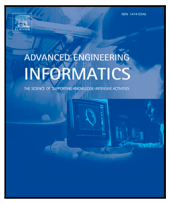
Advanced Engineering Informatics 62 (2024) 102909

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



Full length article

Explainable and interpretable bearing fault classification and diagnosis under limited data

L. Magadán [a,*], C. Ruiz-Cárcel [b], J.C. Granda [a], F.J. Suárez [a], A. Starr [b]

[a] Department of Computer Science and Engineering, University of Oviedo, Gijon, 33204, Spain
[b] School of Aerospace, Transport and Manufacturing, University of Cranfield, Bedford, MK43 0AL, United Kingdom

ARTICLE INFO

Keywords:
Industry 4.0
Fault diagnosis
Fault classification
Rotating machinery
Dynamic time warping
Stacked autoencoder
Explainable AI

ABSTRACT

Rotating machinery plays an essential role in various industrial processes such as manufacturing, power generation, and transportation. These machines, which include turbines, pumps, motors, compressors, and many others, are the heartbeats of numerous industries. The seamless operation of these machines is critical for the efficiency and productivity of these sectors. However, over time, these machines degrade and can suffer faults. One of the most critical components are bearings, which can suffer different types of faults. This paper presents a novel approach for bearing fault classification and diagnosis under limited data. A Monotonic Smoothed Stacked AutoEncoder (MS2AE) is used to infer a smoothed monotonic health index from raw bearing acceleration data. The MS2AE is trained using only healthy data, so this approach can also be used with recently comisioned equipment that has not failed yet. Then, using the evolution of the health index, a first faulty point is computed, so two stages are identified in the lifespan of the rotating machinery: healthy and faulty. Correlation matrices are computed to show the relationship of the health index with time-domain and frequency-domain features in order to provide explainability and validate the health index construction process. When the health index is classified as faulty, Dynamic Time Warping is applied between healthy samples and faulty samples to extract differences. Finally, based on a 1/3-binary tree 3 level kurtogram, these differences are filtered using a bandpass filter and converted to the frequency domain, where characteristic harmonics are used to identify the type of bearing fault. The explainability provided in the health index construction process makes the system useful in certain industries where black-box AI models cannot be trusted due to strict regulations. The classification and diagnosis system achieves robustness in fault classification under different working conditions by utilizing multiple bearing fault datsets. Its ability to be trained using only healthy data and the interpretability offered, makes it suitable for recently installed rotating machinery in real industrial facilities, without requiring qualified staff.

1. Introduction

In recent years, the increasing industrialization and digitalization of many companies have led to the need for more sophisticated, efficient and safe machinery. Most industrial processes are operated by mechanical and electro-mechanical systems, of which around 40% are composed of rotating machinery (RM). These components are key to ensuring the effectiveness and safety of the industrial processes, as it is one of the most prone to failure, given its long operating times and working conditions [1].

In electric motors, various types of faults contribute to its overall health. Bearing faults are around 50% of the faults. The stator accounts for around 30%, followed by the rotor with approximately 10%. The remaining 5% of the faults in RM are caused by gear faults and looseness. Bearing faults, in turn, can occur at different locations in the bearing, the most common being the outer race and the inner race. Depending on the place where it occurs, the damage can be more or less severe and can manifest itself in different ways [2,3]. Understanding and effectively diagnosing bearing faults is crucial for ensuring the reliable operation of the RM.

Although early fault diagnosis is a challenging task, the implementation of Internet of Things in the Industry 4.0, along with the constant evolution of Machine Learning (ML) algorithms, makes it affordable for most of the companies. Industrial Internet of Things (IIoT) facilitates the continuous monitoring of RM due to the deployment of sensor networks in industrial facilities.

Vibration is the most used variable for fault diagnosis in bearings. Temperature, acoustic emission or even current signature are also used. Vibrations analysis used to be an expensive procedure requiring the use

* Corresponding author.
E-mail address: magadanluis@uniovi.es (L. Magadán).

https://doi.org/10.1016/j.aei.2024.102909
Received 22 April 2024; Received in revised form 14 October 2024; Accepted 22 October 2024
Available online 5 November 2024

of highly skilled personnel to carry out in-place analysis [4], so it used to be done on a regular basis and most of the times only in critical machinery. However, with the development of wireless sensor networks and micro-electromechanical systems, it is possible to develop cost–effective condition monitoring systems to continuously monitor bearing vibrations. This enables early fault detection, so maintenance teams have enough time to plan shutdowns to replace affected parts, reducing maintenance costs. Magadán et al. [5] present a low-cost continuous monitoring system that collects the vibrations and temperatures of low-power electric motors and pumps using a multisensor module, sending them to the gateway where they are preprocessed, filtered and finally sent to the cloud. In the work presented by Dameshghi et al. [6], three phase current signals are collected and then processed to extract 8 features per current signal, using them for rotor fault diagnosis in wind turbines. Another example of IIoT system is the one presented by Nirwan et al. [7], where acoustic emissions and vibrations of a cylindrical style roller bearing are gathered by a magnetic mount sensor and a vibration analyzer.

Traditionally, classic signal processing techniques have been used for fault detection. However, the constant evolution of ML algorithms and their ability to make early fault diagnosis efficient using deep learning techniques, has made ML a very popular research field in recent years [8]. The authors in [9] use a Convolutional Neural Network (CNN) for fault diagnosis. Multilevel features are extracted from vibration signals and integrated by a module to merge these features based on their correlations. A smoothing mechanism is also included to avoid overfitting. Another work that uses CNNs for fault classification is the one presented by Peng et al. [10], where vibrations are measured using images. In this work, the phase difference between images obtained from a video of the RM is performed before applying the CNN, removing the signal-to-image transformation process. Meng et al. [11] propose a transfer learning method for fault diagnosis in rolling bearings using multi-scale CNNs and local central moment discrepancy. This method effectively transfers fault knowledge across different conditions by mapping vibration data to a shared space, using source domain labels and target domain pseudo-labels to categorize the subspace, and employing local central moment discrepancy for alignment. A novel fault diagnosis method of rolling bearings using subdomain adaptation with an improved vision transformer network is presented by Liang et al. [12], using the local maximum mean discrepancy as the alignment method and enhancing the traditional vision transformer network with deformable convolution modules and recurrent neural networks. In the work presented by Kaya et al. [13], continuous wavelet transform is used for extracting time–frequency color scalogram images, using 2D CNNs to conduct fault size prediction. The main drawback of these proposals is that deep learning algorithms usually require large amounts of data for fault classification and diagnosis [14].

Autoencoders have taken on a very important role in the area of fault diagnosis. In the work presented by Zhang et al. [15], a stacked autoencoder is used to build a health index representing the condition of the rotating machinery. First, engineering features are manually extracted from the raw vibration samples and then combined by means of the stacked autoencoder. The health index obtained is later smoothed using exponentially weighted moving average. A stacked autoencoder is also used by Tian et al. [16] to fuse four selected features into the health index. These features are firstly smoothed, normalized between 0 and 1 and then selected according to their monotonicity. Once the health index is obtained, the health prognostics of the rotating machinery is carried out. Fan et al. [17] avoid the smoothing process during feature extraction by adding a Lowess filter after each hidden layer to perform denoising operations in the stacked autoencoder. However, all these works require both healthy and faulty data to train the model and handcrafted engineering features.

The authors in [18], use one-dimensional local binary pattern (1D-LBP) method to analyze vibration signals collected at different motor speeds from bearings with diverse intentionally induced faults. This method involves transforming vibration signals into the 1D-LBP plane and extracting statistical features, which are subsequently classified using the gray relational analysis model. Local Binary Pattern is also used in the work of Kaplan et al. [19], where texture features are extracted from gray-scale images to classify the bearing vibration signals using various machine learning models. These studies highlight the method's potential for precise and early bearing fault detection and its versatility across different signal types. Despite its high accuracy and applicability to real-world data, there is still a need of both healthy and faulty data to train the models, also requiring manually feature extraction to perform the classification tasks.

In the work proposed by Meng et al. [20], a Generative Adversarial Network (GAN) is used to generate synthetic data to train the model, using the Wasserstein distance to alleviate the vanishing gradient problem. Furthermore, an attention mechanism is applied on the blocks obtained by the convolutions. However, GANs are highly susceptible to overfitting, leading to a repetitive generation of similar elements. To avoid this, Liu et al. [21] present a Variational Information Constrained GAN, where data synthesis is enhanced by incorporating an encoder into the discriminator, while stable training and convergence is guaranteed through a variational information constraint technique. In the work presented by Li et al. [22] an event data augmentation method is used to introduce variations of event patterns. The event records are transformed into typical data samples by using a vibration event representation, which is later processed by a deep convolutional neural network model. Finally, a clustering method is used to improve the pattern recognition performance. However, these methods are sensitive to hyperparameter configuration, as an inadequate balance between the generator and discriminator can either hinder learning or lead to a persistent generation of similar samples.

Although ML algorithms obtain competitive results, most of them are black-box algorithms, in some cases difficult to understand the reasoning behind their decisions, and consequently providing results that are not useful for practitioners. Interpretable machine learning aims to create models that have inherent interpretability, while explainable machine learning seeks to offer retrospective explanations for black box models [23]. The development of new ML algorithms where explainability is prioritized is essential, not only to enhance model interpretability but also to identify errors or biases in the model, providing the end-user with enough information on how the results have been obtained, so non-qualified staff can comprehend them in an easy way. Explainabillity is also key for highly regulated sectors which are safety critical and require all processes to be transparent and traceable for certification. The authors in [24] propose an aircraft engine monitoring based on variational encoding, consisting of a recurrent encoder and a regression model, where the latent space generated by the encoder is used as a map for providing insights about the data and visual interpretability of the decisions taken by the model. Another work in the context of explainable fault diagnosis is the one presented by Sinan Li et al. [25]. In this work, a Multilayer Gradient-weighted Class Activation Map is used, leveraging gradients from multiple convolutional layers in order to generate activation maps, which are then combined using layer-weighted summation to create a main comprehensive activation map. In the work presented by Yang et al. [26], the understanding of mechanical signal processing in fault diagnosis using deep learning is explored. By visualizing the diagnostic knowledge learned by deep neural networks through neuron activation maximization and saliency map methods, the study provides an intuitive observation of discriminative features for different machine health conditions. Experimental investigations on two datasets confirm the relationship between data-driven methods and conventional fault diagnosis knowledge, highlighting the effectiveness of deep learning in capturing and interpreting critical diagnostic features. Another work where explainability is provided is the one by Xu et al. [27], where a Copula network deconvolution-based framework for explainable fault diagnosis in semiconductor wafer fabrication is proposed. It addresses

the complexity of the manufacturing system by constructing a complex network correlation diagram with parameters as nodes. The framework includes a nonlinear correlation metric model based on adaptive Copula function selection and a network deconvolution-based fault diagnosis method to identify direct correlations. Physics-informed denoising loss is incorporated to improve the model's robustness and interpretability by the authors in [28]. A novel interpretable waveform segmentation model for bearing fault diagnosis is presented, using nested U-Net to enable pixel-level extraction of fault-related information through signal segmentation technology. Lovász-Softmax loss is employed to manage imbalanced pixel distribution and enhance spatial recognition. Although explainability is provided, the previous works have some limitations as they need healthy and faulty data for training, requiring balanced datasets.

To cover these gaps, this paper proposes a novel, explainable and interpretable bearing fault classification and diagnosis system that only requires healthy data for training and parameter tuning. It consists of a Monotonic Smoothed Stacked Autoencoder (MS2AE) that receives the raw vibration data gathered from the monitored RM to compute a Health Index (HI) value, which determines the stage of degradation. Once the HI value is filtered, a subsequent processing stage based on Dynamic Time Warping (DTW) is used to determine the type of fault and the level of degradation by looking at the differences between a newly acquired faulty sample and an existing healthy sample baseline. The performance of this system has been analyzed using two of the most common bearing fault datasets (IMS [29] and XJTU-SY [30]), proving the robustness of the proposed system under different operating conditions.

This work presents significant advancements in the field of fault diagnosis for rotating machinery by introducing the Monotonic Smoothed Stacked Autoencoder model, which addresses several key challenges in the domain. First, the model demonstrates efficacy in scenarios with limited data, requiring only healthy data for its training phase, while other models need healthy and faulty data for training. This reduces the dependency on extensive fault datasets, making it practical for real-world applications. Second, the model eliminates the need for manual feature extraction, as it directly processes raw vibration data samples and integrates them into a comprehensive HI. This streamlines the data processing pipeline and enhances usability. Third, the explainability of the MS2AE model ensures that the derived HI values have a strong correlation with common engineering features, thereby enhancing their reliability and acceptance by practitioners. Four, the model exhibits robustness across various operating conditions and motor loads, ensuring consistent performance in diverse environments. Lastly, the interpretability of the results obviates the need for highly specialized personnel, as the diagnosed faults are presented in a clear and understandable manner. These contributions collectively advance the state of the art in fault diagnosis, offering a more efficient, reliable, user-friendly solution.

The rest of the paper is organized as follows. Section 2 presents the bearing faults analyzed in this work. The data and the methods used in the developed system are presented in Section 3. Section 4 describes in detail each of the steps of the proposed fault classification and diagnosis system. A description of the tests carried out and the results obtained by the proposed fault classification and diagnosis system are shown and discussed in Section 5. Finally, the concluding remarks and future work are outlined in Section 6.

2. Bearing faults

A RM is a device usually composed of a stationary part and a rotating part. The bearings are the elements enabling the motion between the rotating and the stationary parts, so they are essential for the correct performance of the RM. A typical ball bearing comprises an outer race, inner race, balls, and a cage. The outer race is stationary, while the inner race rotates. The balls roll between the races, guided by the cage, reducing friction and facilitating smooth motion. Fig. 1 shows the parts of a bearing. In this work, the bearing faults considered are outer-race, inner-race, ball and cage bearing faults. Each of them are explained in more detail below.

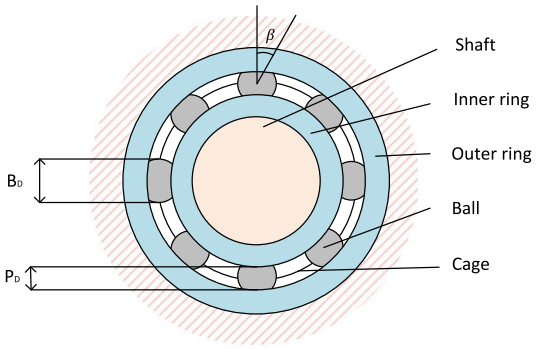


Fig. 1. Ball bearing scheme.

- Outer-race bearing faults: they correspond to faults that occur in the outer ring of the bearing, which encloses the rolling elements to support and guide the rotating shaft [31]. These faults are usually caused by excessive load, improper lubrication or even misalignment. Noise, excessive vibrations and a reduction in the productivity and the efficiency of the RM are the most common consequences. Furthermore, if the fault is not promptly addressed, the fault can be extended to the housing or the shaft of the RM, also damaging external elements. This fault can be detected by analyzing the spectrum of the vibration signal, which will show harmonic peaks in the Ball Pass Frequency Outer (BPFO) or sidebands separated by the BPFO from a central frequency component due to modulation. BPFO is computed as indicated in Eq. (1), where N_B corresponds to the number of balls of the bearing, B_D is the ball diameter, P_D is the pitch diameter, β is the contact angle and F is the shaft frequency.

$$BPFO = F \cdot \frac{N_B}{2} \cdot \left(1 - \frac{B_D}{P_D} \cdot \cos(\beta)\right) \tag{1}$$

- Inner-race bearing faults: in this case, the fault occurs in the inner ring of the bearing, which is in contact with the rolling elements for smooth rotation [31]. As in the case of outer-race bearing faults, excessive load, fatigue, misalignment and improper lubrication are the main causes. These faults cause excessive vibrations in the RM, increasing the noise, causing damage to other components such as the shaft or housing of the RM and reducing its performance. The spectrum of a signal where an inner-race bearing fault exists is characterized by several harmonic peaks in the Ball Pass Frequency Inner (BPFI), which is calculated as shown in Eq. (2).

$$BPFI = F \cdot \frac{N_B}{2} \cdot \left(1 + \frac{B_D}{P_D} \cdot \cos(\beta)\right) \tag{2}$$

- Ball bearing faults: they refer to breakdowns in the metal balls located between the inner and the outer races [31]. Excessive vibrations, noise and heat are the main consequences of this fault, leading to a reduction in the performance of the RM, and in some cases increasing its energy consumption. The spectrum of a ball bearing faulty signal is characterized by harmonics in the Ball Spin Frequency (BSF) or sidebands separated by the BSF from a central frequency component due to modulation. Eq. (3) shows how to calculate BSF. This fault is usually accompanied by outer- or inner-race faults.

$$BSF = F \cdot \frac{P_D}{B_D} \cdot \left[1 - \left(\frac{B_D}{P_D} \cdot \cos(\beta)\right)^2\right] \tag{3}$$

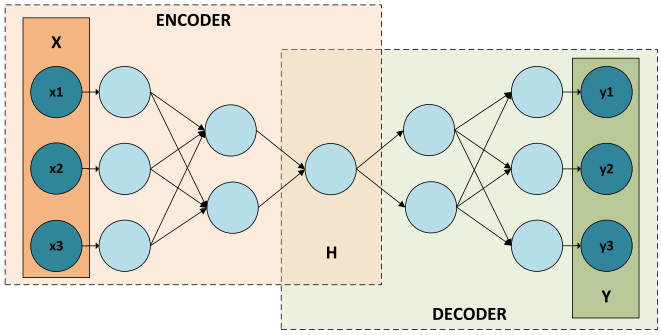


Fig. 2. Autoencoder diagram.

- Cage bearing faults: these imply malfunctions in the cage of the bearings. Its function is to maintain the proper spacing and alignment of rolling elements between the inner and outer races [31]. Wear of the cage, detachment of some of the cage elements or friction with other parts of the bearing are some of the reasons for this kind of fault, leading to excessive vibration, noise and also reducing the performance of the RM. The spectrum of a cage bearing faulty signal is characterized by the existence of harmonics in the Fundamental Train Frequency (FTF), and usually accompanied by outer- or inner-race faults. FTF can be calculated as shown in Eq. (4)

$$FTF = F \cdot \frac{1}{2} \cdot \left(1 - \frac{B_D}{P_D} \cdot \cos(\beta)\right) \tag{4}$$

3. Materials and methods

In this section, the main methods used in the proposed fault classification and diagnosis system are explained. Descriptions of the datasets used are also included.

3.1. Stacked autoencoder

An autoencoder (AE) is an unsupervised three-layer neural network composed of an encoder and a decoder. The former converts the high-dimensional input into a low-dimensional output, while the latter reconstructs the original signal from the encoder output [32]. They are usually used for feature extraction, especially for fault diagnosis and health prognostics in different domains. Thanks to its ability to reduce the dimensionality of the input data, they are also useful for denoising the input signal.

Stacked Autoencoders (SAEs) are composed of multiple AE layers, where the output of a layer is used as the input of the next layer. The weights of each of the layers are connected and fine-tuned in order to get the SAE weights [33]. Unlike other dimensional reduction techniques such as principal component analysis, SAEs enable modeling nonlinearities in the projection from the initial dimensional space. Fig. 2 shows an schematic diagram of an autoencoder network, where the high-dimensional input (X), encoder, low-dimensional output (H), decoder and reconstructed signal (Y) can be observed. The encoder and decoder have symmetric dimensions.

3.2. Dynamic time warping

Dynamic Time Warping (DTW) is an algorithm that is commonly used to measure the similarity between time-dependent sequences, computing the warping path between two different signals, which can be of different length, obtaining as a result the path warping values and the distance between both signals [34]. Its main objective is to achieve optimal alignment between two vector sequences by iteratively adjusting the time axis. This iterative process transforms the time axis to align the two signals, acting in this way as a linear mapping of the time axis.

Given two signals X and Y with lengths N and M respectively, DTW finds the optimal alignment by warping in time X to Y, trying to minimize the distance between them. For achieving this, a $N \times M$ distance matrix D is constructed, where $D(i,j)$ contains the local similarity between $X(i)$ and $Y(j)$, usually computed as the euclidean distance. Once the D matrix is computed, the warping path is selected as the one that minimizes the total distance between signals X and Y. An example of this is shown in Fig. 3.

3.3. Kurtogram

In real environments, raw vibration signals are often complex to analyze, as they contain a mixture of frequencies related to normal machinery operation, possible faults and noise. Filtering these signals is vital for effective analysis, so it is necessary to know the frequency bands in which the relevant information related with faults is located.

A Kurtogram is the representation of the values of Spectral Kurtosis (SK) for different frequencies and window lengths [35]. SK is commonly used to detect temporary changes in a signal, computing the kurtosis of the signal at different frequency ranges, in order to highlight frequency bands containing hidden non-stationary signals. SK is directly related with the impulsivity of the signal, increasing as the impulsivity of the signal increases. Highlighting a high kurtosis is crucial as it indicates a significant presence of non-stationary signals, often associated with faults. This is particularly useful for the detection of bearing faults during early stages of degradation, as the impulse-nature of the signals produced is very characteristic but tends to get hidden behind other vibration components. Thus, SK may detect the frequency bands where the signal has higher impulsivity. Therefore, focusing on the range with the highest kurtosis becomes imperative in fault diagnosis [36].

Kurtograms are commonly used in bearing fault classification and diagnosis before applying envelope analysis to the raw signal in order to determine the frequency range where the kurtosis value is higher. An example of 1/3-binary tree 3-level kurtogram can be observed in Fig. 4. The level indicates the scale in which the main signal has been decomposed using the wavelet transform. In the example presented in Fig. 4, the frequency range where the kurtosis is higher can be observed at level 2.6, which corresponds to the range $\left[\frac{2}{6}f_n, \frac{1}{2}f_n\right]$, where f_n is the Nyquist frequency.

3.4. Datasets

In the context of fault classification in rotating machinery, the quality of the data used plays an important role. Publicly available datasets with run-to-failure bearing fault vibration data facilitate the replication of the results by the research community. These datasets emulate bearing load and speed conditions in real industrial environments. The three most popular bearing datasets are IMS [29], XJTU-SY [30] and PRONOSTIA [37] bearing datasets. However, the latter is not useful in the context of this work as it does not specify the kind of bearing fault of each test. For this reason, only IMS and XJTU-SY bearing datasets are used. The following subsections include descriptions of these datasets.

3.4.1. IMS dataset

Four bearings installed on a shaft powered by an electric motor were used in the experiments of this dataset. The rotational frequency was kept constant at 33.33 Hz while applying a radial load of 26.7 kN. In each of the bearings two PCB 253B33 accelerometers were installed, measuring the vibrations in G's at a sampling frequency of 20.48 kHz [29]. The IMS dataset is divided into three datasets, each one containing vibration signals from a single run-to-failure experiment, with samples of 1 s duration. A summary of the datasets is provided below.

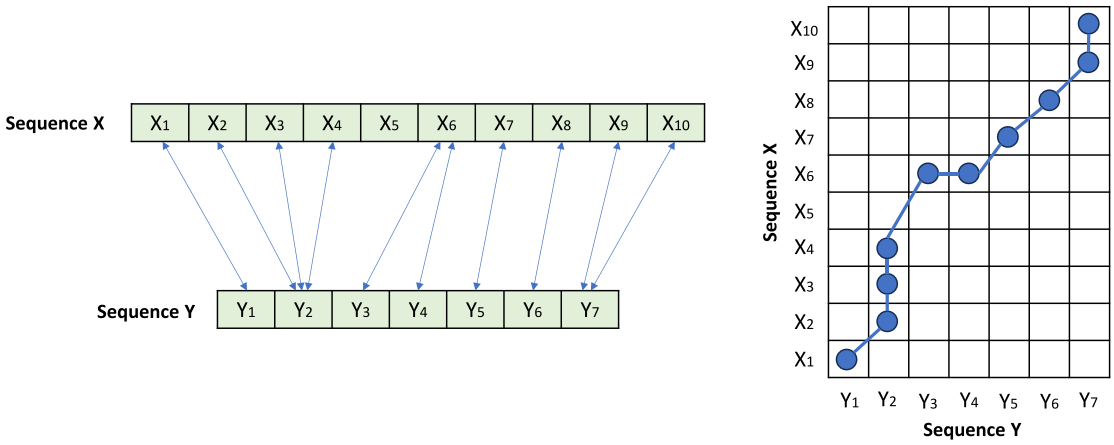


Fig. 3. Warping path obtained with DTW.

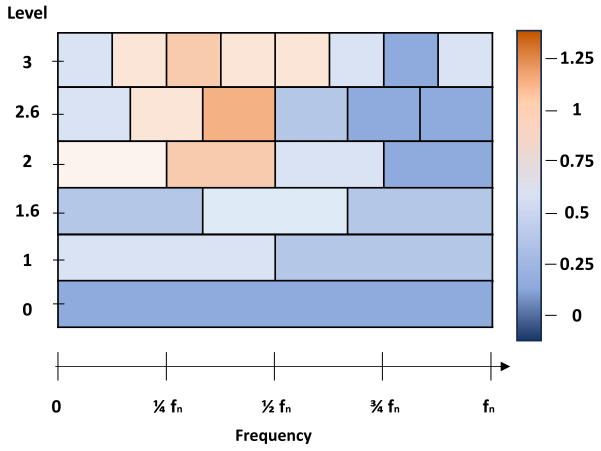


Fig. 4. 1/3-binary tree 3 level kurtogram.

- IMS-1: It consists of a total of 2156 samples separated 10 min, except for the first 43 samples that were taken every 5 min. A total of 14 days 20 h of continuous operation is recorded during this experiment. At the end of the experiment, an inner-race bearing fault has occurred in the third bearing, while a ball bearing fault occurs in the fourth bearing.
- IMS-2: It consists of a total of 984 samples separated 10 min. This experiment corresponds to a total of 6 days 20 h, showing an outer-race bearing fault at the end of the experiment in the first bearing.
- IMS-3: 6324 samples have been collected in this experiment separated 10 min. This corresponds to a total of 43 days 22 h of continuous operation of the RM. At the end of the experiment an outer-race bearing fault occurs in the third bearing.

3.4.2. XJTU-SY bearing dataset

The test bench used in this dataset is composed of an electric motor, a pair of bearings and some elements that allow adjusting the load of the system or the radial force applied to the bearings. It has a total of 15 experiments, where vibration data is collected from the drive-end bearing in G's at different working conditions using two PCB 352C33 accelerometers. Experiments are conducted at 1-minute intervals. Each experiment includes 1-second vibration samples collected at a sampling frequency of 25.6 kHz, resulting in a total of 25,600 raw acceleration values per sample [30]. Three working conditions have been considered. The first working condition corresponds to an electric motor where the shaft frequency is 35 Hz and the load supported 12 kN. In the second working condition the shaft frequency has been increased to 37.5 Hz and the load reduced to 11 kN. Finally, in the third working condition the shaft frequency is increased to 40 Hz and the load reduced to 10 kN. The length of these experiments is quite shorter than the ones of IMS dataset. For this reason, only the experiments with a length longer than 8 h have been selected as they are more realistic. A summary of the selected experiments is presented below.

- XJTU 2-1: 491 samples corresponding to 8 h 11 min of continuous operation have been recorded. The electric motor is working at 37.5 Hz with a load of 11 kN. At the end of the experiment an inner-race bearing fault occurs in the bearing.
- XJTU 2-3: It consists of 533 samples corresponding to a total of 8 h 53 min of continuous operation. As in the previous case the shaft frequency is 37.5 Hz and the load supported 11 kN. In this case a cage bearing fault occurs at the end of the experiment.
- XJTU 3-1: A total of 2538 samples have been collected. They correspond to 42 h 18 min of continuous operation. In this experiment, the shaft frequency is 40 Hz while the load supported is 10 kN. At the end of the experiment, an outer-race bearing fault occurs.
- XJTU 3-4: This experiment consists of 1515 samples corresponding to 25 h 15 min of continuous operation. The working conditions are the same as in the XJTU 3-1 experiment: a shaft frequency of 40 Hz and a load of 10 kN. In this case, an inner-race bearing fault occurs.

4. Proposed fault classification and diagnosis system

A novel explainable and easily interpretable bearing fault classification and diagnosis system has been developed. Fig. 5 shows the three steps in which this fault classification and diagnosis system is divided: health index construction, health stage division and bearing fault classification and diagnosis. These steps are described in more detail in the following subsections.

4.1. Health index construction

The first step consists in constructing the Health Index (HI) value, which determines the health stage of the monitored RM. The lower the HI, the better the health condition of the RM. To avoid the hand-crafted feature extraction process at this stage, a Monotonic Smoothing Stacked Autoencoder (MS2AE) has been developed. The MS2AE is characterized for including two customized layers. The first layer reduces the dimensionality by obtaining monotonic features (monotonicity layer), as degradation is monotonic and so must be the HI. The second layer is used to smooth the features after reducing the dimensionality of the

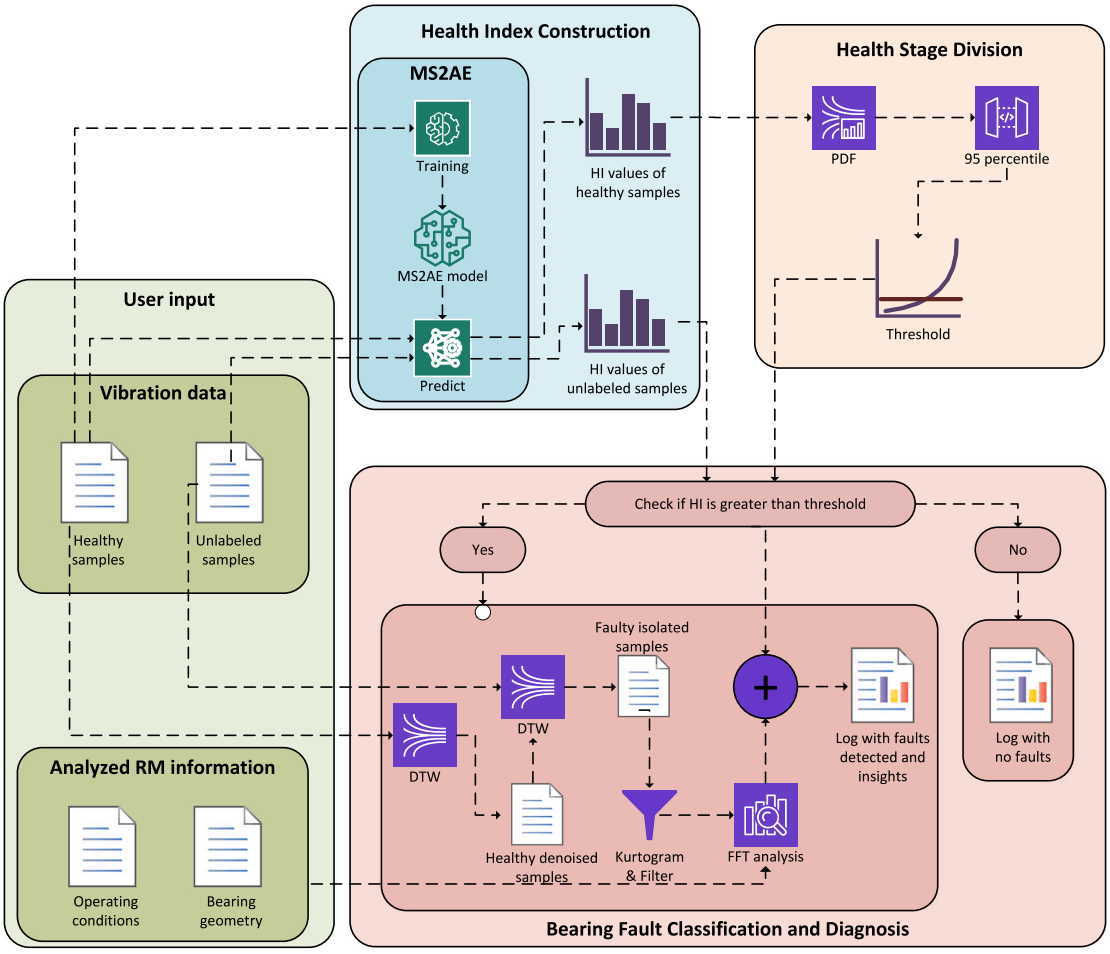


Fig. 5. Proposed bearing fault classification and diagnosis system.

input signal (smoothing layer). A dense layer is also included between the monotonicity and the smoothing layers.

The MS2AE network is trained using only healthy data, so no faulty data is needed, removing the need of balanced datasets, and consequently making the proposed technique useful to be applied with machinery where no faulty data is available. To build the MS2AE network TensorFlow and the Keras framework have been used. The structure and hyperparameters of the MS2AE network have been selected by grid search optimization and are included in Table 1. The structure followed is IS-3500-700-200-1-200-700-3500-IS, where IS indicates the size of the input sample, 3500 nodes in the monotonicity layer, 700 nodes in a dense layer to reduce the monotonicity features dimensionality, 200 nodes in the smoothing layer and 1 node to output the health index of the input sample. The size of the first and last layers of the MS2AE network varies depending on the dataset used, being 20 480 data points in the IMS dataset and 25 600 data points in the XJTU-SY dataset.

After training the MS2AE model, the HI values of the healthy samples and also the ones from the samples that are going to be analyzed (unlabeled) are predicted by the MS2AE network. Fig. 6 shows an example of the HI values obtained for a synthetic signal, where the green line corresponds to the known healthy samples and the blue line to the unlabeled samples.

Table 1
MS2AE structure and hyperparameters.

MS2AE structure	IS-3500-700-200-1-200-700-3500-IS
Optimizer	adam
Learning rate	0.001
Epochs	5
Batch size	64

4.2. Health stage division

Once the HI values are obtained for both healthy and unlabeled samples, the threshold that determines when a sample is healthy and when it is faulty is calculated. The HI values of the healthy samples are assumed to follow a normal distribution, so the Power Density Function (PDF) is calculated, and the 95th percentile is selected as the threshold. To cope with outliers exceeding the threshold, not all the HI values that exceed the threshold are considered faulty samples. As degradation is monotonic and non-reversible, five consecutive HI values must exceed the threshold to be considered as a faulty sample. In the case of the datasets used in this work, the first sample that is classified as a faulty sample is known as First Faulty Point (FFP). An example of health stage division and the FFP identification are shown in Fig. 7. The green line corresponds to the HI values of healthy samples, the blue line to HI values of unlabeled samples, the red line to the threshold and the red circle to the FFP. The green area corresponds to the sample area that has been determined as healthy, while the red corresponds to faulty samples.

4.3. Bearing fault classification and diagnosis

The last of the steps consists in diagnosing the bearing faults of the RM. If the HI value is classified as faulty, a deeper analysis is carried out

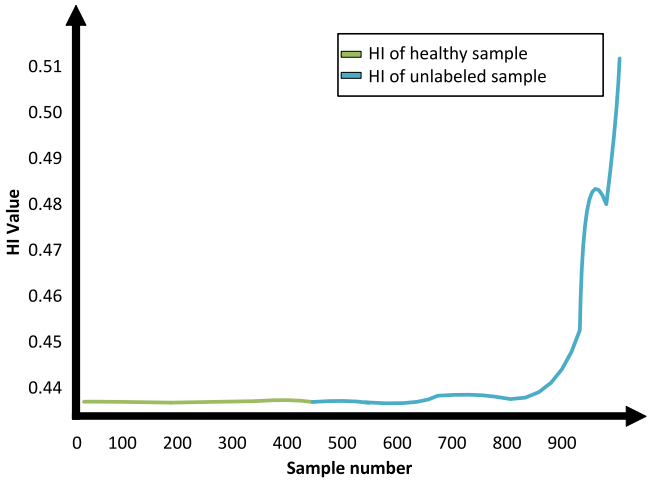


Fig. 6. Health index construction.

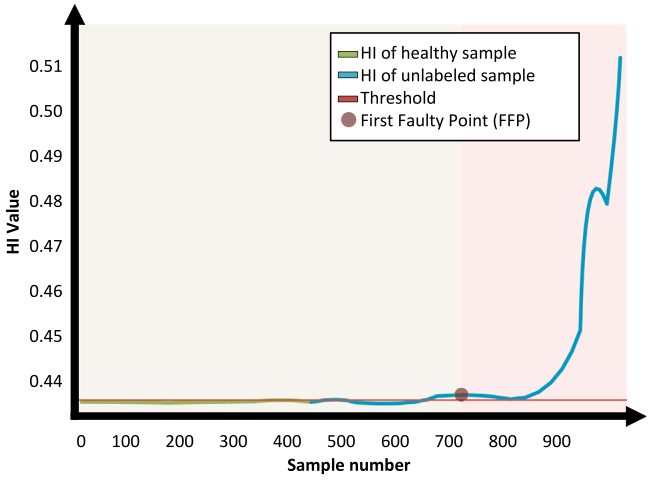


Fig. 7. Health stage division.

in order to determine the kind of fault that is causing the degradation of the RM and its level.

Firstly, DTW is recursively applied to the healthy samples, keeping the common patterns between them, and consequently generating a single denoised healthy sample. The obtained denoised healthy sample is then used as a comparison baseline to isolate the fault. DTW is applied again between the faulty classified sample and the denoised healthy sample, but in this case keeping the difference between them and isolating the faulty component, looking for the differences rather than similarities between signals.

Secondly, a filtered version of the isolated faulty sample is obtained. Incipient bearing faults usually manifest at high frequencies [38]. Therefore, a bandpass filter is used to preserve relevant frequencies. A 1/3-binary tree 3-level kurtogram from the isolated faulty sample is analyzed to determine the frequency range that will be used as the bandpass filter. Levels 0, 1 and 1.6 from the kurtogram are not taken into account as the width of the frequency ranges determined by the kurtogram is not small enough. From levels 2, 2.6 and 3 the first and last frequency ranges are discarded as they correspond to too low or too high frequencies. The frequency range with the highest kurtosis within the considered levels is selected. A fourth order butterworth bandpass filter using this frequency range is then applied and the envelope of the signal obtained.

Thirdly, once the envelope of the filtered isolated faulty sample is obtained, the FFT is applied to convert it from the time domain to the frequency domain. The FFTs of the non-filtered and the envelope of the filtered faulty samples are analyzed checking if the harmonics corresponding to 1X, 2X, 3X, 4X, 5X and 6X BPFO, BPFI, BSF and FTF frequencies appear. This FFT analysis is carried out in both non-filtered and filtered samples because depending on the stage of degradation, bearing faults manifest themselves in different ways: sidebands due to modulation at early stages, then bearing fault components and multiple harmonics appear as the fault progresses. Also, the bearing geometry or even the operating conditions of the analyzed RM change the frequencies where these specific harmonics appear. The SK-based filtering isolates the small impact-like components of the signal that are causing the modulation, making it sensitive to early stage degradation. Based on the number of harmonics that appear in the non-filtered and filtered samples, and the HI value of the faulty sample, the kind of bearing fault (outer race, inner race, ball or cage) and the level of degradation (early, medium or last) of the analyzed sample are diagnosed. In order to classify the stage of degradation the HI value of the analyzed sample is compared with two empirically obtained thresholds, one that determines the beginning of a medium degradation stage ($Threshold_{MD}$) and the other for the beginning of a last degradation stage ($Threshold_{LD}$), which are computed as shown in Eqs. (5) and (6) respectively, where P_{95} corresponds to the 95th percentile and $HI_{healthy}$ to the health index values of the healthy samples. Those equations have been determined after observing a common pattern in the degradation of the datasets used in this study. If the HI value does not exceed the value of $Threshodld_{MD}$ it is classified as an early degradation stage.

$$Threshold_{MD} = \left(max\left(HI_{healthy}\right) - P_{95}\left(HI_{healthy}\right)\right) \cdot 50 \tag{5}$$

$$Threshold_{LD} = \left(max\left(HI_{healthy}\right) - P_{95}\left(HI_{healthy}\right)\right) \cdot 100 \tag{6}$$

At the end of this procedure, a log where the kind of fault detected, indicating the harmonics that have appeared in both non-filtered and filtered samples, together with its level of degradation is generated. This log also contains correlation matrices obtained during the health index construction and the health stage division stages. These matrices provide explainability about the operation of the MS2AE network showing the correlation of the computed HI value with classical time and frequency features such as root mean square (RMS), Kurtosis, Skewness or the amplitude of some frequency components as the fault progresses. The log also includes the kurtograms and FFTs of the non-filtered and filtered samples, where the harmonics that have been detected can be easily observed. Interpretability is provided, so non-qualified staff can understand the outcome of the system. The strength of this method is that a single HI obtained automatically can encompass features that are designed to be sensitive at different stages of degradation, while ideally improving detection time.

5. Results and discussions

This section is divided into two subsections. Section 5.1 contains an example of the application of the proposed bearing fault classification and diagnosis system on the IMS-2 dataset, where an outer-race bearing fault occurs at the end of the experiment. Section 5.2 summarizes the results obtained with the rest of datasets.

5.1. IMS-2 dataset analysis

IMS-2 dataset is composed of 984 samples. For training purposes, the first 300 samples have been selected as healthy samples in order to train the MS2AE network, while the remaining 684 samples have been considered as unlabeled samples. No feature extraction is performed, so the raw healthy samples are used directly without any preprocessing to train the model. The HI values obtained for both healthy (green line) and unlabeled (blue line) samples are shown in Fig. 8. Before further processing, it is necessary to ensure that the HI values of healthy

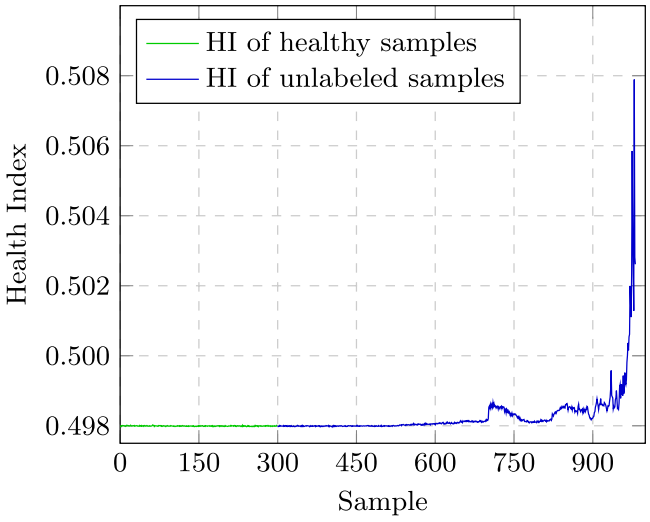


Fig. 8. Health index values of IMS-2 dataset.

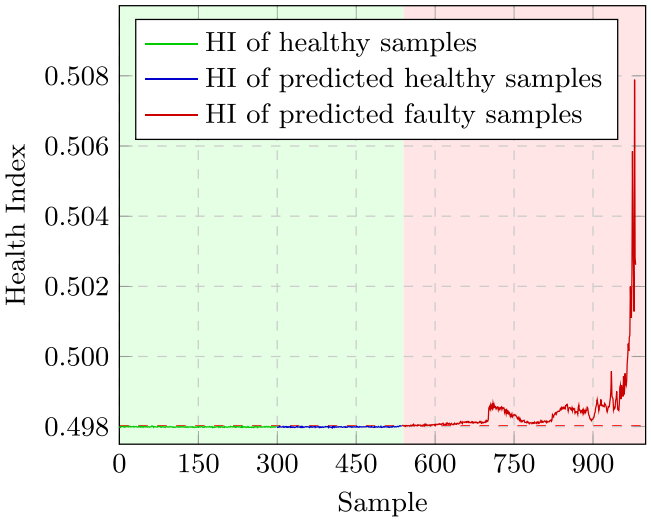


Fig. 10. Health stage division of IMS-2 dataset.

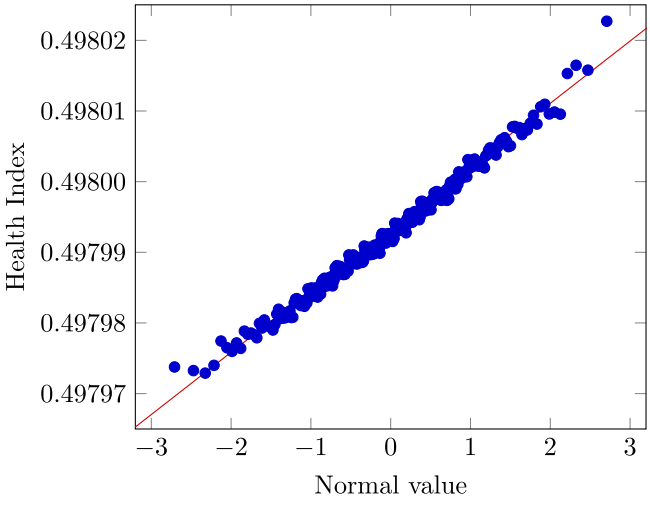


Fig. 9. Normal Q-Q plot diagram of healthy samples from IMS-2 dataset.

samples that have been used follow a normal distribution. This is done by computing the normal Q-Q plot of the healthy samples, which is shown in Fig. 9. According to the figure, it can be assumed that the HI values of the healthy samples is normally distributed.

The PDF of the HI values of the healthy samples is computed and the 95th percentile is selected as the threshold. As stated in Section 4, the FFP is determined as the first five consecutive values that are over the threshold to cope with outliers. Fig. 10 shows the health stage division after applying this procedure. In this case, the threshold is 0.49802 and the FFP is sample #536.

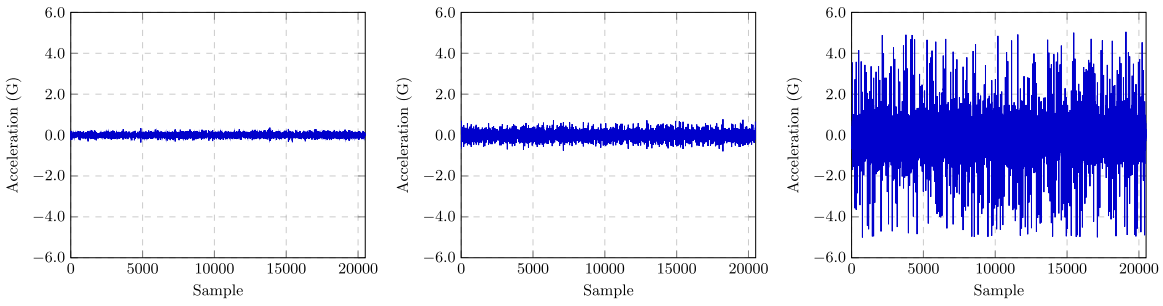
The last part of the procedure is focused on fault diagnosis, indicating which harmonics are contributing to the fault and the degradation stage of the RM. A recursive DTW is applied to all the healthy samples, removing the noise from these samples and only keeping relevant information. Next the DTW is applied between each faulty sample and the denoised healthy sample to isolate the fault signature in the faulty sample. To provide different examples of this, three faulty samples have been selected: sample #536, which corresponds to the FFP, sample #871, which corresponds to a sample in a medium degradation stage, and sample #979, corresponding to one of the last samples that are in the last degradation stage. After isolating the fault, the kurtogram is computed for each of the samples to obtain the frequency ranges for the bandpass filters and envelope analysis is applied.

Fig. 11 shows the isolated faulty samples for sample #536 (a), sample #871 (b) and sample #979 (c). The kurtograms of the isolated faulty samples are shown in Fig. 12, highlighting the frequency ranges that should be used for filtering. From Fig. 12(a) it can be seen that the frequency range that should be used for filtering sample #536 is [3413.33 Hz–5120 Hz]. For sample #871, the frequency range that should be used as bandpass filter is [2560 Hz–5120 Hz], as seen in Fig. 12(b). Fig. 12(c) shows that the bandpass filter for sample #979 should be applied in the frequency range [6826.67 Hz–8533.33 Hz].

Finally, the FFT is applied to the non-filtered and filtered isolated faulty samples, in order to analyze the frequency spectrum and search for the presence of harmonics which are directly related with the developing fault. In the case of the three samples analyzed, the FFTs of the non-filtered samples do not clearly show the harmonics related with the outer-race bearing fault. However, the harmonics corresponding to the outer-race bearing fault appear in the frequency spectrum of the filtered samples. Fig. 13(a) shows the frequency spectrum of sample #536, where 1X and 2X BPFO harmonics can be observed. 1X, 2X, 3X, 4X and 5X BPFO harmonics can be observed in Fig. 13(b), where the frequency spectrum of sample #871 is shown. 1X, 2X, 3X, 4X and 6X BPFO harmonics are present in the frequency spectrum of sample #979 (see Fig. 13(c)).

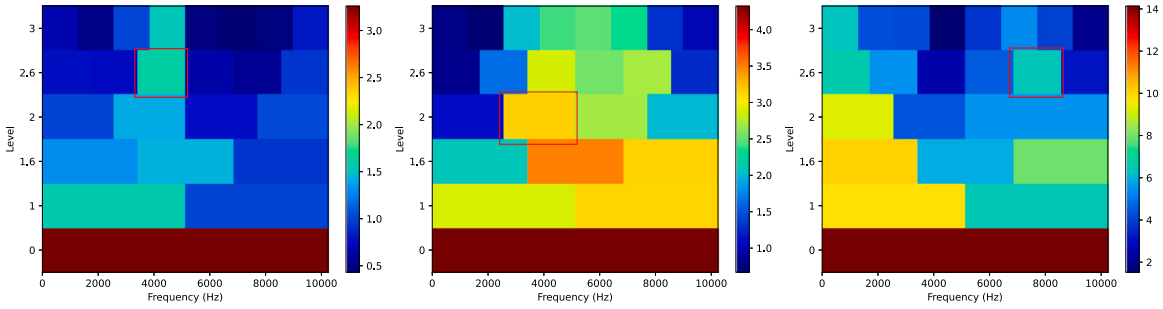
Based on the HI values and the frequency spectra from the filtered samples, the proposed fault classification and diagnosis system determines the level of degradation and the main fault that is causing the degradation. The level of degradation is determined comparing the HI value of the analyzed sample with the thresholds shown in Eqs. (5) and (6) . The system determines that sample #536 corresponds to an early degradation stage, sample #871 to a medium degradation stage, while sample #979 to a last degradation stage, all of them caused by outer-race bearing faults.

Apart from fault classification and diagnosis, the proposed system provides explainability about the health index. The correlation matrix between the HI value and time-domain and frequency-domain features is computed for all the samples in the IMS-2 dataset. The time-domain features analyzed are RMS, skewness (Sk), kurtosis (K), crest factor (CF), shape factor (SF), impact factor (IF) and margin factor (MF), while the frequency-domain features are amplitudes of the 1X fundamental frequency, 1X BPFO, 1X BPFI, 1X BSF and 1X FTF from the sample (Fund nf, BPFO nf, BPFI nf, BSF nf and FTF nf), and a filtered version of the same sample (Fund f, BPFO f, BPFI f, BSF f and FTF f). The filtered sample is obtained by applying a fourth order butterworth bandpass filter in the frequency ranges determined by the kurtogram



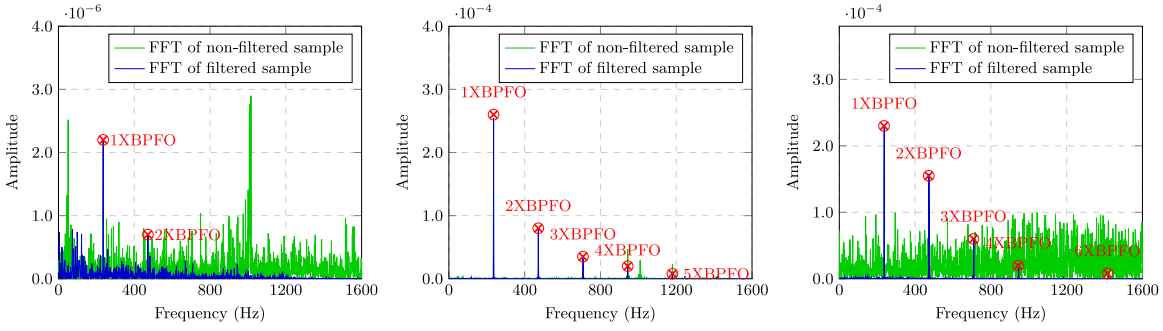
(a) Faulty Sample #536 (b) Faulty Sample #871 (c) Faulty Sample #979

Fig. 11. Isolated faulty samples of IMS-2 dataset.



(a) Faulty Sample #536 (b) Faulty Sample #871 (c) Faulty Sample #979

Fig. 12. Kurtograms of isolated faulty samples of IMS-2 dataset.

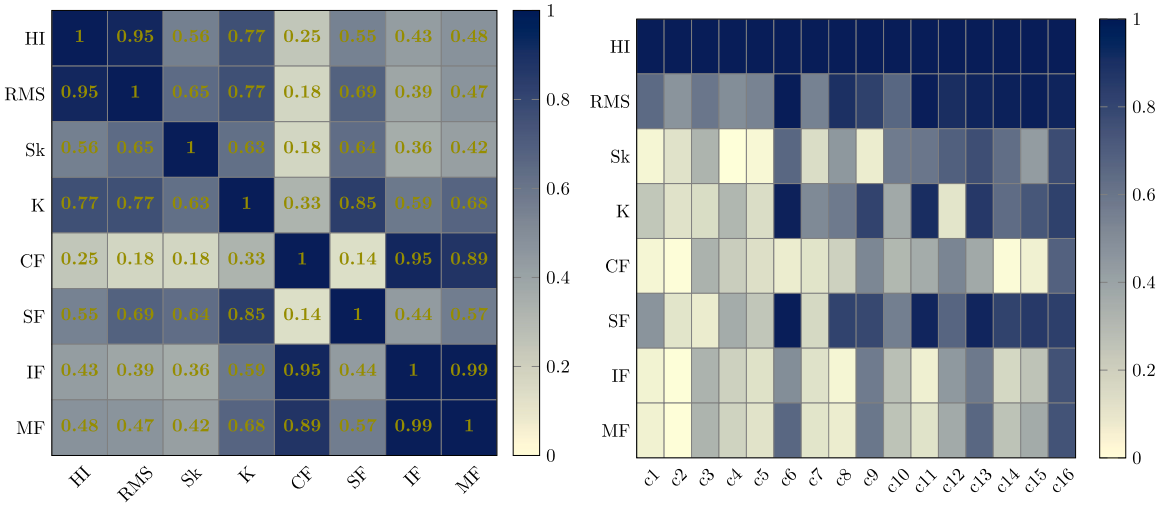


(a) Faulty Sample #536 (b) Faulty Sample #871 (c) Faulty Sample #979

Fig. 13. FFT of the filtered isolated faulty samples of IMS-2 dataset.

analysis of the FFP sample (sample #536) that is [3413.33 Hz–5120 Hz] and then obtaining its envelope, as early degradation stages bearing faults are usually manifested in high frequencies. Fig. 14(a) shows the average correlation of the HI values with the time-domain features. It can be observed that the HI value obtained by the MS2AE network has high correlations with RMS (0.95) and kurtosis (0.77). The correlation matrix between the HI value and the frequency-domain features is included in Fig. 15(a). High correlations between 0.76 and 0.88 can be observed with the BPFO and BPFI of the filtered samples (f), and BPFI, BSF and FTF of the non-filtered samples (nf). As can be seen, the HI value shows the highest correlation with the BPFO of the filtered samples, which is directly related with the kind of fault manifested in this dataset (outer-race bearing fault). Furthermore, the obtained HI enables earlier detection than other traditional analysis approaches in literature, being the FFP obtained similar to the one obtained using other ML algorithms. In addition, the correlation analysis explains where the HI values obtained by the MS2AE network are coming from, as the HI is correlated to different traditional features, showing more sensitivity and flexibility than these features.

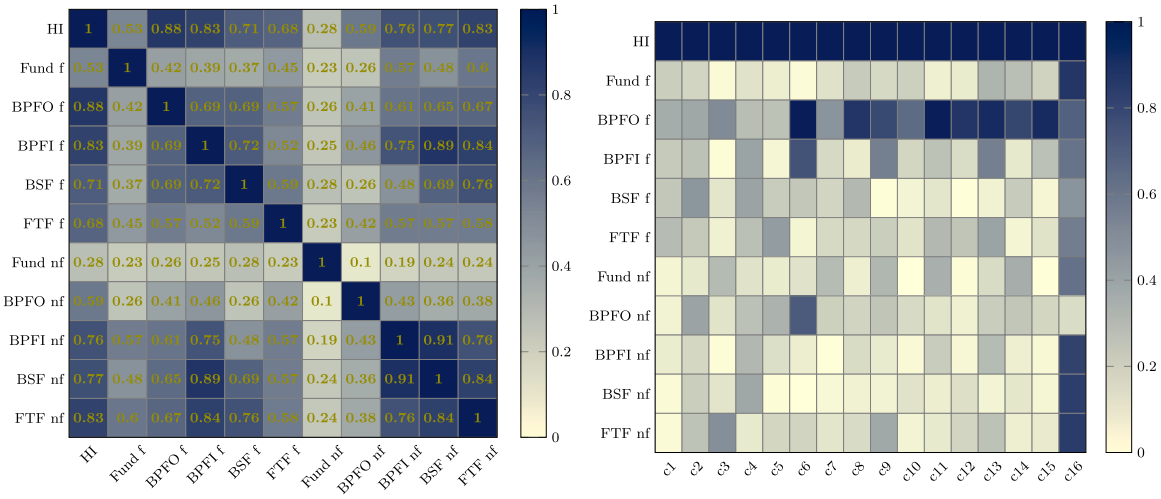
A deeper correlation matrix analysis is carried out, in this case, grouping the original samples into clusters, so the correlation between HI value and time-domain and frequency-domain features over time can be easily observed. This analysis has been made only in the samples predicted as faulty, which have been grouped into 16 clusters. Fig. 14(b) shows the correlation over time between the HI value and time-domain features. It can be observed how the correlation of the HI value and the RMS is almost 1 after the eleventh cluster. Kurtosis and shape factor have also high correlations between the thirteenth and sixteenth cluster. A similar analysis is carried out between the HI value and frequency-domain features. The obtained correlation matrix is shown in Fig. 15(b), where it can be observed how the correlation of 1X BPFO of the filtered samples is the highest between the eighth and sixteenth cluster. In addition, 1X BPFI of the filtered samples and 1X BPFI, BSF and FTF of the non-filtered samples show high correlations



(a) Correlation matrix between HI value and time-domain features

(b) Correlation matrix over time between HI value and time-domain features

Fig. 14. Correlation matrices between HI value and time-domain features in IMS-2 dataset.



(a) Correlation matrix between HI value and frequency-domain features

(b) Correlation matrix over time between HI value and frequency-domain features

Fig. 15. Correlation matrices between HI value and frequency-domain features in IMS-2 dataset.

in the last cluster as expected, as the RM at this cluster is severely damaged, causing an increase in all the frequencies.

5.2. Results with other datasets

The proposed fault classification and diagnosis system has been used with other datasets apart from the IMS-2 dataset. The results obtained with these datasets are presented in Table 2, including relevant information such as the three features with the highest correlation with the HI value, first faulty point (FFP), threshold that separates healthy and faulty stages, the samples selected for being analyzed, frequency range used as bandpass filter in each of the samples and also the diagnosis provided by the system. In all of them, the first 300 samples (5 h of continuous operation) have been considered as healthy, while the remaining samples are considered as unlabeled. It is feasible to assume that the samples gathered during the first 5 h of operation of a RM would correspond to a healthy state.

From the results of Table 2 it can be seen that the diagnosis performed is correctly, classifying the stage of degradation and determining the harmonics that are associated with the kind of fault that is causing the bearing degradation. In the case of sample #451 of XJTU 2-1, which corresponds with the FFP, the proposed fault classification and diagnosis system classifies it as medium stage degradation. This is because the degradation in this experiment is sudden and rapid, causing this sample to be classified as medium degradation.

The HI values of the samples have an average correlation over 0.9 with the RMS in all the datasets. The HI values of the datasets where an inner-race bearing fault is causing the degradation (IMS-1, XJTU 2-1 and XJTU 3-4) have average correlations of 0.82, 0.72 and 0.93 respectively with the shape factor. Besides, in the case of IMS-1 and XJTU 3-4 the amplitudes of the BPFI of the filtered samples have 0.8 and 0.92 of correlation with the HI values respectively, while in XJTU 2-1 the amplitudes of BPFI from the non-filtered samples have an average correlation of 0.88 with the HI values. In the case of the datasets where an outer-race bearing fault is occurring (IMS-2, IMS-3 and XJTU 3-1) the amplitudes of the BPFO from the filtered samples have high correlation with the HI value, being of 0.88, 0.85 and 0.55 respectively.

In addition, the proposed MS2AE network determines in some cases the FFP earlier than other works in the literature. A comparison between the FFPs predicted in this work using 300 samples as healthy samples for training and those predicted in other works is included in Table 3.

Table 2
Summary of fault classification and diagnosis system results.

Dataset	Top 3 correlated features	FFP	Threshold	Sample	Bandpass filter range	Diagnosis
IMS-1	RMS (0.97) Shape factor (0.82) BPFI filtered (0.80)	1857	0.50036	1857 2138 2155	2560 Hz–5120 Hz 2560 Hz–5120 Hz 6826.67 Hz–8533.33 Hz	Early stage - 1X BPFI - 1X FTF Medium stage - 1X BPFI - 3X BPFO - 1X, 2X, 3X FTF Last stage - 1X BPFI - 6X FTF
IMS-2	RMS (0.95) BPFO filtered (0.88) FTF non-filtered (0.83)	536	0.49802	536 871 979	3413.33 Hz–5120 Hz 2560 Hz–5120 Hz 6826.67 Hz–8533.33 Hz	Early stage - 1X, 2X BPFO - 1X, 5X, 6X FTF Medium stage - 1X, 2X, 3X, 4X, 5X BPFO Last stage - 1X, 2X, 3X, 4X, 6X BPFO
IMS-3	RMS (0.97) BPFO filtered (0.85) Fundamental filtered (0.84)	5967	0.49799	5967 6178 6319	2560 Hz–5120 Hz 3413.33 Hz–5120 Hz 1280 Hz–2560 Hz	Early stage - 1X BPFO - 2X, 4X FTF Medium stage - 1X, 2X, 3X, 4X BPFO Last stage - 1X, 2X, 3X, 4X BPFO - 4X, 6X FTF
XJTU 2-1	RMS (0.95) BPFI non-filtered (0.88) Shape factor (0.72)	451	0.50449	451 460 486	2133.33 Hz–4266.67 Hz 1600 Hz–3200 Hz 1600 Hz–3200 Hz	Medium stage - 1X, 2X BPFI - 1X BSF - 5X FTF Medium stage - 1X BPFI - 1X BSF - 5X FTF Last stage - 4X BPFI - 2X, 5X FTF
XJTU 2-3	RMS (0.98) BPFI filtered (0.76) Fundamental filtered (0.76)	301	0.50158	301 419 532	8533.33 Hz–10666.67 Hz 2133.33 Hz–4266.67 Hz 3200 Hz–4800 Hz	Early stage - 1X BSF - 1X, 2X, 3X, 4X, 5X, 6X FTF Medium stage - 1X, 3X BPFO - 1X FTF Medium stage - 2X BPFO - 3X BPFI - 3X BSF - 1X FTF
XJTU 3-1	RMS (0.92) BPFO filtered (0.55) Skewness (0.53)	2347	0.51112	2347 2445 2533	9600 Hz–11200 Hz 9600 Hz–11200 Hz 1600 Hz–3200 Hz	Early stage - 1X, 2X BPFO - 1X, 3X BSF - 3X FTF Medium stage - 1X, 3X BPFO - 4X, 6X FTF Last stage - 1X, 2X, 3X, 4X, 5X BPFO - 5X BPFI - 3X BSF
XJTU 3-4	RMS (0.98) Shape factor (0.93) BPFI filtered (0.92)	1416	0.50718	1416 1453 1505	3200 Hz–4800 Hz 3200 Hz–4800 Hz 3200 Hz–4800 Hz	Early stage - 1X BPFI - 1X BSF - 5X FTF Medium stage - 1X, 2X BPFI - 1X BSF - 5X FTF Last stage - 1X BPFI - 1X BSF - 2X, 5X FTF

Table 3
First Faulty Point comparison.

Dataset	Number of healthy samples	Training dataset	Proposed solution	[39]	[40]	[41]	[42]	[43]	[44]	[45]
IMS-1	300	13.66%	1857		1917	**1833**	1971			
IMS-2	300	30.48%	536	934	632	533	**530**			
IMS-3	300	4.74%	5967			**5952**				
XJTU 2-1	300	61.09%	451		458			455	452	**450**
XJTU 2-3	300	56.28%	**301**					327	302	313
XJTU 3-1	300	11.82%	2347	2348	2407			**2344**	2346	2376
XJTU 3-4	300	19.80%	**1416**	1418	1446			1418	1417	1430

Table 4
Best FFP varying healthy samples size.

Dataset	Number of healthy samples	Training dataset	Best FFP
IMS-1	300	13.66%	1857
IMS-2	300	30.48%	536
IMS-3	300	4.74%	5967
XJTU 2-1	50	10.18%	451
XJTU 2-3	50	9.38%	131
XJTU 3-1	50	1.97%	2347
XJTU 3-4	150	9.90%	1416

On one hand, the FFPs in XJTU 2-3 and XJTU 3-4 are slightly improved compared with the rest of works. On the other hand, the FFPs in IMS-1, IMS-2 and IMS-3 do not improve all the works but they are close to the best ones. The same happens with the FFPs in XJTU 2-1 and XJTU 3-1. However, most of the works determine the FFP only focusing on one dataset while this works focuses on two different datasets. This work also improves the FFPs in Guo et al. [40], where both IMS and XJTU datasets are analyzed at the same time, while providing accurate fault diagnosis.

The number of healthy samples used for training (300 samples) was chosen empirically to be robust to different datasets and working conditions. The performance of the model was studied while determining the first faulty point by varying the number of healthy samples used for training between 50 and 400. Table 4 shows the smallest number of healthy samples used for training that obtains the best result for each dataset.

Comparing the results in Tables 3 and 4, it can be seen that using 300 training samples allows determining the best FFP in all datasets except for the XJTU 2-3 dataset. In this case, when using 50 training samples, the faulty point is determined in sample #131. This would mean that when using 300 samples for training, samples that already show signs of fault would be considered as healthy. Given that the datasets were obtained from experiments with faults induced under accelerated degradation conditions, in real-world scenarios with noisy environments, additional samples could be collected for training to enhance robustness.

6. Conclusions

A novel explainable and interpretable bearing fault classification and diagnosis system has been proposed. The system has been tested using two of the most commonly used bearing fault datasets. The use of a Monotonic Smoothed Stacked Autoencoder overcomes two of the main problems in the fault classification and diagnosis field. On one hand, the system can be applied with limited data and imbalanced datasets, as only non-faulty data from the monitored machinery is needed to train the model. On the other hand, no feature extraction is required as the MS2AE model receives the raw vibration data without any preprocessing. The MS2AE model fuses all features into an HI value that is used to identify the first faulty point. The FFPs obtained in all the datasets are comparable with other state-of-the-art methods, making the proposed HI value accurate for being used for fault classification and diagnosis.

Furthermore, the correlation matrices between the HI value and the time-domain and frequency-domain features show that the HI value is highly correlated with some engineering features such as RMS or the amplitudes of some of the harmonics directly related with the developing fault. It has also been observed that in the case of inner-race bearing faults, the HI value obtained is highly correlated with the shape factor and the 1X BPFI amplitudes of the filtered and non-filtered samples. Something similar happens with outer-race bearing faults,

where the HI value is highly correlated with 1X BPFO amplitudes. All this provide explainability of the operation of the MS2AE model and makes it reliable for being used by practitioners.

The results obtained with relevant bearing fault datasets show that the proposed fault classification and diagnosis system works correctly under different operating conditions and motor loads. The system is autonomous and only requires human intervention to provide the healthy samples, the unlabeled samples and some relevant information of the RM analyzed, such as the shaft frequency, bearing related frequencies and sample frequency. For this reason, the system is suitable for being used in real environments with limited data of the monitored machinery. No qualified staff is necessary, as the diagnosed faults are clearly explained with no ML knowledge needed. Moreover, this fault classification and diagnosis system is extremely flexible, what makes it easily applicable to any RM. The system can also be easily adapted to support other well-known faults such as imbalance, misalignment and gear mesh.

To sum up, the main contributions of the developed fault classification and diagnosis system are:

- It correctly works with limited data as it only requires healthy data for training the MS2AE model.
- It does not require manually feature extraction, as the MS2AE model receives the raw vibration data and fuses them into an HI value.
- Explainability of the MS2AE model ensures that the generated HI is highly correlated with some engineering features, making it reliable for being used by practitioners.
- It is robust under different operating conditions and motor loads.
- Interpretability of the results avoids the need of qualified-staff, as diagnosed faults are clearly explained.

Human intervation is still required for parameter tuning, model structure selection, separation threshold calculation and providing information from the geometry and operating conditions of the RM. The selection of traditional features with which to perform correlation analysis oriented to explainability and interpretability is also done manually.

Future work will focus on developing an API for bearing fault classification and diagnosis, along with a corresponding web tool to enhance accessibility, integration, and practical application. An API allows seamless integration with various industrial systems and applications, facilitating real-time monitoring and diagnostics across different platforms, significantly improving maintenance strategies and operational efficiency by enabling automated fault detection and predictive maintenance. The web tool will serve as a user-friendly interface that showcases the API's functionality, offering practitioners a practical demonstration and allowing them to test the application with their own real datasets. Additionally, the deployment of the fault classification and diagnosis system in real industrial facilities will be prioritized, enabling non-qualified staff to include samples for analysis and automatically generating a comprehensive diagnostics log. Finally, the applicability of the system to other scenarios, such as gears or combustion motors, will also be explored.

CRediT authorship contribution statement

L. Magadán: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **C. Ruiz-Cárcel:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **J.C. Granda:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Formal analysis. **F.J. Suárez:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **A. Starr:** Writing – review & editing, Supervision, Project administration, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: F. J. Suarez reports financial support was provided by Spanish National Plan of Research, Development, and Innovation, project EDNA (PID2021-124383OB-100). L. Magadan reports financial support was provided by Severo Ochoa Program, Principado de Asturias (PA-22-BP21-120). The rest of the authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was partially funded by the Spanish National Plan of Research, Development, and Innovation under project EDNA (PID2021-124383OB-100), the University of Oviedo and the University of Cranfield. L. Magadán is supported by the Severo Ochoa program (PA-22-BP21-120).

Data availability

Data will be made available on request.

References

[1] A.G. Nath, S.S. Udmale, D. Raghuwanshi, S.K. Singh, Improved structural rotor fault diagnosis using multi-sensor fuzzy recurrence plots and classifier fusion, IEEE Sens. J. 21 (19) (2021) 21705–21717.
[2] H. Hashemian, Wireless sensors for predictive maintenance of rotating equipment in research reactors, Ann. Nucl. Energy 38 (2–3) (2011) 665–680.
[3] L. Magadán, F.J. Suárez, J.C. Granda, F.J. delaCalle, D.F. García, A robust health prognostics technique for failure diagnosis and the remaining useful lifetime predictions of bearings in electric motors, Appl. Sci. 13 (4) (2023) 2220.
[4] Y. Yan, Q. Liu, X.q. Gao, Motor fault diagnosis algorithm based on wavelet and attention mechanism, J. Sens. 2021 (1) (2021) 3782446.
[5] L. Magadán, F. Suárez, J. Granda, D. García, Low-cost industrial IoT system for wireless monitoring of electric motors condition, Mob. Netw. Appl. 28 (2023) 97–106.
[6] A. Dameshghi, M.H. Refan, Combination of condition monitoring and prognosis systems based on current measurement and PSO-LS-SVM method for wind turbine DFIGs with rotor electrical asymmetry, Energy Syst. 12 (2021) 203–232.
[7] N.W. Nirwan, H.B. Ramani, Condition monitoring and fault detection in roller bearing used in rolling mill by acoustic emission and vibration analysis, Mater. Today: Proc. 51 (2022) 344–354.
[8] S. Zhang, S. Zhang, B. Wang, T.G. Habetler, Deep learning algorithms for bearing fault diagnostics—A comprehensive review, IEEE Access 8 (2020) 29857–29881.
[9] Y. Xu, K. Feng, X. Yan, R. Yan, Q. Ni, B. Sun, Z. Lei, Y. Zhang, Z. Liu, CFCNN: A novel convolutional fusion framework for collaborative fault identification of rotating machinery, Inf. Fusion 95 (2023) 1–16.
[10] C. Peng, H. Gao, X. Liu, B. Liu, A visual vibration characterization method for intelligent fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 192 (2023) 110229.
[11] Z. Meng, W. Cao, D. Sun, Q. Li, W. Ma, F. Fan, Research on fault diagnosis method of MS-CNN rolling bearing based on local central moment discrepancy, Adv. Eng. Inform. 54 (2022) 101797.
[12] P. Liang, Z. Yu, B. Wang, X. Xu, J. Tian, Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network, Adv. Eng. Inform. 57 (2023) 102075.
[13] Y. Kaya, F. Kuncan, H.M. ERTUNÇ, A new automatic bearing fault size diagnosis using time-frequency images of CWT and deep transfer learning methods, Turk. J. Electr. Eng. Comput. Sci. 30 (5) (2022) 1851–1867.
[14] L. Magadán, J. Roldán-Gómez, J.C. Granda, F.J. Suárez, Early fault classification in rotating machinery with limited data using TabPFN, IEEE Sens. J. (2023) 1.
[15] C. Zhang, Y. Zhang, Q. Huang, Y. Zhou, Intelligent fault prognosis method based on stacked autoencoder and continuous deep belief network, in: Actuators, Vol. 12, (3) MDPI, 2023, p. 117.
[16] T. Han, J. Pang, A.C. Tan, Remaining useful life prediction of bearing based on stacked autoencoder and recurrent neural network, J. Manuf. Syst. 61 (2021) 576–591.
[17] F. Xu, Z. Hao, C. Zhou, Y. Deng, Bearing condition monitoring via an unsupervised and enhanced stacked auto-encoder, J. Braz. Soc. Mech. Sci. Eng. 46 (6) (2024) 367.

[18] M. Kuncan, An intelligent approach for bearing fault diagnosis: Combination of 1D-LBP and GRA, IEEE Access 8 (2020) 137517–137529.
[19] K. Kaplan, Y. Kaya, M. Kuncan, M.R. Minaz, H.M. Ertunç, An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis, Appl. Soft Comput. 87 (2020) 106019.
[20] Z. Meng, Q. Li, D. Sun, W. Cao, F. Fan, An intelligent fault diagnosis method of small sample bearing based on improved auxiliary classification generative adversarial network, IEEE Sens. J. 22 (20) (2022) 19543–19555.
[21] S. Liu, H. Jiang, Z. Wu, Y. Liu, K. Zhu, Machine fault diagnosis with small sample based on variational information constrained generative adversarial network, Adv. Eng. Inform. 54 (2022) 101762.
[22] X. Li, S. Yu, Y. Lei, N. Li, B. Yang, Intelligent machinery fault diagnosis with event-based camera, IEEE Trans. Ind. Inform. (2023).
[23] R. Marcinkevičs, J.E. Vogt, Interpretability and explainability: A machine learning zoo mini-tour, 2020, arXiv preprint.
[24] N. Costa, L. Sánchez, Variational encoding approach for interpretable assessment of remaining useful life estimation, Reliab. Eng. Syst. Saf. (ISSN: 0951-8320) 222 (2022) 108353.
[25] S. Li, T. Li, C. Sun, R. Yan, X. Chen, Multilayer grad-CAM: An effective tool towards explainable deep neural networks for intelligent fault diagnosis, J. Manuf. Syst. 69 (2023) 20–30.
[26] H. Yang, X. Li, W. Zhang, Interpretability of deep convolutional neural networks on rolling bearing fault diagnosis, Meas. Sci. Technol. 33 (5) (2022) 055005.
[27] H.-W. Xu, W. Qin, J.-H. Hu, Y.-N. Sun, Y.-L. Lv, J. Zhang, A copula network deconvolution-based direct correlation disentangling framework for explainable fault detection in semiconductor wafer fabrication, Adv. Eng. Inform. 59 (2024) 102272.
[28] H. Li, J. Lin, Z. Liu, J. Jiao, B. Zhang, An interpretable waveform segmentation model for bearing fault diagnosis, Adv. Eng. Inform. 61 (2024) 102480.
[29] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, J. Sound Vib. 289 (4–5) (2006) 1066–1090.
[30] B. Wang, Y. Lei, N. Li, N. Li, A hybrid prognostics approach for estimating remaining useful life of rolling element bearings, IEEE Trans. Reliab. 69 (1) (2018) 401–412.
[31] T. Williams, X. Ribadeneira, S. Billington, T. Kurfess, Rolling element bearing diagnostics in run-to-failure lifetime testing, Mech. Syst. Signal Process. 15 (5) (2001) 979–993.
[32] Q. Qian, Y. Qin, Y. Wang, F. Liu, A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis, Measurement 178 (2021) 109352.
[33] M. Yu, T. Quan, Q. Peng, X. Yu, L. Liu, A model-based collaborate filtering algorithm based on stacked AutoEncoder, Neural Comput. Appl. (2022) 1–9.
[34] Y. Permanasari, E.H. Harahap, E.P. Ali, Speech recognition using dynamic time warping (DTW), J. Phys. Conf. Ser. 1366 (1) (2019) 012091.
[35] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mech. Syst. Signal Process. 21 (1) (2007) 108–124.
[36] J. Antoni, R.B. Randall, The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines, Mech. Syst. Signal Process. 20 (2) (2006) 308–331.
[37] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier, PRONOSTIA: An experimental platform for bearings accelerated degradation tests, in: IEEE International Conference on Prognostics and Health Management, PHM'12, IEEE Catalog Number: CPF12PHM-CDR, 2012, pp. 1–8.
[38] B. Zhang, Y. Miao, J. Lin, Z. Liu, A new two-stage strategy to adaptively design and finely tune the filters for bearing fault-related mode decomposition, Measurement 210 (2023) 112470.
[39] H. Wang, X. Zhang, X. Guo, T. Lin, L. Song, Remaining useful life prediction of bearings based on multiple-feature fusion health indicator and weighted temporal convolution network, Meas. Sci. Technol. 33 (10) (2022) 104003.
[40] W. Guo, X. Li, X. Wan, A novel approach to bearing prognostics based on impulse-driven measures, improved morphological filter and practical health indicator construction, Reliab. Eng. Syst. Saf. (2023) 109451.
[41] T. Yan, D. Wang, J.-Z. Kong, T. Xia, Z. Peng, L. Xi, Definition of signal-to-noise ratio of health indicators and its analytic optimization for machine performance degradation assessment, IEEE Trans. Instrum. Meas. 70 (2021) 1–16.
[42] Y. Zhou, A. Kumar, C. Gandhi, G. Vashishtha, H. Tang, P. Kundu, M. Singh, J. Xiang, Discrete entropy-based health indicator and LSTM for the forecasting of bearing health, J. Braz. Soc. Mech. Sci. Eng. 45 (2) (2023) 120.
[43] H. Wei, Q. Zhang, Y. Gu, Remaining useful life prediction of bearings based on self-attention mechanism, multi-scale dilated causal convolution, and temporal convolution network, Meas. Sci. Technol. 34 (4) (2023) 045107.
[44] Y. Deng, S. Du, D. Wang, Y. Shao, D. Huang, A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines, IEEE Trans. Instrum. Meas. 72 (2023) 1–15.
[45] Z. Li, P. Xu, X.-B. Wang, Online anomaly detection and remaining useful life prediction of rotating machinery based on cumulative summation features, Meas. Control 56 (3–4) (2023) 615–629.