DiVA
http://www.diva-portal.org

This is the published version of a paper published in *Journal of Intelligent Manufacturing*.

Citation for the original published paper (version of record):

Iunusova, E., Gonzalez, M K., Szipka, K., Archenti, A. (2023)
Early fault diagnosis in rolling element bearings: comparative analysis of a knowledge-based and a data-driven approach
*Journal of Intelligent Manufacturing*, : 1-21
https://doi.org/10.1007/s10845-023-02151-y

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-336688

Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-023-02151-y

# Early fault diagnosis in rolling element bearings: comparative analysis of a knowledge-based and a data-driven approach

Eleonora Iunusova[1] · Monica Katherine Gonzalez[1] · Károly Szipka[1] · Andreas Archenti[1]

Received: 28 June 2022 / Accepted: 17 May 2023
© The Author(s) 2023

**Abstract**
The early identification of a defect that is developing in a bearing is crucial for avoiding failures in rotating machinery. Frequency domain analysis of the vibration signals has been shown to contribute to a better understanding of the nature of a developing defect. Early signs of degradation might be more noticeable in certain frequency bands. The advantages in identifying and monitoring these bandwidths are several: prevention of serious machinery damages, reduction of the loss of investments, and improvement of the accuracy in failure predicting models. This paper presents and compares two approaches for the diagnosis of bearing faults. The first approach was knowledge-based. It relied on principles of mechanics to interpret the measured vibration signals and utilized prior knowledge of the bearing characteristics and testing parameters. The second approach was data-driven whereby data were acquired exclusively from the vibration signal. Both approaches were successfully applied for fault diagnosis by identifying the frequencies of the vibration spectra characteristic for the bearing under study. From this, bandwidths of interest for early fault detection could be determined. The diagnostic abilities of both approaches were studied to analyze and compare their individual strengths regarding the aspects of implementation time, domain knowledge, data processing associated knowledge, data requirements, diagnostic performance, and practical applicability. The advantages, apparent limitations as well as avenues for further improvement of both approaches are discussed.

**Keywords** Fault diagnosis · Data-driven · Knowledge-based · Rolling elements bearings · Vibration · Degradation

## Introduction

Any small defect that is not noticed and eliminated in time, carries the risk of an eventual catastrophic failure in turn leading to the shutdown of the entire mechanical system. Therefore, early fault diagnosis is crucial for well-timed maintenance actions to ensure the safe and efficient operation of rotating machinery. The task of early fault diagnosis requires regular monitoring of the state of the system and judicious interpretation of essential information that the data resulting from the monitoring contains. After detecting signs of degradation and diagnosing the potential failure mode, a model estimating the Remaining Useful Life (RUL) of the affected component or machine needs to be created so that appropriate timings for servicing and maintenance can be scheduled (Lacey, 2008). After nuts and bolts, rolling element bearings are the most widely used machine elements: more than 50 billion bearings are operating in the world at any one time (McQueen, 2010). It is therefore not surprising that investigations into and discussions of bearing failures are extensive in the research discipline of Condition-Based Maintenance (CBM) (Jardine et al., 2006; Wei et al., 2019).

Vibration analysis is an established method for tracking machine operating conditions and trends of deterioration (Renwick & Babson, 1985). The data from the vibration signal are widely used for CBM of bearings. This is due to the fact that a vibration impulse is generated whenever a defect or an irregularity of the raceway surfaces or of the rolling elements comes into contact with another surface of the bearing

✉ Eleonora Iunusova
iunusova@kth.se

✉ Monica Katherine Gonzalez
mgon@kth.se

Károly Szipka
szipka@kth.se

Andreas Archenti
archenti@kth.se

[1] Department of Production Engineering, KTH Royal Institute of Technology, Brinellvägen 68, 11428 Stockholm, Sweden

Published online: 16 June 2023

(Tandon & Choudhury, 1999). The vibration impulses occur periodically at a specific rate called the bearing characteristic frequency of the defect, which is unique for each fault source. These bearing characteristic frequencies are an essential source of information for diagnosing a potential failure mode of a bearing. Identification of these frequencies is not a trivial task, especially in the case of limited information about the geometric dimensions of a bearing or of the test parameters.

The biggest difficulties for diagnosing the early degradation of a bearing are (i) weakness of the vibration impulses; (ii) measurement noise and (iii) other vibration sources (Gautier et al., 2015). Considerable attention has been devoted to the high-frequency resonance technique as this approach appears to separate the vibration of the faulty bearing from the vibrations generated by the other sources (Cui et al., 2016; McFadden & Smith, 1984; Mishra et al., 2021). However, most of the data available for analysis have been recorded with accelerometers that are unable to capture these higher vibration frequencies.

The extensive research in the field of diagnostics and prognostics of bearing defects has relied on the implementation of knowledge-based and data-driven approaches (Cerrada et al., 2018; Jieyang et al., 2022; Tandon & Choudhury, 1999). The comparison of different methods used within each of the approaches is presented, for example, in the papers (Cubillo et al., 2016; Gazizulin et al., 2018) for the knowledge-based approach and (Akpudo & Hur, 2020; Singh et al., 2023; Vargas-Machuca et al., 2020) for data-driven approach. However, there is still a lack of a comprehensive comparison of the diagnostic abilities of the two approaches considering aspects other than prediction accuracy. Significant growth was observed in research in the field of combining knowledge-based and data-driven approaches (Jimenez et al., 2020; Liao & Kottig, 2014; Zhou et al., 2019). However, the way to combine these approaches to achieve the highest quality of diagnosis performance is not always straightforward.

In this paper, two methods for the detection and diagnosis of early faults are described and compared. The first method uses a knowledge-based approach whereas the second uses a data-driven approach. The ultimate purpose of both approaches was to reliably identify the bearing characteristic frequencies within the vibration spectra and to determine the bandwidth of interest for condition monitoring. The individual diagnostic capabilities of both approaches were analysed and compared regarding implementation time, domain knowledge, data processing associated knowledge, data requirements, diagnostic performance, and practical applicability. Furthermore, taking into account the individual strengths and limitations of both approaches, avenues for possible superior diagnostic power obtained by a combination of both approaches are proposed.

The remaining sections of this paper are organized as follows. Section Bearing fault diagnosis contains the background and theoretical information about fault identification in bearings. In Sect. Methodology, the data sets used for this investigation, as well as the knowledge-based and data-driven approaches are described. The results of the implementation of these approaches, a comparison between them, and a discussion about their advantages and limitations are included in Sect. Results and discussion. Finally, the last Section presents the conclusions from this investigation.

## Bearing fault diagnosis

According to ISO-13379-1 (Swedish Standards Institute, 2012), two main approaches can be used for diagnosing the condition of the machine, namely the Knowledge-Based (KB) and the Data-Driven (DD) approach. This ISO standard also mentions that these approaches may overlap or that solutions can rely on several different approaches.

### Knowledge-based approach for bearing fault diagnosis

The knowledge-based approach relies on an explicit representation of the fault or its associated symptoms (Swedish Standards Institute, 2012). This approach is extensively used in engineering as it applies principles of mechanics to model and interpret phenomena under observation. Even if these models initially are simplified abstractions of reality, they usually increase in complexity to ever more accurately represent the physical interactions. However, their main advantage still is that they can be easily interpreted and generalized (Chao et al., 2020).

The knowledge-based approach traditionally relies on an analysis of measurements of either vibrational or acoustic signals in order to detect and diagnose emerging defects in bearings. Regarding vibration signals resulting from localised defects, three different techniques exist to measure and analyze the vibration response, namely time-domain techniques, frequency-domain techniques and time-frequency techniques (Randall & Antoni, 2011; Tandon & Choudhury, 1999).

Within time-domain techniques, the vibration signal is analyzed to extract several statistical features. The most important of these features are amplitude, Root Mean Square (RMS), skewness, kurtosis, crest factor, and entropy, among others. The existence of a fault is conventionally determined once the monitored feature surpasses a predefined acceptable threshold (Li et al., 2017). The main limitations of time-domain analysis are the definition of the thresholds and the selection of the monitoring features. The former is in most cases highly dependent on the type and the specific config-

uration of the machine under analysis (Li et al., 2017). The latter is related to the problems associated with the analysis of non-Gaussian signals whereby important features have a tendency to vary from sample to sample (Attoui et al., 2017).

Frequency-domain techniques are used to identify the characteristic frequency for a certain type of fault (or a combination of faults). These techniques can use different types of spectrum analysis, i.e., Fourier spectrum, cepstrum analysis or envelope spectrum analysis. The main advantage of these techniques is their intuitiveness because they are based on the derivation of the characteristic frequencies of bearing faults, which have a direct relationship to the shaft rotational frequency (Li et al., 2017). In most cases, bearing fault frequencies are masked by more dominant frequency components or noise. In such cases, envelope analysis is a suitable approach to demodulate the bearing faults from the high-frequency resonances emerging from the bearing or from the machine's natural frequencies (Reuben & Mba, 2014). Nevertheless, frequency domain techniques are very sensitive to non-stationary conditions (Attoui et al., 2017).

As the vibration signals are mostly non-stationary, time-frequency-domain techniques have received more attention recently (Reuben & Mba, 2014). Among various approaches, continuous wavelet analysis, short-time Fourier transform, Wigner-Ville distribution, and resonance demodulation techniques have been successfully applied for monitoring the condition of bearings by detecting both the components of the constituent frequencies and the properties of their time variability. However, their higher computational complexity is still the main disadvantage discouraging their widespread utilization (Attoui et al., 2017).

### Bearing characteristic frequencies

Defects in bearings produce a series of periodic impacts resulting in a particular vibration signal depending on the defect. These Bearing Characteristic Frequencies (BCFs) (Saruhan et al., 2014) are associated with the bearing elements and are dependent on the geometric dimensions and arrangements and on the shaft rotational frequency.

The amplitude of BCFs is a sign of the severity of the defect. The presence of harmonics is an indication of the origin of the defect (Rai & Upadhyay, 2016). A machine with a defective bearing can generate any of the following BCFs (Taylor, 2003):

- Ball Pass Frequency of the Outer Race (BPFO): rate at which the ball/roller passes over a single point (defect) on the outer race.
- Ball Pass Frequency of the Inner Race (BPFI): rate at which the ball/roller passes over a single point (defect) on the inner race.
- Two times Ball Spin Frequency (BSF) or Ball Defect Frequency (BDF): frequency of each rolling element as it spins. When one or more balls or rollers have a defect, this defect impacts both the inner and outer race each time one revolution of the rolling element is made. Therefore, the defect vibrational frequency is twice the BSF (Poddar & Chandravanshi, 2013).
- Fundamental Train Frequency (FTF) or cage frequency: This frequency is not encountered very often. The BSF caused by a ball/roller fault will be amplitude-modulated by the FTF (Konstantin-Hansen, 2003).

The following equations provide a theoretical estimate of the BCFs expected when a defect on a bearing element appears (Mais, 2002):

$$BPFO = \frac{nf_s}{2}\left(1 - \frac{d}{D}\cos\alpha\right) \tag{1}$$

$$BPFI = \frac{nf_s}{2}\left(1 + \frac{d}{D}\cos\alpha\right) \tag{2}$$

$$BSF = \frac{Df_s}{2d}\left\{1 - \left(\frac{d}{D}\cos\alpha\right)^2\right\} \tag{3}$$

$$FTF = \frac{f_s}{2}\left(1 - \frac{d}{D}\cos\alpha\right) \tag{4}$$

where $d$ is the ball diameter, $D$ is the pitch diameter, $n$ is the number of balls, $\alpha$ is the contact angle, and $f_s$ is the shaft rotation rate in Hertz. These formulas are theoretical estimates and discrepancies can be expected due to the bearing carrying significant thrust loads or due to the occurrence of slippage (McInerny & Dai, 2003; Randall & Antoni, 2011).

The vibration spectrum pattern changes as the bearing deteriorates. It usually passes through four different stages. The spectrum of each of the four stages of degradation can be further subdivided into four distinct zones (Fig. 1) (Scheffer & Girdhar, 2004).

- Zone A: machine rpm and harmonics
- Zone B: bearing defect frequencies or BCFs
- Zone C: natural frequencies of the bearing components
- Zone D: high-frequency defects

The different stages of degradation, presented in Fig. 1, can be categorized as follows (Scheffer & Girdhar, 2004):

- Stage 1: Some indications of bearing wear are present in the ultrasonic frequency region (Zone D), which can only be detected using high-frequency techniques such as the spike energy or the shock pulse method. The raceways or rolling elements may not show any visible defects.
- Stage 2: The raceways begin to develop pits and to "ring" according to the natural resonance frequencies of the bearing components. An increase in energy in Zone C and D is also noticeable. Depending on the severity of

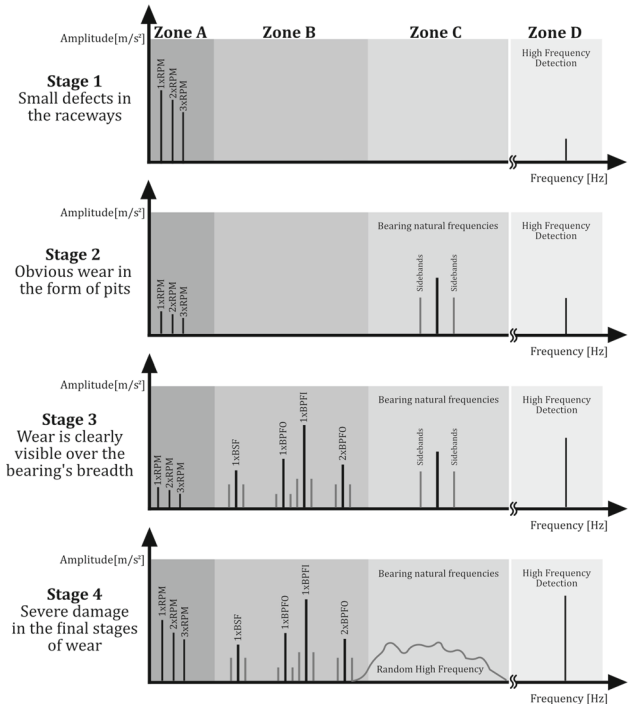


**Fig. 1** Evolution of the bearing vibration spectrum under the different stages of degradation. Adapted from Golafshan et al. (2018) and Scheffer and Girdhar (2004)

the defects, side-band frequencies (corresponding to the BCFs) might appear surrounding the natural frequency peak. Stage 2 of degradation is considered as the early fault diagnosis and is determined as the target stage within the scope of this paper.

- Stage 3: The BCFs start to become visible in the spectrum of Zone B. Harmonics may be present depending on the number of defects and their dispersion on the races. These harmonic frequencies might be side-banded by the shaft speed.
- Stage 4: The pits merge with each other creating rough tracks, thereby generating increased heat and noise. The BCFs and the components' natural frequencies will merge into a random noise floor. The signal levels of Zone D will temporally decrease due to the flattening of pits (becoming spalls) but just prior to total failure the signals will present extreme amplitudes. If the bearing runs further, the cage will break and the rolling elements will go loose, and there will be serious damage to the shaft area under the bearing.

A suitable frequency analysis of the vibration signal in Stage 3 can directly identify the BCFs in the spectrum. However, only the theoretical estimates of the BCFs are known. Due to ageing or speed fluctuations of the rotating machine, the actual frequency components can deviate from those which are estimated from theory (Taylor, 2003).

When defects are in their early stages, the BCFs amplitudes are very small and are usually hidden by measurement noise and other vibration sources (e.g. shaft imbalance). This causes difficulties for detecting the defect frequencies and, consequently, for the localization of the bearing faults (Attoui et al., 2017).

Moreover, the natural frequencies of the components of the bearing are amplitude-modulated at the BCFs. By demodulating these resonances, an indication of the bearing condition can be obtained. In practice, the vibration signal is band-pass filtered around one of the resonant frequencies to eliminate unwanted contributors from other sources. This filtered signal is demodulated by the enveloping technique or by the high-frequency resonance technique. The spectrum of the envelope signal is used to identify the periodic components associated with a fault in a bearing component or BCF (Tandon & Choudhury, 1999).

## Data-driven approach for the diagnosis of bearing faults

The data-driven approach is widely used when there is a significant amount of historical data for the investigation of degradation patterns (Goebel et al., 2008). Usually, a failure follows a certain degradation process, the investigation of which allows to identify potential failures in advance and predict the Remaining Useful Life (RUL) of a machine or of one of its components. Prediction of the RUL can be done through damage estimation followed by calculating the intersection between a suitable extrapolation of the trajectory of the damage progression and the failure criterion (Goebel et al., 2008).

RUL prediction is fundamentally based upon an estimation of the current state of the functional bearing. The prediction typically involves the constant tracking of a state indicator curve, which is usually calculated from signal features. One way to continuously determine the condition of a bearing is by following the behaviour of a particular feature, such as kurtosis or RMS (Ahmad et al., 2018; Li et al., 2015). Different features can indicate different specific problems and may allow an estimation of their severity (Lacey, 2008; Li et al., 2012). Monitoring only one particular feature may lead to losing important information about potentially arising or already existing problems of a bearing when the tracked parameter is not sensitive to that particular condition.

Several methods for constructing a state indicator for the purposes of monitoring the condition of bearings have been proposed. In particular, fusing several signal features into a single Degradation Curve (DC) makes it possible to take into account all potentially relevant parameters (Yan et al., 2004). In the method presented by Wang et al. (2008), a multivariate linear regression model was used to calculate a state indicator designating the degradation stage of a bearing. In this work,

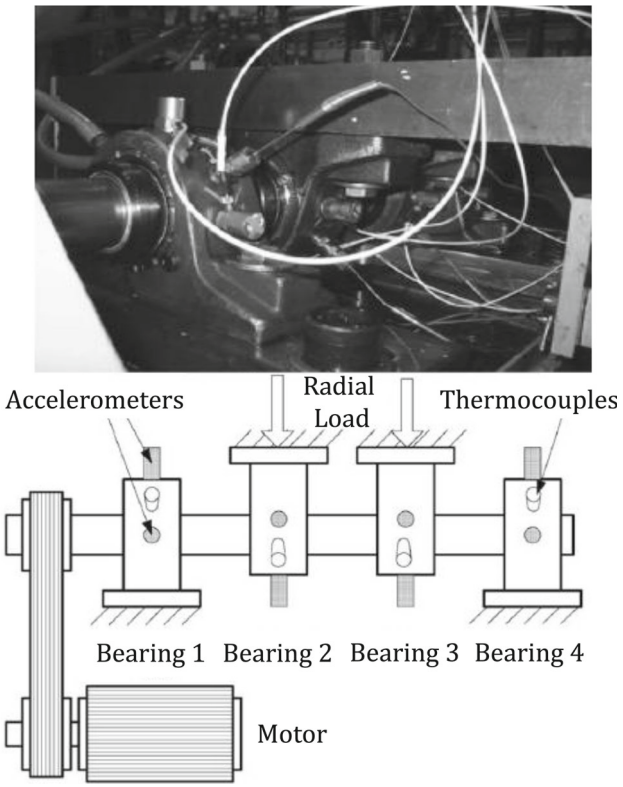


**Fig. 2** Setup of the test rig: Photography and schematic representation. Modified from Lee et al. (2007)

Wang's method is applied to create a degradation curve scaled from 0 to 100% of degradation.

Depending on which features were considered, the shapes of constructed DCs will differ. The accuracy of a single fused curve highly depends on the proper execution of the features generation step where a few important considerations need to be taken into account. First, the raw vibration signal typically contains noise originating either from the operating process itself or from its surroundings. Preliminary signal filtering is necessary in many cases. Second, careful execution of the features engineering process has to be performed (Kuhn & Johnson, 2013). The features engineering process usually contains two main parts: features generation and features reduction.

Publicly available libraries are frequently used to facilitate the features generation task (Barandas et al., 2020; Christ et al., 2018). Usually, the variety of the features generated by those libraries is too wide and not case-specific. Using all the generated features for further analysis is undesirable since high-dimensional data increase interpretation problems and can negatively affect predictive models (Kuhn & Johnson, 2013).

Reducing the number of features can be performed by two different methods, namely, dimensionality reduction and features selection (Niu, 2016). Dimensionality reduction is the linear or nonlinear transformation of the high-dimensional original feature space into a meaningful representation of reduced dimensionality (Van Der Maaten et al., 2009). Furthermore, it is an efficient technique with regard to computational resources. However, it completely erases the traceability of the features and therefore reduces the potential to gain domain knowledge from the analysis. On the other hand, features selection does not change the features themselves and consequently allows for a more detailed interpretation of the influence of individual features on the modelling process (Niu, 2016). The traceability maintained by the use of feature selection methods makes it a preferred approach for this study.

To build a DC, the selected features must be aggregated together. In the case of analysing a data set from a run-to-failure experiment, construction of the curve can be done by creating a linear regression model as proposed by Wang et al. (2008).

## Methodology

This section covers a description of the data set utilized for the proposed analysis and the two different approaches implemented for bearing fault diagnostics. The knowledge-based approach used as input all the available data, namely, the BCFs estimations based on the bearing components, test parameters, and vibration signal recordings. The data-driven approach used only the vibration measured with the accelerometers.

### Data set description

In this work, the data containing a complete record of the natural evolution of a bearing defect were acquired in the course of run-to-failure experiments performed in the NSF I/UCR Center for Intelligent Maintenance Systems (IMS) with support from Rexnord Corporation (Lee et al., 2007).

#### Test rig setup

In the IMS's test rig, four Rexnord ZA-2115 double row bearings were installed on the shaft (see Fig. 2). The rotation speed was kept constant at 2000 RPM, and a radial load of 6000 lbs was applied to the bearings and shaft. The vibration data was collected during the three independent tests. The vibration data was collected with PCB 353B33 High Sensitivity Quartz ICP accelerometers installed on the bearing housings: two for each bearing for test 1, and one for each bearing for tests 2 and 3.

The bearings installed in the IMS's test rig are of the spherical double row roller type (see characteristics in Table 1).

**Table 1** Bearing characteristics

| | |
|---|---|
| Bearing type | Rexnord ZA-2115 |
| Pitch diameter | 2.815 in |
| Roller diameter | 0.331 in |
| Roller number | 16 |
| Contact angle | 15.17 degrees |

### Data structure

Three data sets contain vibration data from the independent run-to-failure tests. Signals (duration 1 s) were collected every 10 min with a sampling rate of 20 kHz. The data was acquired with a DAQ Card-6062E from National Instruments. Inspections confirmed that Test 1 ended with an inner race failure in bearing 3 and a rolling element failure in bearing 4. Tests 2 and 3 ended with an outer race failure in bearing 1 and 3, respectively Lee et al. (2007).

For the proposed analysis, only the data set corresponding to *Test 2*, shown in Fig. 3, was utilized, as one of the parameters for comparing the different approaches was the diagnostics performance. Thus, it was more straightforward to compare the approaches when only one failure mode is present.

To define and set the thresholds for failure, functional failure, and fault points, the RMS of the vibration signal from *bearing 1 of Test 2* was calculated and visualized in Fig. 4.

- *Failure* For *Test 2*, the failure point (considered as 0% RUL) occurred at the 972nd data collection instance (Fig. 4). This point is encircled in red in Fig. 4d. A sharp increase in RMS indicated an acceleration of degradation preceding failure. Therefore, the total life of bearing was estimated to be 9720 min.
- *Functional failure* According to Moubray (1997), a functional failure is an inability to reach the required level of functionality. The DC, presented in Fig. 4a, is not increasing monotonically. The degradation is slowly increasing until a sudden jump (Fig. 4b). Shortly after that jump, the degradation decreases, stabilizes, and starts slowly increasing again (Fig. 4c). This phenomenon was also observed and described by many researchers (Qiu et al., 2006; Williams et al., 2001; Lei et al., 2018) and can be labelled as a "healing process". This phenomenon can be explained by the nature of the damage. Initially, rough surface defects (sharp-edged cracks or small damaged zones) causing an increase in the RMS are smoothed by a sanding effect of continuous rolling contact thereby resulting in a decrease of the RMS. The definition of the healing process implies that it followed on from previously inflicted damage which resulted in functional failure of limited duration. Therefore, the onset of the functional failure was defined as occurring at the sampling instance preceding the sharp increase of RMS. This point corresponded to the 702nd data collection instance and is shown circled in orange in Fig. 4b.
- *Fault* The immediate identification of the instance when the fault occurred is the ideal outcome in early fault diagnosis.

### Knowledge-based approach

The KB approach relies on a representation of the fault and its symptoms. When referring to vibration signal analysis for bearings, the main idea is to apply different techniques

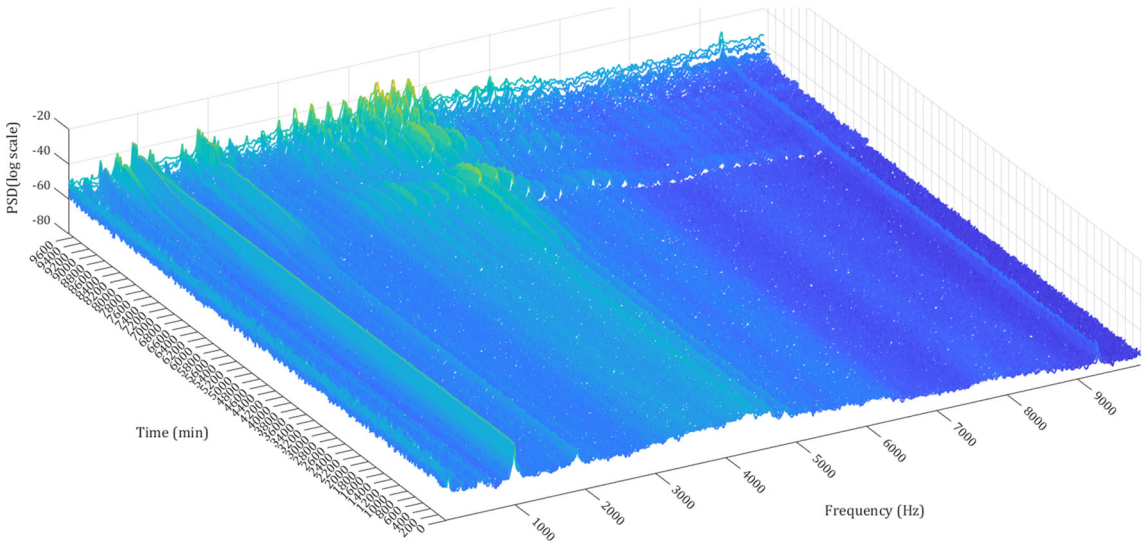


**Fig. 3** Evolution of the power spectral density (PSD) of the vibration signal for the signal from Test 2 (Color figure online)

to identify the changes or variations of the signal either in the time or in the frequency domain. In the case of bearing fault identification and diagnosis, the appropriate technique to apply depends on the bearing degradation stage.

### Identification of the BCFs and their harmonics in the spectrum

As highlighted in Fig. 5, the proposed analysis started under the assumption that the bearing was in the third stage of degradation given that the data collection was not done until total failure but up to a certain threshold (debris accumulation in the magnetic plug) (Qiu et al., 2006).

This assumption was corroborated by the spectra of the last data instances as illustrated in Fig. 6, which clearly shows the presence of the fundamentals and harmonics of the BCFs in Zone B (center) and side-bands in Zone C (right), which matched the description of Stage 3 degradation. Then, the spectrum of the vibration signals was retrospectively inspected to identify which of the BCFs and their corresponding harmonics were present in the signals. The values of the BCFs were obtained following Eqs. 1–4 and are displayed in Table 2.

The spectrum was examined in increments of 5% RUL. This increased the time resolution for the detection of spectral indications of a developing fault. The higher the RUL, the higher the relevance of establishing a fault diagnosis. By simple inspection of the spectrum, the limit at which the BCFs and their harmonics could no longer be clearly identified in Zone B corresponded to approximately 30% RUL. It is not implied that the unfeasible detection at higher RUL confirms the absence of faults (the fault existence is corroborated by the presence of side-bands in Zone C of the spectrum). It was however suggestive that other techniques needed to be applied for earlier fault diagnosis.

### Signal demodulation for early identification of BCFs in the spectrum

For an earlier fault diagnosis (and working under the assumption that the bearing above 30% RUL was in the second stage of degradation as Zone C presented higher energy spectral density), a high-frequency identification technique known as enveloping was used. Enveloping demodulated the defect signal from the structural resonance excited by the defect. After the demodulation, the identification of the BCFs and their respective harmonics could be done by a simple inspection of the envelope spectrum. The entire procedure of the analysis is detailed below and can also be seen in Fig. 5:

- Identification of the bandwidth of interest through Spectral Kurtosis (SK).
- Band-pass filtering of the vibration signal.

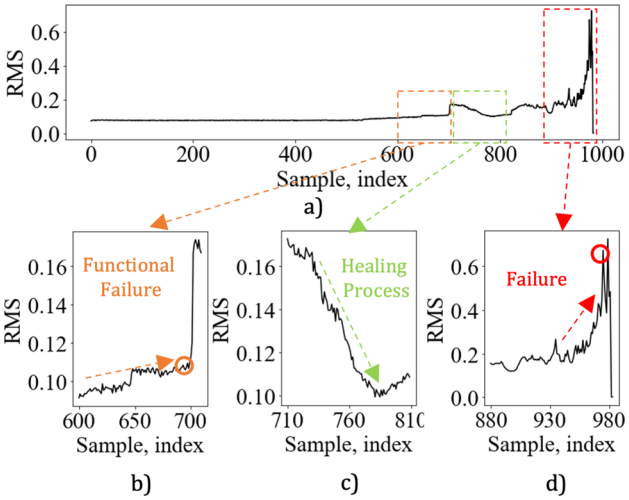


**Fig. 4** The RMS of the vibration signal: **a** the entire recorded data; **b** the functional failure point (encircled in orange); **c** the healing process; **d** the complete failure point (encircled in red) (Color figure online)

- Signal demodulation and identification of the BCFs on the spectrum of the envelope signal.

SK is a statistical parameter that indicates how the impulsiveness of a signal varies with frequency. SK has been considered a fundamental tool for determining the most appropriate band for envelope analysis and obviates the need for historical data (Sawalhi & Randall, 2004). One of the most important properties of SK that has been used in bearing fault diagnosis is that the SK of a non-stationary signal (e.g., the vibration signal of a faulty bearing) with stationary noise has large values at the frequencies where the signal to noise ratio is high (Antoni & Randall, 2006). Thus, the SK of the vibration signals starting from 30% RUL was computed based on the Short-time Fourier Transform STFT $X(t, f)$ (Randall, 2011) as follows:

$$SK(f) = \frac{\langle X^4(t, f)\rangle}{\langle X^2(t, f)\rangle^2} - 2 \tag{5}$$

where $X^2(t, f)$ corresponds to the power spectrum values calculated for each time, and $\langle X^2(t, f)\rangle$ is the average of all these power spectral values, i.e., the power spectrum of the analyzed signal as a whole. When using STFT, parameters such as window length can affect the spectral kurtosis calculation (Leite et al., 2017). To evaluate the proposed methodology, the effect of several window lengths on the resulting SK was investigated and superposed on a Frequency-Spectral Kurtosis plot (Fig. 7). The frequency band and window length that provided the higher but also consistent SK were visually determined. The window length selected for this case study was 64.

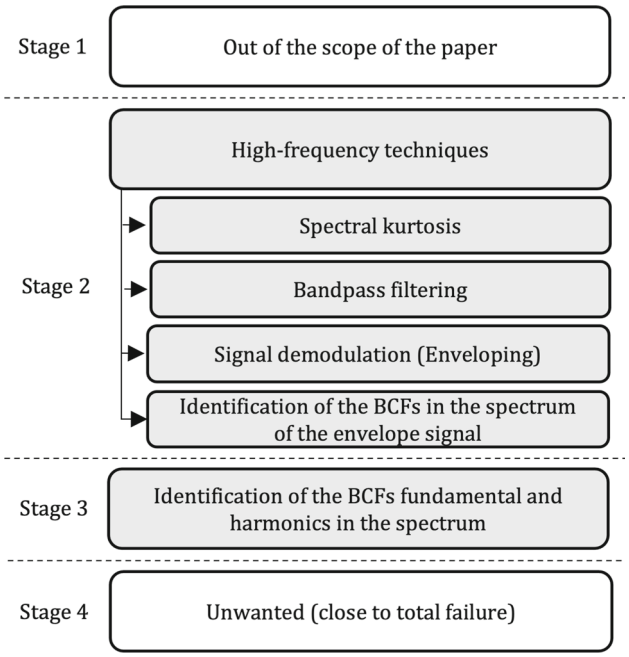


**Fig. 5** Methodology followed for the knowledge-based approach. The steps considered in this work are highlighted in grey

Once the window length was defined, the SK for the vibration signals that corresponded to a RUL higher than 30% was computed and two frequency bands of interest (with the highest kurtosis values) were selected for further analysis: [3300, 4500]Hz and [7800, 9100]Hz, see Fig. 8. These bandwidths were located in Zone C, which corresponded to the zone containing the high-frequency natural resonances of the bearings

**Table 2** Theoretical estimation of BCFs based on the test rig setup Qiu et al. (2006)

| | |
|---|---|
| BPFO | 236Hz |
| BPFI | 297Hz |
| BDF | 280Hz |
| FTF | 15Hz |

(Graney & Starry, 2012), confirming the assumption of Stage 2 degradation. Even if the estimation of the bearing resonant frequencies could be done accurately, it was not feasible to determine which of those was likely to be dominant. Therefore, in the initial stage of the analysis, it was important to have a sufficiently broad frequency range that allowed for a full examination of the spectrum (McFadden & Smith, 1984).

The selected bandwidths of interest were then used to band-pass filter the vibration signals, which shifted the output to a low-frequency region in order for it to be subjected to envelop detection.

The enveloping was performed based on the Hilbert transform, i.e., taking the spectrum of the amplitude of the so-called analytic signal. The analytic signal is a complex time signal whose imaginary part is the Hilbert transform of the real part (Randall, 2007). Its imaginary part is identical to the real part but delayed by $\pi/2$. The demodulation phase ended by obtaining the spectrum of the envelope of the analytic signal that allowed the identification of the BCF at the appropriate passing frequency. Figure 9 shows the different stages of an artificial (synthetic) signal and its envelope analysis.

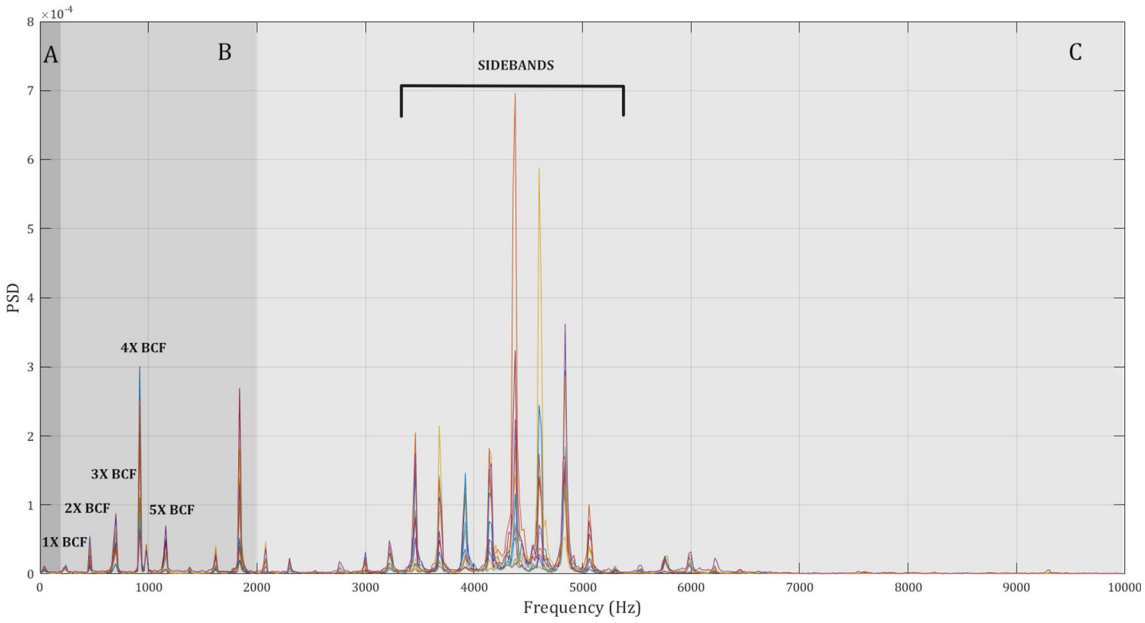


**Fig. 6** PSD of the last ten instances of vibration data collected. Zones A (left), B(center) and C (right) have been highlighted with gray scale (Color figure online)

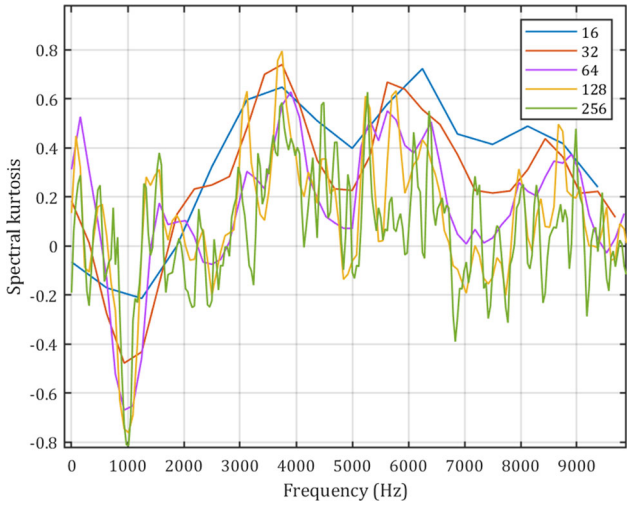


**Fig. 7** SK of the vibration signal for different window lengths (Color figure online)

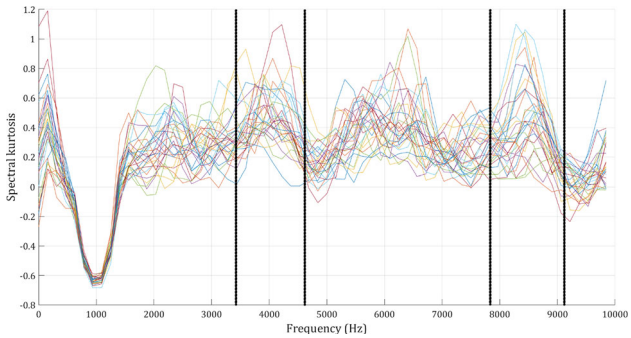


**Fig. 8** Bandwidth identification for band-pass filtering by estimating the spectral kurtosis for 50 instances of the data set analyzed. The vertical black lines show the limits of the bandwidths of interest (Color figure online)

## Data-driven approach

The identification of the BCF and the corresponding failure mode without any information about bearing characteristics and test parameters was accomplished by conceiving and subsequently implementing Workflow 1 (Fig. 10). Different filtering frequency ranges were investigated in order to determine the bandwidth of interest for the monitoring of the bearing condition for early fault diagnosis. The implementation of this investigation was performed by following Workflow 2 (Fig. 10). Steps 1 and 5 differed between workflow 1 and 2 whereas the other steps were identical. These differences are specified and explained in the following subsections.

### Sub-bands isolation with fixed bandwidth size

Defects in various bearing components will result in changes in the vibration signal that are reflected in different frequency

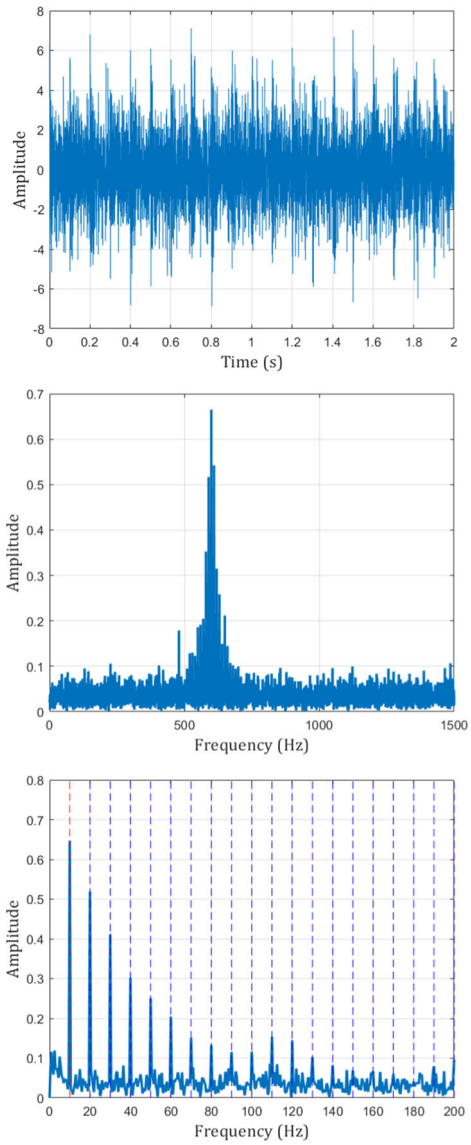


**Fig. 9** Envelope analysis on a synthetic signal: **a** Synthetic signal composed of a transient impact frequency representing a defect (10 Hz), a modulating sinusoidal function (600 Hz), a shaft signal (low frequency sinusoidal at 13 Hz) and Gaussian white noise $N(0, 1)$; **b** spectrum of the synthetic signal; **c** spectrum of the envelope of the analytic signal with dashed lines representing the defect signal fundamental (red) and harmonics (blue) (Color figure online)

bands. Therefore, the raw data was transformed into a group of sub-bands (Fig. 11). Each of these sub-bands contained only the signal components within a particular frequency range. A Butterworth band-pass filter was used for filtering the raw signal. Within Workflow 1, the algorithm of isolation of the frequency sub-bands was implemented from the minimum frequency $f_{\min} = 1$Hz to the maximum frequency $f_{\max} = 10000$Hz. The filtering of the raw data was performed in $BW$ range from $f_{\min}$ to $f_{\max}$ with various filtering $step$ sizes, hereafter called $step$. A lower boundary ($f_{\mathrm{L}}$) and a higher boundary ($f_{\mathrm{H}}$) of the Butterworth band-pass filter

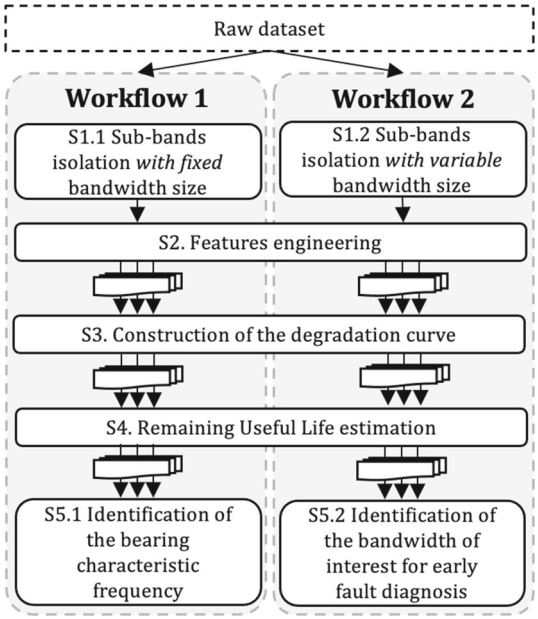


**Fig. 10** The workflow of the data-driven approach

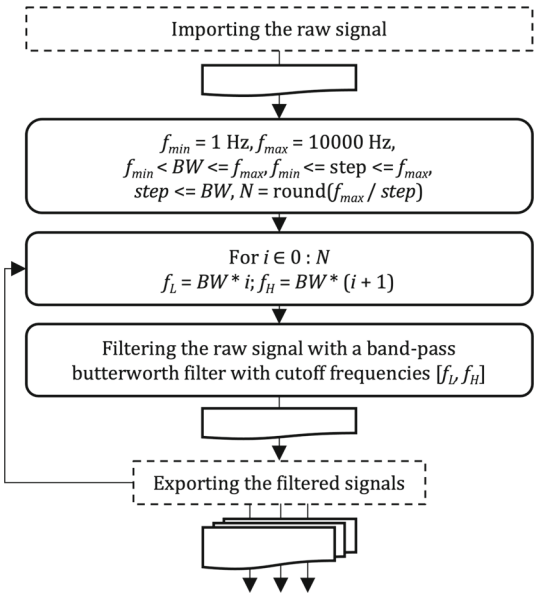


**Fig. 11** Step 1.1: The algorithm of sub-bands isolation with fixed bandwidth size

were calculated for each one of the isolating sub-bands $i$ as shown in Fig. 11.

The choice of values of the $BW$ and $step$ variables affect the overall results and the calculation time of the frequency sub-bands isolation algorithm. A wide $BW$ leads to the blurring of the important information contained in certain frequencies. On the other hand, a narrow $BW$ is sensitive to random noise frequencies. In order to cover the whole frequency range, the values of the variables must follow the rule whereby $step <= BW$. The selected value of the $step$ might be several times smaller than a $BW$. However, the number of sub-bands and, as a consequence, the number of performed calculations depended on the numerical size of the selected $step$. A smaller $step$ size leads to an increase of both aforementioned parameters.

If the real response of the Butterworth filter were to be compared with an ideal response, some differences appear. The real response has a transition zone between a passband and a stopband. The "tail" of the signal will be present outside of the ideal boundaries of the cutoffs (indicated in yellow in Fig. 12b). Furthermore, frequencies with suppressed signals are present within the ideal boundaries of the cutoffs (indicated in orange in Fig. 12b). The selection of $BW$ and $step$ variable values needs to take this characteristic property of the Butterworth filter into account.

While the presence of the tail area only blurs the filtering boundaries, the suppression of the signal within the boundaries leads to a loss of information. In the case of $step = BW$, the boundaries of cutoffs are not overlapping (see Fig. 12a). This information loss can be prevented by selecting a $step$ size significantly smaller than the $BW$, as shown in Fig. 12c. In that case, the filtering boundaries will overlap with the neighbouring boundaries (Fig. 12d). The overlap will obviously lead to a doubling of the information of the neighbouring sub-bands, but it prevents information loss.

In this research work, the $BW$ and $step$ were determined experimentally. Different sizes of $BW$ and $steps$ were compared in a preliminary test the results of which are presented in Sect. Selecting the variables.

### Sub-bands isolation with variable bandwidth size

As was discussed in the previous Section, the presence of the fundamental and its harmonics of the bearing characteristic frequencies in Zone B and present side-bands in Zone C. This corresponds to Stage 3 of degradation. The performances of different filtering ranges have to be compared in order to identify the bandwidth of interest for the early fault diagnosis. Within Workflow 2, all possible combinations of lower boundaries $f_L$ and higher boundaries $f_H$ of the band-pass filter were used for the isolation of the sub-bands. The modification to the algorithm (Workflow 1) kept the $BW$ size as a variable. The modified algorithm is presented in Fig. 13.

### Features engineering

In this step of the analysis, useful features must be extracted from the raw data set. The features engineering algorithm initially generated a large number of various features. This was followed by a multi-step evaluation procedure which only retained the most important features. These retained features were then used for all further calculations in the data-driven workflow for each of the created sub-bands. The

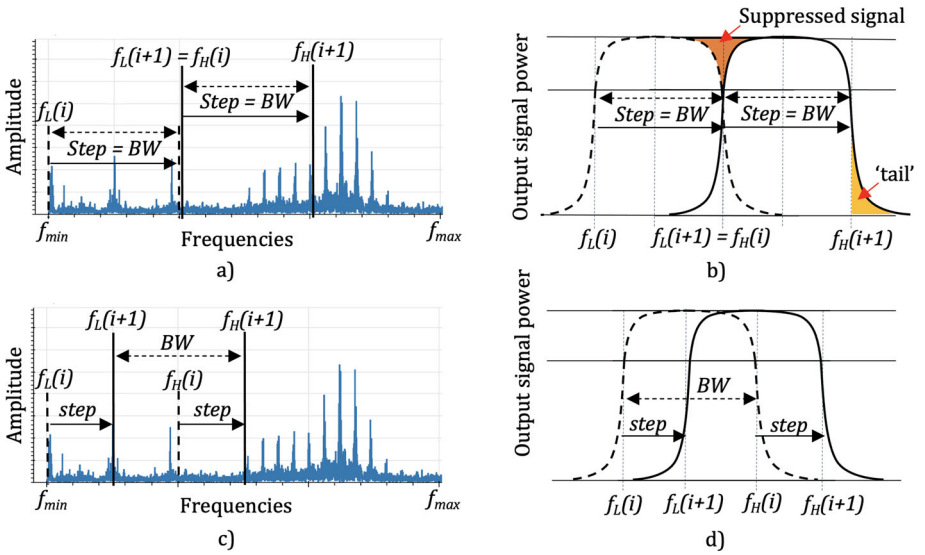


**Fig. 12** Example of sub-bands isolation: **a** filtering boundaries for the case with overlap; **b** filter design for the case with overlap; **c** filtering boundaries for the case without overlap; **d** filter design for the case without overlap (Color figure online)

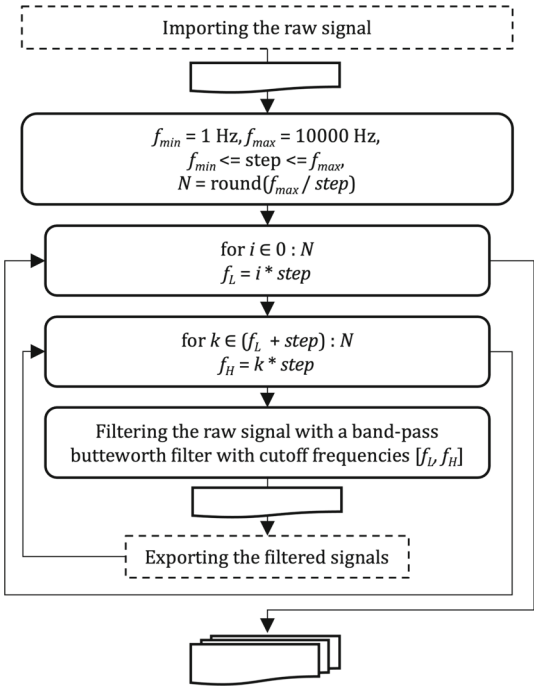


**Fig. 13** Step 1.2: The algorithm of sub-bands isolation with variable bandwidth size

workflow of the features engineering algorithm is presented in Fig. 14.

- *Features generation stage* The majority of the features were created by using two libraries: Time Series Feature Extraction Library (TSFEL) and Time Series Feature Extraction based on Scalable Hypothesis tests (TSFRESH) (Barandas et al., 2020; Christ et al., 2018). Any other features not included in these libraries and that could be appropriate for the specific case were also added at this stage. In addition, signal features that provided general signal-based statistical metrics which can be applied to any kind of signal were manually added from the lists provided by MathWorks [58].
- *Features reduction stage* The features that were most valuable and relevant for the study of the present case

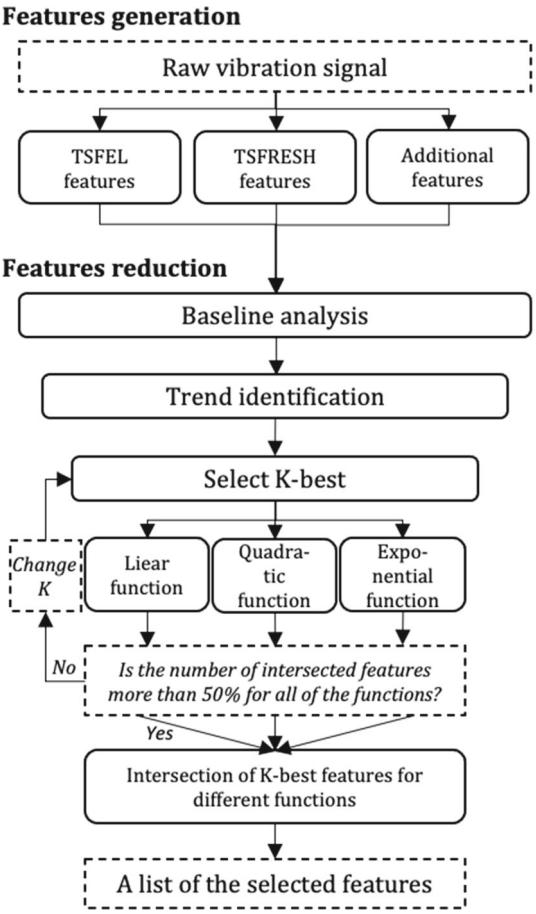


**Fig. 14** Step 2: Features engineering algorithm

were identified and selected for further processing. The selection was carried out in several successive stages, such as (i) baseline analysis, (ii) trend identification, and (iii) supervised selection of top features. *(i) Baseline analysis* This analysis included the identification and removal of duplicates, highly correlated features (more than 95%), and features with low variance. Since variance is range-dependent and different data sets might have different requirements, only zero-variance features were removed. *(ii) Trend identification* The investigated data set contained complete records of the degradation processes of bearings. Therefore, a developing trend of the most valuable features would show the condition of the deteriorating bearing (Lacey, 2008). All the features were normalized by scaling between 0 and 1 for comparability. Calculated slope values of linear fits to features were sorted from the highest to lowest. Features with a slope value lower than the overall median were removed from the features list. *(iii) Supervised selection of top features.* The SelectKBest function in Python's Scikit-learn library selects the top features based on univariate statistical tests [59]. The function refers to a supervised method, which requires a labelled data set. Even though the data set used for this investigation was not labelled, the model representing a theoretical DC had to start with functional integrity (degradation level set at 0%) and end with a failure (degradation level set 100%) since the data set contained the complete record of the run-to-failure experiment. In order to avoid overfitting, a low order function had to be used for the fitting of a DC, so the linear, quadratic and exponential curves were implemented simultaneously. The adaptability to different shapes of DCs was kept by selecting only the intersecting features from the $k$ top selected for each of the fitting functions. If the intersection list contained more than 50% of $k$ top features for each of the fitted functions, then it was used as the final selection. Otherwise, the supervised selection of top features should be repeated with an increased $k$ value.

### Construction of the degradation curve

As was described in the previous Section, the analysis was carried out only for the data obtained until the occurrence of a functional failure of a singular defined type. It means that the degradation reaches 100% on the 702nd data collection instance (Fig. 4).

The DC was constructed by fusing the features selected in Sect. Features engineering step of the data-driven workflow. The DCs were constructed for each of the isolated sub-bands by implementing the method, developed by Wang et al. (2008). In particular, a multiple linear regression model was used to build a condition indicator for the progression of the degradation. The resulting degradation curves were smoothed in order to eliminate errors due to the noise (Fig. 15).

### Remaining useful life estimation

The shapes of DCs varied depending on the input features and filtering bands. Therefore, different DCs will have different remaining times between the point of reaching a particular degradation level and the functional failure. The higher the RUL, the better the capability of the sub-band to identify a developing functional failure earlier. The RUL values of different sub-bands have to be calculated at the same level of degradation to be comparable. The particular degradation level can be set in the range between 0 and 100%, where 0% represents the healthy state and 100% represents the occurrence of a functional failure. The result of calculations of RULs of two DCs for two different sub-bands at a level of degradation of 50% is presented in Fig. 15.

The aim of Workflow 1 was to compare the RULs of different frequency ranges (Fig. 15). At low levels of degradation (0–25%), the difference between the DCs was insignificant. Furthermore, the slow increase over time of the degradation curve leads to there being considerable stretches of time without any noticeable change in the health stage. The curves at a level of degradation of 75–100% did not significantly differ as well. Contrasting their characteristics at the early stage, the DCs at the late stage of degradation are usually very steep. The fast increase over time of the DC leads to an extensive difference in the health stage occurring within a relatively short period of time. The largest relative difference of performances of multiple DCs were found at a level of around 50% of degradation. Therefore, the results for 50% of degradation are considered as the most representative values. The results are presented in Sect. Identification of the BCF and the corresponding failure mode, as RUL over the frequency sub-bands.

In Workflow 2, the aim was to identify the specific sub-bands that indicate a developing failure at an earlier stage of degradation. Therefore, the results for 10% of degradation level were used for comparison of the RULs and presented in Sect. Identification of the bandwidth of interest for the early fault diagnosis .

### Identification of the BCF and the corresponding failure mode

As the last step of Workflow 1, the identification of the BCF and the corresponding failure mode was performed. It is clearly visible in Fig. 6 that harmonics of the BCF can be easily spotted as equidistant peaks of the amplitudes over frequencies. The same ratio should be reflected in the results,

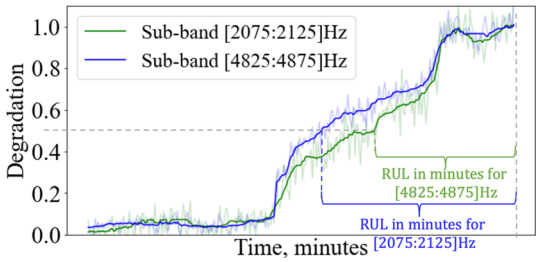


**Fig. 15** Examples of the calculation of the RULs at 50% degradation level for two different sub-bands (Color figure online)

showing the RUL over frequency sub-bands. Averaging over the integer multiples of base frequencies makes it possible to identify the BCF, since it results in a larger average RUL than averaging the integer multiples of other frequencies. The base frequencies in the range between 1Hz and between 1000Hz were used in the calculations as possible options for the BCF. Plotting the averaged RULs over the base frequencies and employing a peak detecting algorithm allows identifying the BCF.

The identification of the corresponding failure mode was made by comparing obtained peak values with calculated peak values (Table 2). It must be mentioned here that this step required inputting information from the results of the knowledge-based approach. An improvement, in this case, might be the implementation of a classification model for distinguishing the different failure modes. However, the availability of sufficient data to train a classification model might be a bottleneck of this solution.

### Identification of the bandwidth of interest for the early fault diagnosis

For Workflow 2, the sub-bands were isolated in the Step 1.2 of the data-driven workflow. The degradation performances of those sub-bands were calculated for the early stage of degradation. The threshold to calculate the RUL values was set at 10% degradation level. The range with the highest performance was defined as the bandwidth of interest for the early fault diagnosis. Those boundaries were used to filter the raw vibration signal and to identify the fault.

## Results and discussion

The results of the knowledge-based and the data-driven approach, as well as their comparison and discussion are presented in this Section.

### Results of the knowledge-based approach

Following sub-sections show how the BCFs were directly identified by inspection of the spectra at a very late stage of degradation and presents the results from the application of enveloping aiming at achieving earlier diagnostics.

#### Identification of the BCFs and their harmonics in the spectrum

The knowledge-based approach was applied to the data set of *Test 2*. As a first step, the BCFs were identified in Zone B of the spectra. The visual inspection of the spectra started at 0% RUL (972nd data collection instance) and was done iteratively (Fig. 16) until approximately 30% RUL (equivalent to 29.1 h). As seen in Fig. 16, prominent peaks in the spectra, particularly in Zone B (inscribed in dashed line), are clearly recognizable and they are noticeably close to the calculated value of the BPFO fundamental and its corresponding harmonics (Table 2).

The calculated BCFs are theoretical estimates that do not consider phenomena such as slippage (McInerny & Dai, 2003; Randall & Antoni, 2011). The actual bearing frequencies can deviate typically around 1–2% (Fig. 16), both as a deviation from the calculated value and as a random variation around the mean frequency (Randall & Antoni, 2011).

As the target is early fault diagnosis, the same visual inspection of the spectra was attempted in data sets retrieved from potentially earlier stages of degradation. However, as can be seen from Fig. 17, above 30% RUL, the harmonics of BPFO could no longer be identified by a simple inspection of the spectra. This might be associated with the amplitude modulation effect, which makes a distinction of the BCFs in the spectra unfeasible due to the large modulation index.

#### Signal demodulation for early identification of BCFs in the spectrum

Figure 17 shows energy concentration frequency bands that could be interpreted as the presence of a failure indicative of a stage 2 degradation. After defining the bandwidths of interest with the help of SK ([3300,4500] Hz and [7800,9100]Hz), the band-pass filtered vibration signals above 30% RUL were demodulated by enveloping.

After demodulation within the bandwidth of interest from 3300 to 4500 Hz, the direct identification of the BPFO and its harmonics (up to the fifth) in the envelope spectrum was feasible. This is shown in Fig. 18 which illustrates the envelope spectrum around 45% RUL (equivalent to 58.7 h). For the present experimental set-up, this instance was the lowest level of RUL at which the fault diagnosis by identification of the BCFs on the envelope spectrum was feasible with the methodology proposed.

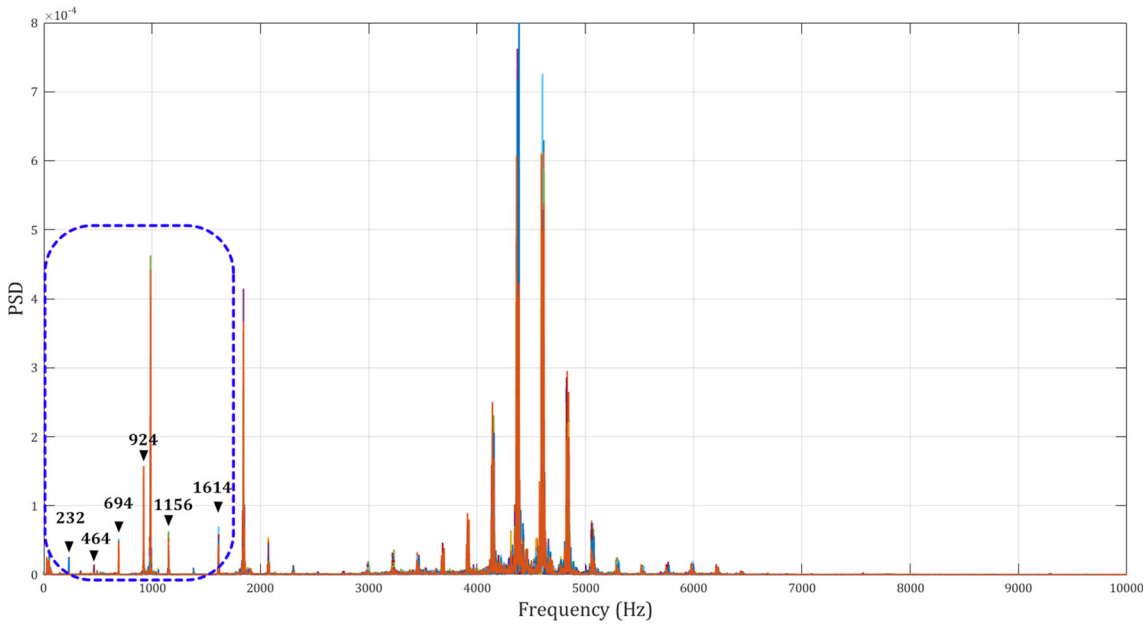


**Fig. 16** Identification of the fundamental BPFO and its corresponding harmonics in Zone B of the power spectra, which corresponds to the description of Stage 3 of bearing degradation (Color figure online)

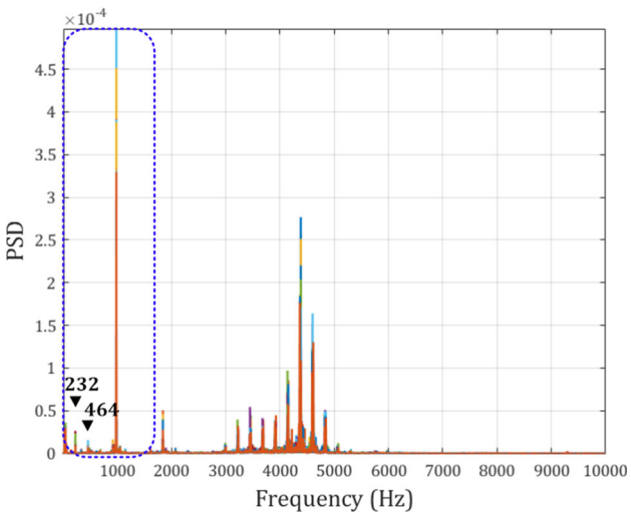


**Fig. 17** PSD of data sets above 30% RUL (Color figure online)

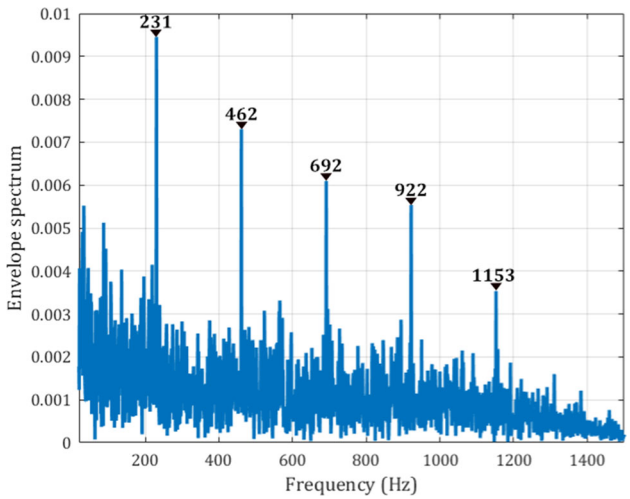


**Fig. 18** Spectrum of the envelope around 45% RUL corresponding to the bandwidth of interest from 3300 to 4500Hz

The bandwidth of interest from 7800 to 9100 Hz made fault diagnosis also feasible (Fig. 19). However, a clear identification of the BCFs in the spectra could not be achieved at levels lower than 32% RUL. As this is similar to the same percentage of RUL used for the visual inspection of Zone B of the spectra, the implementation for this particular bandwidth does not justify the additional effort. Nevertheless, it is arguable that this bandwidth might be capable of achieving an earlier diagnosis, considering that it could be affected more directly by the resonance located at approximately 9300 Hz (potential bearing or system natural frequency, visible throughout the evolution of the spectra, Fig. 3). However, the amplitudes of the vibration signals within this bandwidth are greatly affected by the sensor's frequency range (3 dB) which corresponds to 12 kHz (PCB-Piezotronics).

## Results of the data-driven approach

Following sub-sections contain the results of preliminary tests for differentvalues of BW and step, the identified fun-

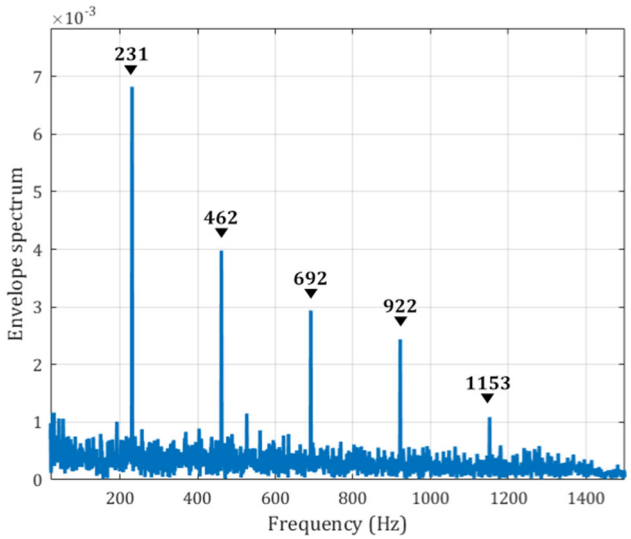


**Fig. 19** Spectrum of the envelope around 32% RUL corresponding to the bandwidth of interest from 7800 to 9100 Hz

damental BCFs with the corresponding failure mode, and the identified bandwidth of interest.

### Selecting the variables

The selection of a BW and of the *step* (Workflow 1, Fig. 10) was made by comparing the results of preliminary calculations. For those preliminary calculations, the signal frequencies within the range, containing the high side-bands were used (from 4200 to 5200 Hz). The procedure for selecting the variables was divided into two stages. In stage (i), a comparison was made of the results for different sizes of bandwidths without overlapping filtering boundaries while keeping the ratio at $step = BW$. In the subsequent stage (ii), a comparison was made of the results for different $step$ sizes of $BW$ [selected in stage (i)] with overlapping filtering boundaries while keeping the ratio at $step < BW$.

- At stage (i), BWs equal to 100 Hz, 50 Hz, and 10 Hz were compared (Fig. 20a–c). The $BW = 100$ Hz was too wide and most of the predominant frequencies were smoothed (Fig. 20a). Furthermore, the RUL was significantly lower in comparison to the other two cases. The $BW = 50$ Hz was able to distinguish the predominant frequencies (Fig. 20b). The RUL increased with decreasing BW size. The $BW = 10$ Hz showed the highest RUL, which meant that it was more sensitive to developing degradation patterns (Fig. 20c). However, calculations using a small BW are very time-consuming and might be too sensitive to the random noisy frequency amplitudes. Decreasing BW below 10 Hz would lead to an increase in computational time and unnecessary noisy results. Consequently, the $BW = 50$ Hz was used in the final calculations.
- At stage (ii), a filtering $step$ size was selected, by comparison of the results for the values 25 and 10 Hz. The results of the preliminary calculations (Fig. 20d–f) show that there was no significant difference in performance between the different $step$ sizes. However, the smaller $step$ size 10Hz has the disadvantage of increasing the computational time to more than double compared to 25 Hz (Fig. 20e, f). Therefore, in the final calculations, the $step$ size equal to 25 Hz was used in combination with the $BW$ size equal to 50 Hz (Fig. 20e).

### Identification of the BCF and the corresponding failure mode

After assigning the variables, the calculation in the range encompassing the whole of the recorded frequencies was performed. The results for the 50% of degradation level are presented in Fig. 21a as the RUL values of each of the calculated sub-bands. It is clear that the highest RULs are in the range between 4000 and 5000 Hz.

The average values of RULs of base frequencies are presented in Fig. 21b. The average of all the RULs is equal to 3865 Hz which corresponds to base frequency 1Hz in the X-axes of the figure. The most prominent peaks of the average values were identified and marked with red dots in Fig. 21b. The first peak was identified at 241Hz and four out of five identified peaks were exact multiples of that value. All the values, representing the harmonics of the first peak are shown by green vertical lines in Fig. 21b. It is obvious that each of these harmonics corresponds to a peak value strongly suggesting that defining the frequency equal to 241 Hz as the $BCF$ was a correct approach.

According to Table 2, the closest ball pass frequency to the identified $BCF$ was the theoretical estimation of the frequency corresponding to an outer race failure BPFO which is equal to 236 Hz.

### Identification of the bandwidth of interest for the early fault diagnosis

To identify the bandwidth of interest for the early fault diagnosis, all possible combinations of the lower boundaries ($f_L$) and higher boundaries ($f_H$) of the band-pass filter were compared. The calculations were performed with the $step = 500$ Hz.

The RULs thus obtained are presented in Fig. 22 as a heatmap, where the x-axes represent $f_L$ and the y-axes the $f_H$ of the bandwidth. The colour map represents the calculated RULs for the sub-band with the corresponding value of $f_L$ and $f_H$, normalized by scaling between 0 and 1. The darkest green corresponds to the highest normalised RUL

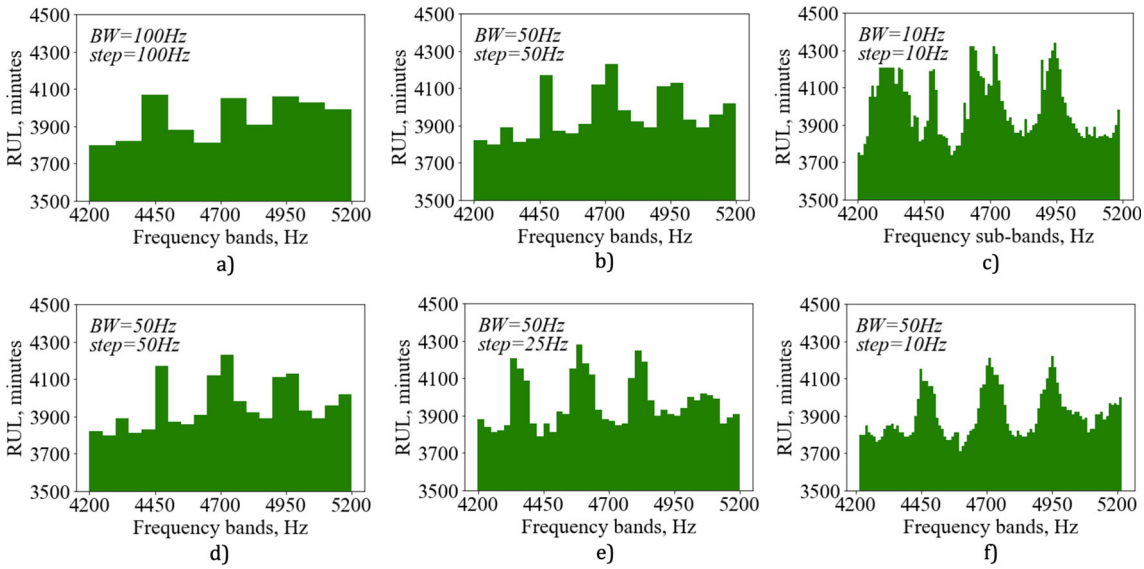


**Fig. 20** Results of preliminary calculations with various BWs and step sizes

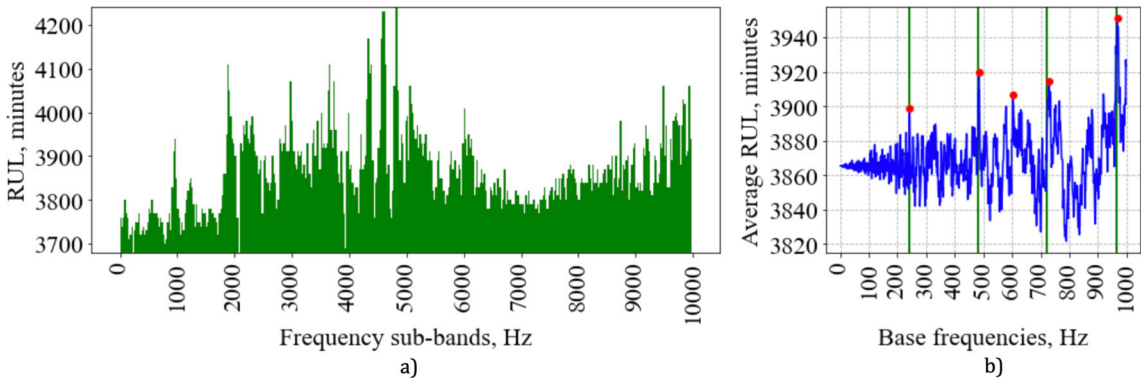


**Fig. 21** The results of Workflow 1: **a** RULs over the frequency sub-bands; **b** averaged RULs over the base frequencies (Color figure online)

value equal to 1 and the white colour corresponds to the lowest normalised RUL value equal to 0. The higher the RUL, the earlier the deterioration of the corresponding degradation level is detected.

In Fig. 22, the results are presented for three stages of degradation (10%, 50%, and 70%). It can be seen, that at the different levels of degradation, the zones with high RULs vary (encircled in orange in Fig. 22). At the early stage of the degradation, the highest RUL is in the range of the $f_L$ =8000Hz and $f_H$ =9000 Hz. The RUL of 10% degradation level identification was 5960 min (60% of RUL). When the degradation reaches 50%, the highest values of performances begin to appear in the zone [4000,5000] Hz where the side-bands were defined. Increasing degradation leads to an increase of the RULs in almost all of the other ranges.

From the present analysis of the results, the bandwidth of interest recommended for real-time monitoring for early fault diagnosis is [8000,9000] Hz for band-pass filtering. Using those boundaries for the filtering identified 10% of the degradation at 64% RUL (equivalent to 258 h).

## Comparison of results

Both the KB and DD approaches allowed to reach the targeted aims which can be summarized as (i) identification of the BCF in the vibration spectra, and (ii) identification of the bandwidth of interest for early failure detection and diagnosis. To further compare and evaluate the capabilities of the two different approaches, parameters such as implementation time, domain knowledge, knowledge associated with data processing, data requirements, diagnostic performance, and practical applicability were analyzed. The summary of these comparisons is presented in Table 4.

- *Implementation and computational time* For the KB approach, the implementation time consists predomi-

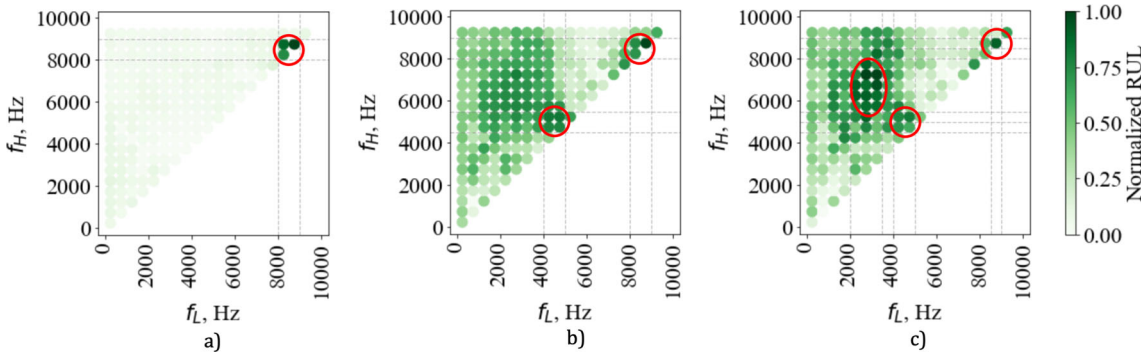


**Fig. 22** Comparison the RULs of different sub-bands: **a** for 10% of degradation; **b** for 50% of degradation; **c** for 70% of degradation (Color figure online)

**Table 3** Summary of diagnostic performance of both approaches

| Approach | KB | DD |
|---|---|---|
| Identification of the BCF | 232 Hz | 241 Hz |
| Identification of the failure mode | Identified | Requires additional information |
| Identification of the bandwidth of interest | 3300–4500 Hz | 4000–5000 Hz |
| RUL | 45% | 46% |

nantly of algorithm programming in which the high-frequency techniques are the most time-consuming. Once the algorithm (programmed in MATLAB) was prepared and given that it was not computationally demanding, the processing time was directly dependent on the amount of data sets to be processed. In the case of the present analysis, for 972 data sets containing 20,480 data points each, the processing time was less than 10 min on a standard laptop for both Stages 3 and 2. In the case of the DD approach, the overall computational time was significantly higher. This is due to the fact that even though the implemented data analysis algorithms were basic, the structure of the workflow containing a large amount of steps was time-consuming. Thus, calculation times were strongly dependent on the selected calculation parameters, such as the size of the $BW$ and the $step$. As an example, the resulting calculation times of Workflow 1 with $BW$ =50 Hz, and $step$ = 25 Hz was more than two hours on a standard laptop.

- *Domain knowledge* The KB approach requires a considerable understanding of rolling element bearings, fault detection and diagnostics techniques, and vibration signal analysis. On the other hand, the DD approach was constructed in a way that no domain knowledge was required.
- *Data processing associated knowledge* Both the KB approach and DD approaches required knowledge of digital signal processing. Additionally, basic analytic tools were required in the DD analysis for the features engineering process and modelling of the degradation curve.
- *Data requirements* This evaluation parameter covers both the vibration signals and any additional information that was provided as input to the analysis (e.g., test rig setup and testing parameters). For the identification of the BPFO in the spectra using the KB approach, the bearing characteristics and testing parameters were required. However, the approach does not require data from the full life cycle of a bearing. The DD approach only used the raw vibration signal and can be used for cases when the information on the system parameters and components is not available. However, the approach requires the full life cycle data of a bearing for training purposes since it has no inherent predictive capability. The current DD algorithm requires modifications if it will be applied to cases that have not yet failed.
- *Diagnostic performance* The agreement between the theoretical estimate of the BPFO (236 Hz) and the experimentally identified values with both the KB and DD approaches (232 Hz and 241 Hz) are within an admissible relative error of 2.1% and 2% respectively, making both approaches essentially comparable. The results also showed important energy concentrations between 3300 and 4500 Hz which could be used for fault diagnosis up until 45% RUL for the KB approach. For the DD approach, a comparable range was identified (between 4000 and 5000 Hz with 46% RUL). Furthermore, for the DD approach the range between 8000 and 9000 Hz showed a slightly higher RUL value equal to 47%. How-

**Table 4** Comparison of the KB and DD approaches

| Parameter | Performance of identification of the BCFs | | Performance of identification of the bandwidth of interest | |
|---|---|---|---|---|
| | KB | DD | KB | DD |
| Time | Low | High | Medium | High |
| | *Low—less than 30 min;* | | | |
| | *Medium—between 30 and 120 min;* | | | |
| | *High—more than 120 min;* | | | |
| Domain Knowledge for early fault diagnosis | Medium | N/A | high | N/A |
| | *Low—basic mechanical knowledge;* | | | |
| | *Medium—vibration and signal analysis knowledge;* | | | |
| | *High—considerable diagnostics knowledge;* | | | |
| Data processing knowledge | Low | Medium | Medium | Medium |
| | *Low—basic knowledge;* | | | |
| | *Medium—general knowledge;* | | | |
| | *High—expert knowledge;* | | | |
| Data requirements | Medium | Low | Medium | Low |
| | *Low—only the vibration signal;* | | | |
| | *Medium—vibration signal and bearing characteristics;* | | | |
| | *High—all the above and indicative of the type of defect;* | | | |
| Diagnostic performance | High | Medium | High | Medium |
| | *Low—fault detection;* | | | |
| | *Medium—fault isolation;* | | | |
| | *High—fault identification;* | | | |
| Practical applicability | Medium | medium | medium | medium |
| | *Low—same type of defect;* | | | |
| | *Medium—other types of defects but not simultaneous;* | | | |
| | *High—multiple types of defects and simultaneous;* | | | |

ever, this range looks less attractive since the amplitudes of the vibration signals within this bandwidth are greatly affected by the sensor's frequency range. The summary of the diagnostic performance is presented in Table 3.

- *Practical applicability* As the theoretical estimates of the BCFs were already calculated when using the KB approach, it seems feasible to utilize the same algorithm for the identification of other modes of bearing failures (single failure mode per data set). It is arguable that the algorithm might also be applicable for the diagnosis of multiple failure modes combined in the same data set. However, this requires further experimentation because the superposition of different frequencies associated with "failures" in bearing components might require the implementation of additional techniques for discrete signal separation and enhancement of the signals associated with failure. Additionally, changes in the test rig setup can also be handled by the algorithm by modifying the input parameters for the BCFs estimation. The DD approach might be implemented in cases with different testing parameters such as variation of rotating speed. It is however recommended that, within the investigated data set, the parameters should stay unchanged. In general, the excessive time required for processing, and the low level of traceability of the DD approach make this approach less attractive for implementation in cases where there exists enough information to implement the KB approach. It is essential that the DD approach performs equally well as the KB approach for those cases analysing and diagnosing other failures resulting from singular damages. However, the proposed DD approach is not implementable for cases where multiple failures develop simultaneously. For these conditions, further investigations are required.

## Conclusions

Identification and monitoring of bearing characteristic frequencies is an essential step for early fault identification in bearings. Various defect modes such as outer race, inner race, cage, and rolling elements failures will present their own characteristic frequencies and corresponding harmonics. Therefore, monitoring the changes in a particular frequency range might give an understanding of the developing failure mode which can enhance Remaining Useful Life and optimize scheduling maintenance interventions.

In this paper, two approaches (knowledge-based and data-driven) were investigated to solve two specific problems, namely, the identification of the Bearing Characteristic Frequencies in the spectra, and the identification of the bandwidth of interest for early fault detection and diagnosis. Performances of each of the approaches are compared in order to identify their strengths and limitations. The comparison takes into account theoretical and practical aspects such as implementation time, domain knowledge, data processing associated knowledge, data requirements, diagnostic performance, and practical applicability. In addition, the ability of the data-driven method to achieve the desired result with limited information about the bearing parameters was investigated.

The knowledge-based approach requires considerable domain knowledge and might require other diagnostic techniques for the detection and diagnosis of complex, combined failures, but is straightforward for single failure cases and has the advantage of short implementation time. The data-driven approach is time-consuming and not traceable but can be used for single failure cases when not enough information about the system is available. The essential disadvantage of the data-driven approach is that it requires data from the entire life cycle of a bearing.

This paper describes an early state of research and is not yet intended to be run in the surveillance of the safety critical system components. Further verification and validation of the methods implemented within the knowledge-based and data-driven approaches are needed to build robust and reliable solutions.

In order to further improve detection and diagnosis, a hybrid approach combining knowledge-based and data-driven approaches could be advantageous. For instance, the knowledge-based approach could benefit from the features engineering process. On the other hand, the data-driven approach could benefit from the time-reduction and interpretability that the knowledge-based approach provides. Future work will pursue the implementation of these methodologies and the evaluation of their individual performance for the diagnosis of ball screws.

**Acknowledgements** The authors are grateful for the financial support provided by Vinnova (Grant Agreement ID: 2018-05033) and research centers KTH DMMS and XPRES.

**Author Contributions** EI: conceptualization, methodology, formal analysis and investigation, writing—original draft preparation, writing—review and editing. MKG: conceptualization, methodology, formal analysis and investigation, writing—original draft preparation, writing—review and editing. KS— conceptualization, methodology, writing—review and editing. AA—writing—review and editing, funding acquisition, resources, supervision.

**Funding** Open access funding provided by Royal Institute of Technology.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material

in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References


Ahmad, W., Ali Khan, S., & Jong-Myon, K. (2018). A hybrid prognostics technique for rolling element bearings using adaptive predictive models. *IEEE Transactions on Industrial Electronics (1982), 65*(2), 1577–1584.

Akpudo, U. E., & Hur, J.-W. (2020). Towards bearing failure prognostics: A practical comparison between data-driven methods for industrial applications. *Journal of Mechanical Science and Technology, 34*(10), 4161–4172.

Antoni, J., & Randall, R. B. (2006). The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing, 20*(2), 308–331.

Attoui, I., Fergani, N., Boutasseta, N., Oudjani, B., & Deliou, A. (2017). A new time-frequency method for identification and classification of ball bearing faults. *Journal of Sound and Vibration, 397*, 241–265.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., & Gamboa, H. (2020). Tsfel: Time series feature extraction library. *SoftwareX, 11*, 100456.

Cerrada, M., Sánchez, R. V., Li, C., Pacheco, F., Cabrera, D., de Oliveira, J. V., & Vásquez, R. E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing, 99*, 169–196.

Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2020). Fusing physics-based and deep learning models for prognostics. *Reliability Engineering and System Safety, 217*, 107961.

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh-a python package). *Neurocomputing (Amsterdam), 307*, 72–77.

Cubillo, A., Perinpanayagam, S., & Esperon-Miguez, M. (2016). A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. *Advances in Mechanical Engineering, 8*(8), 1687814016664660.

Cui, H., Qiao, Y., Yin, Y., & Hong, M. (2016). An investigation of rolling bearing early diagnosis based on high-frequency characteristics and self-adaptive wavelet de-noising. *Neurocomputing (Amsterdam), 216*, 649–656.

Gautier, G., Serra, R., & Mencik, J.-M. (2015). Roller bearing monitoring by new subspace-based damage indicator. *Shock and Vibration, 1–11*, 2015.

Gazizulin, D., Klein, R., & Bortman, J. (2018). Physics based methodology for the estimation of bearings' remaining useful life: Physics-based models, diagnostic methods and experiments. *Fourth European Conference of the PHM Society*.

Goebel, K., Saha, B., Saxena, A., Mct, N., & Riacs N. (2008) . A comparison of three data-driven techniques for prognostics. In: *62nd meeting of the society for machinery failure prevention technology (mfpt)*, (pp. 119–131).

Golafshan, R., Jacobs, G., & Berroth, J. K. (2018). Investigation of rolling bearing condition monitoring techniques based on long term run-to-failure vibration data. *Bearing World Journal, 3*, 107–118.

Graney, B., & Starry, K. (2012). Rolling element bearing analysis. *Materials Evaluation, 70*, 78.

Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483–1510.

Jieyang, P., Kimmig, A., Dongkun, W., Niu, Z., Zhi, F., Jiahai, W., Liu, X., & Ovtcharova, J. (2022). A systematic review of data-driven approaches to fault diagnosis and early warning. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-022-02020-0

Jimenez, J. J., Schwartz, S., Vingerhoeds, R., Grabot, B., & Salaun, M. (2020). Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems, 56*, 539–557.

Konstantin-Hansen, H. (2003). Envelope analysis for diagnostics of local faults in rolling element bearings. Technical report.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Lacey, S. J. (2008). An overview of bearing vibration analysis. *Maintenance & Asset Management, 23*(6), 32–42.

Lee, J., Qiu, H., Yu, G., & Lin, J. (2007). and Rexnord technical services. Bearing data set. http://ti.arc.nasa.gov/project/prognosticdata-repository

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. *Mechanical Systems and Signal Processing, 104*, 799–834.

Leite, V., da Silva, J., Torres, G., Veloso, G., da Silva, L., Bonaldi, E., & de Lacerda, L. (2017). Bearing fault detection in induction machine using squared envelope analysis of stator current. In *Bearing technology*, chapter 5. IntechOpen.

Liao, I., & Kottig, F. (2014). Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability, 63*(1), 191–207.

Li, N., Lei, Y., Lin, J., & Ding, S. X. (2015). An improved exponential model for predicting remaining useful life of rolling element bearings. *IEEE Transactions on Industrial Electronics (1982), 62*(12), 7762–7773.

Li, W., Qiu, M., Zhu, Z., Jiang, F., & Zhou, G. (2017). Fault diagnosis of rolling element bearings with a spectrum searching method. *Measurement Science and Technology, 28*(9), 95008.

Li, R., Sopon, P., & He, D. (2012). Fault features extraction for bearing prognostics. *Journal of Intelligent Manufacturing, 23*(2), 313–321.

Mais, J. (2002). *Spectrum analysis: The key features of analyzing spectra*. Technical report: SKF USA Inc.

Mathworks Help Center. Signal features. https://se.mathworks.com/help/predmaint/ug/signal-features.html.

McFadden, P. D., & Smith, J. D. (1984). Vibration monitoring of rolling element bearings by the high-frequency resonance technique—A review. *Tribology International, 17*(1), 3–10.

McInerny, S. A., & Dai, Y. (2003). Basic vibration signal processing for bearing fault detection. *IEEE Transactions on Education, 46*(1), 149–156.

McQueen, A. (2010). Stathis Ioannides, devoted to the fight against friction. *Evolution: The Business and Technology Magazine from SKF*, (pp. 18–20).

Mishra, S. K., Shakya, P., Babureddy, V., & Ajay Vignesh, S. (2021). An approach to improve high-frequency resonance technique for bearing fault diagnosis. *Measurement: Journal of the International Measurement Confederation, 178*, 109318.

Moubray, J. (1997). *Reliability-centred maintenance: [RCM II]* (2nd ed.). Butterworth Heinemann.

Niu, G. (2016). *Data-driven technology for engineering systems health management: Design approach, feature construction, fault diagnosis, prognosis*. Springer Singapore Pte.

PCB-Piezotronics. Model: 353B33 | Accelerometer, ICP. https://www.pcb.com/products?model=353b33.

Poddar, S., & Chandravanshi, M. L. (2013). Ball bearing fault detection using vibration parameters. *International Journal of Engineering Research and Technology, 2*, 1239.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration, 289*(4), 1066–1090.

Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International, 96*, 289–306.

Randall, R. B. (2007). *Noise and vibration data analysis*. Wiley.

Randall, R. B. (2011). *Vibration-based condition monitoring: Industrial, automotive and aerospace applications*. Wiley.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics-A tutorial. *Mechanical Systems and Signal Processing, 25*(2), 485–520.

Renwick, J. T., & Babson, P. E. (1985). Vibration analysis—A proven technique as a predictive maintenance tool. *IEEE Transactions on Industry Applications, IA–21*(2), 324–332.

Reuben, L. C. K., & Mba, D. (2014). Bearing time-to-failure estimation using spectral analysis features. *Structural Health Monitoring, 13*, 219–230.

Saruhan, H., Sandemir, S., Cicek, A., & Uygur, I. (2014). Vibration analysis of rolling element bearings defects. *Journal of Applied Research and Technology, 12*(3), 384–395.

Sawalhi, N., & Randall, R.B. (2004). The application of spectral kurtosis to bearing diagnostics. In: *Acoustics*, (pp. 393–398).

Scheffer, C., & Girdhar, P. (2004). *Practical machinery vibration analysis and predictive maintenance*. Elsevier.

Scikit learn API. Selectkbest class for extracting best features of given dataset. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html.

Singh, V., Gangsar, P., Porwal, R., & Atulkar, A. (2023). Artificial intelligence application in fault diagnostics of rotating industrial machines: A state-of-the-art review. *Journal of Intelligent Manufacturing, 34*(3), 931–960.

Swedish Standards Institute. Condition monitoring and diagnostics of machines—Data interpretation and diagnostics techniques—Part 1: General guidelines (iso 13379-1:2012, idt).

Tandon, N., & Choudhury, A. (1999). A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International, 32*(8), 469–480.

Taylor, J. I. (2003). *The vibration analysis handbook: A practical guide for solving rotating machinery problems*. VCI.

Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: A comparative review. *Journal of Machine Learning Research, 10*(66–71), 13.

Vargas-Machuca, J., Garcia, F., & Coronado, A. M. (2020). Detailed comparison of methods for classifying bearing failures using noisy measurements. *Journal of Failure Analysis and Prevention, 20*(3), 744–754.

Wang, T., Yu, J., Siegel, D., & Lee, J. (2008). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In *2008 international conference on prognostics and health management, IEEE*, (pp. 1–6).

Wei, Y., Li, Y., Xu, M., & Huang, W. (2019). A review of early fault diagnosis approaches and their applications in rotating machinery. *Entropy (Basel, Switzerland), 21*(4), 409.

Williams, T., Ribadeneira, X., Billington, S., & Kurfess, T. (2001). Rolling element bearing diagnostics in run-to-failure lifetime testing. *Mechanical Systems and Signal Processing, 15*(5), 979–993.

Yan, J., Koc, M., & Lee, J. (2004). A prognostic algorithm for machine performance assessment and its application. *Production Planning & Control, 15*(8), 796–801.

Zhou, Q., Yan, P., Liu, H., & Xin, Y. (2019). A hybrid fault diagnosis method for mechanical components based on ontology and signal analysis. *Journal of Intelligent Manufacturing, 30*(4), 1693–1715.

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