Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 12, No. 1, March 2024, pp. 1~14
ISSN: 2089-3272, DOI: 10.52549/ijeei.v12i1.4612

Diagnosis and Monitoring Method for Detecting and Localizing Bearing Faults

Saida Dahmane[1], Fouad Berrabah[2], Mabrouk Defdaf[3], Saad Salah[4]

[1,2,3]*Department of Electrical Engineering, Faculty of Technology, University of M'sila, 28000, Algeria*
[1]*Electrical Engineering laboratory (LGE), M'sila University, Algeria.*
[4]*Department of Electromecanical, Badji Mokhtar Annaba university, Annaba, 23000, Algeria*
[2,3,4]*LSELM laboratory, Badji Mokhtar Annaba university, Annaba 23000 Algeria*
* Corresponding author's Email: saida.dahmene@univ-msila.dz

Article Info

Article history:

Received Mar 19, 2023
Revised Dec 16, 2023
Accepted Dec 29, 2023

Keyword:

Induction motor
Bearing fault
DWT
Spectral ENV
Random Forest

ABSTRACT

Induction motors in modern industry are becoming more and more functional and complex. Unfortunately, these machines are not free from damages what make their fault diagnosis the most critical aspect of system monitoring and maintenance. Vibrational signal data yields relevant information about the state of the entire system, as well as specifically about one of its components that makes its analysis quite interesting. For this effect, the current paper aims to propose an automatic diagnosis and monitoring method for detecting and locating bearing faults in an induction motor based on vibration signal processing. The suggested method combines the discrete wavelet transform (DWT) with the envelope spectrum (ENV) as advanced signal processing, incorporating a machine learning algorithm based on random forest classifier. The discrete wavelet transforms (DWT), using the Haar wavelet, decomposes the vibrational signal to provide both approximations and details. Each detail is then reconstructed to avoid any missing of information. To precisely select the reconstructed detail ($Recd_k$) that provides pertinent information about bearing faults, a statistical study is conducted. This study involves calculating four indicators (Root mean square (RMS), correlation coefficient (CC), energy coefficient (EC) and peak to peak (P2P) factor) is performed for each ($Recd_k$). These indicators are compared with threshold indicators, and this criterion is met by the reconstructed details 1 and 3. The obtained reconstructed details are then subjected to the spectral envelope analysis to detect the fault frequencies, which are considered as new features entering the random forest classifier model. This combination of approaches allows better feature extraction and structuring of the dataset, leading to improved accuracy of the random forest classifier, achieving a higher classification rate of more than 99,53 %. The proposed DWT-ENV-RF method indicates well its efficiency when compared to other recent works, and the attained results are all confirmed by the experimental tests conducted in the CWRU laboratory.

Corresponding Author:

Saida Dahmane,
Department of Electrical Engineering, Faculty of Technology,
University of M'sila, 28000, Algeria,
Email: saida.dahmene@univ-msila.dz

1. INTRODUCTION

With the rapid development of technology and production fields, induction motors have played a crucial role in industrial sectors, due to their countless advantages. Notably, those that stem from their high reliability, low cost, reduced time of maintenance and dynamic performance. However, owing to an unfavorable operational environment, such as abrasion, unbalanced loads or overload, these motors are not impervious from damages that can affect the stator, the rotor and the bearings. Usually, bearings are unable to

sustain heavy loads due to which they represent, according to statistics, more than 41% of all other fault's occurrences [1]. This supports the significance of bearing faults diagnosis in the condition monitoring of rotating machines.

Eventually, for fault diagnosis three basic approaches are considered corrective, preventive and predictive maintenance. If a defect has already occurred, corrective procedures are implemented, resulting in high replacement costs of components, sudden downtime, and, obviously, a substantial cost associated with production shutdown. Preventive procedures aim to plan and manage periodic industrial maintenance interventions before any breakdown or dysfunction occurs. However, this approach involves various resources compared to low efficiency outcomes. In the other hand, the predictive procedure is developed based on accurate monitoring that can dramatically reduce maintenance cost and machines failure [2].

To monitor failure symptoms, various physical quantities can be measured, including stator current, lubrification analysis, noise, temperature and vibration analysis [3, 4, 5 and 6]. Among these, vibration analysis is the most often used as a diagnostic signal of bearing fault [7].

Furthermore, bearing faults prediction at an early stage can be succussed by identifying dynamic characteristics and performing appropriate signal processing. Thus, this process seems like difficult for the redundant failures and the big-collected data, with which automatic diagnosis is required.

Raw data alone cannot reveal any information about the state of the bearing element that makes its identification neither classification impossible. Among various signal processing techniques, DWT and envelope analysis (ENV) are well-known and utilized for identifying faults, particularly in non-stationary signals [8]. The DWT serves to decompose the original signal, providing approximations and details rich in fault information, while the ENV analysis permits to get only the necessary information corresponding to the fault existence in terms of frequency series.

Pertinent fault identification leads to better feature extraction, which leads eventually to improve system classification performance and reliability [9]. Several machine learning classifiers are used for fault diagnosis automation, among which naïve-bays, extreme gradient boosting (XGBoost), K- nearest neighbors (KNN), and Random forests (RF) are the most robust algorithms in terms of highest accuracy and performance. According to this, the efficiency and the accuracy of the random forest classifier have been proven by many researchers [10]. The built of the random forest has mainly based on an ensemble of decision trees that refers to the name “forest”. The term “random” belongs to the fact that these trees are then trained arbitrarily based on a training dataset.

In the realm of bearing element fault diagnosis, scientific literature has seen known several research endeavors that support many procedures and approaches, enriching the automatic diagnosis of bearing faults. In [11], authors have developed a novel feature extraction method for bearing fault detection, in which the inputs features are the vibration signals entering the convolutional neural network (CNN), getting a varying accuracy from 88% to 99%. Experimentally results in [8] show the effectiveness of the proposed method when using DWT for feature extraction and ensemble machine learning classifiers for bearing fault classification. A comparative study between the random forest algorithm and the XGBoost classifier is performed that give a highest accuracy of 99,14 % and 99,30 %, respectively, for the RF and the XGBoost classifiers. In more recent research [12], both of deep learning and machine learning classifiers are used to monitor bearing faults based on vibration signals provided by the CWRU. Among three ensemble classifiers, such as artificial neural network, random forests, and a deep learner autoencoder, the autoencoder excels the other classifiers with a high classification rate of 91%. However, using raw data without any previous signal analysis is impractical in term of big data and obvious loss of fault information, in how a dimensionality reduction is required without any losing of information.

In the current paper, two signal processing approaches are investigated, the DWT and the envelope analysis. The DWT is used to decompose a bearing vibrational signal into a set of sub-signals forming different approximations and details that they will be reconstructed again to keep the necessary fault information. Based of statistical study applied for every single reconstructed detail, a selection of reconstructed details is done. The envelope analysis is then performed to the selected reconstructed detail that provides frequency features set. The core idea here is to circle only the band zone of the fault components and reduce the dimensionality of the feature matrix. Finally, the random forest classifier is trained through the obtained frequency features to evaluate the performances and attain a robust automatic model for bearing fault diagnosis. It is to mention that this proposed process is validated by a hole experiment test of the CWRU laboratory that includes vibrational signals of four bearing conditions: healthy bearing, defective bearing with the outer race fault, defective bearing with the ball fault and defective bearing with the inner race fault.

The next parts of this paper are reserved for: a constitution of the bearing and specification of defect frequencies are presented in section 2. The third section presents the different approaches followed for the bearing fault detection and localization. The experimental test rig and the flowchart considered in this work

are described in section 4. Section 5 is about the application of the DWT and fault features extraction. The next section, is reserved for the results discussion and a brief evaluation of the effectiveness of our results work. The seventh last section ends the paper with a conclusion.

2. CONSTRUCTION OF THE BEARING AND SPECIFICATION OF DEFECT FREQUENCIES

In this section, a brief description of the bearing used in this research and the frequencies associated with their component faults are presented.

2.1. Rolling bearing construction

Rolling bearings consist of two main parts called the outer race and the inner race, both constructed essentially from steel. Between them there is a cage used to separate the rolling elements at specific distances as shown in Figure 1.

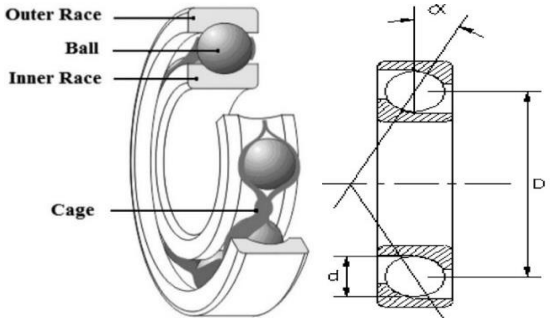

Figure 1. Construction and geometric characteristic of a rolling bearing.

2.2. Frequencies characterizing the bearing defects

When a localized defect appears at a point in one of the components of the bearing, this leads the assembly to vibrate which allows the appearance of vibratory signature of a fault, affecting each one of these components. To calculate the characteristic frequencies, it is necessary to find the rotational frequency f_r , of the motor shaft:

$$f_r = \frac{n}{60} \tag{1}$$

Where: *n* represents the rotation speed of the motor shaft.

Then, the characteristic harmonic frequencies for failures of individual bearing elements can be determined based on the Eq. (2)- (5) below [13]:

$$f_{or} = \frac{n_b}{2} f_r (1 - \frac{d}{D} cos\, \alpha) \tag{2}$$

$$f_c = \frac{1}{2} f_r (1 \pm \frac{d}{D} cos\, \alpha) \tag{3}$$

$$f_{re} = \frac{D}{d} f_r \left[(1 - (\frac{d}{D} cos \quad \alpha)^2 \right] \tag{4}$$

$$f_{ir} = \frac{n_b}{2} f_r (1 + \frac{d}{D} cos \quad \alpha) \tag{5}$$

Where: n_b represents the number of rolling elements (balls); *d*: is the rolling element's diameter; *D*: is the bearing pitch diameter;α: is the bearing working angle (0° for rolling bearing);$f_{or}, f_c, f_{re}, f_{ir}$: belong to the specific frequencies for the given failures, respectively: outer race, bearing cage, rolling element (ball) and inner race.

3. BEARING FAULT DETECTION AND LOCATION METHOD

As shown in the following subsections, this strategy is devided into two parts: the first covers bearing fault identification techniques, and the second part focuses on fault classification methods.

3.1. Bearing fault identification based on Discrete Wavelet Transform (DWT)

The Wavelet, as its name indicates, contains small oscillating waves with time and amplitude bounds. The wavelet transform decomposes the input signal into a series of wavelet functions, denoted as $\Psi_{a,b}(t)$ [14], which are descended from a mother function called $\Psi(t)$ provided by dilatation and translation

operations [15]. The wavelet coefficients $C_{a,b}$, resulting from this transformation, reveal information about the analyzed signal $x(t)$ at different scales and are given by:

$$C_{a,b} = \int_{-\infty}^{+\infty} x(t)\Psi_{a,b}(t)dt \quad (6)$$

Where: $\Psi_{a,b}(t)$ is a wavelet function, which can be defined as:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad (7)$$

With: *a* is the translation parameter and *b* is the scale (frequency) parameter.

The wavelet transform can be categorized into three types: Continuous Wavelet Transform, Discrete Wavelet Transform, and Wavelet Packet Transform. In this paper, Discrete Wavelet Transform (DWT) is the selected tool, which involves decomposing the temporal signal into other signals by means of two complementary filters: a low-pass filter (L) and a high-pass filter (H), as shown in Figure 2. subsequently, two vectors A_k and D_k , are obtained, defined by the following relations:

$$A_{k-1} = A_k + D_k \quad (8)$$

$$X = A_k + \Sigma_{i\leq k} D_k \quad (9)$$

A_k , D_k represent, respectively, the approximations corresponding to the lowest frequencies and the details corresponding to the highest frequencies [14]. From the obtained approximations and details, it is possible to reconstruct the original temporal signal without missing any components. This reverse procedure is often named the inverse wavelet transform or signal reconstruction.

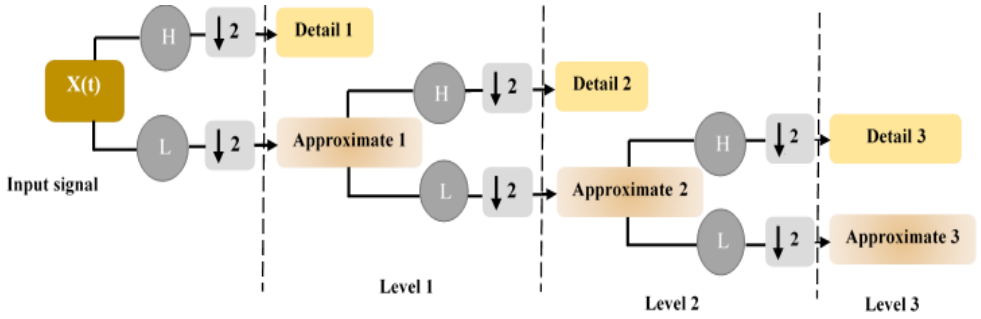

Figure 2. Example of three-level DWT decomposition.

However, two criteria are taken into account before beginning this process: the selection of the mother wavelet and the calculation of the number of decomposition level

3.1.1 Selection of the mother wavelet type

Following the development of numerous tests demonstrates that multiple types of wavelet families can produce the satisfactory results [16]. In our case, the mother wavelet selected for the DWT analysis is the Haar wavelet.

Haar wavelet: is a very fast transform, which was first introduced by Alfred Haar in 1909, who developed a basics that is considered today as the foundation of wavelet theories. The Haar wavelet is expressed by the function $\Psi(t)$ defined in the equation (10), as shown below[17]:

$$h(t) = \begin{cases} 1\ pour & 0 \leq t \leq \frac{1}{2} \\ -1\ pour & \frac{1}{2} \leq t \leq 1 \\ 0 & Elesewhere \end{cases} \quad (10)$$

3.1.2 Calculation of the number of decomposition levels

The procedure of this decomposition is repeated N times, where N is the maximum number of levels. The number of DWT decomposition levels implemented for vibration signal analysis is determined by the following relation:

$$N = 2 * log\left(\frac{f_s}{4}\right) \quad (11)$$

Where fs: is sampling frequency.

3.2. Statistical Study Based On RMS, Correlation Coefficient, Energy Coefficient and Peak-to-Peak factor

In this section, a statistical analysis based on RMS, CC, EC and Peak-to-Peak factor is applied for each reconstructed detail $\mathrm{Re}cd_k$ in order to select the appropriate detail including the necessary information about the state of the bearing.

- **Root Mean Square (RMS)**

RMS is used to monitor the condition of the machine, as it is associated with vibration energy. Consequently, it aids in detecting abnormally high energy dissipation that occurs concurrentely with the birth of a defect. This indicator is given by the following formula [18]:

$$RMS_k = \sqrt{\frac{\sum_{i=1}^{N}(Re\ cd_k(t))^2}{N}} \quad (12)$$

- **Correlation Coefficient (CC)**

The correlation coefficient is an influential indicator used to figure out the linear relationship between the raw signal x(t) and the reconstructed details obtained by the DWT. This indicator is determined by the equation below[19]:

$$CC_k = \sum_{i=1}^{N} \frac{x(t)\,Re\,cd_k(t)}{\left(\sqrt{\sum_{i=1}^{N} x^2(t)}\sqrt{\sum_{i=1}^{N} Re\,cd_k^2(t)}\right)} \quad (13)$$

- **Energy Coefficient (EC)**

The variation of the signal energy serves as a crucial indicator for detecting bearing faults. The energy coefficient of each reconstructed detail is expressed by the following relation [20]:

$$E_k = \frac{1}{N}\sum_{i=1}^{N}(Re\ c\ d_k(t))^2 \quad (14)$$

Where: N is the total number of samples of the reconstructed detail, and k is the number corresponding to the reconstructed detail.

- **Peak-to-Peak factor (P2P)**

This factor is more adapted to represent a signal introduced by impulsive forces, like bearing defects, allowing for early detection of these faults, and it is given by the following expression:

$$P2P_k = max|Re\ c\ d_k(t)| - min|Re\ c\ d_k(t)| \quad (15)$$

3.3. Envelope spectrum analysis

Envelope analysis is a widely used technique for diagnosing rotating machines by early detection of defects of the shock type. For the temporal signal, the envelope represents the instantaneous amplitude of a time-varying signal. Generally, to obtain the envelope of a time-domain signal, the expression of an analytic signal $x_a(t)$ is first derived. Then, it will be composed of a real part, $x(t)$,which represents the original signal, and an imaginary part, $x_h(t)$, which is calculated based on the Hilbert transform, eliminating negative frequency components [17]. Mathematically, the analytic signal is defined as:

$$x_a(t) = x(t) + jx_h(t) \quad (16)$$

The formula of the Hilbert transforms of a signal $x(t)$, is given by:

$$x_h(t) = HT\{x(t)\} = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x(t)}{(t-\tau)}d\tau \quad (17)$$

To generate the envelope, expressed by *E*, the modulus of the analytic signal is calculated as follows:

$$E = \sqrt{x(t)^2 + x_h(t)^2} \quad (18)$$

3.4. Bearing elements faults locating by means of random forest (RF) algorithm

In our research, the fault classification of IM was explored using the Random Forest (RF) model. The RF classification algorithm is formed by a collection of decision tree, where the decision of the output depends of the main rules obtained from the results of these trees. The random forest is affected by significant

parameters such as the number of trees, number of sample leaves, number of features in each split and the maximum depth. The RF algorithm is summarized by the following steps [21]:
Step (1): Create a sample feature set with a total of N samples, each of which has L feature attributes, by collecting the training sample set and then extracting features for each sample;
Step (2): Execute independent repeated sampling using the sample feature set S; choose n samples as 1 sample set; repeat K times; and collect K sample sets;
Step (3): Obtain K sample sets with l feature attributes by selecting randomly l from all L feature attributes in each sample set;
Step (4): Use 1 feature attribute as the decision tree's input such that K sample sets can produce K decision trees to create a RF;
Step (5): Diagnose the test samples using the created decision trees, and then use integrated voting to determine the diagnosis and recognition of the RF based on the outcomes of all decision trees' recognition.

3.4.1. Random forest model evaluation

Better features extraction process enhanced the performance and the efficiency of the classification model. here in, among various evaluation parameters, the accuracy, sensitivity, specificity, precision, g-mean, and F1-Score are taken as the performance criterion in this study. The considered evaluation metrics are calculated through of the equations listed in Table 1 [22].

Table 1. Model evaluation metrics.

Parameters definition	Equations	
Accuracy: represents the proportion of correctly classified events.	$Acc = \frac{T_p + T_N}{T_p + T_N + F_p + F_N}$	(19)
Sensitivity: is the capacity to achieve a positive outcome when a defect exists.	$Sens = \frac{T_p}{T_p + F_N}$	(20)
Specificity: is the possibility to achieve a negative outcome once the defect is absent.	$Spec = \frac{T_N}{T_N + F_p}$	(21)
Precision: is the proportion of expected positive events that actually are positive.	$\Pr ec = \frac{T_p}{T_p + F_p}$	(22)
G-mean: represents the geometric mean of sensitivity and precision.	$G - mean = \sqrt{Sens + \Pr ec}$	(23)
F1-Score: summarises the precision and recall values in a single metric and can be simplified and expressed directly from the components of the confusion matrix.	$F1 - score = \frac{T_p}{T_p + \frac{1}{2}(F_p + F_N)}$	(24)

Where T_P, T_N, F_P and F_N represents respectively, true positive, true negative, false positive, and false negative.

4. EXPERIMENTAL TEST RIG AND FLOWCHART DESCRIPTION

4.1. Experimental test rig description

In order to apply the proposed methods, vibration signals (acceleration) are collected using the test bench of the Case Western Reserve University (CWRU, bearing data center) [22]. The schematic diagram of rotor-bearing system, illustrated in Figure 3, consists of an electric motor, a torque transducer and encoder, and a dynamometer. Vibration signals were measured using accelerometers attached to the motor with magnetic bases, placed in the vertical direction on the bearing housing of the drive end in the motor. The faulty bearings are re-installed in the motor, and the sampling frequency was set at 12 kHz. Thus, the measured vibrational signals were carried out with four different loads (0 Nm, 1 Nm, 2 Nm and 3 Nm), at four motor speeds corresponding to 1797, 1772, 1750 and 1730 RPM, respectively.

The data set of the measured vibrations in the drive end bearing contains about 64 signals, divided into normal and damaged states of bearings. four cases among them belong to the normal bearings, while the remaining cases represent the faulty bearings with defects in the outer race, the rolling element and the inner race. The faults in the rolling bearings are artificially created separately at the outer race, the rolling element and the inner race using Electro-Discharge Machining (EDM), a method of treating hard metals or mechanical components that cannot be penetrated with conventional methods. In this case, the defects are related to a central hole, and it is determined by their diameter. The defect diameters discussed herein correspond to 0,007",

0,014", 0,021" and 0,028" fault diameter for the three bearing elements: outer race, rolling element and inner race. The ball bearings utilized in the experiment are of type 6205-2 RS JEM SKF, and their dimensions are listed in Table 2.

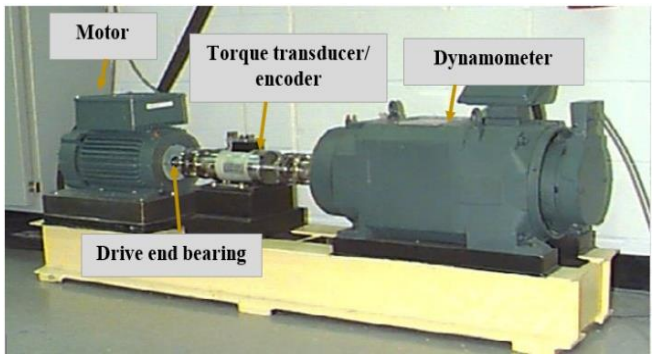

Figure 3. CWRU experimental test-bed [23].

Table 2. Parameters of the bearing 6205-2 RSJEM SKF.

Bearing characteristics	Dimensions
Inside diameter	25mm
Outside diameter	52mm
Intermediate diameter	39mm
Thickness	15mm
Ball diameter	8mm
Rolling element number	9
Contact angle	0 rad

4.2. Flowchart presentation of bearing faults diagnosis

Figure 4 presents the strategy followed in this paper for bearing fault detection and location. The algorithm for this bearing fault diagnosis uses the raw vibrational signals for analyzing and processing based on DWT.

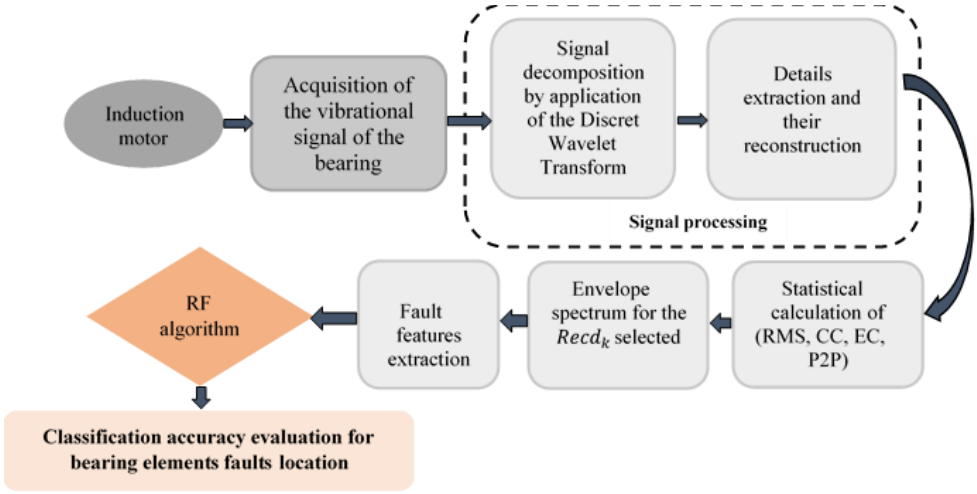

Figure 4. Flowchart of the bearing elements faults detection and classification.

The goal of this process is to extract the features of the signal through several transform methods as structured in sections bellow.

5. VIBRATION SIGNAL TIME DOMAIN ANALYSIS

Figure 5 (a-d) shows, respectively, the temporal vibrational signals of the following cases: undamaged bearing, a defective bearing with the outer race, a defective bearing with the ball, and a defective bearing with the inner race. These signals correspond to the motor with a rotational speed of 1797 RMP under a fault diameter of 0,007".

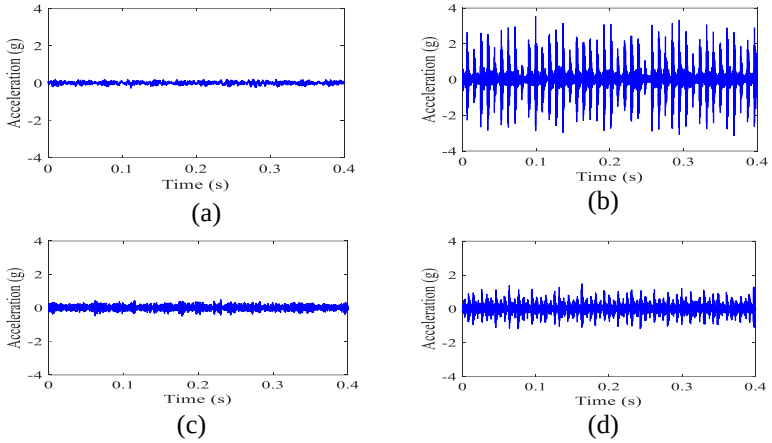

Figure 5 (a-d). Temporal vibrational signals. (a): Undamaged bearing. (b): A defective bearing (outer race). c): A defective bearing (ball). (d): A defective bearing (inner race).

5.1. Discrete wavelet transform application

The raw vibrational signals collected are decomposed into seven details by using the Eq. 11, with the "Haar wavelet" selected as the mother wavelet. Then, they are reconstructed from their wavelet coefficients at each level of decomposition to avoid any missing of the signal information. In our application, the sampling frequency f_s is in the order of 12000 Hz, allowing for a number of decomposition levels equal to:

$$N = 2 * log\left(\frac{12000}{4}\right) \rightarrow N = 7 \quad (25)$$

Figure 6 (a-d) depicts the first three reconstructed details obtained by the DWT transform for the four cases of bearing: healthy bearing, faulty bearing withouter race, faulty bearing with ball, and faulty bearing with inner race.

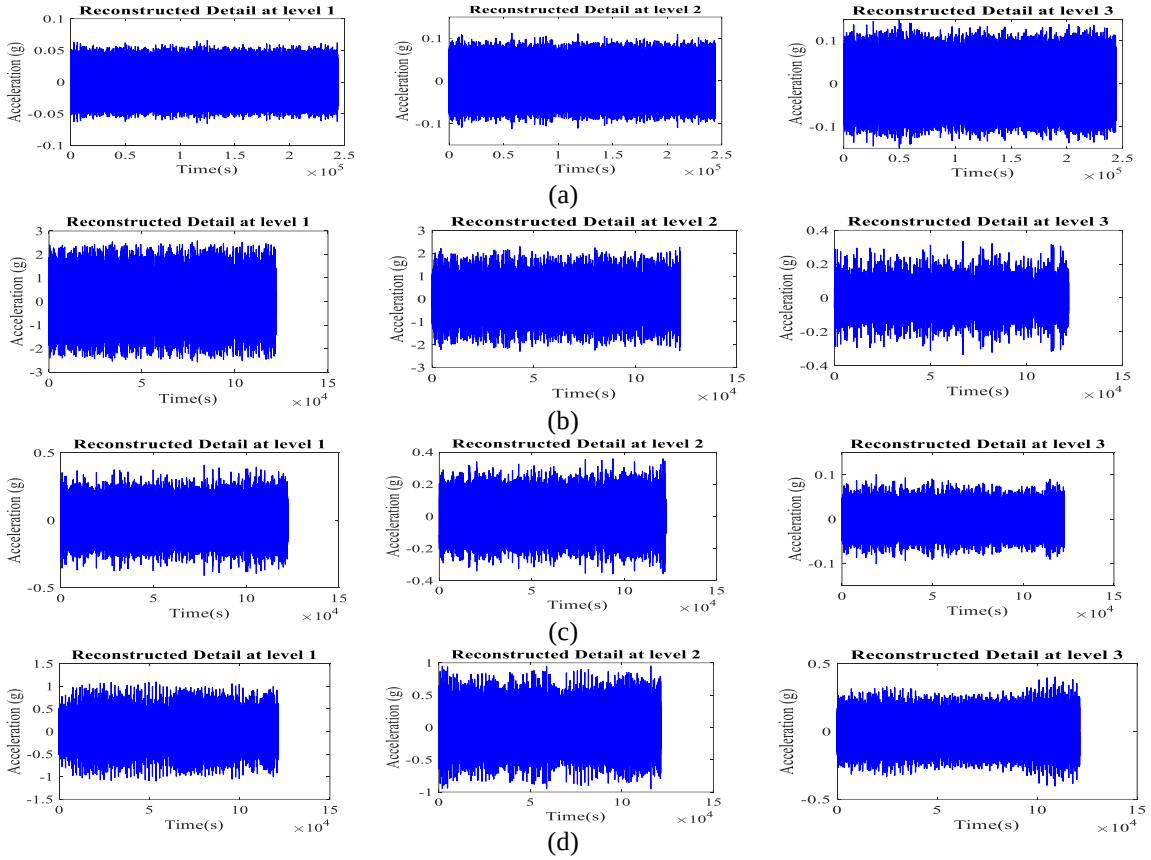

Figure 6 (a-d). The first three reconstructed details for (a): Undamaged bearing. (b): A defective bearing (outer race). c): A defective bearing (ball). (d): A defective bearing (inner race).

5.2. Statistical factors calculation

Four statistical factors are calculated for the four cases performed in this study: the healthy case, the outer race fault, the ball fault, and the outer race fault. The results are listed in Table 3.

Table 3. The calculated factors for each reconstructed detail corresponding to each case.

		Reconstructed details							
Case	Factor	Rd_1	Rd_2	Rd_3	Rd_4	Rd_5	Rd_6	Rd_7	F_{Tresh}
Healthy bearing	*RMS*	0.0186	0.0327	0.0447	0.0255	0.0227	0.0205	0.0141	0.0255
	CC	0.2523	0.4437	0.6060	0.3454	0.3075	0.2784	0.1908	0.3463
	EC	-20.9155	-51.5885	-46.8578	-42.7702	-42.9384	-40.8366	-39.6561	-40.7947
	P2P	0.0659	0.1123	0.1484	0.0876	0.0971	0.0762	0.0636	0.0930
Outer race fault	*RMS*	0.5034	0.4287	0.0584	0.0686	0.0437	0.0192	0.0097	0.1617
	CC	0.7519	0.6403	0.0872	0.1024	0.0653	0.0286	0.0144	0.2415
	EC	-19.5856	-21.2528	-22.4372	-21.4824	-23.1869	-24.0380	-25.9149	-22.5568
	P2P	2.5791	2.2956	0.3350	0.3516	0.2033	0.0815	0.0310	0.8396
Ball fault	*RMS*	0.0989	0.0891	0.0252	0.0236	0.0150	0.0088	0.0032	0.0377
	CC	0.7102	0.6399	0.1807	0.1698	0.1076	0.0629	0.0231	0.2706
	EC	-20.7499	-21.0820	-23.6602	-24.2737	-25.8788	-27.2031	-25.0408	-23.9841
	P2P	0.4101	0.3589	0.1010	0.0985	0.0485	0.0280	0.0256	0.1529
Inner race fault	*RMS*	0.1936	0.1877	0.0887	0.0561	0.0270	0.0086	0.0061	0.0823
	CC	0.6642	0.6438	0.3042	0.1926	0.0924	0.0587	0.0209	0.2824
	EC	-19.4820	-20.5816	-20.5920	-22.2446	-22.4666	-23.1741	-28.0650	-22.3723
	P2P	1.0975	0.9523	0.4030	0.1757	0.1098	0.0821	0.0236	0.4063

Figure 7 (a-d) shows clearly the variation of different factors calculated for each detail.

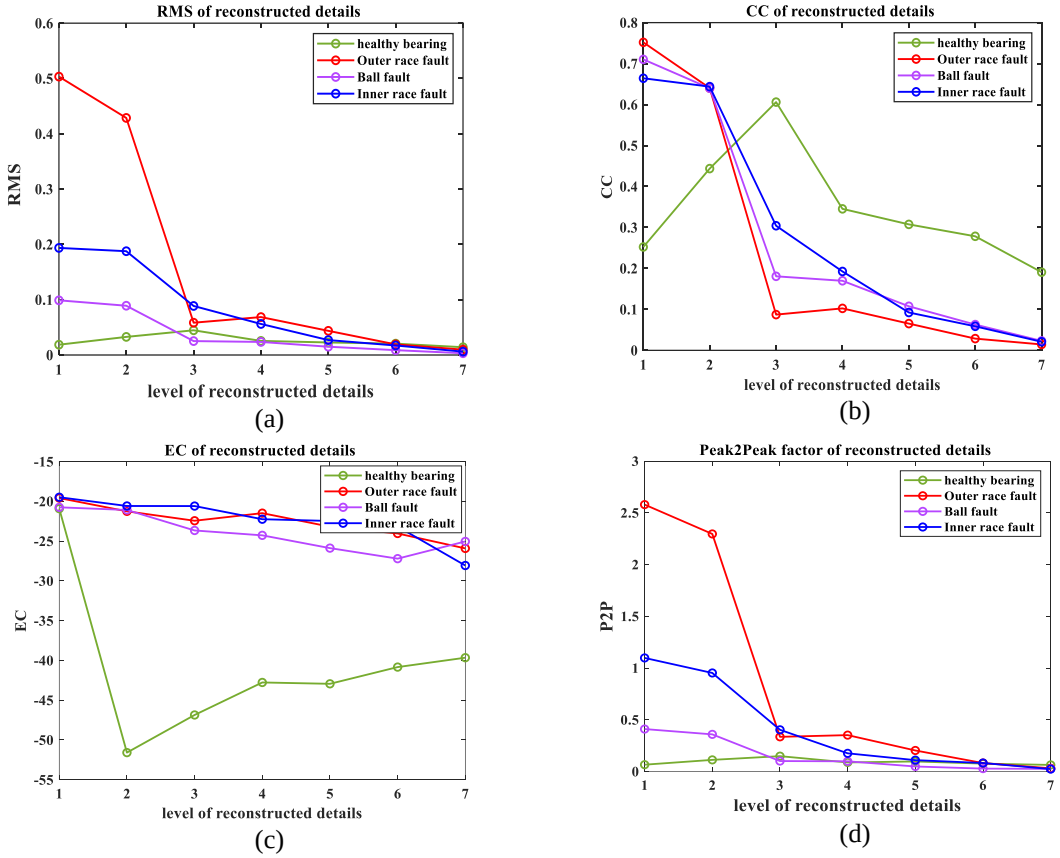

Figure 7 (a-d). Statistical coefficients for each detail corresponding to the healthy case and the other faulty cases.

To select the relevant reconstructed detail ($\mathrm{Re}cd_k$) containing the necessary information about the fault frequencies and their harmonics, the following condition must be released, the RMS value, the CC value, the EC value and the P2P value of the reconstructed detail $\mathrm{Re}cd_k$ have to be the greatest values comparing them, respectively, with the average value of the RMS, the CC, the EC and the P2P factor. A pertinent analysis

of the results yielded from each calculated factor (Table 3 and Figure 7 (a-d)) shows that for the healthy bearing case the $\mathrm{Re}cd_3$ fulfill well at the condition and for the other cases, the outer race fault, the ball fault and the inner race fault we find that the $\mathrm{Re}cd_1$ is well suitable to the given condition.

The time view of the measured signal and the reconstructed details cannot allow extracting any signature indicating the state of the bearing elements. For this reason, the application of envelope spectrum analysis has taken place.

5.3. Fault features extraction based on envelope spectrum analysis

The fault frequencies theoretically calculated using the equations (1)-(5) listed above are summarizing in Table 4 (for the rotational speed: 1797 RPM).

Table 4. Theoretical calculated bearing faults frequencies.

Rotation frequency	Inner race frequency	Outer race frequency	Rolling element frequency	Bearing cage frequency
29.95	162.40	107.12	139.86	11.90-18.04

The results obtained from the application of the envelope spectrum clearly demonstrate the effectiveness of the reconstructed details selection done above. The different fault frequencies and their harmonics are illustrated in Figure 8 (a-d) below.

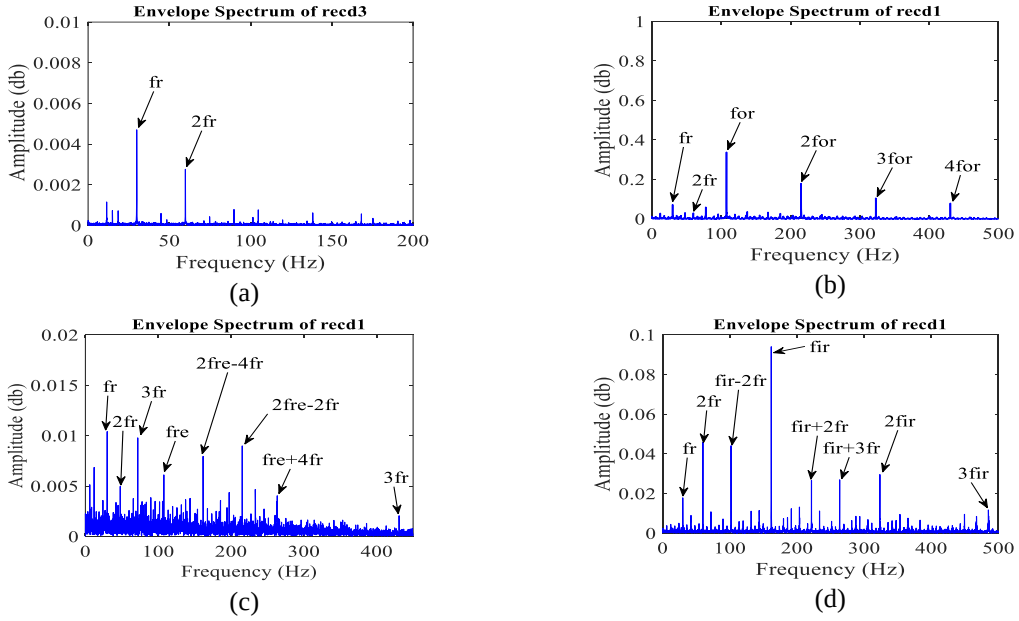

Figure 8 (a-d). Envelope spectrum of the reconstructed details. (a): undamaged bearing. (b): defective bearing with the outer race. (c): defective bearing with the ball. (d): defective bearing with the inner race.

Observing the spectrum of the reconstructed details shows clearly a good agreement between the observed results and the theoretically expected results for bearings with a motor speed of 1797 RPM under a fault diameter of 0.007 ". Figure 8 (a) displays the spectrum of the reconstructed detail 3 for the healthy bearing case, the rotation frequency f_r=29.95 Hz and its multiplicity $2f_r$=59,9 Hz. These values are similar to those calculated theoretically, and there is no frequency of fault, confirming well the normal state of the bearing. Figure 8 (b) shows the spectrum of the reconstructed detail 1 for the outer race fault, containing the following frequencies: f_{or}=107 Hz, $2f_{or}$ =214 Hz, $3f_{or}$ = 322 Hz, $4f_{or}$ =428 Hz. These frequencies belong to the outer race fault frequency and their harmonics, which demonstrates well the presence of a defect in the bearing's outer race. The spectrum of the reconstructed detail 1 corresponding to ball fault shown in the Figure 8 (c) includes the mentioned frequencies: f_{re}=139.50 Hz, $2f_{re} - 4f_r$=160 Hz, $2f_{re} - 2f_r$=220 Hz, $f_{re} + 4f_r$= 259.70 Hz, $3f_{re}$=419.70 Hz.

These results closely align with the theoretical findings, clearly defining the abnormal state of the bearing's ball element. For the last one, Figure 8 (d) depicts the spectrum of the reconstructed detail 1

corresponding to the inner race fault and containing the following frequencies: f_{ir}=162 Hz, $f_{ir} - 2f_r$= 101 Hz, $f_{ir} + 2f_r$=222.50 Hz, $f_{ir} + 3f_r$= 253 Hz, $2f_{ir}$= 325 Hz and $3f_{ir}$=487.50 Hz. These frequencies are closely the same to those calculated above, definitively confirming the presence of a defect in the inner race of the bearing.

5.4. Random forest training model

In our application, the automation of bearing fault monitoring involves extracting features that characterize the bearing's fault. Consequently, the process carried out in the preceding sections was replicated while altering the motor load torque from 0 load to 1, 2, 3 loads. These loads correspond, respectively, to motor speeds of 1797 RPM, 1772 RPM, 1750 RPM, and 1730 RPM, with fault diameters of 0.007", 0.014", 0.021", and 0.028". The analysis of the obtained results indicates that the features extracted in each repeated operation are as follows: f_r, $2f_r$, f_{or}, $2f_{or}$, $3f_{or}$, $4f_{or}$, f_{re}, $2f_{re}$, $f_{re} + 4\,f_r$,f_{ir}, $2f_{ir}$ and $3f_{ir}$. These features are chosen carefully and are also used as new features to train a machine learning model based on the random forest algorithm. Furthermore, the dataset utilized to train the RF algorithm forms a matrix of (64x12), where 64 represents the total number of signals that include the healthy bearing, the bearing with the outer race fault, the bearing with the ball fault, and the bearing with the inner race fault. The number 12 signifies the features exctracted for each signal.

This data is then divided following a split ratio of 80:20, as determined by the findings obtained in [21]. The subdivision results in a training set holding 80% (51 signals) of the total data and a test set holding 20% (13 signals) of the data. This ratio seems like more appropriate for developing and proving the classification algorithms' capacity for fault diagnosis, especially in the classification of bearing defects [8].

The structure of the random forest corresponding to the features selected is illustrated in the Figure 9 below:

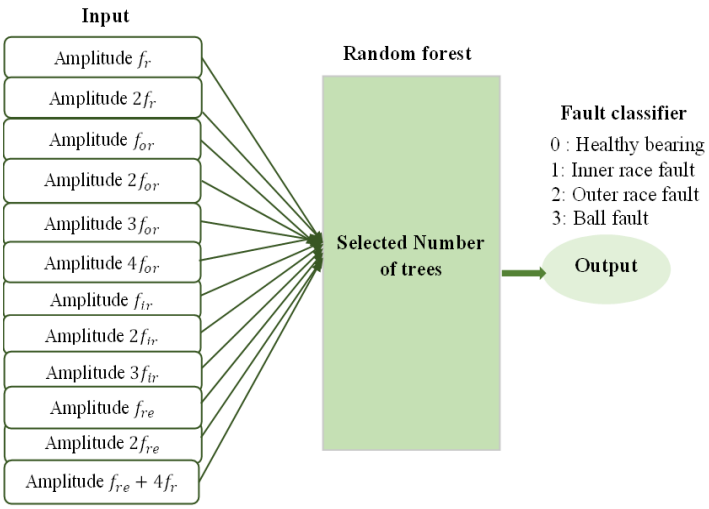

Figure 9. The schematic structure of the Random Forest

6. RESULTS AND DISCUSSION

Figure 10 shows the accuracy as a function of the trees number.

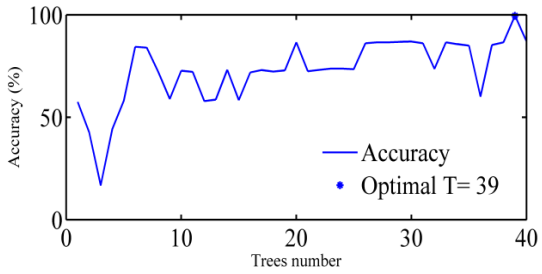

Figure 10. Accuracy as a function of trees number.

Although the database is limited, the sensitivity shows a very high rate of 97.70%. This indicates a very high capacity to obtain a true positive result when the bearing is faulty. In addition, the specificity tends towards a very high rate of 94.66%, meaning the probability of obtaining a negative result when the bearing is

healthy is very high. Moreover, the classifier is highly precise with a rate of 97.82%, indicating a high ability to predict a true positive result from all cases predicted positive by the classifier. Another evaluation metric that combines the precision and recall scores of a model is the F1-score, yielding a rate of 99.23%.

Otherwise, the G-Mean rate is an indication of good or bad performance in the classification of the positive cases, even if the negative cases are correctly classified as such. Herein, the G-mean value has a high rate of 98.01%, denoting well-classified positive cases. Finally, the accuracy or the rate of correctly classified cases equal is 99.53%.

Figure 11 shows the related confusion matrix. According to the confusion matrix, the classifier is quite successful in categorizing the errors, since the false positive and false negative values are low.

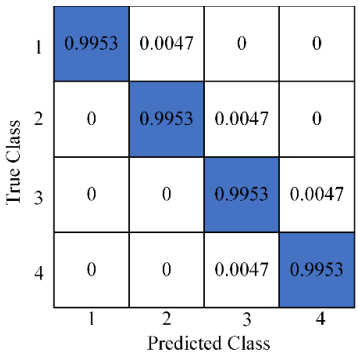

Figure 11. Confusion matrix of the built Random Forest model.

6.1. A brief comparison with some recent works

In table 5, we list some recent works to facilitate a comparison between them and our proposed method.

References	Method	Accuracy
[24]	Multi-domain-entropy-random forest	93,75%
[25]	Convolutional Neural Network-Support Vector Machine	98,75%
[26]	Cross-domain- deep generative neural networks (DGNN)	97,81%
[8]	DWT- RF/XGBoost	99,14 % /99,30 %
[12]	Random Forest, ANN, and Autoencoder	91%
Proposed method	**DWT-ENV-Random Forest**	**99,53%**

As evident from the results, the proposed method consistently achieves higher diagnostic accuracy compared to the other recent works. The findings underscore the superiority of RF over the default classifier. This affirms that the proposed model effectively addresses the challenge of extracting in-depth features from diverse data using DWT-ENV signal analysis methods and successfully meets the demanding sample requirements for RF training, showcasing remarkable stability.

7. CONCLUSION

In this research, a novel strategy based on a new feature set extracted using a bearing vibration signal analyzed by a combined technique DWT-ENV spectral and random forest algorithm, was particularly illustrated for an automatic IM's bearing faults diagnosis. In the first stage, raw vibrational signals of bearing conditions such as healthy, outer race, ball, and inner race were decomposed by the DWT to obtain a certain number of approximations and details using the Haar wavelet function. The DWT assures both of the decomposition while the denoising of the signal, which necessitates the reconstruction of the provided details to avoid any missing of information. Then, only relevant reconstructed details were selected based on the following statistical factors (RMS, CC, EC and P2P factor) to reduce de dimensionality of the original structure of the vibration signal.

The presence of the bearing faults frequencies of the outer race, the ball, and the inner race was proved by appliying ENV spectral analysis that allows to extract a new frequency-feature set. These features were then used as inputs for training the random forest classifier, achieving an accuracy of over 99.53%. This underscores the quality of the feature structure, as well as the efficiency and robustness of the proposed strategy for both the detection and localization of bearing defects. Moreover, this automatic bearing fault diagnosis supports an optimal level of predictive maintenance by ensuring early detection of defects, potentially reducing

redundancy. Finally, the effectiveness of the obtained results was compared with some recent works using the same dataset. The proposed DWT-ENV-RF method emerged as a valuable model for bearing fault localization, attaining higher accuracy while requiring a simpler architecture and less computational complexity compared to other methods.

REFERENCES

[1] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification.," *Third Eur. Conf. Progn. Heal. Manag. Soc. 2016*, no. Cm, pp. 152–156, 2016.

[2] R. A. Rahman, M. F. Erikyatna, and A. F. Hery, "Study on Predictive Maintenance of V-Belt in Milling Machines Using Machine Learning Study on Predictive Maintenance of V-Belt in Milling Machines Using Machine Learning," no. November, 2022, doi: 10.17977/um016v6i22022p085.

[3] A. Picot, Z. Obeid, J. Régnier, S. Poignant, O. Darnis, and P. Maussion, "Statistic-based spectral indicator for bearing fault detection in permanent-magnet synchronous machines using the stator current," *Mech. Syst. Signal Process.*, vol. 46, no. 2, pp. 424–441, 2014, doi: 10.1016/j.ymssp.2014.01.006.

[4] M. Ye, J. Zhang, and J. Yang, "Bearing Fault Diagnosis under Time-Varying Speed and Load Conditions via Observer-Based Load Torque Analysis," *Energies*, vol. 15, no. 10, 2022, doi: 10.3390/en15103532.

[5] A. Allouche *et al.*, "A PLL based mechanical faults detection in PMSM at variable speed," vol. 51, no. 24, pp. 1445–1451, 2018, doi: 10.1016/j.ifacol.2018.09.534.

[6] A. D. Nembhard, J. K. Sinha, A. J. Pinkerton, and K. Elbhbah, "Fault diagnosis of rotating machines using vibration and bearing temperature measurements," *Diagnostyka*, vol. 14, no. 3, pp. 45–51, 2013.

[7] M. A. Jamil and S. Khanam, "Fault Classification of Rolling Element Bearing in Machine Learning Domain," *Int. J. Acoust. Vib.*, vol. 27, no. 2, pp. 77–90, 2022, doi: 10.20855/ijav.2022.27.21829.

[8] R. N. Toma and J. M. Kim, "bearing fault classification of induction motors using discrete wavelet transform and ensemble machine learning algorithms," *Appl. Sci.*, vol. 10, no. 15, 2020, doi: 10.3390/APP10155251.

[9] G. Georgoulas, T. Loutas, C. D. Stylios, and V. Kostopoulos, "Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition," *Mech. Syst. Signal Process.*, vol. 41, no. 1–2, pp. 510–525, 2013, doi: 10.1016/j.ymssp.2013.02.020.

[10] P. Thanh Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors (Basel).*, vol. 18, no. 1, 2017, doi: 10.3390/s18010018.

[11] S. M. Yadavar Nikravesh, H. Rezaie, M. Kilpatrik, and H. Taheri, "Intelligent fault diagnosis of bearings based on energy levels in frequency bands using wavelet and support vector machines (SVM)," *J. Manuf. Mater. Process.*, vol. 3, no. 1, 2019, doi: 10.3390/jmmp3010011.

[12] P. Kamat *et al.*, "Bearing Fault Detection Using Comparative Analysis of Random Forest, ANN, and Autoencoder Methods," *Lect. Notes Networks Syst.*, vol. 204, no. June, pp. 157–171, 2021, doi: 10.1007/978-981-16-1089-9_14.

[13] D. Wu *et al.*, "An automatic bearing fault diagnosis method based on characteristics frequency ratio," *Sensors (Switzerland)*, vol. 20, no. 5, pp. 1–12, 2020, doi: 10.3390/s20051519.

[14] S. Adamczak and W. Makieła, "Analyzing variations in roundness profile parameters during the wavelet decomposition process using the MATLAB environment," *Metrol. Meas. Syst.*, vol. 18, no. 1, pp. 25–34, 2011, doi: 10.2478/v10178-011-0003-6.

[15] L. Souad, B. Azzedine, C. B. D. Eddine, B. Boualem, M. Samir, and M. Youcef, "Induction machine rotor and stator faults detection by applying the DTW and N-F network," *Proc. IEEE Int. Conf. Ind. Technol.*, vol. 2018-Febru, no. February, pp. 431–436, 2018, doi: 10.1109/ICIT.2018.8352216.

[16] C. Uyulan and T. Erguzel, "Comparison of Wavelet Families for Mental Task Classification," *J. Neurobehav. Sci.*, vol. 3, no. 2, p. 59, 2016, doi: 10.5455/jnbs.1454666348.

[17] S. Seninete, M. Mimi, B. D. Eddine Cherif, and A. Ould Ali, "Vibration Signal Analysis for Bearing Fault Diagnostic of Asynchronous Motor using HT-DWT Technique," *Proc. - 2019 6th Int. Conf. Image Signal Process. their Appl. ISPA 2019*, no. February 2020, 2019, doi: 10.1109/ISPA48434.2019.8966801.

[18] A. A. Bengharbi, S. Laribi, T. Allaoui, and A. Mimouni, "Photovoltaic system faults diagnosis using discrete wavelet transform based artificial neural networks," *Electr. Eng. Electromechanics*, vol. 2022, no. 6, pp. 42–47, 2022, doi: 10.20998/2074-272X.2022.6.07.

[19] A. Rabah and K. Abdelhafid, "2777 . Rolling bearing fault diagnosis based on improved complete ensemble empirical mode of decomposition with adaptive noise combined with minimum entropy deconvolution," pp. 240–257, 2018, doi: 10.21595/jve.2017.18762.

[20] N. Bessous, S. E. Zouzou, W. Bentrah, S. Sbaa, and M. Sahraoui, "Diagnosis of bearing defects in induction motors using discrete wavelet transform," *Int. J. Syst. Assur. Eng. Manag.*, vol. 9, no. 2, pp. 335–343, 2018, doi: 10.1007/s13198-016-0459-6.

[21] B. D. Eddine Cherif, S. Seninete, and M. Defdaf, "a Novel, Machine Learning-Based Feature Extraction Method for Detecting and Localizing Bearing Component Defects," *Metrol. Meas. Syst.*, vol. 29, no. 2, pp. 333–346, 2022, doi: 10.24425/mms.2022.140038.

[22] R. Costache *et al.*, "Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques," *Remote Sens.*, vol. 12, no. 1, 2020, doi:

10.3390/RS12010106.
[23] Bearing Data Center - Case Western Reserve university, "http://csegroups.case.edu/bearingdatacenter/pages/welcome-case western-reserve-university."
[24] J. Tian, L. Liu, F. Zhang, Y. Ai, R. Wang, and C. Fei, "Multi-domain entropy-random forest method for the fusion diagnosis of inter-shaft bearing faults with acoustic emission signals," *Entropy*, vol. 22, no. 1, p. 57, 2020, doi: 10.3390/e22010057.
[25] L. Yuan, D. Lian, X. Kang, Y. Chen, and K. Zhai, "Rolling Bearing Fault Diagnosis Based on Convolutional Neural Network and Support Vector Machine," *IEEE Access*, vol. 8, pp. 137395–137406, 2020, doi: 10.1109/ACCESS.2020.3012053.
[26] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5525–5534, 2019, doi: 10.1109/TIE.2018.2868023.

BIOGRAPHY OF AUTHORS

Saida DAHMANE received her B.Sc. degree in industrial maintenance in 2017, her M.Sc. in Management and engeneering of industrial maintenance in 2020 from the national high school of technology in Algies-Algeria, and is currently a Ph.D. student in Industrial maintenance field. Her research interests include signal processing, machine learning, electrical machine and drive modeling and analysis, electrical machine and drive control and converters as well as electrical machine and drive faults diagnosis and tolerance.

Fouad BERRABAH was born in m'sila, in 1979. He received the B.Sc., M.Sc. and Ph.D. degrees in Electromechanical Engineering from Badji-Mokhtar Annaba University Algeria in 2002, 2008 and 2017 respectively. He is a lecturer senior at the University of M'sila in Algeria. His research interests include electrical machine and drive modeling and analysis, electrical machine and drive control and converters as well as electrical machine and drive fault diagnosis and tolerance.

Mabrouk DEFDAF received his B.Sc., M.Sc. and Ph.D. degrees in Electromechanical Engineering from Badji-Mokhtar Annaba University Algeria in 2000, 2007 and 2018 respectively. He is a lecturer senior at the University of M'sila in Algeria. His research interests include electrical machine and drive modeling and analysis, electrical machine and drive control and converters as well as electrical machine and drive fault diagnosis and tolerance.

Saad SALAH was born in Batna, Algeria, in 1958. He received the Engineer degree in electromechanical applied to mining fields from the University of Badji-Mokhtar, Annaba, Algeria, in 1983, and Ph.D. degree from Nottingham University, U.K., in 1988. Since 1988, he has been a Lecturer, a Senior Lecturer, and a Professor with the University of Badji-Mokhtar. He has supervised many graduated and postgraduate student theses. He has conducted many research projects in power electronics applications, electrical ac and dc drives, and diagnosis and faults detection in ac machines. He has authored or co-authored many journal and conference papers. He has co-authored a book in the field of signal processing published in Algeria in 1992. His research interests are mainly in the area of power electronics, such as harmonics elimination by active filters, PWM and space vector modulation control, multilevel inverters, new converter topologies, and vibration sensors.