



# Gears and bearings combined faults detection using optimized wavelet packet transform and pattern recognition neural networks

Issam Moumene<sup>1</sup> · Nouredine Ouelaa<sup>1</sup>

Received: 12 October 2021 / Accepted: 19 January 2022 / Published online: 15 March 2022  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

## Abstract

A great percentage of breakdowns in the rotary machines are caused by faulty gears and bearings generating high costly downtime. The diagnosis of gears and bearings combined faults using conventional frequency methods is difficult because of the presence of a high level of noise in the signal or a disturbance emanating from other sources in/out of the machine. This manuscript develops an improved approach for gears and bearings simple/combined faults automated diagnosis using an optimized wavelet packet transform (OWPT), for signal denoising and fault features extraction, combined with pattern recognition neural networks (PRNN) for fault classification. We applied this improved approach on acceleration signals acquired from a test rig simulating an industrial rotary machine; then, these acquired signals were processed using OWPT, based on the best choice of the wavelet packet and the level of decomposition using minimum Shannon entropy and maximum energy to Shannon entropy ratio criteria. The results gave us the bi-orthogonal 3.1 (bior3.1) as the best mother wavelet among 41 wavelet families being tested and the 6<sup>th</sup> level as the best decomposition level. These two found parameters are used for the extraction of time-domain features, which are used as input data to train neural networks after data normalization. The results showed that the proposed method could detect any type (simple/combined) and size (small “incipient,” mean and large) of gears and bearing faults in non-stationary operations with the presence of high noise. In addition, the automation of the classification process could be generalized to any kind of faults in the machine such as misalignment, electrical faults, and rotor damage.

**Keywords** Maximum energy to Shannon entropy ratio · Optimized wavelet packet transform · Simple/combined fault classification · Rolling element bearings · Spur gears · Features extraction · Machine learning · Pattern recognition ANN

## Abbreviations and nomenclature

$a_i$	The output of the neuron	FFT	Fast Fourier transform
$b$	Bias	$g(k)$	Wavelet High-pass filter. Impulse responses of the quadrature mirror filters (QMF)
bior3.1	Bi-orthogonal 3.1		
$C_i(n)$	The $i^{\text{th}}$ wavelet coefficient of $n^{\text{th}}$ scale	$h(k)$	Wavelet Low-pass filter. Impulse responses of the quadrature mirror filters (QMF)
CWT	Continuous Wavelet Transform		
CE	Cross Entropy Error	$j$	The level of decomposition. Integer: 1, 2, 3...j
$d_{j,n}$	Wavelet coefficients for $j^{\text{th}}$ level and n Sub-band. Integer: 1, 2, 3...j	$m$	Number of Wavelet Coefficients
DWT	Discrete Wavelets Transform	OWPT	Optimized wavelets packets transform
EMD	Empirical Mode Decomposition	PRNN	Pattern recognition neural networks
$f(u_i)$	Activation Function	$p_i$	The energy probability distribution for the $i^{\text{th}}$ wavelet coefficient
		STFT	Short-time Fourier Transform
		$S(n)$	Time Signal
		TF	Time Features
		$u_i$	The weighted linear sum of the input vectors ( $x^i$ )

✉ Issam Moumene  
moumene.issam@univ-guelma.dz; i\_moumene@yahoo.fr  
Nouredine Ouelaa  
ouelaa.nouredine@univ-guelma.dz; n\_ouelaa@yahoo.fr

<sup>1</sup> Mechanics & Structures Laboratory, University 8 May 1945 - Guelma, PB 401, 24000 Guelma, Algeria

$V_{SDmin}$ and $V_{SDmax}$	The minimum and maximum values of the normalization interval
$w_i$	The synaptic weights
WRA	Wavelets Multiresolution Analysis
WPT	Wavelets Packet Transform
$x_i$	Inputs Vectors
$\sum x_i$	Summation Function
$x_{ND}$	The Normalized Data Value for the Extracted Feature
$x_{SDmin}$ and $x_{SDmax}$	The Minimum and the Maximum Values in the Feature Vector in the Input Data Matrix
X1	The Acquired Data from the 1st Accelerometer $m/s^2$
X2	The Acquired Data from the 2nd Accelerometer $m/s^2$
X3	The Acquired Data from the 3rd Accelerometer $m/s^2$
$\eta(n)$	The Maximum Energy-to-Shannon Entropy ratio

## 1 Introduction

All industries need an efficient predictive maintenance plan with great performance in order to optimize the management of resources and improve the economy of the plant by reducing the extra costs and increasing the safety level and the plant process quality. Machine condition monitoring and fault diagnosis are the most important challenges in the manufacturing industry; based on them, we can reduce the maintenance costs by increasing machines' availability and improving productivity and safety. Vibratory analysis, oil analysis, thermography, and acoustic analysis are the principal methods of industrial rotating machines monitoring, but the problem of which method we can use for any application is the main question. To answer this question, Zani [1] gave selection criteria between such methods or techniques and their field of application, including the advantages and limitations of each method.

Among the techniques of monitoring previously cited, the most important, reliable, and powerful method especially for chocks, machine vibration, and the phenomena of wear and fatigue is the *vibratory analysis* [2–5], the technique that had the largest field of applications.

The classification of vibration monitoring methods is done based on the existence of a mathematical model for the machine/equipment surveyed or not; for that, they are divided into two parts: *model-based methods* and *non-model-based methods*. The model-based methods need an analytical model based on a deeper understanding of the internal process of the machine system, which should be built first. Any variation between the established model

(model output) and the experimental results is a fault indicator; this kind of method is also called the *residual-based method* [6].

On the other hand, we have three types of non-model-based methods applied in the diagnosis of industrial systems, including *artificial intelligence*, *statistical methods*, and *signal processing methods*. Artificial intelligence and statistical methods play an important role in fault diagnosis, usually applied to constitute a pattern classifier for discriminating different types of faults. From those methods, we can rise *artificial neural networks* [7–10] and *support vector machines* [11–15]. For signal processing method, in the literature, there are many signal processing tools for vibration analysis, such as power spectrum or fast Fourier transform (FFT) and the Cestrum, time, and frequency domain averaging, adaptive noise cancellation, demodulation technique, or envelope analysis and order analysis for stationary signals, and Winner-Ville and short-time Fourier transform (STFT), as well as empirical mode decomposition (EMD) and wavelet transform in different forms and recently deal with non-stationary signals. According to [16, 17], we can find different versions of wavelets analysis like the continuous version (CWT) in this works [18, 19], as a discrete form (DWT) and its fast algorithm called Wavelets multiresolution analysis (WRA) in [20–22], and its generalized version named the wavelet packet transform (WPT) in [23, 24], which will be used in this study, allowing us to extract more features than the previous version of wavelets.

The big challenge to using the WPT is the wavelet family and decomposition level selection. Several approaches in the literature are treated in this subject, and they are divided into *qualitative and quantitative criteria*. For the qualitative criteria, we have the variance criterion [25], where the sum of wavelet coefficient variance values is calculated and the base wavelet that maximizes this value will be the best choice. For the quantitative criteria, we have the cross-correlation coefficients [26], in the case where the base wavelet maximizes the correlation coefficients value and can be selected as the mother wavelet for the study. Another quantitative criterion that will be used in this paper is the *minimum Shannon entropy and the maximum energy* [27] in which the wavelet family and the decomposition level that maximize the energy and minimize the Shannon entropy are the best choices for signal denoising. The maximum energy to Shannon entropy ratio is also a powerful tool for mother wavelet and best decomposition level selection [28, 29].

Automated classification of gears and bearings faults in the rotary machines using neural networks is an open domain for research. Many approaches are developed in this issue, but most of them are focused on gear [30] or bearing [31] faults and treated each component separately [32], without entering in the combination of both types of faults. Therefore, few papers have been published related

to the combined faults, including the size and type of fault [33, 34]. In our study, the wavelet packet transform is used for signal processing, optimized by the use of minimum Shannon entropy and maximum energy to Shannon entropy ratio of the wavelet coefficients, which is one of the main contributions in this work. For combined fault classification, we used the pattern recognition artificial neural network, which has become in recent decades the outstanding method exploiting their non-linear pattern classification properties, offering advantages for automatic detection and identification of gearboxes failures including gears and bearings faults, without need for the behavior deep knowledge of the surveyed systems. In the first stage, acceleration signals were acquired using Pulse.18 system, from a test rig simulating an industrial rotary machine constructed in the LMS laboratory in the University of Guelma; then, these signals were pre-processed using an improved WPT called in this paper optimized wavelet packets transform “OWPT,” based on the wavelet family optimization using the minimum Shannon entropy and maximum energy to Shannon entropy The ratio for the choice of the best level of decomposition, which was used as input to train the neural networks to classify the different types (simple/combined) and sizes (small, mean and large) of gears and bearings faults.

This paper is organized as follows: Sect. 1 is an introduction to different research works on the same subject. Section 2 presents the proposed flow chart for gears and bearings combined faults diagnosis. Followed by three big part Sects. 3, 4, and 5, the theory part, the experimental parts, and the results and discussion part. In the theory part, we include the wavelets packets transform definition including the wavelet family choice and the decomposition level selection using the minimum Shannon entropy and the maximum energy to Shannon entropy ratio criteria, followed by a summary about the artificial neural networks and their application pattern recognition. In the experimental part, we include a description of the test rig used for different types of gears/bearings faults and their location, the equipment used for signal acquisition, and the experimental plan. In the results and discussion part, we include the application of the proposed approach on measured signals and a results validation with other test rigs from the INSA de Lion laboratory. Section 6 is reserved for general conclusions.

## 2 Gears and bearings combined fault diagnosis proposed approach

Using spectral analysis for gears and bearings combined fault diagnosis has some limitations concerning time and frequency resolutions. This problem is solved using time–frequency analysis tools. Wavelet packet transforms (WPT) overcome these limitations, but it has a problem of

spectral leakage, which is related to the choice of the wavelet family and the mother wavelet used in the analysis. In order to minimize these errors, we proposed an approach to select the most suitable wavelet family, the mother wavelet, and the best decomposition level, using the *Shannon entropy criterion and the maximum energy to Shannon entropy ratio*.

The OWPT is used to extract different features that are used as input data for gears and bearings simple and/or combined faults classification using *pattern recognition neural networks*.

The steps of faults classification proposed in this manuscript are described below:

1. Signal acquisition and data preparation
2. Data processing using wavelet packet and the maximum energy to Shannon entropy ratio for WPT decomposition level choice (level, which has the maximum energy to Shannon entropy ratio, will be used as input in PRNN)
3. Feature extraction using Table 1
4. Input data value normalization
5. Feature selection
6. Fault classification and the output vector determination: The output vector is a matrix of total classification that represents the training target of a neural network, where each class is represented by one (1), which means the presence of a fault, and the other column by zero (0), which means the absence of a fault.
7. Performance function error calculation (cross-entropy error “CE”): We compare the output matrix with the target matrix, and we determine the classification process performance.

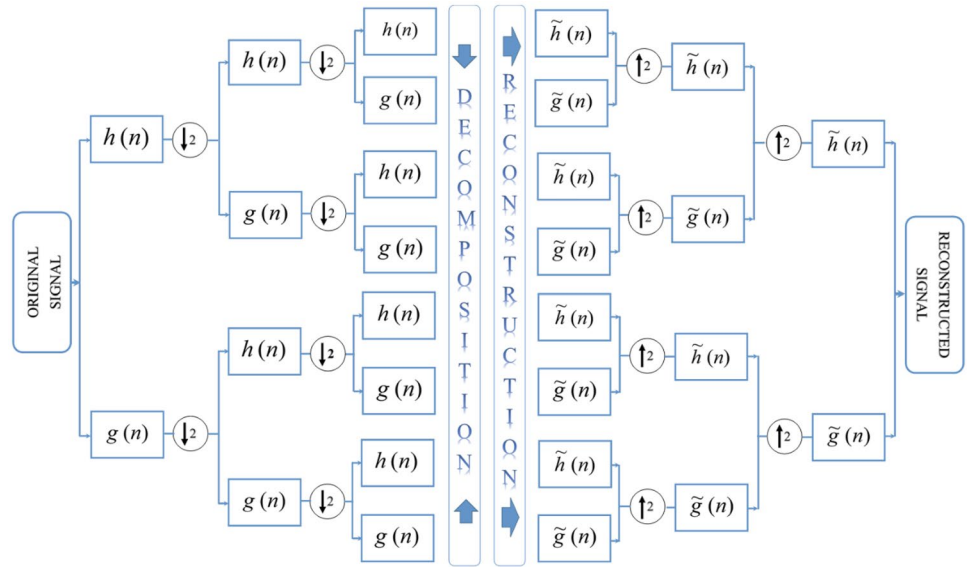
## 3 Theory part

### 3.1 Wavelet packet transform

#### 3.1.1 Wavelet packet transform definition

Wavelet transform is able to process stationary and non-stationary signals in time and frequency domains simultaneously. Wavelet transform could be found in continuous “CWT” and discrete “DWT” forms; the discrete form is faster and easier for application than continuous form, so DWT has been widely applied in the field of condition monitoring using its fast algorithm called wavelets multiresolution analysis. A general form for this transform is called wavelet packet transform “WPT.” Kumar et al. in [35] have used the multiscale slope feature extraction method to compare the two transforms DWT and WPT, and they concluded that WPT is more effective in bearing condition classification when compared to DWT and it’s considered superior to DWT in classifying

**Fig. 1** Wavelet packet decomposition tree: Schematic for signal decomposition/reconstruction



bearing vibration signals for different conditions. WPT is a generalization of wavelet decomposition that offers more possibilities of reconstructing the signal from the decomposition tree. The transformation can be achieved by implementing a pair of wavelet filters,  $h(k)$  low-pass filter, and  $g(k) = (-1)^k h(1 - k)$  high-pass filters. These high-pass wavelet filters are constructed using the wavelet function  $\psi(t)$  and its corresponding scaling function  $\phi(t)$ . Wang et al. in [36] have well-described WPT calculations and used equations for more details. The schematic of WPT decomposition has been shown in (Fig. 1).

In this study, the row signal will be decomposed at the sixth level ( $j = 6$ ), where 64 coefficients were obtained.

**3.1.2 Wavelet packet transform optimization**

Several criteria for mother wavelet and decomposition level selection are mentioned in Sect. 1; the best of them and the most widely used for mother wavelet selection is the *minimum Shannon entropy* and the *maximum energy*, in which the wavelet family that minimizes the Shannon entropy and maximizes the energy will be the best for the signal decomposition process (Fig. 2).

The energy of the “ $m$ ” number of wavelet coefficients  $C_i(n)$  at each resolution scale “ $n$ ” of the signal is written as:

$$Energy(n) = \sum_{i=1}^m |C_i(n)|^2 \tag{1}$$

$C_i(n)$  is the  $i^{th}$  wavelet coefficient of  $n^{th}$  scale.

Shannon entropy that measures the uncertainty of signal wavelet coefficients is defined by:

$$Shannon\ Entropy(n) = - \sum_{i=1}^m p_i \log p_i \tag{2}$$

where  $p_i$  is the distribution of the energy probability for each wavelet coefficient with  $\sum_{i=1}^m p_i = 1$ , and in the case of,  $p_i = 1$  for some  $i$  the value of  $p_i \log p_i$  is taken as zero, given by:

$$p_i = \frac{|C_i(n)|^2}{Energy(n)} \tag{3}$$

There are some situations when the wavelet family will not satisfy both criteria. To resolve such conflict, the *maximum energy to Shannon entropy ratio*  $\eta(n)$  will be used [37, 38].

$$\eta(n) = \frac{Energy(n)}{Shannon\ Entropy(n)} \tag{4}$$

The best wavelet family is selected in this case that has the maximum energy to Shannon entropy ratio. It has the same principle for decomposition level selection.

**3.2 Time domain features extraction**

In order to characterize the time information of gears and bearings defects signals, we propose to use the time domain statistical features because of their simplicity of implementation and the computational time reduction.

**3.3 Input data normalization**

Input data value normalization between [0, 1] using the linear mapping improves the learning speed and the sensitivity of the neural network. In most cases, the input data values mapped were between  $-1$  and  $1$  for the *hyperbolic tangent function* and between  $0$  and  $1$  for the *Sigmoid function* using the following linear mapping generalized formula:

**Table 1** Time domain feature extraction

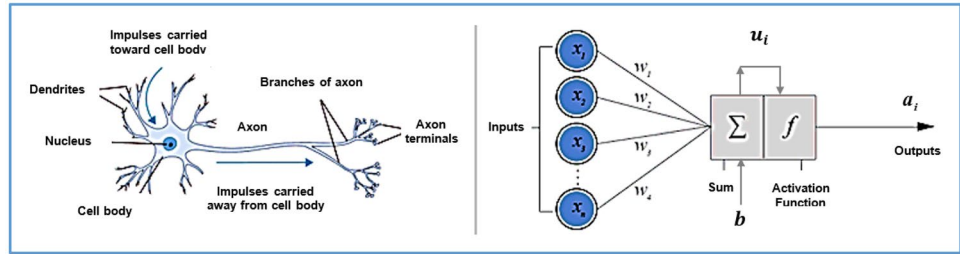
Features Notation	Features Description	Features Equation	Features Notation	Features Description	Features Equation
$TF_1$	Peak Value	$Max S(n) $	$TF_{11}$	Median	$\frac{n+1}{2}$
$TF_2$	Root Mean Square	$\sqrt{\frac{1}{N_e} \sum_{n=1}^{N_e} (S(n))^2}$	$TF_{12}$	Kurtosis Factor	$\frac{Max S(n) }{\sqrt{\frac{1}{N_e} \sum_{n=1}^{N_e} (S(n))^2}}$ $\left(\frac{1}{N_e} \sum_{n=1}^{N_e} \sqrt{ S(n) }\right)^2$
$TF_3$	Square Root Mean	$\left(\frac{1}{N_e} \sum_{n=1}^{N_e} \sqrt{ S(n) }\right)^2$	$TF_{13}$	Clearance Factor	$\frac{Max S(n) }{\left(\frac{1}{N_e} \sum_{n=1}^{N_e} \sqrt{ S(n) }\right)^2}$
$TF_4$	Absolute Mean Value	$\frac{1}{N_e} \sum_{n=1}^{N_e}  S(n) $	$TF_{14}$	Mean Value	$\frac{1}{N_e} \sum_{n=1}^{N_e} S(n)$
$TF_5$	Kurtosis	$\frac{\frac{1}{N_e} \sum_{n=1}^{N_e} (S(n) - \bar{S})^4}{\left[\frac{1}{N_e} \sum_{n=1}^{N_e} (S(n) - \bar{S})^2\right]^2}$	$TF_{15}$	Standard Deviation	$\frac{1}{(N_e - 1)} \sum_{n=1}^{N_e} (S(n) - \bar{S}(n))^2$
$TF_6$	Crest Factor	$\frac{Max S(n) }{\sqrt{\frac{1}{N_e} \sum_{n=1}^{N_e} (S(n))^2}}$	$TF_{16}$	Signal to Noise Ratio	$\frac{\frac{1}{N_e} \sum_{n=1}^{N_e}  S(n) }{\frac{1}{(N_e - 1)} \sum_{n=1}^{N_e} (S(n) - \bar{S}(n))^2}$
$TF_7$	Skewness	$\frac{N_e}{(N_e - 1)(N_e - 2)} \sum_{n=1}^{N_e} \left(\frac{S(n) - \bar{S}}{STD}\right)^3$	$TF_{17}$	Signal Energy	$\sum_{n=1}^{N_e} [S(n)]^2$
$TF_8$	Impulse Factor	$\frac{\max S(n) }{\frac{1}{N_e} \sum_{n=1}^{N_e}  S(n) }$	$TF_{18}$	Signal Power	$\frac{1}{N_e} \sum_{n=1}^{N_e}  S(n) $
$TF_9$	Margin Factor	$\frac{\max S(n) }{\left(\frac{1}{N_e} \sum_{n=1}^{N_e} \sqrt{ S(n) }\right)^2}$	$TF_{19}$	Variance	$\frac{1}{N_e} \sum_{n=1}^{N_e} (S(n) - \bar{S}(n))^2$
$TF_{10}$	Shape Factor	$\frac{\sqrt{\frac{1}{N_e} \sum_{n=1}^{N_e} (S(n))^2}}{\frac{1}{N_e} \sum_{n=1}^{N_e}  S(n) }$			

$$x_{ND} = \frac{(x_{SD} - x_{SDmin})(V_{SDmax} - V_{SDmin})}{(X_{SDmax} - X_{SDmin})} + V_{SDmin} \tag{5}$$

'  $x_{ND}$  is the normalized data value for the extracted feature,  $V_{SDmin} = 0$  or  $-1$  and  $V_{SDmax} = 1$  are the minimum and maximum values of the normalization interval.  $x_{SDmax}$  and  $x_{SDmin}$  are respectively the minimum and the maximum values in the feature vector in the input data matrix. The sigmoid function is most useful for training a set of data that is saturated between 0 and 1.

The last step before classification is *feature selection* which is the most important task to get a high-performance intelligent fault diagnosis system; after the feature extraction step, we will have a large input vector or matrix with irrelevant and redundant information and this too many features can result in a long time calculation caused by a large dimensionality due to the fact that the number of training samples increases exponentially with the number of features in order to have an accurate training of the neural network model. Several kinds of research took into consideration this

**Fig. 2** Biological neuron versus McCulloch and Pitts’ artificial neuron model



issue and proposed many procedures for feature selection, such as the genetic algorithms (GAs) [39], decision tree [40], Fisher discriminant analysis (FDA) [41], and the principal component analysis (PCA) [42] to optimize and reduce the dimension of the input data. For this study, the process of features selection is carried out manually, and the basis for the selection is the neural network performance using the method of elimination. This means that we eliminate an indicator and we will notice its effect on the network performance. The features selection process is done manually, due to its minor effect on the network’s outputs. Therefore, we used all the 19 features in Table 1.

Automatic classification (or discrimination) is an area of application of neural networks based on the specific (selected) features; therefore, the last step is fault classification using artificial neural networks, which will be illustrated in the next section.

**3.4 Pattern recognition neural networks**

The main idea of using artificial neural network ANN which is a subset of artificial intelligence in condition-based main-

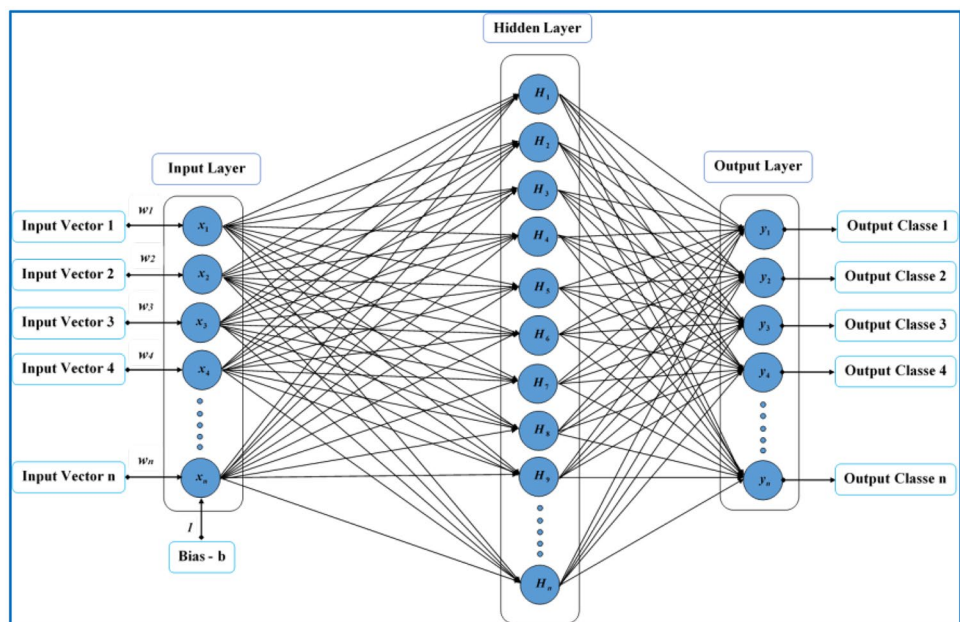
tenance is to create an expert system to simulate a simple model of the human brain with maintenance field experience in order to use it as a monitoring system for mechanical faults diagnosis. ANN is a high-value and low-cost tool in modeling, simulation, automation, condition monitoring, and fault diagnosis of different systems including all types of rotating machinery. ANNs have a variety of applications such as prediction and validation, clustering and optimization, classification and pattern recognition, function approximation, and time series analysis.

The basic element of neural networks is the artificial neuron (nodes), as presented in Fig. 2 [43].

In Fig. 2, each neuron consists of two functions: *net function*  $\sum(x_i)$  and *activation function*  $f(u_i)$ .

The net function provides the weighted linear sum of the input vectors ( $x_i : 1 \leq i \leq N$ ), with an additional constant term called *bias* and noted ( $b$ ) (or threshold) its connection weighted by one (1). It is mathematically given [44, 45] by:

**Fig. 3** The artificial neural network typical architecture



$$u_i = \sum_{i=1}^N w_i x_i + b \tag{6}$$

$w_i : 1 \leq i \leq N$ : The *synaptic weights* represented the effects of the synapses.

The output of the neuron  $a_i$  is related to the network input  $u_i$  via a linear or nonlinear transformation noted  $f$  and called activation (transfer) function, mathematically given by:

$$a_i = f(u_i) = f\left(\sum_{i=1}^N w_i x_i + b\right) \tag{7}$$

Several activation functions (linear and non-linear) are used in the ANN applications, such as Sigmoid function, which is used in this study, hyperbolic tangent, inverse tangent, threshold, Gaussian radial basis (exponential function), used for radial basis neural network, and the linear function.

As shown in Fig. 3, ANN basic or the linear architecture consists of 03 layers of neurons: input, hidden, and output layers. A great variety of network topologies can be imagined, under the sole constraint that the graph of connections is acyclic. However, the vast majority of neural network applications implement multilayer networks.

## 4 Experimental part

### 4.1 Experimental setup and data acquisition

We presented in this part the test rig configuration, the experimental setup for data acquisition and collection, and the number and the details of all the experiments carried out

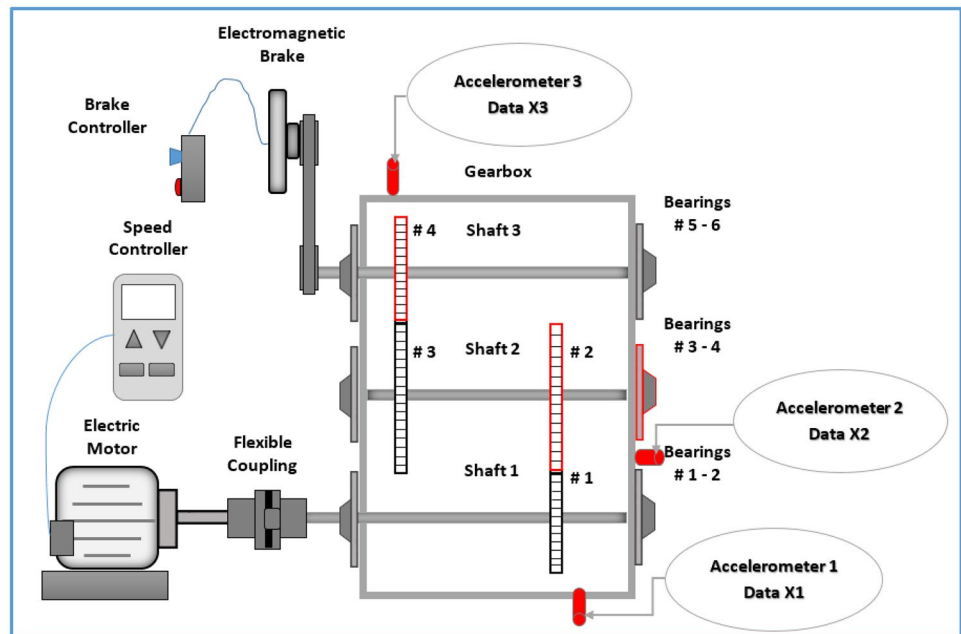
on the test rig. All data used in this work are collected from a test rig (Fig. 4), constructed in the laboratory of mechanics and structures LMS, in the University 8 May 1945 Guelma. The test rig is composed of an electrical motor controlled by a speed regulator which allows the choice of rotational frequency/speed of the engine (Fig. 5); this regulator can order engines of various powers until 4 Kw/50 Hz, and the rotational motion has been transmitted to the gearbox drive shaft using a flexible coupling. The gearbox by its turn is composed of 04 spur gears mounted on three shafts (drive/middle/driven), guided in rotation by 06 housing bearing type 6002, fixed by bolts on a steel box, a structure made up of steel plates. The top face is made up of Plexiglas allowing us to create the gears’ faults even we can change the gears and put their lubrication (lubricant oil). In order to apply a load on the gearbox’s transmission, an electromagnetic brake connected to the third shaft in the gearbox by a geared belt is used to generate a torque which is considered a load. The four gears (1–2–3–4) that composed the gearboxes have a mechanical module of 2 mm and have respectively 42, 50, 65, and 46 teeth. Gears 2 and 4 are the tested wheels.

The vibration data were acquired under the same load which is generated by the brake and two different motor speeds 900 and 1500 r/min. The sampling frequency is 32768 Hz, and the number of data points in each sample was 16,384 Hz and time recordings of 1-s duration.

Twenty (20) sets of data were obtained from the experimental setup; more details about the experiments will be presented in the next section.

All the experiments carried out in the following conditions:

Fig. 4 LMS laboratory test rig for gears and bearing fault simulation





**Fig. 5** (a) Experiment's test rig, (b) PULSE system for signal acquisition, post-processing, and data analysis type Brüel & Kjær 18th version with four (04) inputs and two (02) output channels. All the experiments are carried out in the following conditions: Driveshaft

- Drive shaft rotational frequencies: 15 and 25 Hz;
- Frequency bandwidths: [0 – 6400 Hz] and [0 – 12800 Hz];
- Operation Mode: in the presence of load and grease.

We present in Table 2 the detailed experimental plan, including the following:

rotational frequencies, 15 and 25 Hz; frequency bandwidths, [0 – 6400 Hz] and [0 – 12,800 Hz]; operation mode: in the presence of load and grease

- I - Gears and bearings *simple* faults
- II - Gears and bearings *combined* faults

To create different gears and bearings defects, we have removed a part from the tooth contact surface (small defect: 15% of tooth contact surface material was removed which

**Table 2** Experiments plan

Number of experiments	Gears and bearings faults combinations	Code
<b>I. Gears and bearings simple faults</b>		
1	Without faults	WF
2	Bearing small fault on bearing 2	BSF2
3	Bearing mean fault on bearing 2	BMF2
4	Bearing large fault on bearing 2	BLF2
5	Gear small fault on gear 2	GSF2
6	Gear mean fault on gear 2	GMF2
7	Gear large fault on gear 2	GLF2
<b>II. Gears and bearings combined faults</b>		
8	Gear Small Fault on gear 2 Combined with bearing small fault on bearing 2	GSF2 + BSF2
9	Gear small fault on gear 2 Combined with bearing mean fault on bearing 2	GSF2 + BMF2
10	Gear small fault on gear 2 Combined with bearing large fault on bearing 2	GSF2 + GLF2
11	Gear mean fault on gear 2 Combined with bearing small fault on bearing 2	GMF2 + BSF2
12	Gear mean fault on gear 2 Combined with bearing mean fault on bearing 2	GMF2 + BMF2
13	Gear mean fault on gear 2 Combined with bearing large fault on bearing 2	GMF2 + GLF2
14	Gear large fault on gear 2 Combined with bearing small fault on bearing 2	GLF2 + BSF2
15	Gear large fault on gear 2 Combined with bearing mean fault on bearing 2	GLF2 + BMF2
16	Gear large fault on gear 2 Combined with bearing large fault on bearing 2	GLF2 + GLF2



is equivalent to 0.3 mm in the module, the same for mean defect:  $30\% = 0.6$  mm and the large defect:  $45\% = 0.9$  mm) depending on the experiment needed. The relationship and the difference between the artificial defects and the defects formed in the actual working conditions are explained next.

#### Bearing defects

- Bearing small fault (BSF) is a small groove (indentation) made through the ball's raceway on the outer race of bearing 2, using a high-speed electric grinder, actually simulating small chipping.
- The same goes for bearing mean fault (BMF) and bearing large fault (BLF) respectively, just increasing the width of the groove (indentation).

#### Gears defects

- The gear small fault (GSF) is the tearing part of the material all along the contact surface of a tooth of the gear 2 using a high-speed electric grinder, simulating progressive wear on the gear teeth.
- The same goes for the gear mean fault (GMF) and gear large fault (GLF) respectively, just increasing the amount of material torn from the gear tooth surface contact.

## 5 Results and discussion part

In this section, we applied the proposed approach to the measured signals, following different steps described in Sect. 2. We started by wavelet packet transform optimization and features extraction, followed by the classification process using pattern recognition neural networks; the latter is divided into 04 networks (ANN1, ANN2, ANN3, and ANN4) as shown in the algorithm in Fig. 7) in which ANN1 is dedicated to whether the system components (gears and bearings) are faulty or healthy. ANN2 and ANN3 are used to classify gears *or* bearings *simple* faults after we confirm that one of the components (gears *or* bearings) is faulty using ANN1. ANN4 is used to classify gears *and* bearings *combined* faults after we confirm that both gears *and* bearings are faulty using ANN2 and ANN3. Another network named ANN5 is used to generalize the classification process including all previous networks for whole system classification.

### 5.1 Wavelet family and decomposition level optimization

Energy to Shannon entropy ratio was used to select an appropriate mother wavelet, and the vibration signals in different

conditions are decomposed at the *sixth* level using WPT where both total Shannon entropy and the ratio total energy to total Shannon entropy are calculated for each wavelet family and decomposition level; then the level and the wavelet having maximum energy to Shannon entropy ratio are considered for gear and bearing fault diagnosis.

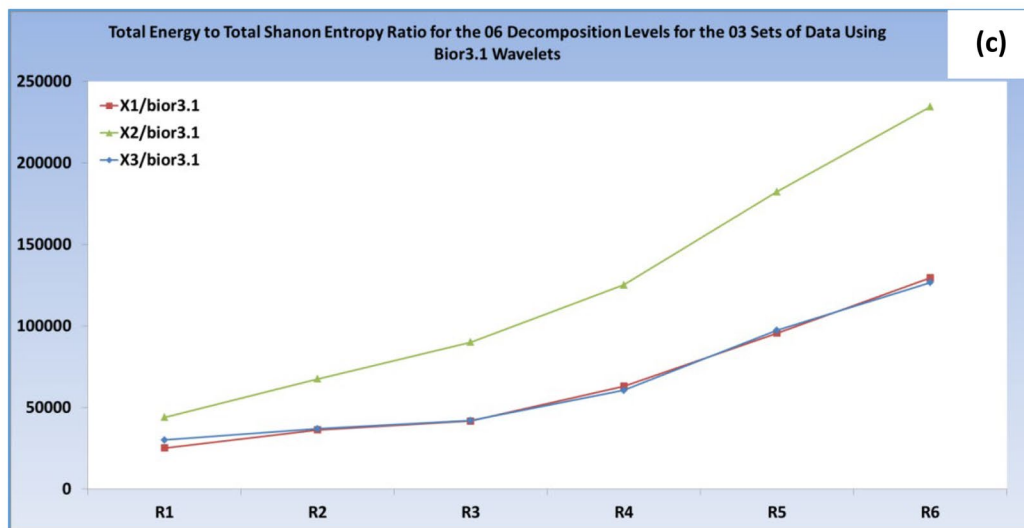
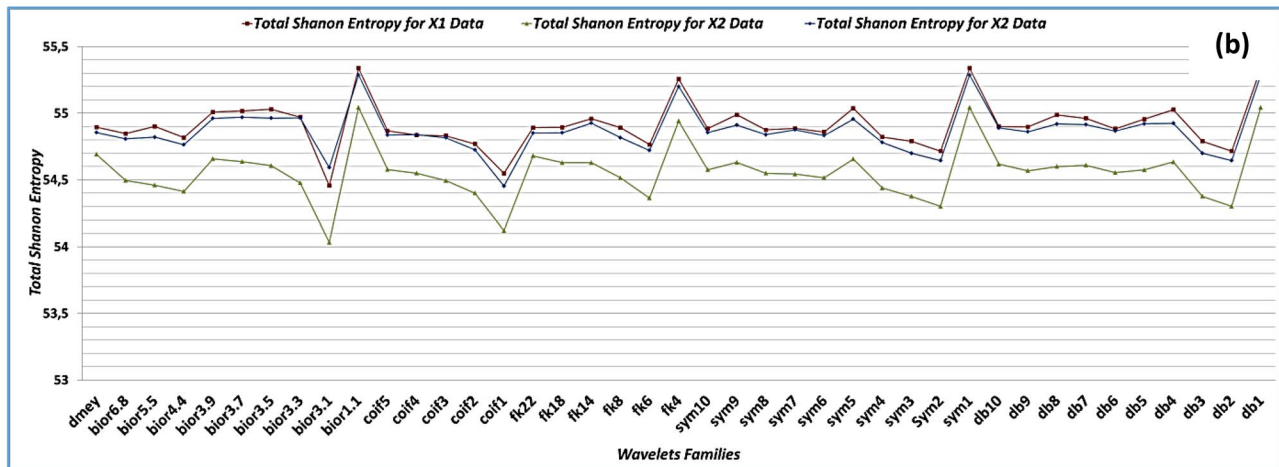
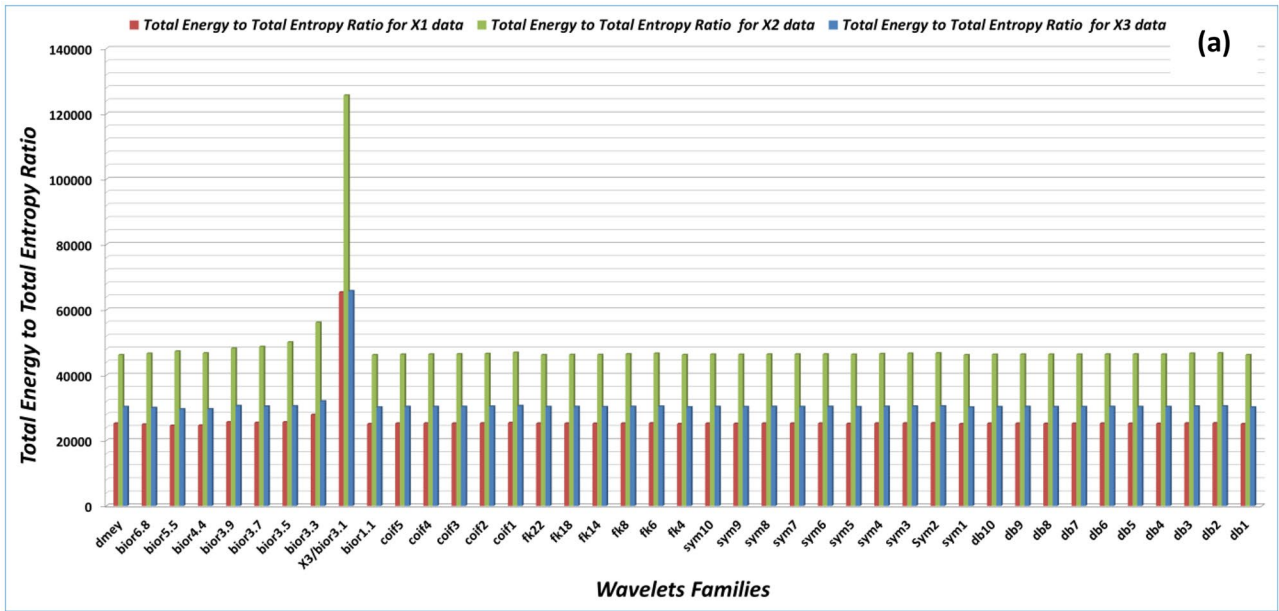
From Fig. 6a, we find that the wavelet  $\psi$  that has the maximum energy to Shannon entropy ratio is the “*bior3.1*,” and to confirm that we pass to the second (Fig. 6b), we remark that the wavelet which has the minimum entropy is the same wavelet family which is *bior3.1*. However, we conclude that the best mother wavelet  $\psi$  to be selected for decomposition process is the *Bi-orthogonal* (“*bior3.1*”), and the same for the level of decomposition from the third (Fig. 6c), we remark that the decomposition level that has the maximum energy to Shannon entropy ratio is the last one which is the *6<sup>th</sup> level*; thus, the best decomposition level that will be selected is the *sixth* (6) *level*. For more accuracy, we applied the above criteria on all data that has been acquired from the 03 sensors on the test rig for 6400 Hz as maximal frequency; thus X1, X2, and X3 are respectively the data acquired from the sensors or accelerometers 1, 2, and 3.

The feature vector is a matrix that contains all the extracted features. If the number of features extracted from one signal is 19 as mentioned in Table 1, and each signal will be decomposed at the 6 levels with “*bior3.1*” wavelet packet, and the chosen level is the 6<sup>th</sup> with 64 coefficients, then the number of features, in this case, will be  $19 \times 64$ . If we have 10 signals in each class, the matrix dimension of these features will be  $19 \times 64 \times 10$ , which will be  $19 \times 640$  for each class, such as we took 40 samples from each class for tests. In this next section, we present the results of classification process with different faults classes in the case of simple and/or combined faults. As shown by the algorithm presented in Fig. 8, we generally have four (04) networks: ANN1, ANN2, and ANN3 with 03 classes for each, and ANN4 with 09 classes, in addition to ANN5 that regroups all the previous networks. The details of each network will be presented next in Tables 3, 4, and 5.

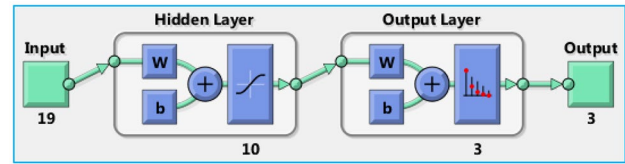
### 5.2 Gears and bearing fault classification process using PRNN

The classification process for all the networks was done using the Pattern Recognition Neural Network toolbox, in the MATLAB software environment. In Fig. 7, we represent the algorithm used for gears and bearings simple/combined fault classification and the detailed classes for each neural network.

The architecture of the artificial neural network is as follows: network type: forward neural network trained with feed-forward backpropagation; transfer function: Sigmoid



**Fig. 6** (a) Total energy to total Shannon entropy ratio of 41 wavelets families with the acquired data from the 03 sensors X1, X2, and X3. (b) Total Shannon entropy for 41 wavelets families. (c) Total energy to total Shannon entropy ratio for 06 decomposition level with the acquired data from the 03 accelerometers X1, X2, and X3



**Fig. 8** ANN1, ANN2, and ANN3 architecture

transfer function in hidden and output layer; training function: TRAINSCG (scaled conjugate gradient); adaption learning function: LEARNGDM; performance function: cross-entropy (CE); number of hidden layers: 02; and 10 neurons in the hidden layer for ANN1, ANN2, and ANN3, and 30 neurons in the hidden layer for ANN4 and ANN5. All these pieces of information are summarized next in Table 6.

**NB** The number of neurons in the hidden layer has been chosen manually based on the number of classes (input data) and the different ANN performance, which will be increased with the increase in the number of neurons in the hidden layer until 30; then, it will be decreased.

**5.2.1 Gears and bearing simple faults classification**

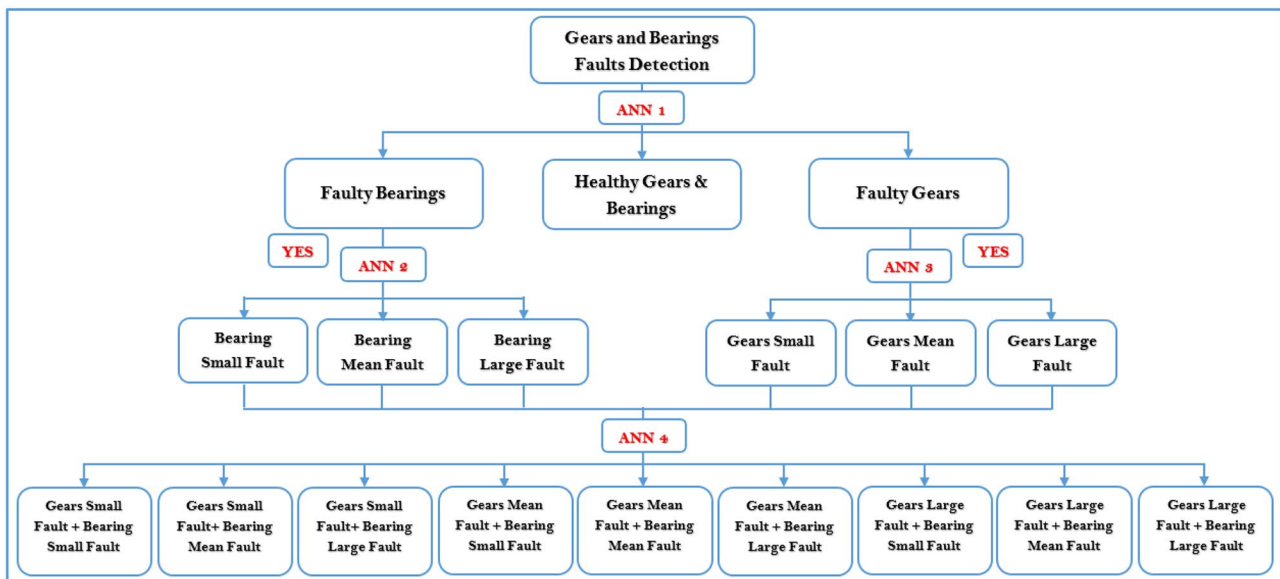
ANN1, ANN2, and ANN3 are 03 networks that have the same architecture, which is presented in Fig. 8. ANN1 is used to check the status of the system whether it is healthy or with a faulty gear and/or bearing, without providing information on the type and/or size of the defect. The

classes used in this network are (a) healthy gears and bearings (without faults) (WF), (b) faulty bearing (BMF2), and (c) faulty gear (GMF2). After confirming the presence of faulty gears or bearing using ANN1, we pass to ANN2 to classify the simple faults of bearing or ANN3 to classify the simple faults gears, with information on the fault sizes. The classes that have been used in these networks are (a) BSF2, (b) BMF2, and (c) BLF2 for ANN2 and (d) GSF2, (e) GMF2, and (f) GLF2 for ANN3.

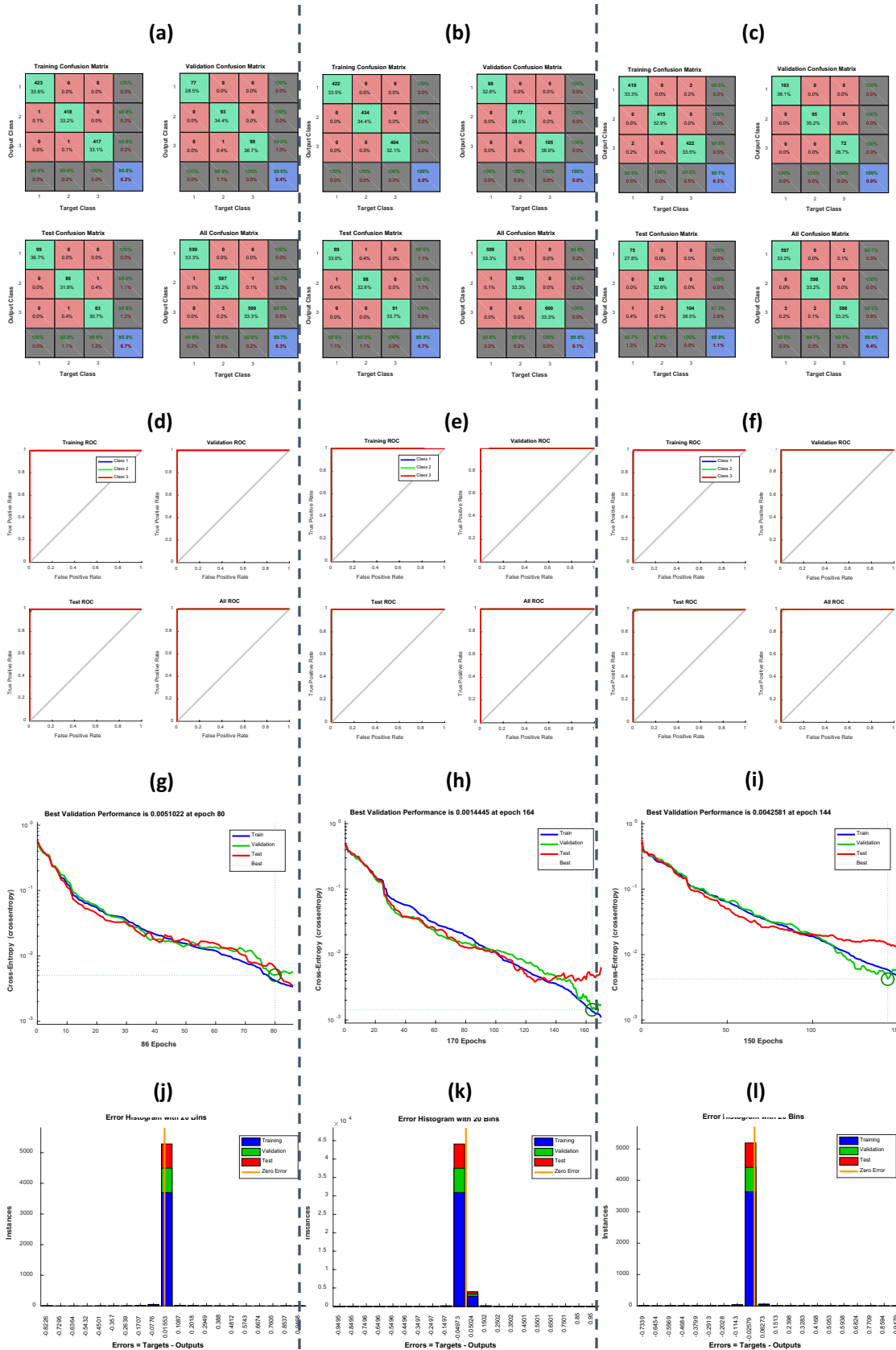
As per default configuration in the pattern recognition neural network toolbox 9.0, the input data samples for ANN1, ANN2, and ANN3 are divided as below:

We have 1800 samples divided into the following:

- Training set with 1260 samples (70%) used by the network during training, and the network is adjusted according to its error.
- Testing set with 270 samples (15%) provide an independent measure of network performance during and after training.



**Fig. 7** Flow chart for gears and bearings simple/combined faults detection using PRNN



◀**Fig. 9** ANN1, ANN2, and ANN3 classification process results, respectively: (a), (b), and (c) Confusion matrix; (d), (e), and (f) Receiver operating characteristic or ROC plot; (g), (h), and i Performance, and (j), (k), and (l) Error histogram

- Validation set with 270 samples (15%) used to measure network generalization and to stop training when generalization stops improving.

Figure 9 represents the classification process results for ANN1, ANN2, and ANN3. In Fig. 9a, b, and c, we find the confusion matrices for training, testing, and validation and their combination. From the combination of the confusion matrices, we can see that the performance of the 03 ANNs is quite high with, 99.9%, 99.7%, and 99.6% respectively for ANN1, ANN2, and ANN3, which could justify the accuracy of the classification process.

In Fig. 9d-f, we presented the receiver operating characteristic (ROC) graphs of training, test, validation, and their combination for ANN1, ANN2, and ANN3.

We remark in the combination of all ROC graphs for all 03 ANNs that all the classes are in the point (0, 1). This means (0 false-positive rate and 1 true-positive rate) that for all classification process, AUC = 1 represents perfect classification by almost 100% of the true-positive rate values of the 03 classes, which validate the results of the confusion matrices.

In Fig. 9g-i, we presented the ANN1, ANN2, and ANN3 networks' total performance plots, which are equal to 0.0051022 at epoch 80 for ANN1, 0.0014445 at epoch 164 for ANN2, and 0.0042581 at the epoch 144 for ANN3, and they are all converging to zero representing a very good performance, due to the minority of the ANN's errors.

In Fig. 9j-l, we presented the ANN1, ANN2, and ANN3 networks' error histogram, which provides accurate infor-

mation about its error. The main data that can be extracted from the error histogram is the *error value* and the *error frequency*. The negative sign of an error means that the outputs are greater than their targets. As shown in the error histogram for the 03 ANNs, the training, test, and validation data sets, which represent respectively 70%, 15%, and 15% of all input data, their errors are almost zero for ANN1 and ANN3 or converged to zero and 20 bins for ANN2. This means they have a high classification accuracy and a high degree of generalizability.

### 5.2.2 Gears and bearing combined fault classification

In this sub-section, we tried to classify only the combined faults to see the effect of the proposed approach on the classification process. We used ANN4 with the architecture presented in Fig. 10, to classify the 09 classes of gears and bearings combined faults.

We have 5400 samples divided into the following:

- Training set with 3780 samples (70%) used by the network during training, and the network is adjusted according to its error.
- Validation set with 810 samples (15%) used to measure network generalization and to stop training when generalization stops improving.
- Testing set with 810 samples (15%) provide an independent measure of network performance during and after training.

In Table 4, ANN4 is used to find out the combined faults of gears and bearings, with information on the type and size of faults, and the classes used in this network are (a) BSF2 + GSF2, (b) BMF2 + GSF2, (c) BLF2 + GSF2, (e)

**Table 3** Classification for gears or bearing *simple faults*

ANN1 to confirm the presence /absence of gear and bearing faults			
Pattern number	Faulty component	Gears and bearing status	Patterns names/name of classes
01	N/A	Without faults (healthy gears and bearings)	Class 1
02	Bearings	With faults (faulty bearings)	Class 2
03	Gears	With faults (faulty gears)	Class 3
ANN2 for bearing simple fault classification			
Pattern Number	Faulty component	Size of bearing simple faults	Patterns names/name of classes
01	Intermediate shaft bearing #2	Bearing small fault (BSF)/outer race fault	Class 1
02	Intermediate shaft bearing #2	Bearing mean fault (BMF)/outer race fault	Class 2
03	Intermediate shaft bearing #2	Bearing large fault (BLF)/outer race fault	Class 3
ANN3 for gears simple faults classification			
Pattern number	Faulty component	Size of gears simple faults	Patterns names/name of classes
01	Intermediate shaft gear #2	Gear small fault on gear 2 (GSF2)/worn tooth	Class 1
02	Intermediate shaft gear #2	Gear mean fault on gear 2 (GMF2) /worn tooth	Class 2
03	Intermediate shaft gear #2	Gear large fault on gear 2 (GLF2) /worn tooth	Class 3

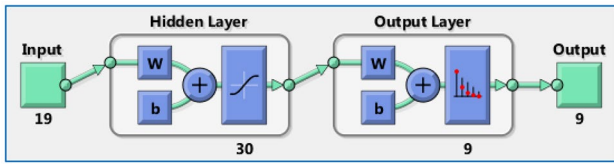
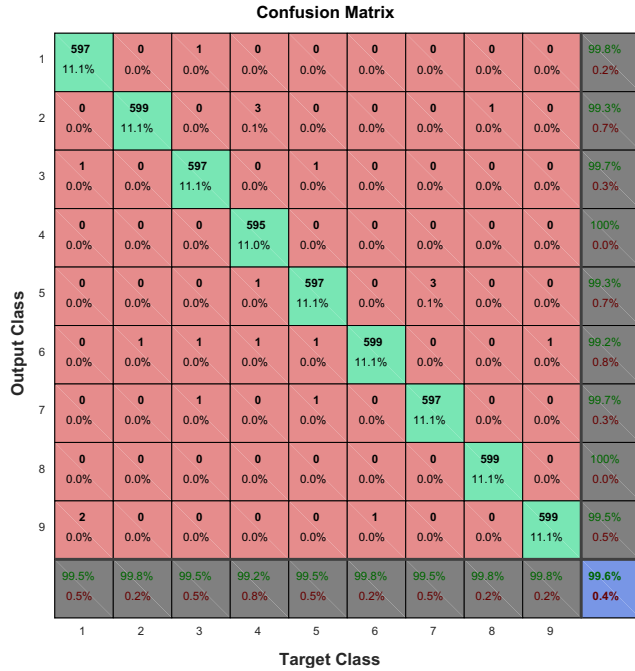


Fig. 10 ANN4 architecture

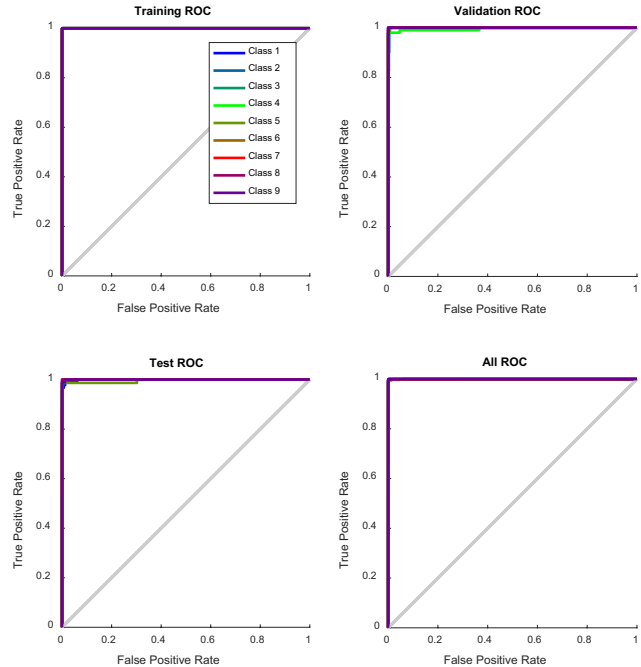
BSF2 + GMF2, (f) BMF2 + GMF2, (g) BLF2 + GMF2 (h) BSF2 + GLF2, (i) BMF2 + GLF2, and (j) BLF2 + GLF2.

ANN4 configuration which is presented in Fig. 10 shows three layers: the input layer with 19 features, the hidden layer with 30 neurons inside, and the output layer with 09 outputs.

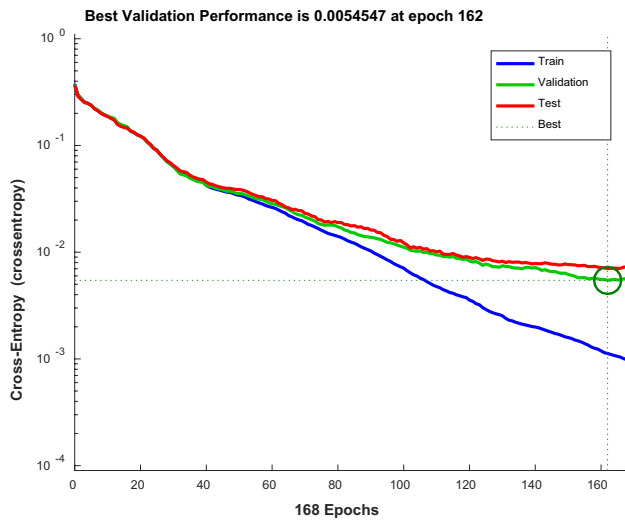
Figure 11 represents the results of the 4th neural network ANN4 with 30 neurons in the hidden layer (Fig. 11a). We find the confusion matrix for the combination of training, testing, and validation matrices. The network outputs are very accurate with a high



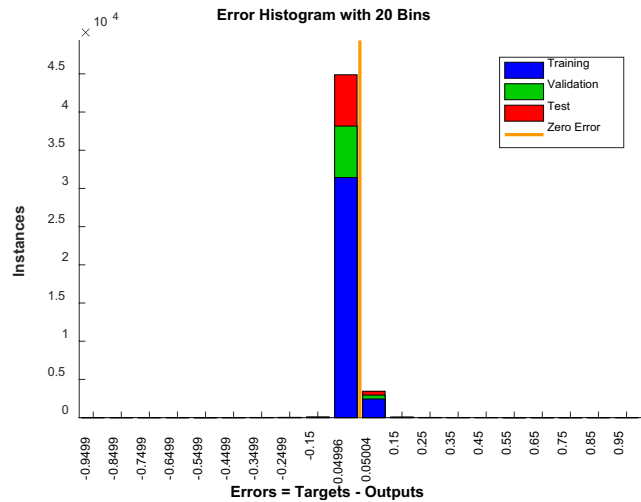
(a)



(b)



(c)



(d)

Fig. 11 ANN4 classification plots: (a) Confusion matrix, (b) receiver operating characteristic or ROC plot, (c) ANN4 performance, and d error histogram

**Table 4** Classification for gear and bearing *combined faults*

Pattern number	Faulty component	Type of gear and bearing combined fault	Patterns names/name of classes
01	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing small fault (BSF2) + gear small fault (GSF2)	Class 1
02	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing mean fault (BMF2) + gear small fault (GSF2)	Class 2
03	Intermediate shaft bearing #2, Intermediate shaft gear #22	Bearing large fault (BLF2) + gear small fault (GSF2)	Class 3
04	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing small fault (BSF2) + gear mean fault (GMF2)	Class 4
05	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing mean fault (BMF2) + gear mean fault (GMF2)	Class 5
06	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing large fault (BLF2) + gear mean fault (GMF2)	Class 6
07	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing small fault (BSF2) + gear large fault (GLF2)	Class 7
08	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing mean fault (BMF2) + gear large fault (GLF2)	Class 8
09	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing large fault (BLF2) + gear large fault (GLF2)	Class 9

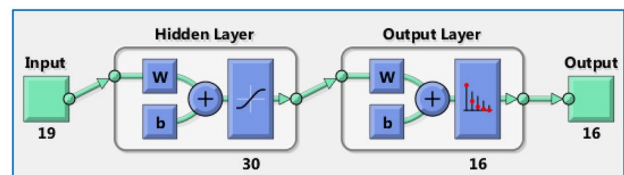
performance of 99.6%. The total confusion matrix indicates that 5379 samples from 5400 samples of the input data set have been correctly classified and only 21 samples from them are misclassified. That means the ANN4 is highly generalizable and can be used for different sizes of gears and bearings combined faults diagnosis and decision-making in the future. Figure 11b presents the *receiver operating characteristics* (ROC) graphs for 09 classes of the ANN4: training, test, validation, and their combination, which are useful for organizing classifiers and visualizing their performance. We remark in the combination of all ROC graphs that all the classes are almost in the point (0, 1) with AUC almost 1. That means 0 false-positive rates and 1 true-positive rate for all classification processes which represent perfect classification by almost 100% of the true positive rate values of the 09 classes, which validate the results of the confusion matrices.

Figure 10c presents the ANN4 network’s total performance which is equal to 0.0054547 at epoch 162 and is converging to zero which represents a very good performance with a minor error. We remark that all plots are descending to epoch 162 and converging to the 0.0054547 error. Figure 11d presents the ANN4 network’s error histogram, which evaluates the error distributions based on the ANN4 classification results. It is shown in the distribution of errors that it occurred near-zero vertical line in the center and is gradually decreasing when moving away from it. This proves that ANN4 has a high classification accuracy and a high degree of generalizability.

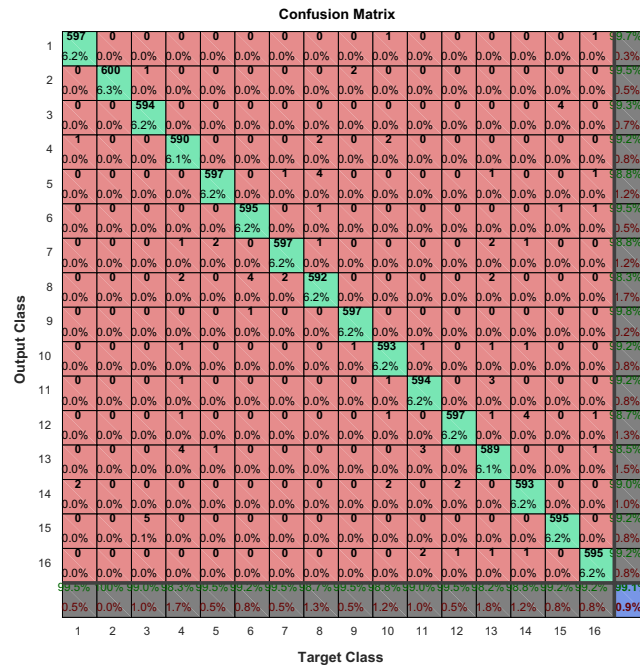
### 5.2.3 Generalization of the classification process on both simple and combined faults

The final step is to confirm if the proposed approach is able to classify the healthy case with simple and combined faults at the same time and with the same performance or not. As a general validation, we tried to classify the entire system conditions with the 16 classes, which are used before in ANN1...ANN4, in one neural network named ANN5 with the architecture presented in Fig. 12. For ANN5, we have 9600 samples divided into the following:

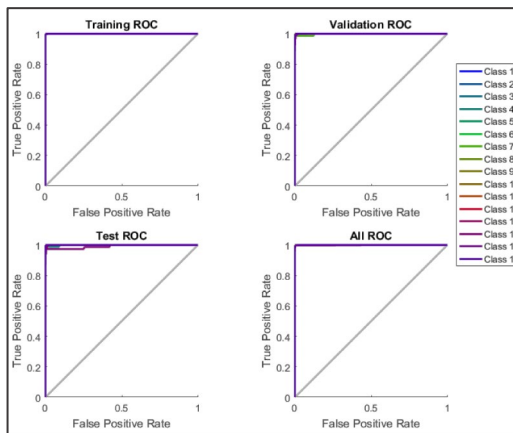
- Training set with 6720 samples (70%) used by the network during training, and the network is adjusted according to its error.
- Validation set with 1440 samples (15%) used to measure network generalization and to stop training when generalization stops improving.



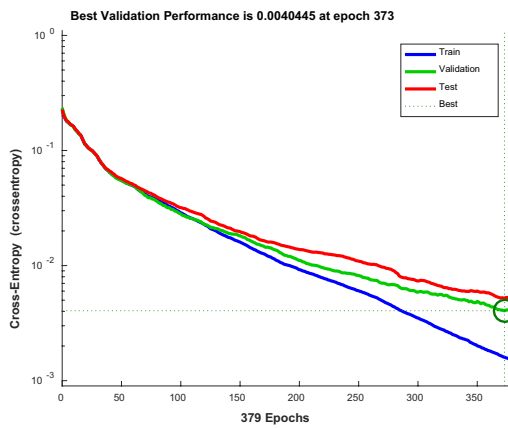
**Fig. 12** ANN5 Architecture



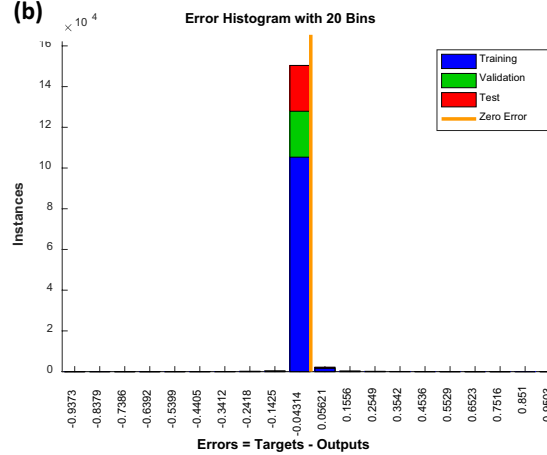
(a)



(b)



(c)



(d)

Fig. 13 ANN5 classification plots: (a) Confusion matrix, (b) Receiver Operating Characteristic or ROC plot, (c) ANN5 Performance, and (d) Error histogram



**Table 5** Gears and bearing condition (healthy, simple, and/or combined faults) classification

Pattern number	Faulty component	Type of gear and bearing simple fault	Patterns names/name of classes
01	N/A	Without faults (healthy gears and bearings)	Class 1
02	Intermediate shaft bearing #2	Bearing small fault (BSF)/outer race fault	Class 2
03	Intermediate shaft bearing #2	Bearing mean fault (BMF)/outer race fault	Class 3
04	Intermediate shaft bearing #2	Bearing large fault (BLF)/outer race fault	Class 4
05	Intermediate shaft gear #2	Gear small fault on gear 2 (GSF2)/worn tooth	Class 5
06	Intermediate shaft gear #2	Gear mean fault on gear 2 (GMF2) /worn tooth	Class 6
07	Intermediate shaft gear #2	Gear large fault on gear 2 (GLF2) /worn tooth	Class 7
08	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing small fault (BSF2) + gear small fault (GSF2)	Class 8
09	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing mean fault (BMF2) + gear small fault (GSF2)	Class 9
10	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing large fault (BLF2) + gear small fault (GSF2)	Class 10
11	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing small fault (BSF2) + gear mean fault (GMF2)	Class 11
12	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing mean fault (BMF2) + gear mean fault (GMF2)	Class 12
13	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing large fault (BLF2) + gear mean fault (GMF2)	Class 13
14	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing small fault (BSF2) + gear large fault (GLF2)	Class 14
15	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing mean fault (BMF2) + gear large fault (GLF2)	Class 15
16	Intermediate shaft bearing #2, Intermediate shaft gear #2	Bearing large fault (BLF2) + gear large fault (GLF2)	Class 16

**Table 6** Classification performance for the 05 neural networks and their networks properties

ANN properties and number	ANN 1	ANN 2	ANN 3	ANN 4	ANN 5
<b>Neural network application</b>	Pattern recognition neural networks for classification				
<b>Neural network connection</b>	Static (feed-forward)				
<b>Neural network topology</b>	Multi-layer perceptron (MLP)				
<b>Learning/training method</b>	Supervised				
<b>Number of neurons on the hidden layer*</b>	10	10	10	30	30
<b>Number of hidden layers</b>	1				
<b>Activation function</b>	Sigmoid				
<b>Performance error function</b>	Cross-entropy (CE)				
<b>Neural network performance</b>	<b>99.7%</b>	<b>99.9%</b>	<b>99.6%</b>	<b>99.7%</b>	<b>99.1%</b>

- Testing set with 1440 samples (15%) provide an independent measure of network performance during and after training.

ANN5 have the same configuration presented in the schematic in Fig. 12:

Table 5 represents the classes used in ANN5, which includes all previous classes used in ANN1, ANN2, ANN3 and ANN4: (a) Without faults (WF), (b) BSF2, (c) BMF2, (d) BLF2, (e) GSF2, (f) GMF2, (g) GLF2, (h) BSF2 + GSF2, (i) BMF2 + GSF2, (j) BLF2 + GSF2, (k) BSF2 + GMF2, (l) BMF2 + GMF2, (m) BLF2 + GMF2, (n) BSF2 + GLF2, (o) BMF2 + GLF2, (p) BLF2 + GLF2.

Figure 13 represents the results of the 16 classes, the 5th network ANN5 which has 30 neurons on the hidden layer (Fig. 13a). We find the confusion matrix of the combination of all training, testing, and validation matrices. We remark that the network outputs are very accurate with a high performance of 99.1%. The training matrix indicates that 9515 samples from 9600 samples of input data set have been correctly classified and only 85 samples from them are misclassified. That means the ANN5 is highly generalizable and can be used for any gears and bearings combined faults diagnosis and for faults detection and decision-making in the future.

In the same figure (Fig. 13b), the ROC graphs of the ANN5 are presented. We remark in the combination of all ROC graphs that all the classes are almost in the point (0, 1) for all classification processes which indicates perfect classification by almost 100% of the true positive rate values of the 16 classes, which validated the results of the confusion matrices. On the same figure (Fig. 13c), the ANN5 network's total performance is presented, which is equal to 0.0040445 at the epoch 373 and is converging to zero that represents a very good performance because the ANN5 error is minor. We remark that all plots are descending to epoch 373 and converging to the 0.0040445 error. In the same figure (Fig. 13d), the ANN5 network's error histogram is presented, which provides accurate information about its error. As shown in the error histogram figure, the errors are in the center and almost zero.

That means the ANN5 has a high classification accuracy and a high degree of generalizability.

\* The choice of the number of neurons on the hidden layer was made according to the number of classes, the size of the input data, and the network performance, which means that we put 10 neurons for the classification of simple faults with 03 classes, and 30 neurons for the combined faults with 09 classes, although we can fix the number of neurons on the hidden layer by 30 for all types of faults and all the networks, in order to avoid any problem during the classification process of the two types of faults (simple/combined) as we did in ANN5. We put 30 neurons on the hidden layer for the classification of the 16 classes including simple and combined faults.

### 5.3 Model generalization and approach validation

To validate our network performance and their classification process accuracy, we used 02 signals (1st signal was for gear simple fault and the 2nd signal was for gears combined faults) that were acquired from another test rig in the Laboratory of Vibration and Acoustic (LVA) of the institute of INSA Lyon- France, similar to the test rig used in our study with gear teeth numbers  $Z1 = 42$ ,  $Z2 = 50$ ,  $Z3 = 65$ , and  $Z4 = 46$  respectively for gears 1, 2, 3, and 4. The gear fault frequencies corresponding to shafts rotational speeds where the gears are mounted are as follows:

For LMS test rig: The gear simple faults (small, mean, and large) are simulated on gear 2 (11.78 Hz) and the gear combined faults on gear 2 (11.78 Hz) and gear 4 (16.93 Hz).

For LVA test rig: The simple faults (small, mean, and large) are simulated on gear 1 (14.45 Hz), and the gear combined fault on gear 1 (14.45 Hz) and gear 2 (11.78 Hz).

Gearmesh frequencies: 588 Hz for the first Gearmesh frequency of the reducer between gear 1 and gear 2 and 761 Hz for the multiplier, between gear 3 and gear 4.

The gear mean fault of the LVA test rig corresponds in size to gear large fault "GLF2" in our study and the gears combined fault (gear 1 small fault + gear 4 large fault) correspond to the combination of gear large fault "GLF2" and mean bearing fault "BMF2" in our study.

The two tested signals were classified in the corresponding class and the network classification performance was more than 97%. This result shows the efficiency and the robustness of the classification method that we have proposed.

## 6 Conclusions

In this paper, gears and bearing isolate/combined faults have been investigated using automated diagnosis, in a time where almost researches focused on gear or bearing simple fault diagnosis only.

- The WPT optimization using minimum Shannon entropy and the maximum energy to Shannon entropy ratio criteria gives us the *bior3.1* as the best wavelet family among 40 wavelets families, and the 6<sup>th</sup> level as the best decomposition level with 64 vectors.
- Five (05) neural networks, ANN1, ANN2, ANN3, ANN4, and ANN5, are used in this study. From Table 6, the performance of the five networks is shown as very high, which could justify the accuracy of the classification process; however, the use of pattern recognition neural network with the proposed optimization approach for gears and bearings combined fault automated diagnosis gave us encouraging results, validated by the results of INSA-LVA test rig.
- As a general conclusion, the model could be generalized to other types of faults in the rotary machine.

The following are the perspectives to this work:

- Frequency and time–frequency features will be included in the future, and the proposed approach will be applied to the acceleration signals acquired from a real machine in the manufacturing industry.
- Introduce deep learning (CNN) to combined faults automated diagnosis and develop an expert system in this subject.

**Acknowledgements** This study was completed at the University of Guelma, Algeria, and with financial support from the Algerian Ministry of Higher Education and Scientific Research MHESR (MESRS) and the Delegated Ministry for Scientific Research (MDRS).

## References

- Zani ML (2003) ‘‘Mesures mécaniques: surveillance des machines tournantes’’, guide d’achat, Mesures N° 757, pp.77–84, 2003
- McFadden PD, Smith JD (1984) Vibration monitoring of rolling element bearings by the high-frequency resonance technique — a review, *Tribology International*, Volume 17(1):3–10, ISSN 0301–679X. [https://doi.org/10.1016/0301-679X\(84\)90076-8](https://doi.org/10.1016/0301-679X(84)90076-8), Elsevier Science Ltd
- Wang W, Ismail F, Golnaraghi F (2001) ‘‘Assessment of gear damage monitoring techniques using vibration measurements.’’. *Mech Syst Signal Process*. 15:905–922. Elsevier Science Ltd.
- McFadden PD (1985) Low frequency vibration generated by gear tooth impacts, *NDT International*. 18(5):279–282, October 1985, Butterworth & Co (Publishers) Ltd
- Cempel, C. (1988). Vibroacoustical diagnostics of machinery: an outline. *Mech Syst Signal Process* 2(2):135-151. Elsevier Science Ltd
- Zheng Z, Petrone R, Péra MC, Hissel D, Béchérif M, Pianese C, Sorrentino M (2013) A review on non-model based diagnosis methodologies for PEM fuel cell stacks and systems. *International J Hydrogen Energy* 38(21):8914-8926
- Samanta B, Al-Balushi KR (2003) ‘‘Artificial neural network based fault diagnostics of rolling element bearings using time-domain features.’’ *Mech Syst Signal Process* 17(2):317–328. Elsevier Science Ltd
- Saravanan N, Ramachandran KI (2010) Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert Syst Appl* 37:4168–4181. Elsevier Science Ltd
- Paya BA, Esat II, Badi MNM (1997) Artificial neural networks based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mech Syst Signal Process* 1:751–765. Elsevier Science Ltd
- Sorsa T, Koivo H (1993) Application of artificial neural networks in process fault diagnosis. *Automatica* 29:843–849
- Liu Z, Cao H, Chen X, He Z, Shen Z (2013) Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing* 99:399–410, Elsevier Science Ltd
- Hossein Abadi HZ, Amirfattahi R, Nazari B, Mirdamadi HR, Atashpour SA (2014) GUV-based structural damage detection using WPT statistical features and multiclass SVM. *Appl Acoust* 86:59–70. Elsevier Science Ltd.
- Li N, Zhou R, Hu Q, Liu X (2012) Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine. *Mech Sys and Signal Process* 28:608–621. Elsevier Science Ltd
- Yu Yang, Dejie Yu, Junsheng Cheng, A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM, *Measurement* 40 (2007) 943–950. Elsevier Science Ltd.
- Changqing Shen, Dong Wang, Fanrang Kong, Peter W. Tse. Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, *Measurement* 46 (2013) 1551–1564, Elsevier Science Ltd.
- Junsheng C, Dejie Y, Yu Y (2007) Application of an impulse response wavelet to fault diagnosis of rolling bearings. *Mech Syst Signal Process* 21:920–929. Elsevier Science Ltd
- Yan R, Gao RX, Chen X (2014) Wavelets for fault diagnosis of rotary machines: a review with applications, *Signal Process* 96:1–15. Elsevier Science Ltd
- Liu W, Tang B (2011) A hybrid time-frequency method based on improved Morlet wavelet and auto terms window. *Expert Syst Appl* 38:7575–7581. Elsevier Science Ltd
- Yang WX (2007) A natural way for improving the accuracy of the continuous wavelet transforms, *J Sound Vibration* 306:928–939. Elsevier Science Ltd
- Sanz J, Perera R, Huerta C (2012) Gear dynamics monitoring using discrete wavelet transformation and multi-layer perceptron neural networks, *Appl Soft Comput* 12:2867–2878. Elsevier Science Ltd
- Moumene I, Ouelaa N (2016) Application of the wavelets multiresolution analysis and the high-frequency resonance technique for gears and bearings faults diagnosis. *Int J Adv Manuf Technol* 83:1315–1339. <https://doi.org/10.1007/s00170-015-7436-0>
- Djebala A, Ouelaa N, Hamzaoui N (2008) Detection of rolling bearing defects using discrete wavelet analysis. *Meccanica* 43:339–348. <https://doi.org/10.1007/s11012-007-9098-y>
- Hu Q, He Z, Zhang Z, Zi Y (2007) Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble, *Mech Syst Signal Process* 21:688–705. Elsevier Science Ltd

24. Rajeswari C, Sathiyabhama B, Devendiran S, Manivannan K (2014) Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine. *Procedia Eng* 97:1772–1783, ISSN 1877–7058. <https://doi.org/10.1016/j.proeng.2014.12.329>
25. Ngui WK, Leong MS, Hee LM, Abdelrhman AM (2013) Wavelet analysis: mother wavelet selection methods. *Appl Mech Mater* 393:953–958
26. Ji N, Zhou H, Guo K, Samuel OW, Huang Z, Xu L, Li G (2019) Appropriate mother wavelets for continuous gait event detection based on time-frequency analysis for hemiplegic and healthy individuals. *Sensors* 19(16):3462. <https://doi.org/10.3390/s19163462>.
27. Huang W, Kong F, Zhao X (2018) Spur bevel gearbox fault diagnosis using wavelet packet transform and rough set theory. *J Intell Manuf Springer Science+Business Media New York*
28. Hashim MA, Nasef MH, Kabeel AE, Ghazaly NM (2020) Combustion fault detection technique of spark ignition engine based on wavelet packet transform and artificial neural network. *Alexandria Eng J*, 59(5):3687–3697, ISSN 1110–0168. <https://doi.org/10.1016/j.aej.2020.06.023>
29. Rodrigues AP, D’Mello G, Srinivasa Pai P (2016) Selection of mother wavelet for wavelet analysis of vibration signals in machining. *J Mech Eng Autom (5A)*:81–85. <https://doi.org/10.5923/c.jmea.201601.15>
30. Li H, Lian X, Guo C, Zhao P (2013) Investigation on early fault classification for rolling element bearing based on the optimal frequency band determination. *J Intell Manuf*. <https://doi.org/10.1007/s10845-013-0772-8>, Springer Science+Business Media New York
31. Akbari M, Homaei, H, Heidari, M (2014) An intelligent fault diagnosis approach for gears and bearings based on wavelet transform as a preprocessor and artificial neural networks. *Int J Math Modell Comput* 4(4): 309–329. [http://ijm2c.iauctb.ac.ir/article\\_521870\\_00.html](http://ijm2c.iauctb.ac.ir/article_521870_00.html)
32. Dhamande L, Chaudhari M (2016) Detection of combined gear-bearing fault in single stage spur gear box using artificial neural network. *Procedia Engineering* 144:759–766. <https://doi.org/10.1016/j.proeng.2016.05.082>
33. Medina R, Macancela JC, Lucero P et al (2020) Gear and bearing fault classification under different load and speed by using Poincaré plot features and SVM. *J Intell Manuf*. <https://doi.org/10.1007/s10845-020-01712-9>
34. Schlechtingen M, Santos IF (2011) Review : Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mech Syst Signal Process* 25:1849–1875. Elsevier Science Ltd
35. Kumar HS, Srinivasa Pai P, Sriram NS, Vijay GS (2014) Selection of mother wavelet for effective wavelet transform of bearing vibration signals. *Adv Mater Res*
36. Wang C, Gan M, Zhu CA (2019) A supervised sparsity-based wavelet feature for bearing fault diagnosis, *J Intell Manuf Springer Science+Business Media New York*
37. Yang Yu (2006) YuDejie and Cheng Junsheng, A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J Sound Vib* 294:269–277
38. Sloukia FE, Bybi A, Drissi H (2017) Selection of mother wavelets for analyzing bearing vibration signals. 3rd International Conference on Electrical and Information Technologies ICEIT978–1–5386–1516–4/17/2017 IEEE
39. Unal M, Onat M, Demetgul M, Kucuk H (2014) Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. *Measurement* 58:187–196
40. Rafiee J, Tse PW, Harifi A, Sadeghi MH (2009) A novel technique for selecting mother wavelet function using an intelligent fault diagnosis system. *Expert Syst Appl* 36:4862–4875
41. Li F, Wang J, Chyu MK, Tang B (2015) Weak fault diagnosis of rotating machinery based on feature reduction with Supervised Orthogonal Local Fisher Discriminant Analysis, *Neurocomputing Elsevier Science Ltd*
42. Shao R, Hu W, Wang Y, Qi X (2014) The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement* 54 :118–132. Elsevier Science Ltd
43. Chew L, Leung K. *Learning Paradigms in Neural Networks, Medium, the Startup*, website: <https://medium.com/swlh/learning-paradigms-in-neural-networks-30854975aa8d>.
44. Gérard D (2005) *Neural network: methodology and applications*, Springer Book Springer-Verlag Berlin Heidelberg
45. Worden K, Staszewski WJ, Hensman JJ (2011) Review: Natural computing for mechanical systems research: a tutorial overview, *Mech Syst Signal Process* 25:4 – 111

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