Remaining Useful Life Prediction of Rolling Element Bearings using Minimum Redundancy Maximum Relevance and Extreme Learning Machine

Mouloud Boumahdi[1,2] and Redha Rebhi[1,3*]

[1]Department of Mechanical Engineering, University of Medea, Medea, 26000, Algeria.

[2]Laboratory of Mechanics,Physics, Mathematical modeling (LMP2M), University of Medea,Medea, Algeria

[3]LERM—Renewable Energy and Materials Laboratory, University of Medea, Medea 26000, Algeria

Received: 10-05-2023 Accepted: 28-10-2023 Published: 09-11-2023

Abstract

Estimating the remaining useful life of rolling element bearings is essential to ensure the reliable and efficient operation of rotating machinery, as well as reduce maintenance costs and downtime. In this study, a novel methodology was used to estimate the Remaining Useful Life (RUL) of ball bearings. Data were collected from two platforms: one for testing ball bearings with artificial defects, and another called PRONOSTIA for run-to-failure tests. Noise reduction techniques (Variational Mode Decomposition (VMD), AutoRegression (AR) filtering and Bandpass filtering) were applied to the data to select manualy the features. Using the second platform's data, a Minimum Redundancy Maximum Relevance (MR2) method was used to select automaticaly the features then the extreme Learning Machine (ELM) classification model was constructed. Furthermore, an ELM regression model was developed using the second platform's data to estimate the Remaining Useful Life (RUL). The proposed feature selection method effectively prevents delayed anticipation of failure. The results provide evidence for the effectiveness of the proposed approach in enhancing the accuracy of rolling element bearing Remaining Useful Life (RUL) prediction.

Keywords: Rolling element bearing, Remaining useful life, Minimum Redundancy Maximum Relevance, Extreme learning machine.

Tob Regul Sci. ™ 2022 ;8(2) : 5705-5732
DOI : doi.org/10.18001/TRS.8.2.398

1.Introduction

Rolling element bearings are a critical component in a wide range of rotating machine applications, providing support and enabling smooth rotation of various components. However,

*Author to whom correspondence should be addressed : rebhi.redha@gmail.com

rolling element bearings are subject to wear and damage over time, leading to premature failure and costly downtime. To ensure the reliable and efficient operation of rotating machinery, as well as reduce maintenance costs and downtime, adopting a constructive maintenance strategy is a critical task. The process of detection, diagnosis, and prognosis aims to identify possible faults within a machine or component and assess the type and severity of the detected problems. This process often uses diverse tools and techniques, including vibration analysis, acoustic monitoring, and visual inspection, among others.The general methodology for monitoring rotary machines includes three steps[1]: obtaining data for relevant signals such as vibration and temperature, processing data to analyze acquired data, and making maintenance decisions to select appropriate maintenance methods based on machine operating conditions.

Estimating the remaining life (RUL) of rolling element bearings has emerged as a prevalent concern within the scientific and industrial community. Despite significant advances in RUL estimation techniques, the accurate prediction of RUL in defective rolling element bearings remains a challenging and complex problem.In this framework, the common existing methodologies for predicting the remaining useful life (RUL) of rolling element bearings can be categorized into two main groups: model-based approaches and data-driven approaches [2].Model-based approaches rely on fundamental principles derived from physics. Hidden Markov models [3], Kalman filters[4], Wiener processes [5], and particle filters [6] are common techniques in model-based approaches.The complexity of the degradation of rolling element bearings poses a challenge in creating precise physical models. As a result, data-driven approaches for predicting the remaining useful life (RUL) have become increasingly popular and rapidly developed. These approaches use event and condition monitoring data combined with machine learning techniques for accurate RUL prediction. Various approaches were implemented to train a prediction model and then apply the trained model to estimate the remaining useful life (RUL) of rolling element bearings, such as support vector machine (SVM) [7, 8],relevance vector machine (RVM) [9, 10], Artificial neural networks (ANN)[11, 12], and convolutional neural network (CNN)[13–15]. However, data-driven methods heavily rely on the selection of features. This selection can be done manually, automatically, or through a combination of manual and automated procedures. Manual feature selection is advantageous when there is domain-specific expertise available or when the data is well understood. On the other hand, automated feature selection uses algorithms. In general, the process of feature selection significantly impacts the precision of predictions.

This paper proposes a new approach to predict the Remaining Useful Life (RUL) of ball bearings.

Firstly, data were collected from two platforms. The first platform aimed to test the ball bearings with artificial defects, while the second platform, PRONOSTIA, focused on testing the ball bearings through run-to-failure tests.

Secondly, noise reduction techniques named Variational Mode Decomposition (VMD), AutoRegression (AR) filtering, and Bandpass filtering were applied to the data from the first platform. These techniques were used to extract features and examine their effectiveness in defect detection, isolation, and identification. The relevant features were manually selected from the obtained data.

Thirdly, the extreme learning machine (ELM) algorithm was applied to the data from the second platform to construct an ELM classification model. The method of Minimum Redundancy Maximum Relevance (MRMR or MR2) was used to automatically select features.

Fourthly, the extreme learning machine (ELM) algorithm was again applied to the data from the second platform, this time to construct an ELM regression model. This model was used to estimate the Remaining Useful Life (RUL).

Finally, the proposed approach was validated using performance metrics to evaluate its effectivenessand understanding its limitations.

This paper makes several key contributions.

(i) it includes two types of ball bearing tests: one with artificial defects and the other subjected to run-to-failure testing, which allowed for a combination of manual and automatic feature selection. (ii) it uses a data-driven approach along with the MR2-ELM classification method to assess the impact of noise reduction techniques.

(iii) it uses a data-driven approach with the ELM regression method for predicting the Remaining Useful Life (RUL).

The remaining sections in this paper is as follows: Section 2 provides a brief introduction to the datasets used in this study, along with the feature extraction process to validate the proposed method. Additionally, it provides a brief description of the noise reduction techniques employed. Section 3 elaborates on the detailed methodology for constructing the MR2-ELM model, accompanied by illustrations and the presentation of results for discussion. Finally, Section 4 concludes the paper with our overall findings and conclusions.

2. Methodology for Estimating Remaining Useful Life

Rotating machinery undergoes degradation over time due to operational stresses and dynamic loads. Maintenance is crucial for maintaining reliability throughout its useful life. The general methodology for monitoring rotating machines involves three steps [1]: data acquisition to obtain relevant signals like vibration and temperature, data processing for analysis of the acquired data, and maintenance decision-making to select suitable maintenance methods based on the machine's operating conditions. This methodology enables proactive monitoring, early issue detection, and ensuring optimal reliability and performance.

This study presents a methodology for monitoring rolling element bearing deterioration and estimating Remaining Useful Life (RUL) using an extreme learning machine algorithm (ELM), as depicted in Figure 1. Prior to implementation, it is important to clarify the tools employed in this study.

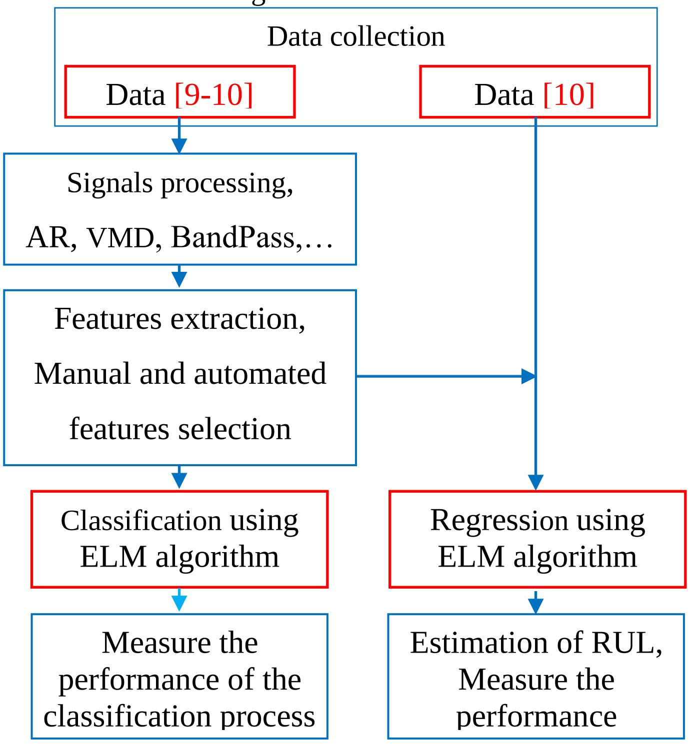


Figure 1. Flowchart of the proposed methodology

2.1 Overview of noise reduction techniques in vibration signals

Accurate diagnosis of fault bearings is crucial for ensuring the reliability and safety of machinery. Vibration signals are commonly used for this purpose but are often contaminated with noise, which can obscure critical fault signatures. Moreover, stochastic noise often obscures impulsive signatures in complex and non-stationary vibration signals (e.g., in ball bearings, vibrations, and noise in the form of periodic impulses produced when the balls pass over a defect in the rings).

To achieve these objectives, using advanced techniques such as AutoRegression (AR) filtering, Variational Mode Decomposition (VMD), and Bandpass filtering is of utmost importance in implementing effective condition monitoring and maintenance strategies. These techniques can be used individually or in combination to remove noise from vibration signals and preserve important features of the signal. In the following sections, we present an overview of the techniques used in this study.

AutoRegression (AR)

The AR model is widely used for denoising signals as it has the ability to capture the underlying dynamics of the signal and isolate it from the noise component. Let us considère a time series $x(n); n=1,\ldots,N$, which is the measured vibration signal. The actual current value can be expressed as the sum of the predicted value and a noise term, as shown in equation 1:

$$x(n) = x_p(n) + e(n) \quad (1)$$

$$x_p(n) = -\sum_{k=1}^{p} a(k)x(n-k) \quad (2)$$

where $x_p(n)$ is the predicted current value obtained through a weighted sum of the *p* previous value, $a(k)$ are the AR coefficientscan be calcultaed using the Yule–Walker equations [18, 19], and the residual error is denoted as$e(n)$.

The AR filter as described by Eq. (2), well predicts the deterministic pattern of the signal but is not capable of adapting to the sudden impulses caused by a localized fault. The benefits of using an AR model arise from its simplicity, efficiency and the low prior requirements for its application.

Although the AR filter as described by Eq. (2), is proficient in predicting the deterministic pattern of the signal, it may not effectively adapt to sudden impulses caused by localized faults. However, the advantages of using an AR model stem from its simplicity, efficiency, and minimal prerequisites for application [20].

Variational Mode Decomposition (VMD)

Variational mode decomposition (VMD), like Empirical Modes Decomposition (EMD), is one of the most recent signal decomposition techniques used to reduce noise. Dragomiretskiy and Zosso proposed variational mode decomposition in 2013 as a new self-adaptive signal decomposing method[21].Since its proposal, VMD has found applications in various fields.

VMD method allows to decompose the original signal into a finite number of signals so-called Intrinsic Mode Functions (IMFs) which is defined as follows.

$$x_n = \sum_{i=1}^{k} u_k(t) + res(t) \qquad (3)$$

Where x_n is the original signal, $\{u_k\} = \{u_1, u_1, \ldots, u_n\}$ are decomposition signals, and res is the residual signal after optimization.

The decomposition procedure involves resolving the following optimization problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \qquad (4)$$
$$\text{subject to } \sum_k u_k = x_n$$

Where x_n is the original signal, $\{u_k\}$ is the IMFs set, $\{\omega_k\}$ is the center frequencies of each $\{u_k\}$, $\delta(t)$ is an impulse function, and *k* is a modal component number.

To minimize equation4, the augmented Lagrangian function is addressed by introducing both a quadratic penalty termand Lagrangian multipliers λ:

$$L(\{u_k\},\{\omega_k\},\lambda) := \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| x_n(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), x_n(t) - \sum_k u_k(t) \rangle \qquad (5)$$

where α is quadratic penalty factor and λ is Lagrange multiplier.

The solution is achieved by optimizing the center frequency and bandwidth of each IMF. The updating procedures of frequency $\hat{u}_k$ can be expressed as [22]:

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{x_n(\omega)\sum_{i\langle k}\hat{u}_k^{n+1}(\omega) - \sum_{i\rangle k}\hat{u}_k^{n}(\omega) + \hat{\lambda}(\omega)/2}{1+2\alpha\left(\omega-\omega_k^n\right)^2} \quad (6)$$

Where $\hat{u}_k^{n+1}(\omega)$, $x_n(\omega)$, and $\hat{\lambda}(\omega)$ represented the Fourier transformations of $u_k^{n+1}, x_n(t)$, and $\lambda(t)$.

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega\left|\hat{u}_k(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k(\omega)\right|^2 d\omega} \quad (7)$$

The VMD process is terminated when the relative error e is less than a convergence tolerance ε.

$$e = \frac{\sum_k \left\|\hat{u}_k^{n+1} - \hat{u}_k^n\right\|_2^2}{\left\|\hat{u}_k^n\right\|_2^2} \langle \varepsilon \quad (8)$$

Bandpass filtering

Bandpass filters are widely used in vibration signal analysis for diagnosis and prognosis purposes. They effectively filter out high and low-frequency noise from the signal by allowing only frequencies within a specific range to pass through while attenuating frequencies outside that range. A bandpass filter can be applied to the vibration signal by selecting two cutoff frequencies to isolate the frequency band associated with a rotating component fault. All frequencies below the low-frequency cutoff and above the high-frequency cutoff are removed in the filtered signal, resulting in a focused frequency band associated with the fault.

When a fault occurs in a rotating component of a machine, such as a bearing, it can significantly impact the rotating machine's natural frequencies. The FFT (Fast Fourier Transform) spectrum is a helpful tool for identifying the cutoff frequencies within the significant resonance range.

2.2Feature selection with MR2

Minimum redundancy maximum relevance (MRMR or MR2) is a feature selection method that combines filter and wrapper approaches to select a subset of features that are both highly relevant to the target variable and minimally redundant with each other.This method was proposed by Peng et al. [23] and has been widely used in machine learning applications.

The determination of redundancy and relevance is based on mutual information, which can be is expressed as:

$$I(X,Y) = \sum_{i,j} p\left(x_i, y_j\right)\log\frac{p\left(x_i, y_j\right)}{p\left(x_i\right)p\left(y_j\right)} \quad (9)$$

Where X and Y are feature vector (features that describe each instance) or class vector (class label of each instance), $p(x_i, y_j)$ is the probability that X takes the value x_i and Y takes the value y_j simultaneously, and, and $p(x_i)$ and $p(y_j)$ are the marginal probabilities of X and Y respectively.

Assume that Srepresents the subset of features being evaluated($S \in X$), $|S|$is the number of features, and Y is theclass variable.

The relevance of a feature set S is defined as the mutual information between S and Y, which can be expressed as:

$$RL(S) = \frac{1}{|S|} \sum_{i,j} I(x_i, Y) \quad (10)$$

The redundancy of a feature set S with respect to the target variable Y is defined as the average mutual information between the features in S, which can be expressed as:

$$RD(S) = \frac{1}{|S|^2} \sum_{i,j} I(x_i, x_j) \quad (11)$$

Wherex_i and x_j are two distinct features in S, and $I(x_i, x_j)$ is the mutual information between features x_i and x_j.

The MRMR algorithm aims to identify the optimal feature set S that maximizes the relevance RL and minimizes the redundancy RD. To evaluate the relevance of features in S, the Mutual Information Quotient (MIQ) is commonly used. The process involves selecting the feature with the largest MIQ value in S and adding the selected feature to the set S. the maximum of MIQ is expressed as:

$$\max MIQ(S) = \frac{\max I(x_i, Y)}{\frac{1}{|S|} \sum_j I(x_j, Y)} \quad (12)$$

2.3 Overview of ELM

Extreme learning machine (ELM) has acquired significant attention due to its simplicity, fast training speed, and excellent generalization performance, making it widely used in various applications [24]. ELM algorithm is proposed by Huang et al. [25]for training single hidden layer feedforward neural networks (SLFNs). In the current section, a brief overview of ELM is presented.

Given a set of N training samples $(x_i, t_i); i = 1,2,...,N$, where x_i represents the input features and t_i represents the corresponding target outputs, the output of an ELM with L hidden node is expressed as:

$$y_j = \sum_{i=1}^{L} \beta_i g_i(x_j) = \sum_{i=1}^{L} \beta_i g(w_i . x_j + b_i), j = 1,2,\dots,N \quad (13)$$

where$L$ is the number of hidden neurons,$x_j$is the vector representation of the jth input sample, and $g_i(x_j)$ is the activation function of the ith hidden neuron,β_i is theoutput layer weight of the ith hidden neuron, w_i is a vector containing the input layer weights,and $b_i$is the bias of the ith hidden neuron.

The Eq. (6) can be written as follows:

$$H\beta = T.(14)$$

$$\text{where} H = \begin{bmatrix} g(w_1 x_1 + b_1) & \cdots & g(w_L x_1 + b_L) \\ . & \cdots & . \\ . & \cdots & . \\ g(w_1 x_N + b_1) & \cdots & g(w_L x_N + b_L) \end{bmatrix}_{NxL} ; T = [t_1, t_2, \ldots, t_N]^T; \beta = [\beta_1, \beta_2, \ldots, \beta_L]^T (15)$$

H is the hidden layer output matrix, T is the target outputs vector of the training samples, and β is the output weights that can be calculated analytically by obtaining the following least square solution:

$$\hat{\beta} = H^{\dagger} T (16)$$

where $H^{\dagger}$ is the Moore–Penrose generalized inverse of matrix H.

2.4 Data set and feature extraction

This study used data collected from two distinct platforms to detect, diagnose, and predict the remaining useful life of rolling element bearings. The first platform is designed specifically for detecting and diagnosing defects in rolling bearings by conducting tests on new bearings or those with artificial defects. The second platform is intended for monitoring the deterioration of rolling bearings throughout their entire operating life, with the aim of predicting the remaining useful life of the ball bearings. These two platforms provide distinct data for assessing the condition of rolling bearings.

Time domain features

During operation, rotating machines generate vibrations, and the elevated level of these vibrations is often indicative of the deteriorating condition of their components. To track the health status of these components, statistical indicators, commonly referred to as features, are used.

In this study, a total of fifteen statistical features were extracted from each vibration signal, which included RMS, Peak, Kurtosis, CF, KF, Rang, mean, STD, VAR, ADEV, Skewness, Margin, RA, IF, and shape. Table 1 provides a mathematical description of these features. The Range (or peak-to-peak), mean, peak, RA, and RMS indicators reflects the vibration amplitude and provides a measure of the overall vibration energy present in the time signal. In contrast, kurtosis, skewness, variance crest factor, shape, std, KF, IF, and margin provide information on other aspects of the vibration signal, such as its distribution, shape, and extreme values, but they do not directly reflect the vibration amplitude and energy in the time domain as the RMS value does. Through time-based analysis of these features, potential failures can be predicted and prevented, resulting in reduced unplanned downtime and minimized maintenance costs.

Table 1. Mathematical description of features.

Features	Mathematical description
Range	$Range = \max(x_i) - \min(x_i)$

Mean	$Mean = \frac{\sum_{i=1}^{N} x_i}{N}$
Peak	$Peak = \max\lvert x_i \rvert$
Root amplitude (RA)	$RA = \left(\frac{\sum_{i=1}^{N} \sqrt{\lvert x_i \rvert}}{N} \right)^2$
Root mean square (RMS)	$RMS = \frac{\sqrt{\sum_{i=1}^{N} x_i^2}}{N}$
Average deviation (ADEV)	$ADEV = \frac{\sum_{i=1}^{N} \lvert x_i - Mean \rvert}{N}$
Variation (VAR)	$VAR = \frac{\sum_{i=1}^{N} (x_i - Mean)^2}{N}$
Standard deviation (STD)	$STD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - Mean)^2}{N-1}}$
Skewness	$Skewness = \frac{\sum_{i=1}^{N} (x_i - Mean)^3}{(N-1).STD^3}$
Kurtosis	$Kurtosis = \frac{\sum_{i=1}^{N} (x_i - Mean)^4}{(N-1).STD^4}$
Crest Factor (CF)	$CF = \frac{Peak}{RMS}$
K Factor (KF)	$KF = Peak.RMS$
Margin	$Margin = \frac{Peak}{RA}$
Shape	$Shape = \frac{RMS}{Margin}$
Impulse Factor (IF)	$IF = \frac{Peak}{Margin}$

Detection and diagnosis of defects

A series of experiments were conducted on the dedicated platform[16] to accurately detect and diagnose defects in rolling bearings.Figure 2 shows a deteriorated rolling element bearing with two artificial elliptical spalls that appear on both the inner race and outer race of a rolling element bearing.The reason for giving an elliptical shape to artificial defects is that defects frequently have an elliptic shape in real-life scenarios. Hence, adopting an elliptical shape for artificial defects can help simulate realistic conditions for experimental testing purposes.

Figure 2. Ball bearing with artificial spalls.

Specifically, SKF6206 single-row ball bearings were mounted and tested on a rolling bearing fault detection test rig. Defects of varying sizes, ranging from 8 mm2 to 20 mm2, were introduced through electroerosion on the inner, outer, or both races simultaneously. The experiments were conducted at a rotational speed of 1000 rpm and with a horizontal radial load of 5000 N generated by a hydraulic cylinder. In addition, acceleration measurements have been performed over a broadband frequency range of 20 kHz. Table 2 summarizes the relevant details of the experiments.

Table 2. Ball bearing characteristics and test details.

Ball bearing characteristics		Operating conditions	
TypeSKF 6206		Sampling frequency	51.2 kHz
Number of balls	9	Rotational speed	1000 rpm
Pitch diameter	46.00mm	Radial loads	5000 N
Diameter of a ball	9.53 mm		

Figure 3 shows signals obtained by an accelerometer that was mounted on a bearing in a horizontal radial direction and measured under identical rotational speeds and loads. These signals show periodic pulses that are typical of rolling defects that arise when a ball passes over a surface discontinuity on race. The distinctive nature of these signal patterns is characteristic of rolling defects and provides a reliable indicator of potential issues.

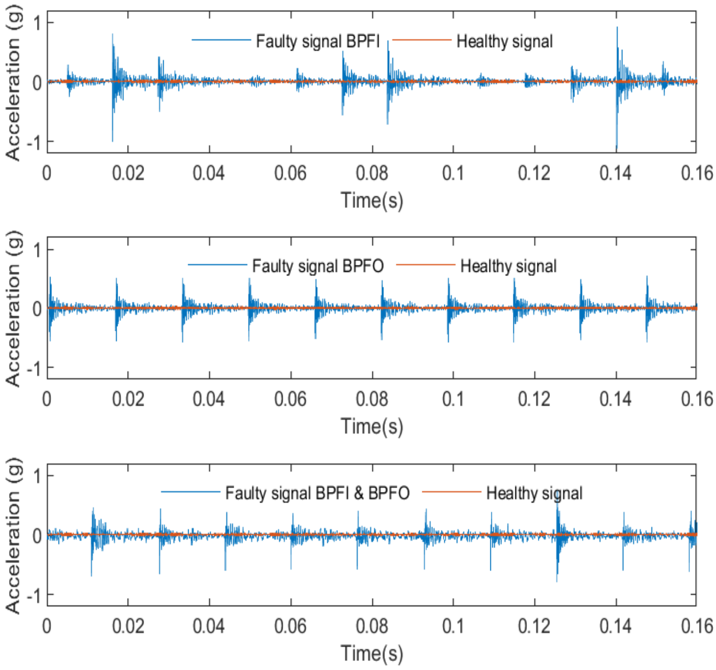


Figure 3. Vibration signals associated with defects in the inner race, outer race, or both races simultaneously.

Prognosis for remaining useful life prediction

A series of experiments were carried out on the PRONOSTIA platform designated [17]to accurately estimate the remaining useful life. The platform's objective is to provide accurate empirical data to monitor and describe the degradation of ball bearings throughout their entire operational lifespan. In addition, run-to-failure experiments were conducted in which radial loads exceeding the permitted limits were applied to the ball bearings to accelerate their degradation.

The PRONOTIA platform was used to conduct tests on NSK 6804DD deep groove ball bearings, which are designed to operate at a maximum speed of 13,000 rpm. The experiments were conducted at a rotational speed of 1800 rpm, with a horizontal radial load of 4000 N generated by a hydraulic cylinder. The pertinent details of the experiments are summarized in Table 3.

Table 3. Ball bearing characteristics and test details.

Ball bearing characteristics		Operating conditions	
TypeNSK 6804DD		Sampling frequency	25.6kHz
Number of balls	13	Rotational speed	1800 rpm
Pitch diameter	25.6mm	Radial loads	4000 N
Diameter of a ball	3.5 mm		

Two high-frequency accelerometers were installed on the bearing housing to collect monitoring data during testing. One accelerometer was positioned horizontally, while the other was placed vertically. Acceleration measurements were taken at a sampling frequency of 25.6

kHz, and vibration data were recorded at 1-second intervals. The objective of the experiment was to gather high-quality data that could be used for predicting potential defects in the bearings.

Figure 4 displays the temporal changes of the RMS feature for four rolling bearings: Bearing 1_1 (B 1_1), Bearing 1_2 (B 1_2), Bearing 1_3 (B 1_3), and Bearing 1_4 (B 1_4). Although the rolling elementbearings are of the same type and were tested under the same operating conditions listed in Table 3.We observe that the progression of deterioration for each rolling bearing differs from one another. This finding indicates the complexity of analyzing the deterioration condition of rolling elementbearings and confirms the need for appropriate tools and techniques to assess rolling bearings conditions effectively.

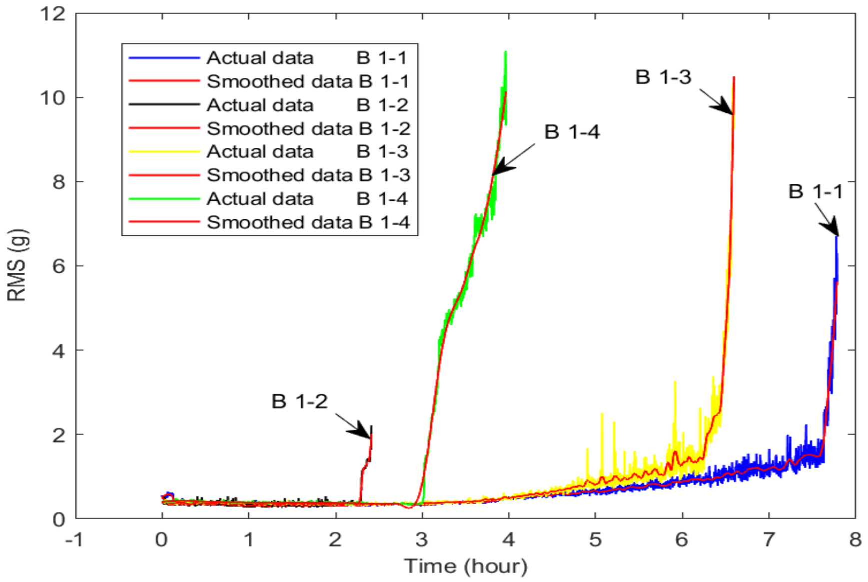


Figure 4. RMS evolution over time for four bearings: B 1_1, B 1_2, B 1_3, and B 1_4.

3.Results and Discussion

This study focuses on two interrelated processes: the detection and diagnosis of defects and the prognosis of the remaining useful life of rolling element bearings. The first process involves analyzing the impact of noise reduction techniques on identifying the condition of bearings, both with or without defects, as well as assessing the validity of proposed parameters. The second process entails using the MR2-ELM algorithm to predict the remaining useful life of the bearings.

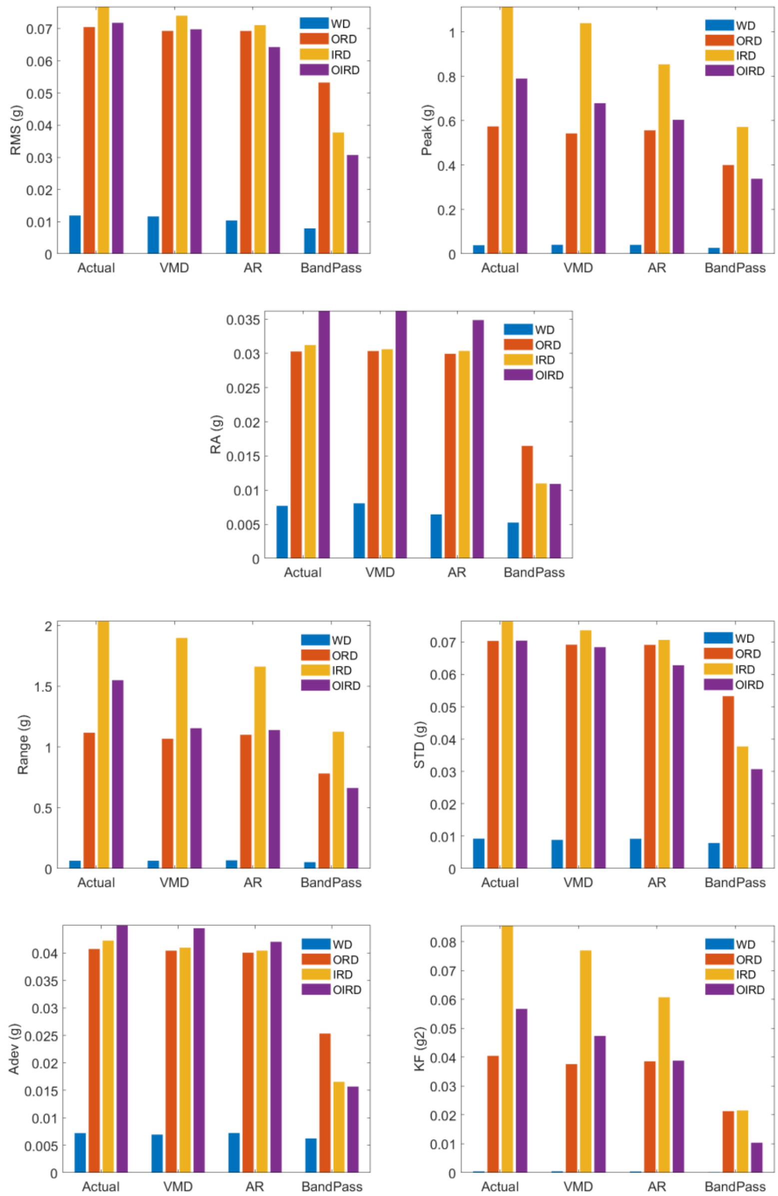
Detection and diagnosis of defects

This section aims to investigate how defect types and noise reduction techniques affect the integrity of signals generated by rolling element bearings. By analyzing the effects of various types of defects and noise reduction methods, we can gain a better understanding of accurate defect detection and diagnosis.

Figure 5 illustrates the impact of noise reduction techniques on feature values for different types of defects. Specifically, we consider four distinct types: rolling element bearings without defects (WD), rolling element bearings with inner race defects (IRD), rolling element bearings with outer race defects (ORD), and rolling element bearings with both inner and outer race defects (OIRD). In addition, we evaluate three noise reduction techniques: VMD, AR, and BandPass.

The figure displays fifteen bar graphs, each representing the features listed in Table 1. After examining the bar graphs depicting the values for RMS, Peak, RA, Range, std, adev, KF, VAR, margin, CF, and kurtosis, we noticed that bearings afflicted with IRD, ORD, and OIRD defects displayed higher levels of vibration amplitude than those lacking any defects. Additionally, we found that inner race defects led to higher vibrations than outer race defects. However, the bar graphs representing the features for skewness, mean, IF, and shape do not provide clear insights into detecting and diagnosing rolling bearing defects. Therefore, we have excluded these features to maintain accuracy in our analysis.

In the next section, we will explore the potential of the remaining eleven features.



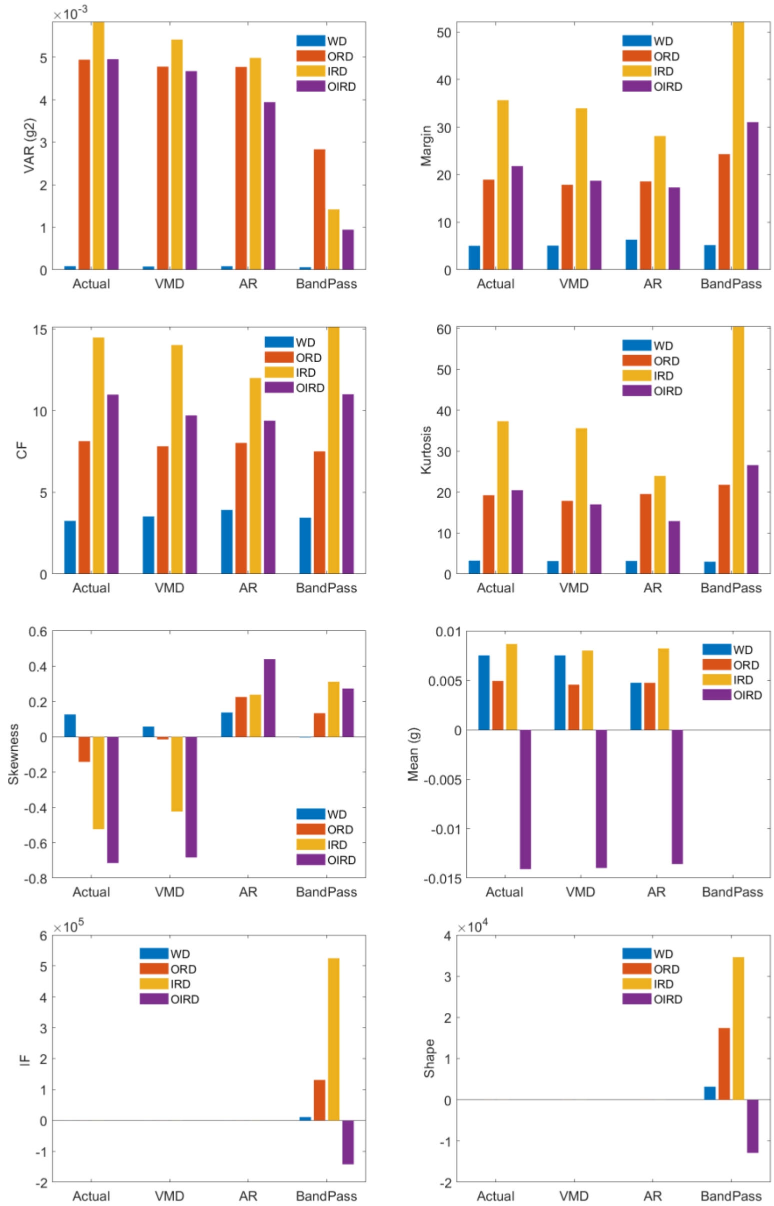


Figure 5. The effectiveness of noise reduction techniques on feature values across different defect types.

Figure 6shows the signals recorded by an accelerometer fixed on the bearing in a horizontal radial direction, with identical rotating speed and loading conditions. These signals exhibit periodic pulses that indicate bearing defects caused by a ball passing over a discontinuity in the bearing raceway. The figure also shows that the noise reduction methods, VMD, AR, and BandPass, did not affect the integrity of the vibration signals. The FFT and envelope spectrum were generated by processing the vibration signals, which reveal peaks corresponding to the characteristic frequencies of bearing defects. Tonote in particular, the envelope spectrum provides a more precise identification of the frequency characteristics of defects compared to the FFT spectrum.

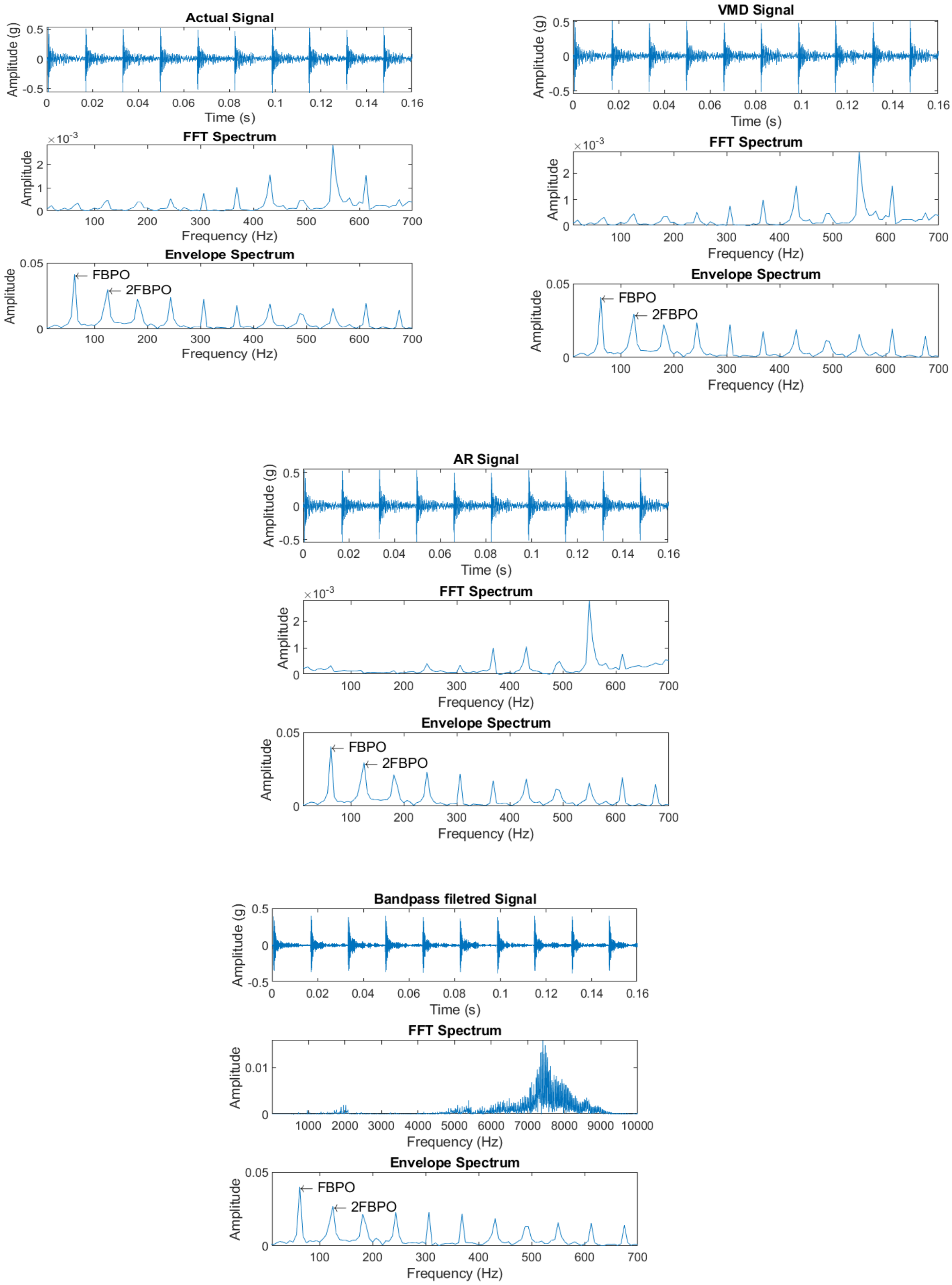


Figure 6. The vibration signals generated by a defect in the outer race of a bearing, its FFT spectrum, and its envelope spectrum.

Prognosis for Remaining Useful Life Prediction

Figure 7 illustrates the temporal progression of feature values (refer to Table 1) over time. It is evident that certain curves exhibit a notable trend, indicating the deterioration of rolling element bearings, while others do not show a significant trend. In Figure 8, the FFT and envelope spectrum of the latest vibration signal recorded from Bearing1_1 are presented. The impact of rolling element bearing deterioration on the natural frequency of the platform can be observed in the frequency range of 500 to 1500 Hz. Furthermore, Figure 8 demonstrates that the envelope spectrum offers a more precise identification of the frequency characteristics of defects in comparison to the FFT spectrum.Figure 9 illustrates the impact of noise reduction techniques on RMS feature for Bearing1_1.

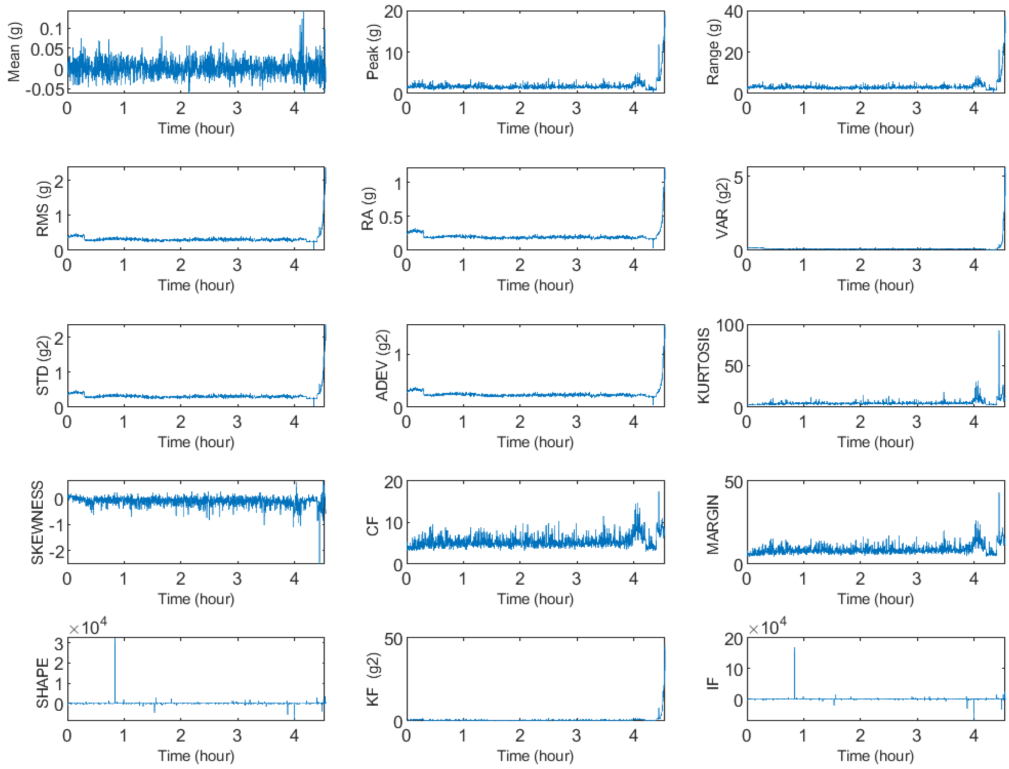


Figure 7. Temporal feature trends for Bearing 1_1.

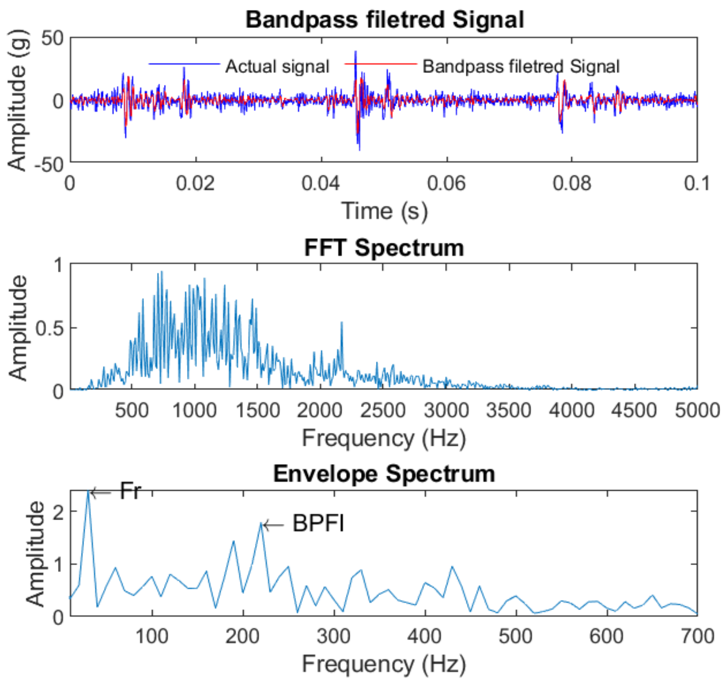


Figure 8. FFT and envelope spectrum of the last vibration signal measured from Bearing1_1.

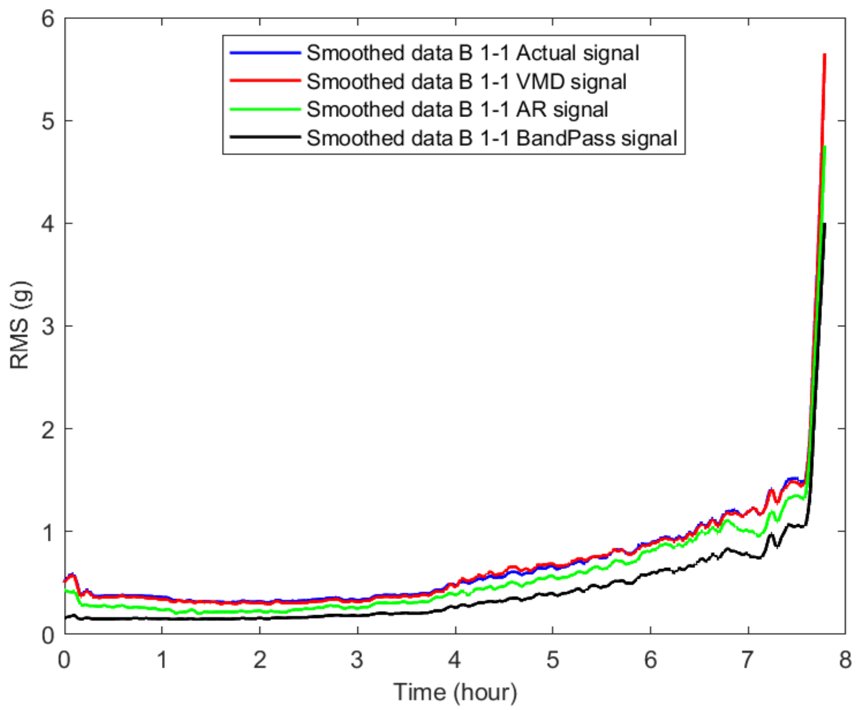


Figure 9. The impact of noise reduction techniques on RMS feature for Bearing1_1.

Selection features using MR2

The Minimum Redundancy Maximum Relevance (MR2) algorithm is a powerful feature selection tool that can help improve the accuracy and efficiency of machine learning models.The application of MR2 algorithm involves the following main steps: firstly, instances are extracted from vibration signals to build the dataset. Each instance is defined by fifteen statistical features (as listed in Table 1), and the data is divided into three classes based on two alarm thresholds, alert and danger thresholds. These classes are referred to as Normal, Abnormal, and Danger.Secondly, the MR2 algorithm is used to conduct correlation-based feature selection, where the features are ranked based on their relevance to the classes. Thirdly, a specific number of top-ranked features are selected. Fourthly, an extreme machine learning model is trained using the selected features. Lastly, the model's performance is evaluated using a suitable metric, such as classification rate or Kappa statistic.

This study examined various features to evaluate their performance and identify the optimal number of features. The results in Table 4 highlight the key features selected by manual and automatic selection for the accurate classification of ball bearing condition classes. Figures 5 and 7illustrate the close relationship between the RMS and STD features. Additionally, the feature STD is highly correlated with the high-ranked feature RMS. Including STD in the model may lead to overfitting or a reduction in interpretability. Consequently, to prevent overfitting, we removed this feature from the model.

So far, we have excluded five of the fifteen features in Table 1. The rest of this study will focus on the remaining ten features, shaping the foundation for our analysis.

ELM classification model

To establish the Extreme Learning Machine (ELM) classification model involves determining the dataset and the appropriate values for key parameters: the optimal number of hidden neurons and the activation function.The optimal number of hidden neurons, determined through Grid Cross Validation (GridCV) algorithm, and the sigmoid activation function were used in this study. Multiple ELM classification models were constructed using four datasets. One dataset was used without any noise reduction, while the other three datasets corresponded to three noise reduction techniques: VMD, AR, and BandPass. According to the number of features used in the classification, each dataset is divided into six sub-datasets. as follow: The first sub-dataset comprised all features, the secondsub-dataset comprised ten manually selected features, while the remaining foursub-datasets consisted of features selected and ranked based on their relevance using the MR2 algorithm (See Table 4).

Performance ofclassification models

In this section, we present a summary discussion of the results obtained through the application of the ELM classification algorithm on the various datasets. The obtained models can be classified into two categories. The first category comprises models established using denoised signals, while the second category includes models established using original signals. A comparison between the performance of the two categories (See Table 4) reveals that the models in the first category demonstrate superior performance compared to the second category. The presence of well-established features used for the identification of bearing faults, namely root mean square (rms), kurtosis, and crest factor, is evident within the chosen set of features. The effectiveness of these renowned features has been substantiated through extensive references[26]. Evaluation metrics, including classification rate, Kappa statistic, MAE, and RMSE, consistently favor the models in the first category, highlighting their better predictive accuracy. The results presented in Table 4 showcase a classification rate range of 93.94%–97.72% and kappa statistics ranging from 0.88 to 0.94, indicating a strong correlation. Furthermore, the mean absolute error falls within the range of 0.003 to 0.006. These findings emphasize the importance of dataset selection and further validate the effectiveness of the models in the first category for the given task. In the confusion matrix (Tables 5–8), the diagonal elements represent the instances that were correctly classified for datasets which correspond to: actual signals, signals denoised with VMD, signals denoised with AR, and signals denoised with BandPas filter, respectively. Conversely, the remaining elements in the matrix indicate the number of instances from each class that were incorrectly classified.

Table 4.Performance of ELM classification models

Data	without selecting features (15 features)	Manual feature selection (10 features)	Number of selected features using MR2 algorithm			
			5	4	3	2
Actual	ONHN: 18 CR: 93.94% KS: 0.88 MAE: 0.06 RMSE: 0.27	ONHN: 20 CR: 96.68% KS: 0.93 MAE: 0.03 RMSE: 0.19 Rms Cf kurtosis range ra peak kf adev marg var	ONHN: 12 CR: 95.68% KS: 0.91 MAE: 0.04 RMSE: 0.22 Rms cf kurtosis range ra	ONHN: 16 CR: 96.36% KS: 0.93 MAE: 0.04 RMSE: 0.20 Rms cf kurtosis range	ONHN: 19 CR: 96.21% KS: 0.92 MAE: 0.04 RMSE: 0.20 Rms cf kurtosis	ONHN: 18 CR: 96.18% KS: 0.92 MAE: 0.04 RMSE: 0.20 Rms cf

AR	ONHN: 19 CR: 97.72% KS: 0.89 MAE: 0.05 RMSE:0. 24	ONHN:17 CR: 95.43% KS: 0.91 MAE: 0.05 RMSE: 0.22 Rms cf kurtosis peak ra range marg kf adev var	ONHN: 20 CR: 95.68% KS: 0.91 MAE: 0.04 RMSE: 0.22 Rms cf kurtosis peak ra	ONHN: 15 CR: 95.18% KS: 0.90 MAE: 0.05 RMSE: 0.22 Rms cf kurtosis peak	ONHN: 20 CR: 96.15% KS: 0.92 MAE: 0.04 RMSE: 0.20 Rms cf kurtosis	ONHN: 20 CR: 95.40% KS: 0.91 MAE: 0.05 RMSE: 0.22 Rms cf
VMD	ONHN: 13 CR: 94.04% KS: 0.88 MAE: 0.06 RMSE: 0.27	ONHN: 14 CR: 95.61% KS: 0.91 MAE: 0.04 RMSE: 0.21 Rms cf kurtosis range ra peak kf adev marg var	ONHN: 20 CR: 96.00% KS: 0.92 MAE: 0.04 RMSE: 0.20 Rms cf kurtosis range ra	ONHN: 14 CR: 95.79% KS: 0.92 MAE: 0.04 RMSE: 0.21 Rms cf kurtosis range	ONHN: 20 CR: 95.93% KS: 0.92 MAE: 0.04 RMSE: 0.20 Rms cf kurtosis	ONHN 20 CR: 96.11% KS: 0.92 MAE: 0.04 RMSE: 0.20 Rms cf
BandPass	ONHN: 17 CR: 94.26% KS: 0.89 MAE: 0.06 RMSE: 0.25	ONHN: 20 CR: 96.32% KS: 0.93 MAE: 0.04 RMSE: 0.19 adev cf range kurtosis Rms peak ra kf marg var	ONHN: 17 CR: 96.75% KS 0.93 MAE 0.03 RMSE 0.18 adev cf range kurtosis Rms	ONHN: 20 CR: 97.18% KS: 0.94 MAE: 0.03 RMSE: 0.17 adev cf range kurtosis	ONHN: 15 CR: 96.25% KS: 0.92 MAE: 0.04 RMSE: 0.19 adev cf range	ONHN: 19 CR: 96.79% KS: 0.94 MAE: 0.03 RMSE: 0.18 adev cf
Note: Number of hidden neuron (ONHN); Classification rate (CR); Kappa statistic (KS); Mean absolute error (MAE); Root mean square error (RMSE)						

Table 5. Confusion matrix correspondent to actual signals for all features.

Classified as →	Normal	Abnormal	Danger
Normal	1571	39	11
Abnormal	92	1006	41
Danger	0	1	42

Table 6. Confusion matrix correspondent to signals denoised with VMD for all features.

Classifiedas →	Normal	Abnormal	Danger
Normal	1574	45	2
Abnormal	91	1010	38
Danger	0	1	42

Table 7. Confusion matrix correspondent to signals denoised with AR for all features.

Classifiedas →	Normal	Abnormal	Danger
Normal	1590	29	2
Abnormal	88	1040	11
Danger	0	7	36

Table 8. Confusion matrix correspondent to signals denoised with BandPass for all features.

Classifiedas →	Normal	Abnormal	Danger
Normal	1543	71	7
Abnormal	61	1044	34
Danger	0	6	37

Figures 10-11, show a comparison of the accuracy and the MAE results of four different ELM algorithms. The horizontal axis represents the different datasets tested (Actual, AR, VMD, BandPass) to construct the ELM models.In term of accuracy,Figure 10shows that the fourth method (denoised signals using BandPass) has the highest median value of a set of accuracy numbers (96.535), followed closely by the first method (actual vibration signals) with a median value of 96.20. The third method (denoised signals using VMD) has a median value of 95.86, and the second method (denoised signals using AR) has the lowest median value of 95.55.In term of mean absolute error (MAE), Figure 11shows that the fourth method has the lowest median value of a set of MAE (0.035), followed by the first and third methods with median values of 0.04 and 0.04, respectively. The second method has a slightly higher median value of 0.05. However, the variation in accuracy and mean absolute error (MAE) among the four methods is insignificant. Hence, it is recommended to consider additional factors outside of these criteria to support the selection of the most appropriate features.

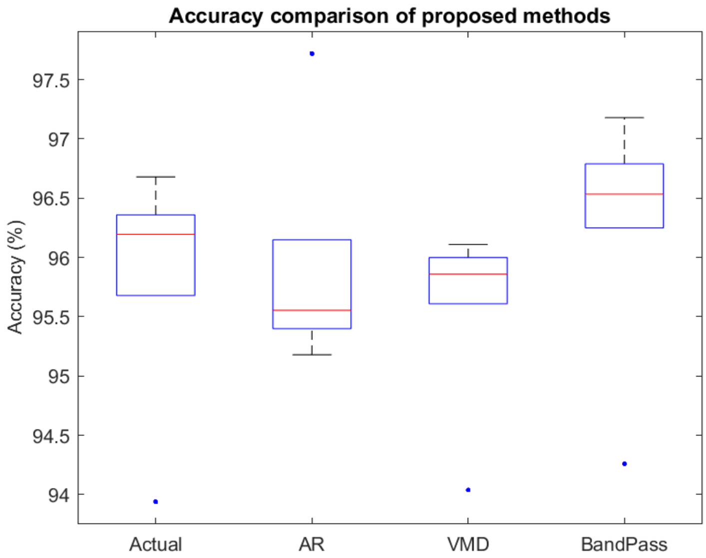


Figure 10. Comparison of Accuracy for four ELM classification models.

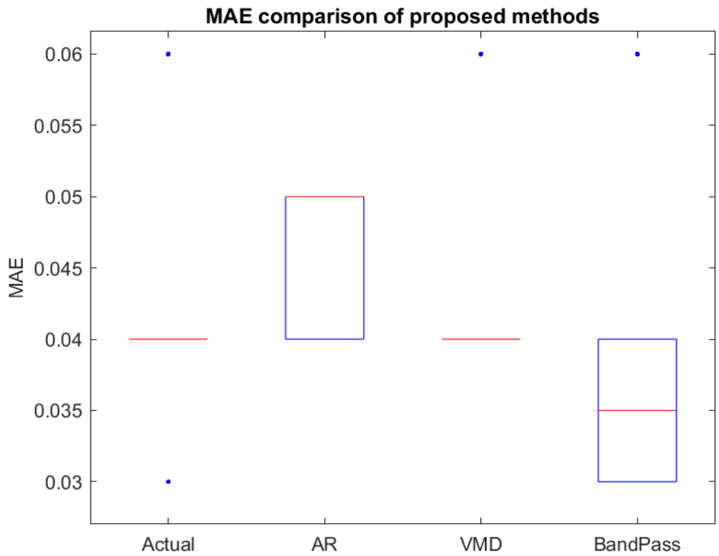
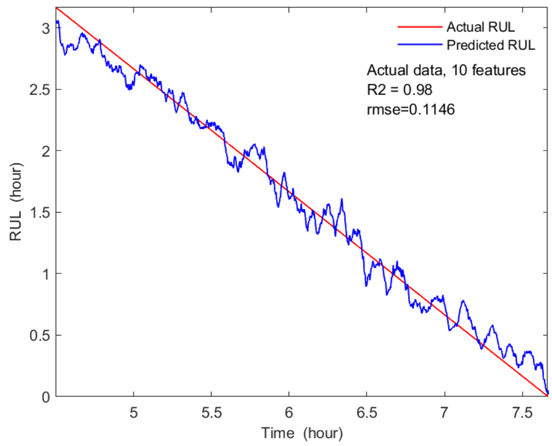


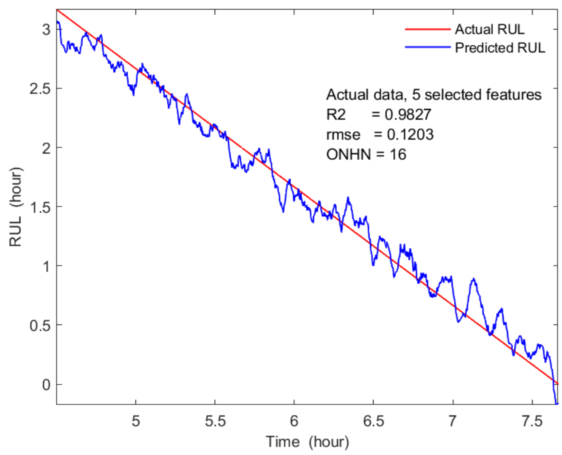
Figure 11. Comparison of MAE for four ELM classification models.

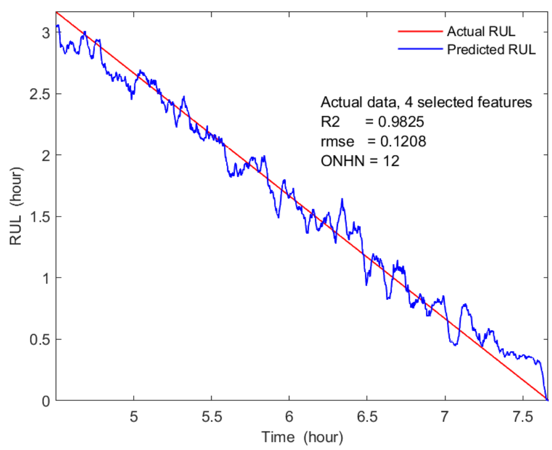
ELM regressionmodel to estimate RUL

This section focuses on estimating the Remaining Useful Life (RUL) using the Extreme Learning Machine (ELM) algorithm. The ELM is applied to construct regression models using the previously mentioned datasets (See Table 4). To evaluate model performance and identify the optimal configuration Cross-validation is employed. Figure 12 summarizes the results obtained from the application of the ELM algorithm, offering valuable insights into the accuracy of RUL estimation and the effectiveness of the ELM approach.

From Figure 12a, it is evident that the RUL in the upper region of the curve is lower than the actual RUL, indicating early anticipation of failure. Conversely, the RUL in the lower region of the curve is higher than the actual RUL, indicating delayed anticipation of failure. In Figures 12b-12f, the RUL in the lower region closely aligns with the actual RUL. This change can be attributed to the use of selected features only.The occurrence of disparities between predicted RUL and actual RUL has been documented in numerous scholarly investigations[8].







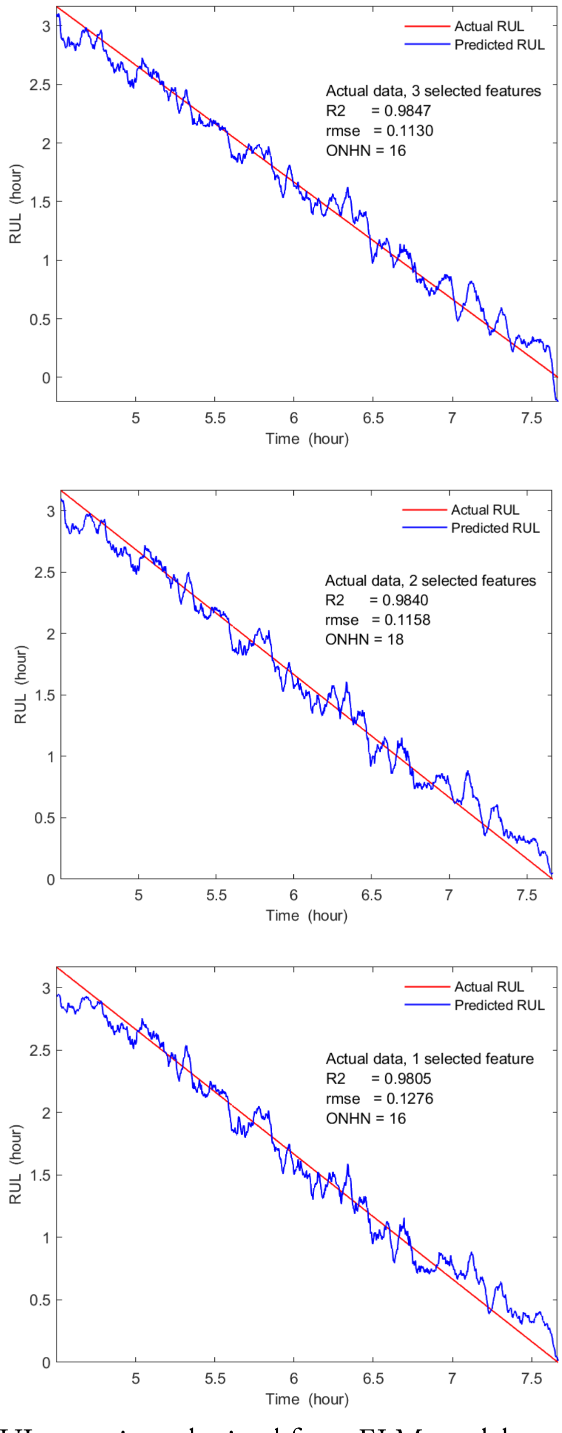


Figure 12. RUL over time obtained from ELM models and actual signals.

Figure 13highlights the varying degrees of performances between the ELM models obtained from different datasets (Actual signals, AR, VMD, BandPass). This analysis reveals that the first method (which is based on actual vibration signals) exhibits the highest median coefficient of determination (R2) value of 98.34, indicating superior predictive performance. The fourth method (which is based on denoised signals using BandPass filter) follows closely with a median value of 97.68, showing good performance. The third method (which is based on denoised signals using VMD) also performs well, with a median value of 97.14. In contrast, the second method (which is based on denoised signals using AR) has the lowest median value of 96.38, suggesting relatively weaker predictive ability.

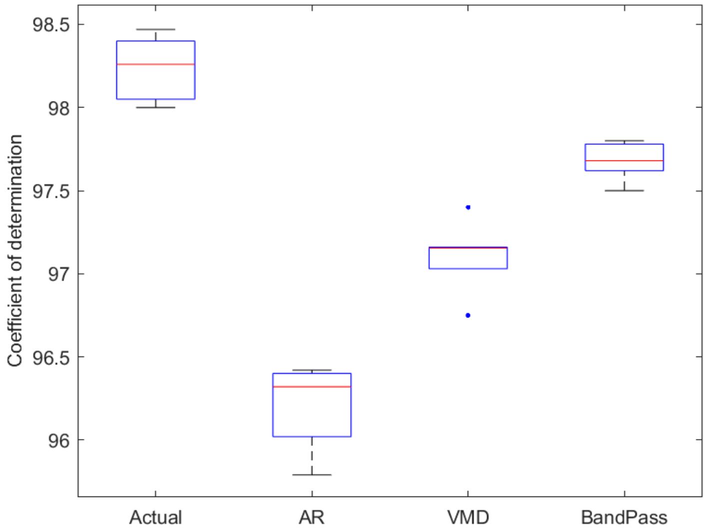


Figure 13. Comparison of coefficients of determination for four ELM regression models.

4.Conclusion

This paper presents a novel methodology for estimating the Remaining Useful Life (RUL) of rolling element bearings. The proposed method uses a data-driven approach and machine learning techniques, specifically MR2 and ELM, to accurately estimate the RUL of degraded rolling element bearings. By combining manual and automatic feature selection methods, the proposed approach effectively avoids delayed anticipation of failure. Additionally, the impact of noise reduction techniques on feature selection is highlighted. The successful application of this methodology in condition monitoring of rotating machines holds significant promise for intelligent decision-making, leading to enhanced efficiency and cost-effectiveness.

While this study showcases notable advancements in RUL estimation, it is important to acknowledge the need for further research to explore the performance of the proposed methodology across different machine-learning techniques. Future work should focus on extending and refining the methodology to incorporate various noise reduction techniques. In conclusion, the novel methodology proposed in this study provides a practical solution for predicting the RUL of defective ball bearings, contributing to the field of condition monitoring and maintenance optimization.

Conflicts of interests

The authors declare that they have no known competing for financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Jardine AK, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process.vol. 20, pp.1483–1510.

https://doi.org/10.1016/j.ymssp.2005.09.012

2. Lei Y, Li N, Guo L, Li N, Yan T, Lin J (2018) Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process.vol. 104, pp.799–834.https://doi.org/10.1016/j.ymssp.2017.11.016
3. Xiahou T, Zeng Z, Liu Y (2020) Remaining useful life prediction by fusing expert knowledge and condition monitoring information. IEEE Trans. Ind. Informatics.vol. 17, pp.2653-2663.DOI: 10.1109/TII.2020.2998102
4. Cui L, Wang X, Wang H, Ma J (2019) Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter. IEEE Trans. Instrum. Meas.vol. 69, pp.2858–2867.DOI: 10.1109/TIM.2019.2924509
5. Zhang Z, Si X, Hu C, Lei Y (2018) Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. Eur. J. Oper. Res.vol. 271, pp.775–796.https://doi.org/10.1016/j.ejor.2018.02.033
6. Li N, Lei Y, Lin J, Ding SX (2015) An improved exponential model for predicting remaining useful life of rolling element bearings. IEEE Trans. Ind. Electron.vol. 62, pp.7762–7773.DOI: 10.1109/TIE.2015.2455055
7. Soualhi A, Medjaher K, Zerhouni N (2015) Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. IEEE Trans. Instrum. Meas.vol. 64, pp. 52–62.DOI: 10.1109/TIM.2014.2330494
8. Benkedjouh T, Medjaher K, Zerhouni N, RechakS (2013) Remaining useful life estimation based on nonlinear feature reduction and support vector regression. Eng. Appl. Artif. Intell.vol. 26, pp.1751–1760.https://doi.org/10.1016/j.engappai.2013.02.006
9. Di Maio F, Tsui KL, ZioE (2012) Combining relevance vector machines and exponential regression for bearing residual life estimation. Mech. Syst. Signal Process.vol. 30, pp.405–427.https://doi.org/10.1016/j.ymssp.2012.03.011
10. Wang B, Lei Y, Li N, LiN (2018) A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Trans. Reliab.vol. 69, pp.401–412.DOI: 10.1109/TR.2018.2882682
11. Ali JB, Chebel-Morello B, Saidi L, Malinowski S, FnaiechF (2015) Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. Mech. Syst. Signal Process.vol. 56, pp.150–172.https://doi.org/10.1016/j.ymssp.2014.10.014
12. Guo L, Li N, Jia F, Lei Y, Lin J (2017) A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing.vol. 240, pp.98–109.https://doi.org/10.1016/j.neucom.2017.02.045
13. Cheng C, Ma G, Zhang Y, Sun M, Teng F, Ding H, YuanY (2020) A deep learning-based remaining useful life prediction approach for bearings. IEEE/ASME Trans. mechatronics.vol. 25, pp.1243-1254. DOI: 10.1109/TMECH.2020.2971503
14. Zhu J, Chen N, Peng W (2018) Estimation of bearing remaining useful life based on multiscale convolutional neural network. IEEE Trans. Ind. Electron.vol. 66, pp.3208–

3216.DOI: 10.1109/TIE.2018.2844856

15. Liu R, Yang B, Hauptmann AG (2019) Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network. IEEE Trans. Ind. Informatics.vol. 16, pp.87–96.DOI: 10.1109/TII.2019.2915536
16. Boumahdi M, Dron JP, Rechak S, Cousinard (2010) On the extraction of rules in the identification of bearing defects in rotating machinery using decision tree. Expert Syst. Appl.vol. 37, pp.5887–5894.https://doi.org/10.1016/j.eswa.2010.02.017
17. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Chebel-Morello B, Zerhouni N, Varnier C (2012) PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In IEEE International Conference on Prognostics and Health Management. pp. 1–8.
18. Randall RB, Antoni J (2011) Rolling element bearing diagnostics—A tutorial. Mech. Syst. Signal Process.vol. 25, pp.485–520. https://doi.org/10.1016/J.YMSSP.2010.07.017
19. Kim S, An D, Choi JH (2020) Diagnostics 101: A tutorial for fault diagnostics of rolling element bearing using envelope analysis in MATLAB. Appl. Sci.vol. 10, pp.1–23. https://doi.org/10.3390/app10207302
20. Sawalhi N, Randall RB, Endo H (2007) The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. Mech. Syst. Signal Process.vol. 21, pp.2616–2633. https://doi.org/10.1016/j.ymssp.2006.12.002
21. Dragomiretskiy K, Zosso D (2013) Variational mode decomposition. IEEE Trans. signal Process.vol. 62, pp.531–544.
22. Huang Y, Lin J, Liu Z, Wu W (2019) A modi fi ed scale-space guiding variational mode decomposition for high-speed railway bearing fault diagnosis. J. Sound Vib.vol. 444, pp.216–234. https://doi.org/10.1016/j.jsv.2018.12.033
23. Ding C, PengH (2005) Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol.vol. 3, pp.185–205.DOI: 10.1142/s0219720005001004
24. Ding S, Xu X, Nie R (2014) Extreme learning machine and its applications. Neural Comput. Appl.vol. 25, pp.549–556.https://doi.org/10.1007/s00521-013-1522-8
25. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing.vol. 70, pp.489–501.https://doi.org/10.1016/j.neucom.2005.12.126
26. Pachaud C, Salvetat R, Fray C (1997) Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces. Mech. Syst. Signal Process.vol. 11, pp.903–916.https://doi.org/10.1006/mssp.1997.0115