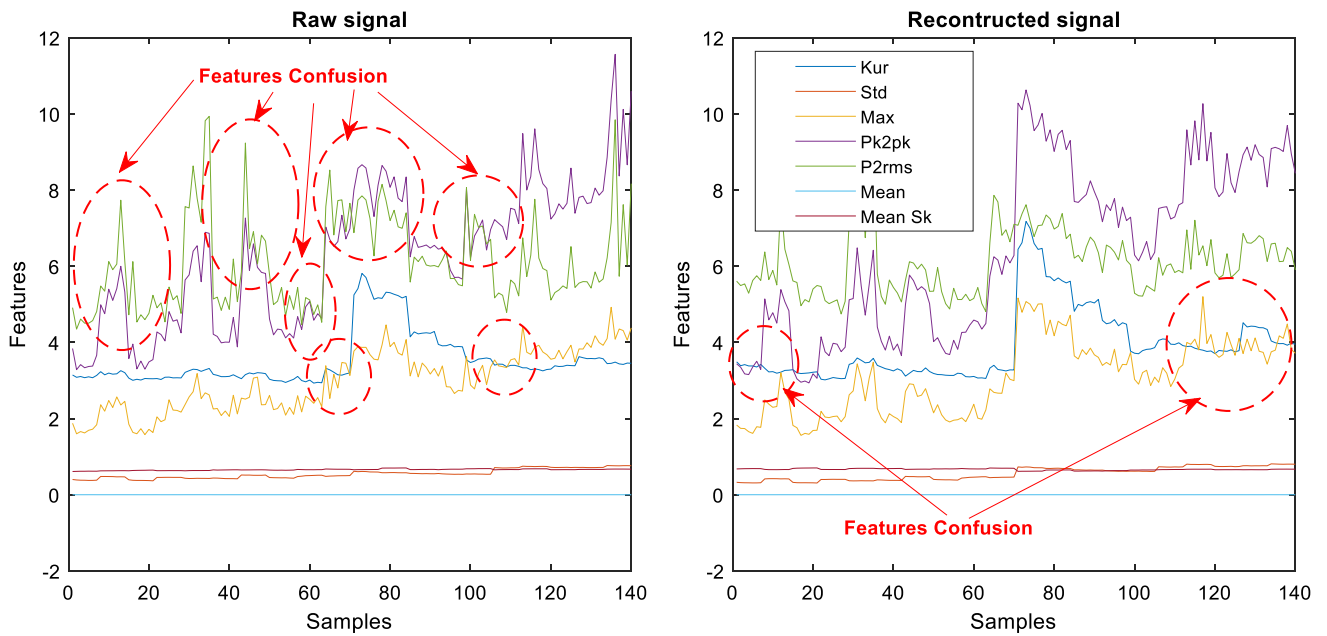
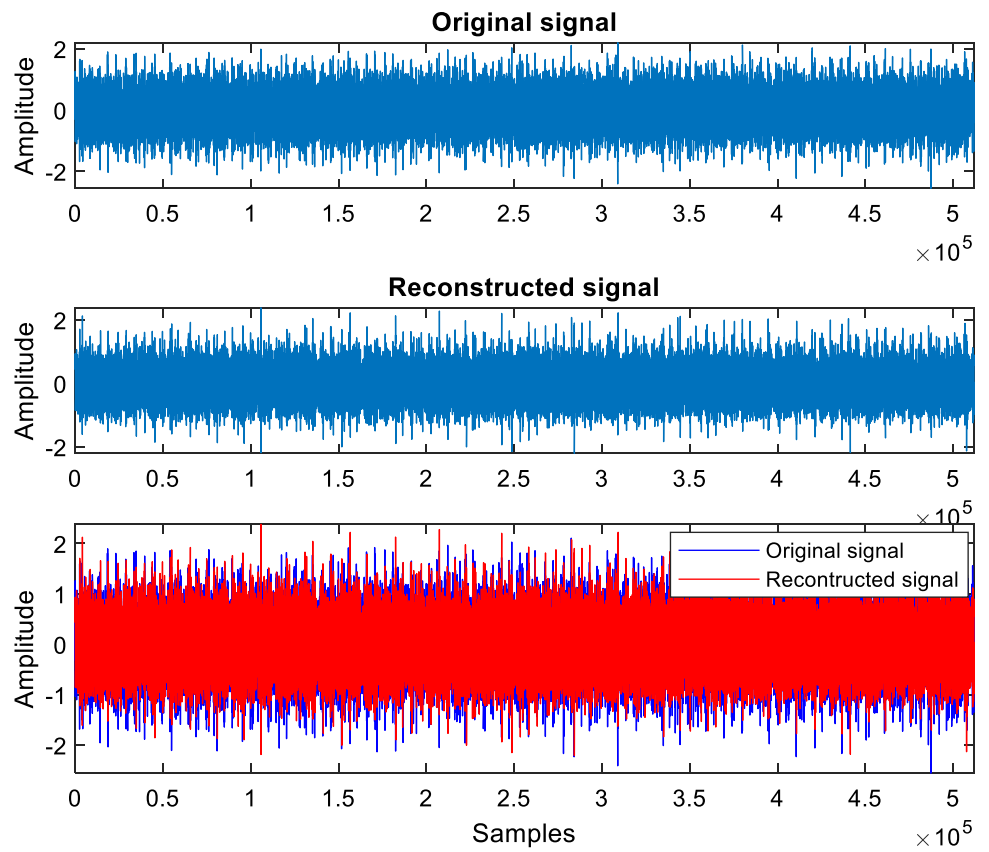


Fig. 13 Feature signatures of different states

Table 5 Bearing fault frequencies

Types fault frequencies	Formula	Simplified formula	Reference frequencies (from testbed)
Inner race	$f_{IR} = \frac{N_b}{2} f_r \left(1 + \frac{D_b}{D_c} \cos \alpha \right)$	$f_{IR} = 5.430 \times f_r$	$f_{IR}(13.05 \text{ Hz}) = 70.86 \text{ Hz}$ $f_{IR}(10.15 \text{ Hz}) = 55.11 \text{ Hz}$
Outer race	$f_{OR} = \frac{N_b}{2} f_r \left(1 - \frac{D_b}{D_c} \cos \alpha \right)$	$f_{OR} = 3.572 \times f_r$	$f_{OR}(13.05 \text{ Hz}) = 46.61 \text{ Hz}$ $f_{OR}(10.15 \text{ Hz}) = 36.25 \text{ Hz}$
Ball	$f_B = \frac{D_c}{D_b} f_r \left(1 - \frac{D_b^2}{D_c^2} \cos^2 \alpha \right)$	$f_B = 4.644 \times f_r$	$f_B(13.05 \text{ Hz}) = 60.60 \text{ Hz}$ $f_B(10.15 \text{ Hz}) = 47.13 \text{ Hz}$

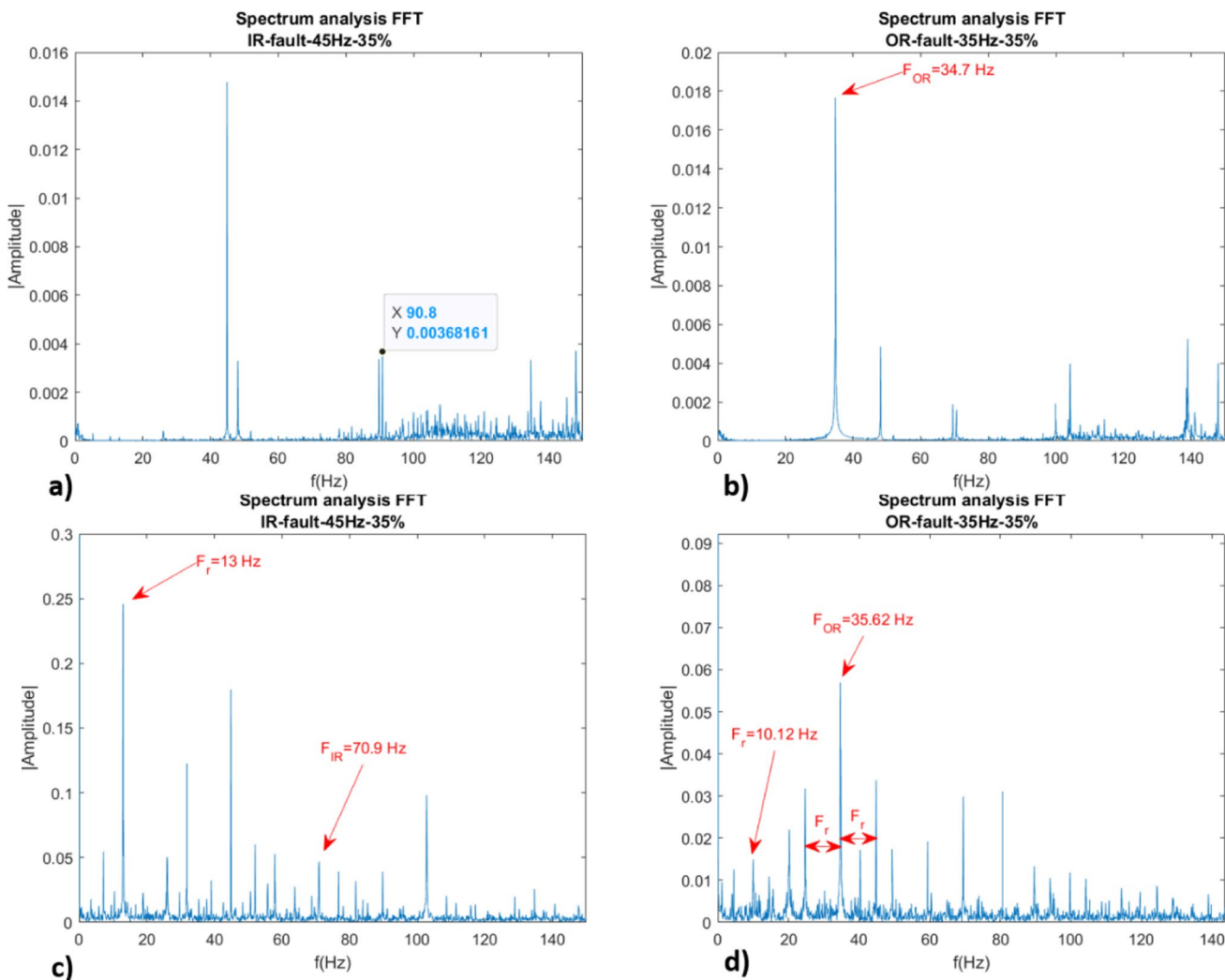


Fig.14 a FFT of raw signal IRF-45 Hz-35%, b FFT of raw signal ORF-35 Hz-35%, c FFT of constructed signal IRF-45 Hz-35%, d FFT of constructed signal ORF-35 Hz-35%

operation, can exhibit specific patterns in these signals when a fault arises. By scrutinizing these patterns in the frequency domain, experts can identify the nature and severity of the fault.

Table 5 shows the characteristic frequencies associated with various bearing faults types. The gear ratio of intermediate shaft on the experimental benchmark is equal to 0.29 (Fig. 14), where the speed frequency (f_r) can be obtained using Eq. 13.

$$f_r = \frac{N_{teeth-G1}}{N_{teeth-G2}} \times f_{r-motor} = 0.29 \times f_{r-motor} \tag{13}$$

To evaluate the efficacy of the proposed reconstruction approach, a thorough comparison was conducted in the frequency domain. This analysis involves applying the fast Fourier transform (FFT) to both reconstructed signal, generated through the proposed method, and the original raw signal. The FFT provides a detailed representation of the signal's frequency components, allowing for a comprehensive assessment of any discrepancies or improvements introduced by the reconstruction process. This comparative study serves to quantify and qualify the performance of the proposed approach in reproducing the original signal within the frequency domain.

In Fig. 14, the spectral analysis extracted using the proposed approach reveals more characteristic frequencies related to the presence of faults and the shaft rotation speed. By comparing the frequencies obtained from the proposed approach with the reference frequencies presented in Table 5, we observe a very close values of both frequencies. This leads to confirm the effectiveness of the proposed approach in extracting fault features from the raw signal.

To avoid overfitting and ensure reliable model training on a large dataset, *K*-fold cross-validation (KFCV) has been used. This technique involves dividing the data into *K* subsets or "folds" and training the model *K* times, each time using a different fold as the test set and the remaining folds as the training set. By averaging the results of the *K* iterations, a more accurate model performance and generalization ability is provided. KFCV reduces the impact of dataset variability and the reliance on a single training–testing split,

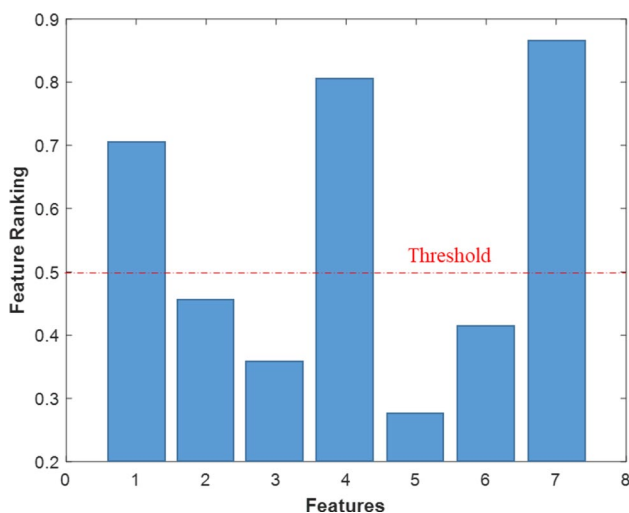


Fig. 15 Feature ranking for selection

Table 6 Parameters tuning of ML

Model	Hyperparameters	
	Parameter	Identification
Ensemble tree (ET)	Min-parent	2
	Min-leaf	1
	Nvar-to-sample	3
	Merge-leaves	‘Off’
	Method	Bag
Random forest(RF)	Num-tree	20
	Min-samples-split	2
	Min-samples-leaf	1
	Bootstrap	‘on’
	Verbose	0
Decision tree(DT)	Min-parent	10
	Min-leaf	1
	Max-cat	10
	Max-split	50
	Nvar-to-sample	‘All’
Multi-support vector machine (MSVM)	Merge-leaves	‘On’
	Kernel-scale	0.63
	Kernal-function	‘Gaussian’
	Method	ErrorCorrecting OutputCodes’E COC’
	Coding-type	‘OneVsOne’

providing an effective approach for ML model training on large datasets, while overcoming the overfitting issue.

The Relief algorithm is a feature selection method commonly used in machine learning to select robust features for classification tasks. Particularly useful for high-dimensional data, it aims to identify the most relevant features that contribute significantly to the classification process. The Relief algorithm operates by assigning a weight or score to each feature based on its ability to distinguish between instances of different classes.

From Fig. 15, the three features generally exhibit a higher level of significance for improving the machine learning model accuracy for classification. The selection of these three features has the potential to improve classification accuracy. The selected features, such as X_{kur} , X_{p2p} , and X_{meanSK} , have distinct characteristics that make them particularly instructive for defect classification.

Confusion matrices are generated for different ML models trained on labeled samples. The models used in this study are multi-support vector machine (MSVM), decision tree (DT), ensemble tree (ET), random forest (RF), etc.

In the training process, we used an automated learning function of Matlab (*fitcensmbe*, *fitctree*, *fitcecoc*) [48–50]. The training process involves the utilization of automatic parameter adjustment to achieve optimal validation results.

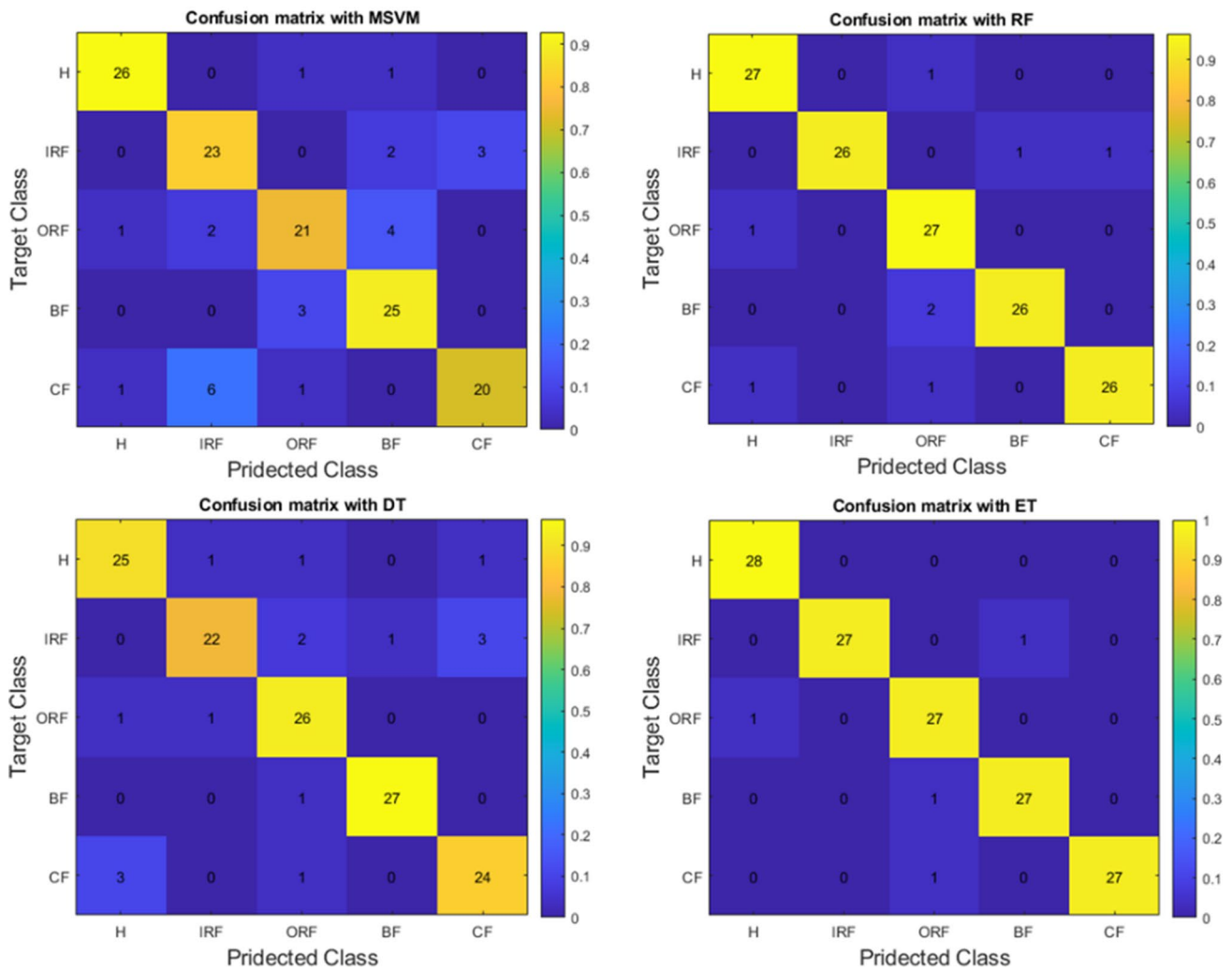


Fig. 16 Confusion matrices of different ML models

Table 7 Metrics evaluation of confusion matrices

	DA	KNN	MSVM	RF	DT	ET
Accuracy	46.4285	71.4285	82.1428	94.2857	88.5714	97.1428
Precision	0.50	64.5161	92.8571	93.1034	89.2857	96.5517
Recall	60.7142	71.4285	92.8571	96.4285	89.2857	100
Specificity	86.6071	97.3214	98.2142	98.2142	96.3963	99.0990

The hyperparameters are well tuned to enhance the optimization of the model's performance [51]. We set three parameters: the learning rate to 0.01, the maximum split to 50, and the number of folds in *K*-fold cross-validation to 10. In Table 6, we present some hyperparameters of used ML tools for classification.

From Fig. 16, confusion matrices allow evaluation of model performance in terms of correctly or incorrectly classified instances. ET, which combines several decision trees, was found to provide more accurate results than the other models. The model's accuracy is determined by evaluating its performance parameters, such as precision, recall, and F1 score, on labeled samples (Table 7).

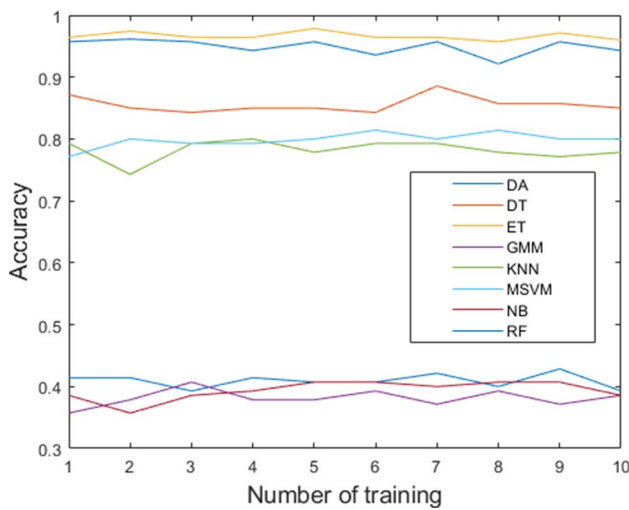


Fig. 17 Accuracy of ML models

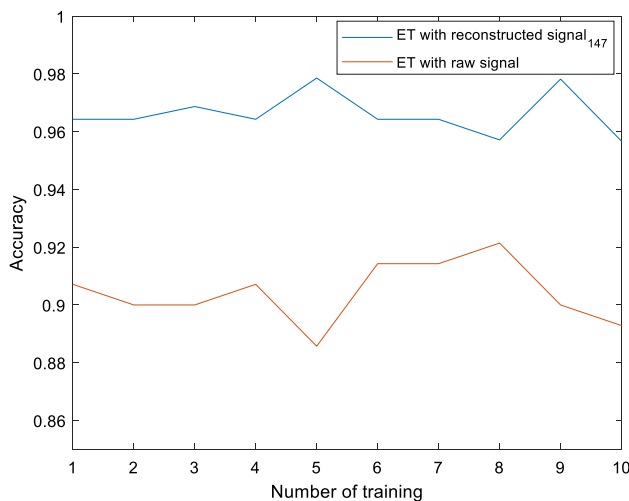


Fig. 18 Classification accuracy of proposed approach and the raw signal

$$\begin{aligned}
 \text{Accuracy} &= \frac{\sum (TP + TN)}{N_{\text{total}}} \\
 \text{Precision} &= \frac{\sum TP}{\sum (TP + FP)} \\
 \text{Recall} &= \frac{\sum TP}{\sum (TP + FN)} \\
 \text{Specificity} &= \frac{\sum TN}{\sum (TN + FP)}
 \end{aligned}
 \tag{14}$$

By focusing on the obtained results, the model can effectively capture the relevant aspects of the data that contribute significantly to the classification process. These features,

selected by R-relief, reduce noise and eliminate redundant or irrelevant information; moreover, they improve the model's ability to be generalized and to make accurate predictions. It's important to note that the specific nature of these features and their impact on classification accuracy may vary depending on the dataset and the specific employed machine learning algorithm. Therefore, it is advisable to conduct further analysis and experimentation to validate the consistent performance of these selected features across different evaluation metrics and models.

Consequently, to assess the ML models' accuracy, the models' stability is also investigated. Figure 12 is used to compare the models' accuracy across ten training iterations. This comparison provided insights into the models' consistency and reliability in producing accurate results. Studying model stability is important for identifying potential issues related to overfitting, data variability, or model sensitivity. By considering the accuracy across multiple training iterations, a more comprehensive evaluation of the models' reliability was achieved, confirms selection of ET model according to the stability and accuracy of bearing faults classification.

From Fig. 17, a remarkably accurate and stable of ET model is observed, leading to very satisfactory fault classification results. In addition, a comparison is made between the proposed approach and the features extracted from the raw signal (Fig. 18), which clearly demonstrates the advantages of our classification method.

12 Conclusion

The presented paper provides a new vibration signal processing-based fault diagnosis method for rotating machines which combines the advantages of MODWPT and covariance techniques. The presented approach is implemented and tested on a real test rig demonstrating and proving its effectiveness in extracting fault characteristics. By leveraging the robust time–frequency representation provided by MODWPT and exploiting the extracted features, an improvement in classification accuracy using confusion matrix is observed. Consequently, the integration of these methods has the potential to enhance the overall performance of bearing fault classification systems, enabling more reliable and accurate diagnosis in various industrial applications.

References

1. LaCava W, Xing Y, Marks C et al (2013) Three-dimensional bearing load share behavior in the planetary stage of a wind turbine gearbox. *IET Renew Power Gen* 7:359–369

2. Wang J, Peng Y, Qiao W (2016) Current-aided order tracking of vibration signals for bearing fault diagnosis of direct-drive wind turbines. *IEEE Trans Ind Electron* 63:6336–6346
3. Li X, Elasha F, Shanbr S et al (2019) Remaining useful life prediction of rolling element bearings using supervised machine learning. *Energies* 12:2705
4. Tama BA, Vania M, Lee S, Lim S (2023) Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. *Artif Intell Rev* 56(5):4667–4709
5. Touzout W, Benazzouz D, Gougam F, Afia A, Rahmoune C (2020) Hybridization of time synchronous averaging, singular value decomposition, and adaptive neuro fuzzy inference system for multi-fault bearing diagnosis. *Adv Mech Eng* 12(12):1687814020980569
6. Afia A, Gougam F, Rahmoune C, Touzout W, Ouelmokhtar H, Benazzouz D (2023) Intelligent fault classification of air compressors using Harris Hawks optimization and machine learning algorithms. *Trans Inst Meas Control* 46:01423312231174939
7. Gougam F, Rahmoune C, Benazzouz D, Varnier C, Nicod JM (2020) Health monitoring approach of bearing: application of adaptive neuro fuzzy inference system (ANFIS) for RUL-estimation and Autogram analysis for fault-localization. In: 2020 prognostics and health management conference (PHM-Besaçon). IEEE. p 200–206
8. Li L, Cai H, Han H, Jiang Q, Ji H (2020) Adaptive short-time fourier transform and synchrosqueezing transform for non-stationary signal separation. *Signal Process* 166:107231
9. Akujuobi CM (2022) Wavelets and wavelet transform systems and their applications. Springer International Publishing, Berlin/Heidelberg, Germany
10. Yu X, Liang Z, Wang Y, Yin H, Liu X, Yu W, Huang Y (2022) A wavelet packet transform-based deep feature transfer learning method for bearing fault diagnosis under different working conditions. *Measurement* 201:111597
11. Afia A, Gougam F, Rahmoune C, Touzout W, Ouelmokhtar H, Benazzouz D (2023) Gearbox fault diagnosis using REMD EO and machine learning classifiers. *J Vib Eng Technol.* <https://doi.org/10.1007/s42417-023-01144-8>
12. Afia A, Rahmoune C, Benazzouz D (2020) An early gear fault diagnosis method based on RLMD, Hilbert transform and cepstrum analysis. *Mechatron Syst Control* 49:115–123
13. Gougam F, Rahmoune C, Benazzouz D, Merainani B (2019) Bearing fault diagnosis based on feature extraction of empirical wavelet transform (EWT) and fuzzy logic system (FLS) under variable operating conditions. *J Vibroeng* 21(6):1636–1650
14. Shanbr S, Elasha F, Elforjani M et al (2018) Detection of natural crack in wind turbine gearbox. *Renew Energy* 118:172–179
15. Rahmoune C, Benazzouz D (2013) Monitoring gear fault by using motor current signature analysis and fast Kurtogram method. *Int Rev Electr Eng* 8:616–625
16. Afia A, Rahmoune C, Benazzouz D, Merainani B, Fedala S (2021) New gear fault diagnosis method based on MODWPT and neural network for feature extraction and classification. *J Test Eval* 49(2):1064–1085
17. Soualhi M, Nguyen KT, Soualhi A, Medjaher K, Hemsas KE (2019) Health monitoring of bearing and gear faults by using a new health indicator extracted from current signals. *Measurement* 141:37–51
18. Soualhi M, Nguyen KT, Medjaher K (2020) Pattern recognition method of fault diagnostics based on a new health indicator for smart manufacturing. *Mech Syst Signal Process* 142:106680
19. Soualhi A, Medjaher K, Zerhouni N (2014) Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE Trans Instrum Meas* 64(1):52–62
20. Gougam F, Rahmoune C, Benazzouz D, Afia A, Zair M (2020) Bearing faults classification under various operation modes using time domain features, singular value decomposition, and fuzzy logic system. *Adv Mech Eng* 12(10):1687814020967874
21. Moshrefzadeh A, Fasana A (2018) The autogram: an effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis. *Mech Syst Signal Process* 105:294–318
22. Afia A, Rahmoune C, Benazzouz D (2018) Gear fault diagnosis using autogram analysis. *Adv Mech Eng* 10(12):1687814018812534
23. Soualhi A, Clerc G, Razik H (2012) Detection and diagnosis of faults in induction motor using an improved artificial ant clustering technique. *IEEE Trans Industr Electron* 60(9):4053–4062
24. Benaggoune K, Yue M, Jemei S, Zerhouni N (2022) A data-driven method for multi-step-ahead prediction and long-term prognostics of proton exchange membrane fuel cell. *Appl Energy* 313:118835
25. Benaggoune K, Meraghni S, Ma J, Mouss LH, Zerhouni N (2020) Post prognostic decision for predictive maintenance planning with remaining useful life uncertainty. In: 2020 prognostics and health management conference (phm-besaçon). IEEE. p 194–199
26. Gougam F, Chemseddine R, Benazzouz D, Benaggoune K, Zerhouni N (2021) Fault prognostics of rolling element bearing based on feature extraction and supervised machine learning: application to shaft wind turbine gearbox using vibration signal. *Proc Inst Mech Eng C J Mech Eng Sci* 235(20):5186–5197
27. Afia A, Rahmoune C, Benazzouz D, Merainani B, Fedala S (2020) New intelligent gear fault diagnosis method based on AUTOGRAM and radial basis function neural network. *Adv Mech Eng* 12(5):1687814020916593
28. Afia A, Hand O, Fawzi G et al. (2022) Gear fault detection, identification and classification using MLP neural network. In: Recent advances in structural health monitoring and engineering structures: Select proceedings of SHM and ES 2022. Springer, Singapore. p 221–234
29. Afia A, Gougam F, Rahmoune C, Touzout W, Ouelmokhtar H, Benazzouz D (2023) Spectral proper orthogonal decomposition and machine learning algorithms for bearing fault diagnosis. *J Braz Soc Mech Sci En.* <https://doi.org/10.1007/s40430-023-04451-z>
30. Li X, Yang Y, Hu N, Cheng Z, Shao H, Cheng J (2022) Maximum margin Riemannian manifold-based hyperdisk for fault diagnosis of roller bearing with multi-channel fusion covariance matrix. *Adv Eng Inform* 51:101513
31. Lai Y, Li R, Zhang Y, Meng L, Chen R (2023) Fault detection of reciprocating plunger pump with fault-free data based on unsupervised feature encoder and minimum covariance determinant. *Meas Sci Technol.* <https://doi.org/10.1088/1361-6501/acde97>
32. Wang N, Jia L, Qin Y, Li Z, Miao B, Geng J, Wang Z (2023) Scale-independent shrinkage broad learning system for wheelset bearing anomaly detection under variable conditions. *Mech Syst Signal Process* 200:110653
33. You K, Qiu G, Gu Y (2023) An efficient lightweight neural network using BiLSTM-SCN-CBAM with PCA-ICEEMDAN for diagnosing rolling bearing faults. *Meas Sci Technol* 34(9):094001
34. Jin Z, He D, Ma R, Zou X, Chen Y, Shan S (2022) Fault diagnosis of train rotating parts based on multi-objective VMD optimization and ensemble learning. *Digit Signal Process* 121:103312
35. Zhong X, Ban H (2022) Crack fault diagnosis of rotating machine in nuclear power plant based on ensemble learning. *Ann Nucl Energy* 168:108909
36. Cheng J, Sun J, Yao K, Xu M, Wang S, Fu L (2022) Development of multi-disturbance bagging Extreme Learning Machine method for cadmium content prediction of rape leaf using hyperspectral imaging technology. *Spectrochim Acta Part A Mol Biomol Spectrosc* 279:121479
37. Charbuty B, Abdulazeez A (2021) Classification based on decision tree algorithm for machine learning. *J Appl Sci Technol Trends* 2(01):20–28

38. Aravinth S, Sugumaran V (2018) Air compressor fault diagnosis through statistical feature extraction and random forest classifier. *Prog Ind Ecol, Int J* 12(1–2):192–205
39. Benkercha R, Moulahoum S (2018) Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system. *Sol Energy* 173:610–634
40. Breiman L (2001) Random forests. *Mach Learn* 45:5–32
41. Pal M (2005) Random forest classifier for remote sensing classification. *Int J Remote Sens* 26(1):217–222
42. Zhou L, Wang Q, Fujita H (2017) One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies. *Inf Fusion* 36:80–89
43. Gougam F, Rahmoune C, Benazzouz D, Zair MI, Afia A (2018) Early bearing fault detection under different working conditions using singular value decomposition (SVD) and adaptatif neuro fuzzy inference system (ANFIS). In: *International conference on advanced mechanics and renewable energy (ICAMRE)*. p 28–29
44. Abellán J (2013) Ensembles of decision trees based on imprecise probabilities and uncertainty measures. *Inf Fus* 14(4):423–430
45. Achirul Nanda M, Boro Seminar K, Nandika D, Maddu A (2018) A comparison study of kernel functions in the support vector machine and its application for termite detection. *Information* 9(1):5
46. Guo Y, Zhang Z, Tang F (2021) Feature selection with kernelized multi-class support vector machine. *Pattern Recogn* 117:107988
47. Soualhi M, Soualhi A, Nguyen T-P, Medjaher K, Clerc G, Razik H (2023) LASPI: Détection et diagnostic des défauts de boîte de vitesses. *LASPI*. <https://doi.org/10.25666/DATAUBFC-2023-03-06>
48. “fitctree” [online] Available at: <https://www.mathworks.com/help/stats/fitctree.html> (Accessed 5 Oct 2019)
49. “fitcecoc” [online] Available at: <https://www.mathworks.com/help/stats/fitcecoc.html> [Accessed 5 Oct 2019]
50. “fitcensemble” [online] Available at: <https://www.mathworks.com/help/stats/fitcensemble.html> (Accessed 5 Oct 2019)
51. Dirvanauskas D, Maskeliunas R, Raudonis V, Damasevicius R (2019) Embryo development stage prediction algorithm for automated time lapse incubators. *Comput Methods Programs Biomed* 177:161–174

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.