



A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: shallow and deep learning

Moussa Hamadache¹ · Joon Ha Jung¹ · Jungho Park¹ · Byeng D. Youn¹

Received: 22 August 2018 / Revised: 8 April 2019 / Accepted: 9 April 2019 / Published online: 16 May 2019
© The Korean Society of Mechanical Engineers 2019

Abstract

The objective of this paper is to present a comprehensive review of the contemporary techniques for fault detection, diagnosis, and prognosis of rolling element bearings (REBs). Data-driven approaches, as opposed to model-based approaches, are gaining in popularity due to the availability of low-cost sensors and big data. This paper first reviews the fundamentals of prognostics and health management (PHM) techniques for REBs. A brief description of the different bearing-failure modes is given, then, the paper presents a comprehensive representation of the different health features (indexes, criteria) used for REB fault diagnostics and prognostics. Thus, the paper provides an overall platform for researchers, system engineers, and experts to select and adopt the best fit for their applications. Second, the paper provides overviews of contemporary REB PHM techniques with a specific focus on modern artificial intelligence (AI) techniques (i.e., shallow learning algorithms). Finally, deep-learning approaches for fault detection, diagnosis, and prognosis for REB are comprehensively reviewed.

Keywords Deep learning · Diagnosis · Fault detection · Rolling element bearing · Shallow learning · Prognostics and health management

1 Introduction

Modern engineered systems are becoming increasingly complex and are operating under harsh and uncertain conditions; thus, systems are now more vulnerable to system breakdowns. On the other hand, in the past 10 years there has been a revolution in the Internet-of-Things (IoT), big data analytics, and artificial intelligence (AI). Through this revolution, ideas such as deep learning have risen from obscurity to provide a collection of innovative new techniques, thus achieving human-level performance in image recognition and gaming [1]. These technical trends will undoubtedly have a profound effect on an emerging discipline, Prognostics and Health Management (PHM), as it becomes a core technology for the 4th Industrial Revolution. An emerging discipline, PHM ensures cost-effective operation and management of

engineered systems by protecting the engineering assets from potential hazards and sudden breakdowns. PHM also increases the efficiency, reliability, and availability of the engineering assets. To this end, PHM is concerned with the presence of faults (fault detection), analysis of fault type and location (diagnosis), and forecast of future health condition and remaining useful lifetime (RUL) (prognosis) [2–10]. PHM also examines decision-making and feedback to provide improved condition-based maintenance (CBM) strategies. A generalized representation of the main PHM processes is shown in Fig. 1.

PHM, which aims to detect machine breakdown and prevent consequent accidents that bring economic losses, is a wide research domain. This paper focuses on reviewing and summarizing contemporary PHM techniques applied to rotating electrical machines (REMs). REMs are at the heart of most engineering processes (due to their relatively low price and operational ease [11], [12]) and REM failures are one of the foremost causes of breakdown in industry, causing high costs of operating maintenance. Furthermore, rolling element bearing (REB) faults account for 45–55% of REM failures [13, 14] and for about 41% of motor faults, followed by stator faults (37%) and rotor faults (10%) [15].

Moussa Hamadache and Joon Ha Jung are co-first authors. Byeng D. Youn is the corresponding author.

✉ Byeng D. Youn
bdyoun@snu.ac.kr

¹ Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul 08826, Republic of Korea

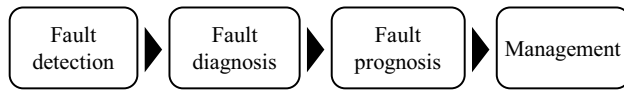


Fig. 1 Main processes of prognostics and health management (PHM)

Existing model (physical/mathematical)-based studies on REB PHM suffer from many difficulties. This is because the noisy and complex working conditions limit the development of the required model as shown in Fig. 2a, with the required precision. In addition, existing models, especially physics-based models, cannot be updated in real time (online) with newly measured data. Data-driven approaches, as opposed to model-based approaches, are gaining in popularity because they are model-free techniques. In addition, there have been significant advances in the development of sensors, sensor networks and computing systems. Existing data-driven techniques require extra information, more data acquisition equipment to implement, and additional measurements, such as vibration, temperature, acoustic emission, sound measurement, oil debris, laser displacement, stator current monitoring [16], or not such rotor speed signal monitoring [17]. The acquired signals contain the fault information and characteristics; signals must be preprocessed first, and then different features are extracted to better understand the REB health status. It is crucial to recognize that those signals often have a low signal-to-noise ratio and non-stationary statistical parameters due to the actual harsh operating conditions in industry (e.g., high mechanical load, time-varying speed, mechanical shocks). These factors make standard data-driven REB PHM methods difficult [4] and limit their effectiveness, performance, and flexibility. Therefore, PHM efforts for REBs have focused

on extending and/or improving the existing standard data-driven REB PHM methods or completely developing other approaches, referred to as smart data-driven approaches. For example, shallow learning-based PHM (SL-based PHM) and deep learning-based PHM (DL-based PHM) techniques, as seen in Fig. 2b, c, have emerged as alternatives to model/physics-based approaches [18–20]. These data-driven approaches have become more and more attractive due to the widespread deployment of low-cost sensors and their connection to internet, introducing the phenomena of big data. Thus, the aim of this paper is to review and summarize the most recent intelligent PHM techniques applied to REB fault detection, diagnosis, and prognosis, providing a reference for further studies on the related topics. Therefore, this paper first discusses and classifies shallow learning algorithms, then it reviews the most advanced techniques, deep learning-based rolling element-bearing fault detection, diagnosis, and prognosis.

Taking this into consideration, this paper will present a brief description of the different bearing failure modes, and a comprehensive description of the different health features (indexes, criteria) used for REB fault diagnostics and prognostics, with the goal of providing an overall platform for researchers, system engineers, and experts to select and adopt the best fit for their applications. This paper is organized as follows: Sect. 2 briefly introduces the different bearing failure modes and their causes, followed by a comprehensive representation of the different health features (indexes and criteria). The different existing shallow-learning algorithms for REB PHM are detailed in Sect. 3. Section 4 provides the most recent investigations and studies that are based on the hottest subfield, deep learning-based REB fault detection, diagnosis, and prognosis. Finally, a summary and concluding remarks are given in Sect. 5.

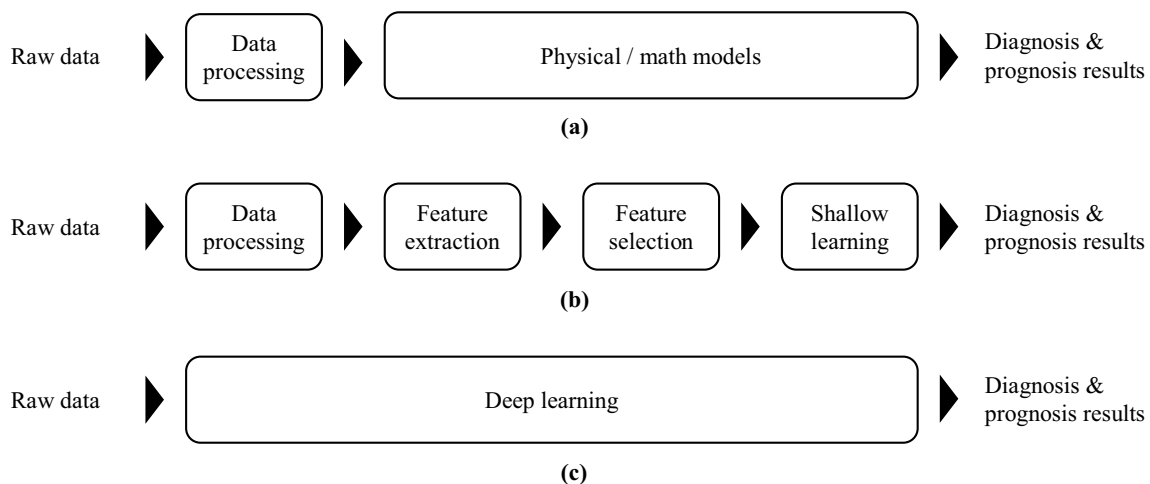


Fig. 2 Comparison between: **a** physics/math model-based PHM technique, **b** shallow learning-based PHM technique, and **c** deep learning PHM technique

A prior survey paper [21] gives a review of the emerging research work related to deep learning and new trends related to its use in machine health monitoring for different applications and systems. In addition, the review paper of Zurita et al. [22] mainly reviewed the state-of-the-art vibration condition-based monitoring of gears and bearings that are based on advanced digital signal processing techniques and artificial intelligence methods. In contrast to these prior works, this paper focuses only on reviewing contemporary learning algorithms (i.e., the shallow learning algorithms and the deep learning algorithm and its variants) for REB fault detection, diagnosis, and prognosis techniques. Contemporary PHM techniques are summarized as follows.

Modern engineering systems are embracing more and more user-friendly data acquisition tools and low-cost sensors that are connected to the internet. Therefore, PHM researchers and practitioners are adopting contemporary techniques, i.e., smart data-driven approaches—SL-based PHM and DL-based PHM techniques—that have been developed in the last decade. These techniques aim to synthesize information available from the acquired data to better represent the system's health condition. Further, the latter (i.e., DL-based PHM) extracts the best-suited features from big data and better represents the system health condition in a hierarchical architecture. With the propagation of acquired data, DL-based PHM techniques model the high-level representation of the complex multivariate nonlinear relationship behind the data without need for a profound understanding of the system physics; this eliminates the need for a significant amount of human labor. In contrast, SL-based PHM methods require a manual feature extraction step, which may require domain knowledge. Thus, these methods can face problems in extracting useful representations from big data.

Availing from the shallow structures (e.g., artificial neural network (artificial NN), support vector machine (SVM), etc.), SL-based PHM techniques were constructed. They consist mainly of four phases, as shown in Fig. 2b: data processing, hand designed features extraction, feature selection (e.g., using principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA), etc.), and model training. This process enables SL-based PHM techniques to achieve decent performance in dealing with fault detection & diagnosis in specific [22]. However, the huge amount of data in the PHM field makes it difficult for these SL-based PHM approaches to know and determine the best-suited features to be designed; further, these methods have other challenges [23]. This limits their performance, since the features must be manually designed and the four phases cannot be optimized simultaneously. Many studies have been performed based on the shallow learning technique to detect, diagnose, and predict the various REB faults, which used to describe conventional

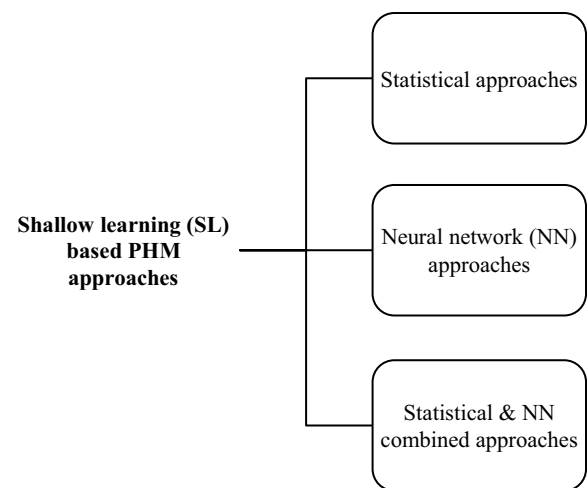


Fig. 3 Classification of shallow learning-based PHM techniques

machine learning methods, as compared with deep learning methods. In this paper, SL-based PHM techniques are classified into statistical methods, neural network methods, and combined approaches, as shown in Fig. 3. Our aim is to provide a state-of-the-art review of SL-based PHM techniques and their application to REB PHM, as given in Sect. 3.

In light of the challenges of SL-based PHM techniques outlined above, DL-based PHM methods are considered to outperform the previous methods. DL-based PHM techniques seek to handle big data by modeling the high-level illustration behind the data. Thus, they automatically extract the best features, which are highly nonlinear and complex, via stacking multiple layers in hierarchical architecture, instead of handcrafting the optimum features using domain knowledge [21, 23]. Thus, they have four main advantages, as compared to others methods: (1) they achieve an end-to-end system, as shown in Fig. 2c, (2) they do not require intensive human labor and knowledge that would otherwise be necessary to handcraft the feature design, (3) they construct models with high hierarchical architectures and a nonlinear combination of multiple layers, (4) compared to SL-based PHM techniques, the DL-based PHM algorithm's parameters are optimized simultaneously. Our classification of the different existing DL-based PHM methods for REB PHM applications follows Zhao et al.'s [21] classification; methods are classified into (deep) convolutional neural network (CNN) approaches, (deep) recurrent neural network (RNN) approaches, restricted Boltzmann machine (RBM)-based deep neural network (DNN) approaches, and autoencoder (AE)-based DNN approaches, as shown in Fig. 4, with the aim of providing a comprehensive review of DL-based PHM methods and their applications to REB PHM, as given in Sect. 4.

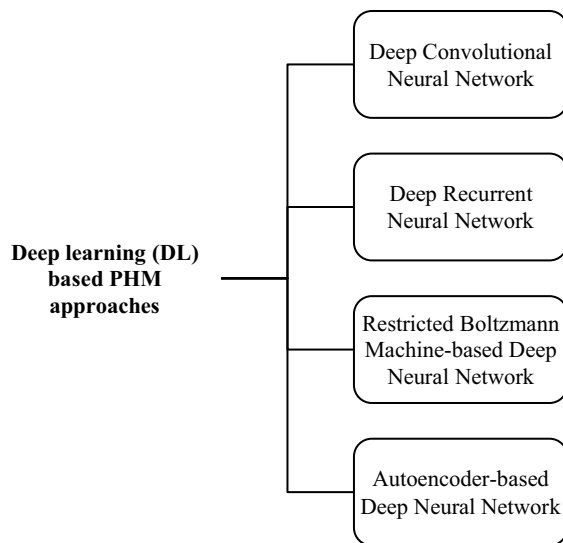


Fig. 4 Classification of deep learning-based PHM techniques

2 Fundamentals of rolling element bearing (REB) prognostics and health management (PHM)

In industry, the health of many machines depends on the robustness and reliability of the REBs. Failures may appear in REBs during operation or before (i.e., during the manufacturing process). From a prior FMECA (failure modes, effects, and criticality analysis) study of servo motors, which are the core component for mechanism control of electrical machinery, bearing faults were shown to have the highest frequency, severity, and criticality [24]. Therefore, detection, diagnosis, and prognosis of these defects are important for prognostics and health management, as well as for quality inspection of bearings [25].

2.1 Bearing failure modes

It does not require severe REB failure to induce vibration, noise, or even sudden breakdown of equipment; tiny faults such as cracks, crushes, wears, indentation, etc. will also cause breakdowns. These different faults can be caused by a wide range of factors. Flaking, pitting, spalling, rusting, corroding, creeping, and skewing can all lead to failure [26]. From ISO 15243 [27], the most common faults are fatigue, wear, corrosion, electrical erosion, plastic deformation, and fracture & cracking. Each is briefly introduced below.

- *Fatigue* begins as a tiny crack on the bearing surface (rollers or races) due to a material structure change, which is caused by repeated stress in the contact areas.

- *Wear* comes from the presence of dirt or foreign particles inside the bearing due to inaccurate sealing or inadequate lubrication (contamination).
- *Electric erosion* is damage (in the form of craters) in one of the bearing parts (rollers or races) due to a passing through the bearing of an electric current.
- *Corrosion* comes from the presence of water or corrosive agents inside the bearing due to damaged seals, acidic lubricants, or a sudden high change of operating temperature.
- *Plastic deformation* generates mainly when the bearing is subject to an excessive load that results in an indentation of the raceways.
- *Fracture and cracking* results from the stress that comes from rough treatment (impacts) or from cyclic stress. Additionally, fracture and cracking can be caused by high heating (thermal).

2.2 REB health features

Rolling element bearing PHM techniques often use different sensors to collect several raw physical signals (vibration, stator current, temperature, rotor speed, etc.); the result is the so-called big data phenomena. Dozens of indices or criterion (i.e., features) are usually extracted from those raw signals in the time, frequency, and time–frequency domains, to detect, diagnose, and predict the health condition of the REB system. This section attempts to provide a complete list of those features in Table 1 with the aim of providing a comprehensive platform for researchers, system engineers, and experts to identify and adopt those that best fit their needs.

In addition, many techniques have been developed and applied. In the frequency domain [45–48], the power spectrum analysis, the fast Fourier transform (FFT), the discrete Fourier transform (DFT), the Welch method, and the noise cancellation techniques can be found. In the time–frequency domain [49–51], well-known techniques are the short-time Fourier transform, the Wigner–Ville distribution, the continuous wavelet transform (CWT), the discrete wavelet transform (DWT), and the Wavelet packet transform (WPT).

3 Shallow learning algorithms for REB PHM

This section presents a state-of-the-art review of SL-based PHM methods and their application to REB PHM. In attempt to organize and classify the diverse SL-based REB PHM techniques, which may originate from the artificial neural network (NN) or may not, three categories are proposed: statistical approaches, NN approaches, and combined methods. Further, the statistical approaches are sub-divided, according to the nature and the task of each algorithm, into LDA-based REB PHM, SVM-based REB PHM, K-nearest

Table 1 Various features used in REB PHM techniques

No.	Features	Definition	Physical meaning
<i>Time domain features</i>			
1	Maximum [29]	$I_{\max} = \max_{k=1 \dots N} (x(k))$	Kinetic energy related
2	Minimum [29]	$I_{\min} = \min_{k=1 \dots N} (x(k))$	Kinetic energy related
3	Absolute maximum [30]	$I_{\max} = \max_{k=1 \dots N} (x(k))$	Kinetic energy related
4	Sum [36]	$I_{\text{sum}} = \sum_{k=1}^N x(k)$	Kinetic energy related
5	Median [36]	$I_{\text{med}} = \text{median}(x(k))$	Kinetic energy related
6	Most frequent value [36]	$I_{\text{mod}} = \text{mode}(x(k))$	Kinetic energy related
7	Mean [28]	$I_{\text{mean}} = \bar{x} = 1/N \sum_{k=1}^N x(k)$	Kinetic energy related
8	Absolute mean [30]	$I_{\text{amean}} = \bar{x} = 1/N \sum_{k=1}^N x(k) $	Kinetic energy related
9	Mean absolute deviation [36]	$I_{\text{mad}} = \text{mad}(x(k))$	Kinetic energy related
10	Harmonic mean [36]	$I_{\text{har}} = N / \sum_{k=1}^N \frac{1}{x(k)}$	Gives the truest average energy
11	Trapezoidal numerical integration [36]	$I_{\text{trap}} = \text{trapz}(x(k))$	None
12	Percentiles [36]	$I_{\text{prc}} = \text{prctile}(x(k))$	None
13	Interquartile rang (IQR) [36]	$I_{\text{IQR}} = \text{iqr}(x(k))$	None
14	Energy quantification related [29]	$I_{\sigma^2} = 1/N \sum_{k=1}^N (x(k) - I_{\text{mean}})^2$	Energy quantify-
<i>Cation related</i>			
15	Root mean square (RMS) [28, 43]	$I_{\text{rms}} = \sqrt{1/N \sum_{k=1}^N (x(k))^2}$	Kinetic energy related
16	RMS error (RMSe) [32]	$I_{\text{rmse}} = \sqrt{1/N \sum_{k=1}^N (x(k) - I_{\text{mean}})^2}$	Kinetic energy related
17	Delta RMS [36]	$I_{\text{drms}} = I_{\text{rms}}^j - I_{\text{rms}}^{j-1}$ where j is the current segment of time record and $j-1$ in the previous segment	Kinetic energy related
18	Energy quantification related [29]	$I_{\sigma} = \sqrt{I_{\sigma^2}}$	Energy quantify-
<i>Cation related</i>			
19	Peak value [28]	$I_{\text{pv}} = 1/2[I_{\max} - I_{\min}]$	Kinetic energy related
20	Peak to peak [28]	$I_{\text{p to p}} = [I_{\max} - I_{\min}]$	Kinetic energy related
21	Peat to RMS [29]	$I_{\text{p to rms}} = I_{\max} / I_{\text{rms}}$	Kinetic energy related
22	Skewness [28]	$I_{\text{sk}} = \frac{1/N \sum_{k=1}^N (x(k) - I_{\text{mean}})^3}{(I_{\sigma})^3}$	Data statistic related
23	Kurtosis [40]	$I_{\text{kur}} = \frac{1/N \sum_{k=1}^N (x(k) - I_{\text{mean}})^4}{(I_{\sigma})^4}$	Data statistic related
24	Crest factor [28]	$I_{\text{cf}} = I_{\text{pv}} / I_{\text{rms}}$	Sinusoidal wave shape related
25	Clearance factor [28]	$I_{\text{clf}} = I_{\text{pv}} / (I_{\text{mean}})^2$	None
26	Impulse factor [28]	$I_{\text{if}} = I_{\text{pv}} / I_{\text{amean}}$	Sinusoidal wave shape related
27	Shape factor [28]	$I_{\text{sf}} = I_{\text{rms}} / I_{\text{amean}}$	Sinusoidal wave shape related
28	Margin factor [34]	$I_{\text{mf}} = I_{\max} / I_{\sigma^2}$	None
29	Coefficient of variance [30]	$I_{\text{cv}} = I_{\text{mean}} / I_{\sigma}$	None
30	Coefficient of skewness [30]	$I_{\text{csk}} = \frac{1/N \sum_{k=1}^N (x(k))^3}{(I_{\sigma})^3}$	None
31	Coefficient of kurtosis [30]	$I_{\text{ckur}} = \frac{1/N \sum_{k=1}^N (x(k))^4}{(I_{\sigma})^4}$	None
32	TALAF [36]	$I_{\text{TALAF}} = \log \left(I_{\text{kur}} + \frac{I_{\text{rms}}}{I_{\text{rms}_{\text{th}}}} \right)$	None

Table 1 (continued)

No.	Features	Definition	Physical meaning
33	THIKAT [36]	$I_{\text{THIKAT}} = \log \left((I_{\text{kur}})^{I_{\text{cf}}} + \left(\frac{I_{\text{rms}}}{I_{\text{rms}_h}} \right)^{I_{\text{fv}}} \right)$	None
34	Normalized sixth central moment [36]	$I_{\text{kur6}} = \frac{1}{N} \sum_{k=1}^N \frac{(x(k) - I_{\text{mean}})^6}{(I_{\sigma})^6}$	None
35	Add factor 1 [30]	$I_1 = \frac{I_{\text{max}}}{I_{\text{sd}} \cdot I_{\sigma^2}}$	None
36	Add factor 2 [30]	$I_2 = \frac{I_{\text{kur}} \cdot I_{\text{cf}}}{I_{\text{sd}}}$	None
37	Fisher criterion [31]	$I_{\text{fisherc}} = \frac{(I_{\text{mean}} - I_{\text{mean}_h})^2}{I_{\text{sd}}^2 + I_{\text{sd}_h}^2}$	None
38	Square root of amplitude [34]	$I_{\text{stra}} = \left(\frac{1}{N} \sum_{k=1}^N \sqrt{ x(k) } \right)^2$	None
39	Euclidian distance [35]	$I_{\text{ed}} = \sqrt{\sum_{k=1}^N (I_h(k) - I_f(k))^2}$	None
40	Sum square error distance [37]	$I_{\text{sse}} = \ I_h - I_f\ ^2$	None
41	Mahalanobis distance [33]	$I_{\text{mahd}} = \sqrt{(I_h - I_f) C^{-1} (I_h - I_f)}$ $I_{\text{mahd}} = \sqrt{(I_f - I_{f_{\text{mean}}}) C^{-1} (I_f - I_{f_{\text{mean}}})}$	None
42	Manhattan distance [39]	$I_{\text{manhd}} = \sum_{k=1}^N I_h(k) - I_f(k) $	None
43	Median error distance [39]	$I_{\text{meded}} = \arg \min \sum_{k=1}^N \ I_h(k) - I_f(k)\ _2$	None
<i>Frequency domain features</i>			
44	Shaft rotational frequency [37]	$I_{\text{srf}} = f_r = N_{\text{rpm}} / 60$	Position change of main frequency
45	Outer-race fault (ORF) frequency [39]	$I_{\text{orf}} = \frac{N_b}{2} f_r (1 - (d_b \cos \beta / d_p))$	Occurrence of fault frequency
46	Inner-race fault (IRF) frequency [39]	$I_{\text{irf}} = \frac{N_b}{2} f_r (1 + (d_b \cos \beta / d_p))$	Occurrence of fault frequency
47	Roller (ball) fault (BBF) frequency [39]	$I_{\text{bbf}} = \frac{d_b}{d_p} f_r (1 - (d_b \cos \beta / d_p)^2)$	Occurrence of fault frequency
48	Cage fault frequency [37]	$I_{\text{cff}} = \frac{N_b}{2} f_r (1 - \frac{d_b \cos \beta}{d_p})$	Occurrence of fault frequency
49	Mean frequency [41]	$I_{\text{meanf}} = \frac{1}{N} \sum_{k=1}^N f_k$	Main frequency position changes
50	Variance [29]	$I_{\sigma^2 f} = \frac{1}{N} \sum_{k=1}^N (f_k - I_{\text{meanf}})^2$	Frequency quantification related
51	RMS frequency [42, 44]	$I_{\text{rmsf}} = \sqrt{\frac{1}{N} \sum_{k=1}^N f_k^2}$	Kinetic frequency related
52	Frequency center [42, 44]	$I_{\text{fcenter}} = \frac{1}{N} \sum_{k=1}^N f_k$	Main frequency position changes
53	Root variance frequency [42, 44]	$I_{\text{root}\sigma^2 f} = \sqrt{I_{\sigma^2 f}}$	Convergence of spectrum power
<i>Envelope spectrum features</i>			
54	RMS frequency of the 1 st harmonic [34]	$I_{\text{rmsf 1h}} = \sqrt{\frac{1}{b_1 - a_1} \sum_{k=a_1}^{b_1} f_k^2}$	Certain frequency range magnitude
55	RMS frequency of the 2nd harmonic [34]	$I_{\text{rmsf 2h}} = \sqrt{\frac{1}{b_2 - a_2} \sum_{k=a_2}^{b_2} f_k^2}$	Certain frequency range magnitude
56	RMS frequency of the 3 rd harmonic [34]	$I_{\text{rmsf 3h}} = \sqrt{\frac{1}{b_3 - a_3} \sum_{k=a_3}^{b_3} f_k^2}$	Certain frequency range magnitude
<i>Statistical features</i>			
57	T^2 statistics [37]	$I_{T^2} = t^T \text{Cov}^{-1} t$	None
58	Q statistics [37]	$I_Q = e^T \epsilon$	None
59	Residual error matrix [32]	$I_{\text{rem}} = \hat{Y} - Y$	None
60	Reconstruction error [38]	$I_{\text{rec error}} = \ x_{\text{new}} - (WW^T)x_{\text{new}}\ ^2$	None

Table 1 (continued)

No.	Features	Definition	Physical meaning
61	Bhattacharyya distance [39]	$I_{\text{Bhattacharyya}} = \frac{1}{8}(\mu_1 - \mu_2)^T \left[\frac{C_1 + C_2}{2} \right]^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \frac{ C_1 + C_2/2 }{ C_1 ^{\frac{1}{2}} C_2 ^{\frac{1}{2}}}$	None

For time domain signal, $x(k)$ is a signal series for $k = 1, 2, \dots, N$, where N in the number of samples

I_h and I_f are the features of the same nature, where the suffix h means that the feature was computed from the healthy state and f means that the feature was computed from a faulty state

C covariance matrix, N_{rpm} rotational speed in rpm (rotation per minute), d_b ball diameter, d_p ptch ball diameter, Nb number of balls, β ball contact angle, a and b two non-negative integers, where $b > a$, t score vector, ε residual vector in residual space, \hat{Y} is the estimated output matrix of the reference model, W weight matrix, μ_i mean vector of class i

neighbor (KNN)-based REB PHM, extreme learning machines (ELM)-based REB PHM, and other non-NN algorithms applicable to REB PHM. The combined methods are the ones that utilize a non-NN algorithm with a NN method, or a NN algorithm with a signal processing technique, or a non-NN algorithm with a signal processing approach.

3.1 Statistical approaches for REB PHM

Several shallow learning algorithms exist that were constructed using a shallow architecture that benefits from the statistical properties of the data and uses this information to classify it to already known group [2]. The following section provides a detailed description of those statistical SL-based REB PHM techniques as applied to REB PHM. The structure typical of each algorithm is briefly introduced, and its application to REB PHM is outlined to highlight its challenges, its pros and cons, and its latest advancements.

3.1.1 LDA-based REB PHM

The LDA algorithm aims to find a linear combination of features that separates different classes well. It helps in the classification process by finding new projection directions in which, when data are projected in those directions, the within-mode distance decreases while the between mode distance increases [52]. Thus, it reduces the dimensionality by maximizing the ratio of between-class scatter and within-class scatter. Therefore, the main objectives of LDA are either to reduce the dimensionality or to perform classification. Figure 5 shows a descriptive example of LDA-based classification. As can be seen, if two classes are presented (red and blue), the LDA will map them into a new feature space following the projection lines ((a) and (b)) as

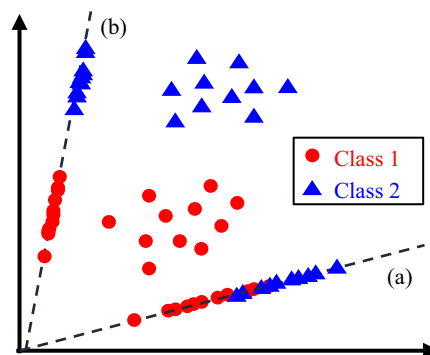


Fig. 5 Example of linear discriminant analysis (LDA)

Table 2 Between- and within-class scatters of Fig. 5

	Projection line	
	(a)	(b)
Between-class scatter	Small	Large
Within-class scatter	Large	Small

can be seen in Table 2 in which they will be more linearly discriminant.

The LDA algorithm has been used to improve classification of ball bearing faults according to their severity level [53]. LDA was also used as a dimensionality reduction technique to find the dimensions of a few features that best discriminate a set of features extracted from raw vibration signals [54]. Zhao et al. [55] proposed a trace ratio version of LDA, which uses the between-class scatter matrix to evaluate the separability of different classes and the within-class scatter matrix to evaluate the compactness within each class. The extended discriminative subspace learning method was

used for dealing with the trace ratio problem in linear discriminant analysis for a REB fault detection and diagnosis problem. A trace ratio LDA algorithm was also introduced by Jin et al. [56] and used to reduce the dimension and then to classify the motor bearing health conditions, which arose from single-point faults and generalized-roughness faults. Another form of LDA, called Δ -LDA, was proposed by Ciabattini et al. [57] to deal with fault data dimension reduction and fault detection issues with application to REB fault detection. Δ -LDA was proposed to overcome the problem of a between-class scatter matrix trace very close to zero, which is the case when detecting different bearing faults. It did indeed improve the classification accuracy when the classes were overlapped. Evaluating the current feature generated by frequency selection in the stator current spectrum by means of LDA algorithm, a fault diagnosis of bearing damage was proposed in [58], in which the fault diagnosis was performed by the Bayes classifier.

3.1.2 SVM-based REB PHM

The classifier in machine learning and statistics learns from the data input given to it and then uses this learning to classify a new observation. The same can be done to detect and then diagnose various bearing faults (outer-race fault (ORF), inner-race fault (IRF), and ball bearing fault (BBF) or cage faults). One of the most used classifier-based PHM techniques is the SVM.

The basic idea of SVM is to first map the input data nonlinearly into a feature space; in this feature space, a linear decision function is constructed. Then, the inner product of the feature space is nonlinearly mapped to the original space [59]. Thus, the main purposes of SVM are classification and estimation. In machine learning, SVM is considered a supervised learning model that is used for classification and regression analysis. For classification, SVM finds the optimal separating hyperplane with maximum margin to build a maximum margin classifier. The margin is defined as the perpendicular distance between the support vectors, as can be seen in Fig. 6, where only those support vectors are used to determine the boundaries.

Numerous researches have modified SVM algorithms for various reasons. Sugumaran et al. [60] used the SVM and proximal-SVM (PSVM) classifiers to find this optimal number of time domain statistical and histogram features of a vibration signal. A hybrid, two-stage one-against-all SVM approach was proposed for REB fault diagnosis in [61] to predict the type of faults more accurately. In the first SVM stage, the vibration signal can be classified into either normal or fault. Then, the fault types are classified in the second SVM stage. In addition, one-class ν -SVM, which use only the normal state data, was used in an automatic bearing fault diagnosis [62]. To fully exploit the advantage of SVM, two

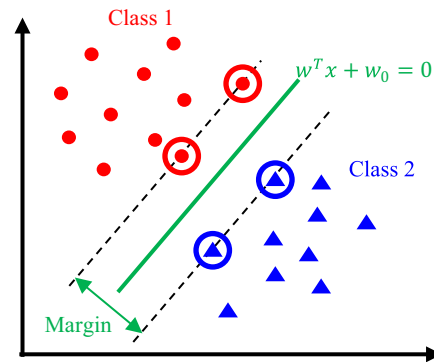


Fig. 6 Example of support vector machine (SVM)

multi-layer kernel learning models, supervised incremental local tangent space alignment (SILTSA)-SVM and supervised linear local tangent space alignment (SLLTSA)-SVM were proposed in [63] and applied to REB fault diagnosis. The proposed method combines the supervised method with the dimension reduction algorithms (ILTSA and LLTSA) [64]. In addition, to optimize the SVM parameters, which have significant impact on classification performance, an improved ant colony optimization (IACO) algorithm was proposed to determine the parameters, and then the IACO-SVM algorithm was applied to rolling element bearing fault detection [65]. More recent studies were performed to further investigate the use of SVM for REB bearing fault detection and diagnosis, including [66–70].

3.1.3 K-nearest neighbor (KNN)-based REB PHM

KNN is a non-parametric (i.e., the model structure is determined mainly from the data without any assumptions on the underlying data distribution), lazy algorithm (i.e., as opposed to an eager algorithm, it does not learn discriminative functions but uses all the training data in the classification step) used for classification, in which the existing (historical) data are grouped into several classes to be used to classify the new data. Thus, the main advantages of KNN are that the learning is very simple and easy to interpret (i.e., it has a physical meaning), and it is an effective classification method for noisy training data and complex target function, which makes it a well-suited algorithm for REB PHM [71]. However, there are also some disadvantages of KNN-based REB PHM. Specifically, since it is a lazy algorithm it needs to store the entire training dataset and thus needs to compare distance values for whole training samples; this is time- and power-consuming.

A descriptive KNN example is shown in Fig. 7. The test data (green multiplication sign) are classified to class 1 (blue circles) if $k=3$, but if $k=6$ the test data are classified to class 2 (red circles). Therefore, determining the parameter k is

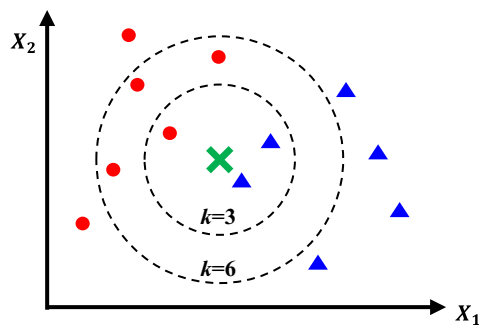


Fig. 7 Example of K -nearest neighbor

critical for accuracy of the KNN-based REB fault detection and diagnosis. However, the best choice of k depends on the data. If k is large, the effect of noise is reduced, but the boundaries between classes are less distinct; whereas, if k is small, strict boundaries can be obtained, but analysis may be vulnerable to noise and outliers (i.e., the overfitting problem may occur); this is the case when dealing with bearing fault detection and diagnosis [71].

KNN was used first for fault detection and diagnosis of low speed (≤ 100 rpm) REBs in the year 1992 [72]. A combination of weighted KNN (WKNN) classifiers was proposed by Y. Lei et al., [73] to overcome the two previously mentioned disadvantages of KNN-based REB fault detection and diagnosis. The KNN was also combined with other classification methods to enhance the REB fault detection and diagnosis capability, such as with SVM [74], kernel PCA (KPCA) [75], the fuzzy C-means method [76], the binary differential evolution algorithm [77], or the K-star classifier [78]. More recently, an optimal KNN model was combined with KPCA to deal with bearing fault detection and diagnosis, in which the KNN was optimized using a particle swarm optimization method [79].

3.1.4 Extreme learning machine (ELM)-based REB PHM

ELM was proposed in 2006 by G. Huang et al. [80] to provide good generalization performance at an extremely fast learning speed. ELM offered improvement over the learning speed of feedforward neural networks (FNNs), which are very slow, especially in real-time applications [80]. The slow learning speed of FNNs arises for two main reasons: the FNNs extensively use slow gradient-based learning algorithms for training, and using such learning algorithms, all network parameters are tuned iteratively [80]. ELM uses a single hidden layer feedforward neural network (SLFNN), as shown in Fig. 8, that randomly chooses hidden nodes and analytically determines the output weights w_i of the SLFNN. Thus, as can be seen the ELM can be represented as a linear system that employing an activation function, $F(\cdot)$, to generate the learned output y .

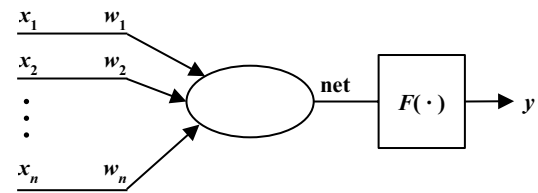


Fig. 8 Basic structure of extreme learning machine (ELM)

To the authors' knowledge, ELM was first applied alone to REB fault diagnosis system by Razavi-Far and Saif [81] to deal with the abilities of incremental learning in non-stationary environments and to detect and diagnose bearing faults under the class imbalance condition. The proposed ELM methods adopted: two state-of-the-art ensemble-based techniques, Learn ++.CDS (Concept Drift with SMOTE) [82], which was used to overcome the class imbalance issue in non-stationary environments, and the Learn ++.NIE (nonstationary and imbalanced environment) [83] to handle class-imbalanced data during the incremental phase in non-stationary environments. A more recent study that used ELM for REB condition monitoring was carried out by W. Mao et al., [84] in which they tried to solve the online imbalanced data problem that occurs when collecting data online in a sequential way and the number of fault data is much less than the number of the normal data.

3.1.5 Other statistical algorithms for REB PHM

Sugumaran et al. [85] investigated the effectiveness of an automatic rule learning-based decision tree for classification when employing a fuzzy classifier. The decision tree was used to select the different extracted statistical features from the vibration signals, and then multiple membership functions based on the generated 'if-then' rules were designed. Finally, a fuzzy inference engine was built and used to classify the REB health conditions based on predefined threshold. Then, they [86] proposed a decision tree based method for the use of the histogram features to improve the previous results in the case of small data points in the data set.

Other different non-NN methods were investigated to detect and diagnose an REB's health state. Yu [87] proposed a supervised-learning-based local and nonlocal preserving projection (SLNPP) method; Kankar et al. [88] used learning vector quantization (LVQ) as a REB fault classifier. In Cao et al. [89], a novel fault diagnosis method based on semi-supervised fuzzy C-means (SFCM) cluster analysis was developed; and more recently, targeting the nonstationary and non-Gaussian characteristics of a vibration signal from a faulty rolling bearing, Han et al. [90] developed a VMD-AR (variational mode decomposition-autoregressive) model and investigated diagnosing REB faults using the random

forest learning (RFL) classifier. The VMD was applied to decompose vibration signals where a series of stationary component signals were obtained, then, an AR model was established for each component mode. The models were used as fault characteristic vectors. Finally, a novel RFL classifier was considered for pattern recognition to diagnose different bearing faults.

Mohsenzadeh et al. [91] introduced a novel sparse Bayesian learning (SBL) algorithm called the relevance sample feature machine (RSFM), which had the capability of choosing the relevant samples and the relevant features simultaneously for regression or classification problems. Further, it was concluded that the RSFM had the advantage of avoiding overfitting, resulting in less system complexity during the testing stage, and better generalization. Wong et al. [92], successfully adopted a novel structure that is based on a pairwise-coupled sparse Bayesian extreme learning committee machine to intelligently and simultaneously diagnose bearing faults.

A bearing fault diagnosis technique was also presented by Shen et al. [93] based on a transfer learning (TL) technique, which was not limited to the same field [94]; it used singular value decomposition (SVD) [95] as its feature extraction tool. The authors describe the main idea of the proposed TL method as [93] “to utilize selective auxiliary data to assist target data classification, where a weight adjustment between them is involved in the TrAdaBoost algorithm for enhanced diagnostic capability. In addition, negative transfer is avoided through the similarity judgment, thus improving accuracy and relaxing computational load of the presented approach.”

Manifold learning (ML) [96] techniques are widely used in cluster analysis, image processing, bio-informatics, etc., [97–99]. However, ML techniques are rarely used for fault diagnosis, and were only used as a nonlinear time series noise reduction method applied to the analysis of gearbox vibration signals with snaggletooth in [100]. Recently, Wang et al. [101] proposed a novel machinery REB fault diagnosis approach based on a statistical locally linear embedding (S-LLE) manifold learning algorithm, which was an extension of LLE [102]. Another study, which applied the ML technique in combination with wavelet packet transform to detect weak transient signals for REB fault diagnosis, was carried out by Wang et al. [103]. This study proposed an extraction method, named waveform feature manifold (WFM), that used the binary wavelet packet transform to obtain the waveform feature space, which was then used to extract the weak signatures.

It should be noted that there are a few remaining learning techniques, such as the Bayesian learning (BL) [104] technique and the Widrow-Hoff learning (WHL) [105] algorithm. The authors did not find any study that applied these techniques to the bearing prognostics and health

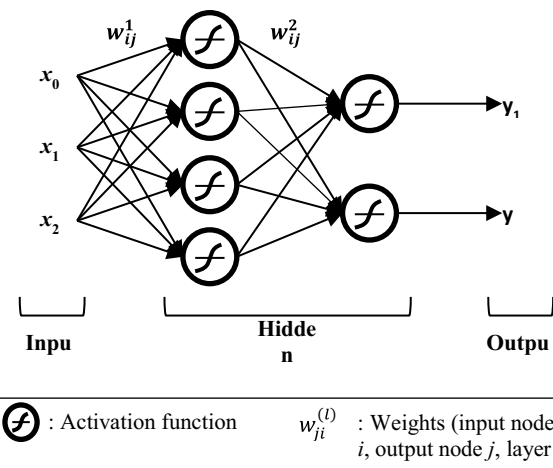


Fig. 9 Example of 2-layer artificial neural network (ANN)

management field, although researchers may consider these techniques in the future.

3.2 Neural network approaches for REB PHM

Other SL-based REB PHM techniques that were constructed using a shallow structure originating from the artificial NN are grouped and reviewed in this subsection. It is worth noting that the deep learning methods originated as an extension of these NN-based techniques.

Artificial NNs are a statistical model inspired by the biological neural networks that constitute the human brain. The NNs typically consist of an input layer, a hidden layer, and an output layer, as shown in Fig. 9. The nodes, x_i , in the input layer represent the normalized features extracted from the acquired signals. The output layer nodes, y_j , are the two nodes that can have only binary levels when dealing only with bearing fault detection (i.e., they represent healthy and faulty bearings). For bearing fault diagnosis, more nodes must be added to the output layer to localize and identify the different bearing faults. The hidden layers are generated from the input layer based on imposing weights $w_{ji}^{(l)}$. Using forward/backward propagation, proper weights, which minimize the cost value, are calculated from the labeled data. The output nodes are activated and their values are defined using activation functions. To train neural networks by gradient descent [106], the activation function should be differentiable; the nonlinear activation function adds nonlinear properties to the neural network. Different linear or nonlinear activation functions exist, such as the sigmoid function, the tanh function, the rectified linear unit (ReLU) function, the exponential linear unit (ELU) function, etc. [107].

As stated above, the weights to be used in the network are calculated using forward/backward propagation. For

the gradient descent method, computing the error gradient with respect to each weight is needed to quantify the influence of each weight on the final error. Backpropagation is an efficient way to compute gradients of the cost function; it is commonly used to train the network [108]. The backpropagation procedure can be defined as follows: first, initialize the weights randomly, apply the forward propagation (through the neural network, to obtain output & cost), then apply the backward propagation (calculate the influence of each weight on cost; error gradient), and finally, update the weights by repeating those steps until the performance of the network is satisfactory.

Neural network (NN) techniques have been applied to the PHM field for different engineered systems [109–111]. One of the earliest works that used NN for motor REB fault diagnosis was performed by Li et al. [112]. Frequency domain features extracted from the vibration signal were first performed (i.e., using FFT), then a NN was trained to emulate the knowledge of the vibration experts, which are very expensive. Thus, motor REB fault diagnosis was achieved more efficiently and at a reduced cost. Another study [113] used time domain features (I_{rms} , I_{σ^2} , I_{sk} , and I_{kur6}) for artificial NN-based bearing fault diagnosis instead of frequency domain features. Pandya et al. [114] used time–frequency domain features for NN-based REB fault diagnosis. They used the wavelet packet decomposition for feature extraction from the measured vibration signal. A comparison study [115] of three types of artificial NNs, the multilayer perceptron (MLP), the radial basis function (RBF) network, and the probabilistic neural network (PNN), for bearing fault detection was also performed. With the goal of automating the process of feature extraction, fault detection and identification was performed for REMs. A matching pursuit analysis was used to extract time–frequency domain features that were used subsequently as inputs to a feedforward neural network (FFNN) to classify the different bearing conditions (healthy, IRF, ORF, and BBF) [116]. Gebraeel et al. [117] proposed a way to predict the residual life from vibration-based degradation signals to estimate the bearing failure time. They developed two classes of models—a single bearing and a clustered bearing neural network—to perform REB fault prognosis. Different combinations of time, frequency, time–frequency domain features with an NN-based approach were also carried out to deal with REB fault detection and diagnosis [118–120]. A non-intrusive artificial NN approach that used stator current signals instead of vibration signals was also previously applied for REB fault detection and diagnosis for a three-phase induction motor [121].

Recently, in the last 2 years, a comparative study was published [122], where NN-based REB fault diagnosis was compared to SVM-based REB fault diagnosis; results showed that the latter gave better results than the former. An assessment study of the effect of the NN structure and parameters

on REB fault diagnosis was carried out in [123] since no formula exists to select the optimal values of these network characteristics. A hybrid fault diagnosis method for a REB fault in the field of gas turbine health management was investigated in [124]. This hybrid technique combined the S-transform algorithm [125] and the artificial NN method. Their results showed that the S-transform could extract good time–frequency domain features from the raw vibration signals for REB fault detection and diagnosis.

3.3 Combined methods for REB PHM

Merging different techniques is a common method of technique development. On this basis, many researchers have combined different SL-based methods to deal with REB PHM. Thus, in this paper, these combined SL-based methods are classified into statistical algorithms with NN methods, NN algorithms with signal processing methods, and statistical algorithms with signal processing methods. Many papers were found [126–152]; selected papers for each group are discussed in the following subsection. It should be noted that the papers discussed are just examples; selection does not indicate a preference or an endorsement by the authors. The only criteria applied for selection is that only the most recent published works were chosen for discussion. However, a summary of the classification of the SL-based REB PHM methods and a list of all reviewed SL-based REB PHM papers with the principle, pros and cons, and their applications, is presented in Table 3.

As a statistical algorithm with an NN method, J. B. Ali et al., [126] combined PCA and LDA with PNN and a simplified fuzzy adaptive resonance theory map (SFAM) neural network for early online diagnosis of naturally progressing bearing degradations. The former was used for feature reduction and the latter were used for classification. An artificial NN method combined with a signal processing technique (i.e., discrete wavelet transform (DWT)) was investigated to detect and diagnose the bearing faults of an industrial robot [127]. An ensemble SVM, as a statistical algorithm, was combined with composite multiscale fuzzy entropy (CMFE), as a signal processing method, for REB fault detection and diagnosis [128]. Other studies [129] and [130], tended to combine the ELM algorithm with other methods to better detect and diagnose the different REB faults. The ELM was used as a classifier and combined with multi-scale intrinsic mode function permutation entropy, which extracted feature parameters, after a preprocessing stage to de-noise the original vibration signals using Wavelet as the pre-filter [129]. Tong et al. [130] proposed a fault diagnosis approach for REBs based on redundant second generation WPT and ELM.

A more recent work that was just published in the current year, 2018, proposed a novel FDD Method for REB

Table 3 Summary of the reviewed SL-based REB PHM methods

SL-based REB PHM algorithms	Principle	Pros	Cons	Application	References
<i>Statistical approaches</i>					
LDA-based	LDA can find a linear combination of features that separates different classes. Its main objectives are either to reduce dimensionality or to perform classification.	Powerful statistical theory Generally outperforms centroid classification Powerful dimensional reduction technique	Difficult to be used for more than two-class classification, i.e., not suitable for REB fault diagnosis Cannot work properly with nonlinear data Has a high misclassification percentage when the number of trained data is small Struggles when dealing with missing data	Fault detection and diagnosis (FDD) Fault prognosis	[52–58] None
SVM-based	The basic idea of SVM is to map the nonlinear input data into a feature space first, then the inner product of this feature space is nonlinearly mapped to the original space via kernels. Thus, the main purposes of SVM are classification and estimation	Uses a well-established model, thus it could eliminate the need for experimental training data with the specific defective bearing Can handle nonlinear data Faster to be trained compared to NNs	Difficult to be used for more than two-class classification, i.e., not suitable for REB fault diagnosis Must deal with the optimization of the kernel functions	FDD Fault prognosis	[59–63], [65–70] None
KNN-based	KNN is a non-parametric, and lazy algorithm used for classification, in which the historical data are grouped into several classes to be used later to classify the new (unknown) data	Learning is very simple Easy to interpret (i.e., has a physical meaning) Can easily deal with more than two classes, i.e., suitable for REB fault diagnosis Effective classification method for noisy training data and complex target function Can handle nonlinear data	Need to store all the training data, i.e., memory-consuming Need to compare distance values for whole training samples, i.e., time- and power-consuming It is not robust to outliers Its fault detection and diagnosis accuracy highly depend on determining the parameter k When applied to REB PHM, overfitting problem may occur	FDD Fault prognosis	[71–79] None
ELM-based	ELM uses single-hidden layer feed-forward neural network (SLFNN), contrary to FNN that randomly chooses hidden nodes and analytically determines the output weights of the SLFNN	Provides a good generalization performance at an extremely fast learning speed Suitable for real-time REB fault diagnosis Has the ability of incremental learning in a non-stationary environment Can deal with the class imbalance issue in a non-stationary environment	Difficult to be extended to a deep architecture since it has basically only two layers The input weights and biases for hidden nodes are randomly selected, which may cause instability in the output nodes	FDD Fault prognosis	[81–84] None
Others	Different algorithms were found and grouped here including fuzzy classifier, decision tree, RFL classifier, clustering method (fuzzy C-means), etc	Fuzzy classifier efficiently handles uncertainty Decision tree is easy to interpret RFL can deal with non-stationary signals	Fuzzy inference engine needs many data points in the data set Fuzzy classifier requires prior knowledge Complexity of the decision tree	FDD Fault prognosis	[85], [86], [88–90], [92], [93], [101], [103] [87]

Table 3 (continued)

SL-based REB PHM algorithms	Principle	Pros	Cons	Application	References
<i>NN approaches</i>	NNs are nonlinear and multivariable models that can be seen as the reference models for the model-based PHM techniques and as a classifier for the SL-based PHM methods	Can handle nonlinear data Can easily deal with more than two classes, i.e., suitable for REB fault diagnosis Relatively easy to use	Slow to train (i.e., time-consuming algorithm) Prior domain knowledge is needed for feature extraction Weak generalization ability Increasing its classification accuracy by a few percent can hugely bump up its scale	FDD Fault prognosis	[107], [112–116], [118–124] [117]
<i>Combined approaches</i>	Merging the above techniques in the attempts to benefit from the Pros of some or eliminate the Cons of the others	Merging the above techniques to better detect and diagnose the REB faults under the highly nonlinear, non-stationary operating conditions Provides an online REB fault detection and diagnosis technique	The complexity of combined method May be difficult to interpret Merging two or methods may result in a time-consuming and/or power-consuming issue	FDD Fault prognosis	[127–139], [141–152] [126], [140]

based on ensemble local characteristic-scale decomposition (ELCD) and the ELM (ELCD-ELM) algorithm [131]. First, numerous intrinsic scale components (ISCs) were obtained by decomposing the vibration signals using ELCD, and then different ISCs (in the time domain, energy, and relative entropy) were calculated to be the inputs to the ELM-based REB FDD. The proposed ELCD-ELM was found to be able to process nonstationary vibration signals and overcome mode-mixing phenomenon of the LCD method.

4 Deep learning for REB PHM

Many SL-based techniques have been applied to the PHM field and investigated to detect, diagnose, and predict (sometimes) rolling element bearing health conditions, as reviewed and summarized in the previous section. Those SL-based techniques achieved decent performance, especially when *detecting* and *diagnosing* REB faults. However, few studies that deal with REB fault *prognosis* were found. Further, from surveying the above-reviewed SL-based REB PHM techniques, it is clear that the performance of those techniques depends greatly on extracting the best-suited features, which were summarized in Table 1. Given the fact that the SL-based PHM techniques manually design and extract the features, in addition to the variety and the large amount of data in the PHM field, it can be concluded that those SL-based PHM techniques will face significant challenges in actually determining the best-suited features to be extracted, especially in the big data scenario. Further, the SL-based PHM techniques have other challenges that come from the big data, such as the high dimensionality of feature space, the proliferation of multimodal data, and multicollinearity among data measurements [23]. Moreover, the four phases of the SL-based PHM technique shown in Fig. 2b cannot be optimized simultaneously (i.e., data processing, feature extraction, feature selection, and model training usually are done successively, not at the same time), which boosts the required processing time (i.e., time-consuming issue) and increases complexity. Therefore, as a technique that has the capability to be a bridge that connects the big data from the machinery and the intelligent machine PHM methods, the DL-based PHM technique is being adapted to the REB PHM field. This DL-based REB PHM method is known as a method that classifies different patterns via stacking multiple layers in hierarchical architectures and can model the high-level representations behind data [21]. Further, the DL-based techniques are gaining popularity even in the PHM field because they can use the raw data directly (without any preprocessing, as shown in Fig. 2c) as an input, i.e., representation learning. They can learn complex and highly nonlinear representations from high-dimensional data [153].

Although the deep learning is not a new concept, it has only recently started to gain more attention and to be successfully applied in different fields, such as computer vision, language and audio processing, and (automatic) recognition [153], [154]. It is only in the last few years that deep learning started to be applied to the PHM field [155–157]. To the authors' knowledge, the deep learning technique was first applied to rolling element bearing prognostics and health management in the year 2015, except for a few works, such as the one from Liu et al. [158]. Liu et al. used sparse coding with a learned dictionary instead of a predefined one for adaptive feature extraction from the vibration signal for REB fault diagnosis; they introduced a natural extension of sparse coding, the shift-invariant sparse coding algorithm. In other work, Verma et al. [159] proposed intelligent condition-based monitoring of REMs using a sparse autoencoder method.

As mentioned in Sect. 3 and following the classification of deep learning methods in Zhao et al. [21], this paper thus classifies and reviews the existing studies on DL-based REB PHM into four groups: (deep) CNN-based, (deep) RNN-based, RBM-based DNN, and AE-based DNN approaches. In addition to briefly introducing the definition and principle of each algorithm with its typical structure, its application to REB PHM is outlined to highlight its challenges, its pros and cons, and its latest advancements.

4.1 (Deep) CNN-based REB PHM approaches

One deep learning technique is the convolutional neural network (CNN) approach. Per its definition, CNN is a feed-forward neural network of multiple layers, which assumes inputs as images [160]. It was inspired by neurons of the human visual cortex that have two features [153]. One is local connections, which means that since images have high correlation within sub-regions, the correlation information is critical in recognizing those images, where the sub-regions in the previous layer are connected to local patches in the feature maps by filters. The other feature is shared weights, where a pattern can appear in various locations in the images, and by convoluting filters across an image, the pattern can be extracted independent of location. In addition, using the same filter across an image, the number of parameters is reduced significantly. Nowadays, many open-sourced CNN models are available (e.g., GoogLeNet, AlexNet) which make them attractive to researchers.

CNN is structured by a series of layers, in which the convolutional and pooling layers come first, and the fully connected layers come last. A descriptive example of the CNN architecture is shown in Fig. 10. The convolutional layer is used to detect local correlation from the previous layer (the raw input). It has a number of hyper-parameters, such as the number of filters, the filter size, the stride, and

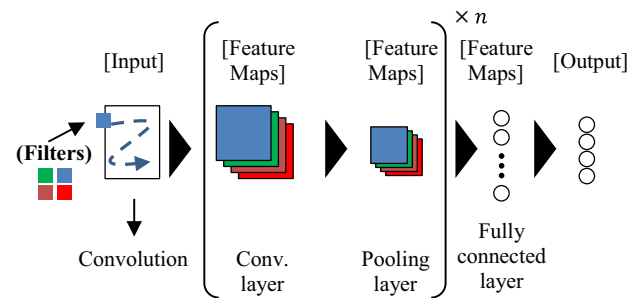


Fig. 10 Descriptive example of convolutional neural network (CNN)

the zero-padding. Usually, an activation function is applied that can be linear or nonlinear, such as the ReLU (the most used one), the sigmoid, etc., to treat the raw input data and to generate invariant local features. The pooling layer is used to pool out the good features among those local features or to merge several features into one. To do that, it has different pooling operation types, i.e., max pooling, average pooling, L2-norm pooling; max pooling is the most-used type. In some studies, the pooling layer is not used, but instead larger strides are considered. Feature learning can be performed in CNN by stacking and switching the convolutional layers and the pooling operations. The fully connected layer contains full connections to all activations in the previous layer (i.e., the pooling layer). Where, after learning the features, the two-dimensional map is converted into a one-dimensional vector within this layer (i.e., fully connected layer) and then fed into a softmax function for model construction [23].

An investigation of the use of the convolutional neural networks (CNN) with a deep structure, from one-layer up to three-layers, on raw signals to test the accuracy of it as a classifier on bearing fault data was proposed in [161], where its effectiveness was investigated when the input signals were corrupted with noise. More works were carried out to deal with REB PHM based on DCNN (deep CNN) [156] or a modified DCNN, i.e., the hierarchical adaptive DCNN [162] and energy-fluctuated multiscale feature learning with deep Convnet for intelligent spindle bearing fault diagnosis in [163]. CNN-based bearing fault detection was proposed in [164], which was considered as a feature-learning model for condition monitoring, so that it can autonomously learn useful features for bearing fault detection from the data itself.

Just last year, 2017, several papers were published, [165–171], that used CNN-based deep learning to deal with detecting and diagnosing REB faults. Thus, it should be noted that there is a clear tendency toward applying such deep learning techniques for REB fault detection & diagnosis tasks; however, no study paper has yet considered the prognostic task—this research still needs to be pursued.

A hybrid method was proposed by You et al. that benefits from the feature-learning capability of the CNN method—as

a deep learning technique—and the generalization ability of the support vector regression (SVR) [165]. In [166] the method was used to detect and diagnose different bearing faults as well as gear faults. The proposed hybrid model, CNN-SVR, was constructed by replacing the top layer of the traditional CNN with an SVR classifier and then the new model was stacked layer-by-layer with convolutional layers and pooling layers inside. The structure of the proposed hybrid model consists of 10 layers totally, including the input layer, three convolutional layers, three pooling layers, two fully connected layers, and a support vector regressive classifier as the top layer. In [167], [168], and [169], the DCNN was applied as a deep learning technique that was combined with other methods to deal with REB FDD. Zhang et al. [167] proposed a novel method named DCNN with wide first-layer kernels (WDCNN); Fuan et al. [168] utilized DCNN with a particle swarm optimization method and the t-distributed stochastic neighbor embedding (t-SNE) technique. Li et al. [169] proposed IDSCNN, which is based on ensemble DCNN and an improved Dempster–Shafer theory based on an evidence fusion technique. Another paper that combined the CNN with a feature extraction algorithm based on EMD method was proposed by Xie and Zhang [170] with attention to extracting distinguishing features (compressed features with spatial information) to solve the nonstationary characteristic in the original vibration signals. Finally, Lu et al. [171] investigated a new hierarchical network of CNN-based deep learning for bearing fault diagnosis under fluctuated working conditions and noisy environments making use of cognitive computing theory.

4.2 (Deep) RNN-based REB PHM approaches

When constructing a deep NN with the same weights that are applied in a recursive way in the network (i.e., the weights are shared across the whole network), a recurrent neural network (RNN) will be generated [172]. RNN is a neural network that has cyclic connections in the hidden units; these can hold past information. Furthermore, RNN processes input data, x_t at time step t , sequentially, and the past data is stored in a state vector, h_t at time step t , implicitly. Therefore, in the RNN, the output, O_t at time step t , of current data depends on all the past data. An RNN can be seen as a very deep network using the same weight as can be seen in Fig. 11a, b, where, W_{in} are the weights of the inputs, W are the weights of the state, and W_{out} are the weights of the output. An RNN is powerful in analyzing sequential information; however, it can face the vanishing gradient problem during backpropagation for model training. Thus, the long-short term memory (LSTM) [173] and the gated recurrent units (GRU) techniques [174] were developed.

Almost all found papers, [175–180] used the RNN as a tool not only for REB fault diagnosis, but also for prognosis,

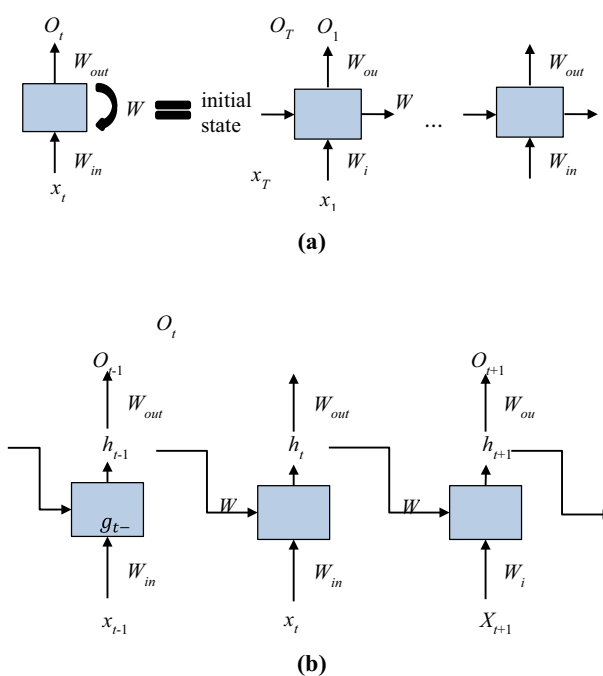


Fig. 11 **a** A general structure and **b** the detailed one of recurrent neural network (RNN)

except Abed et al. [175]. Abed et al. [175] used dynamic recurrent neural networks (DRNNs) that can learn the dynamics of nonlinear systems, whereas conventional static neural networks cannot. The DRNN was fed with the orthogonal fuzzy neighborhood discriminant analysis (OFNDA) features to be applied for real-time REB FDD.

Malhi et al. [176] preprocessed vibration signals from defect-seeded REB using CWT and then used a competitive learning-based approach based on the RNN algorithm for long-term prognosis. Different statistical parameters were utilized as inputs to the RNN, which were clustered based on the principle of competitive learning to effectively represent the bearing defect propagation. The results showed that the RNN did not work well in a short-term prediction case, but for long-term prediction, the RNN did increase the training speed and achieved good prognostic results. Sharma et al. [177] proposed a robust fault analysis method to diagnose and predict the level of fault severity of a REB. They used DWT for feature extraction, and an orthogonal fuzzy neighborhood discriminative analysis (OFNDA) technique for feature reduction. Finally, a DRNN method was used to predict the REB conditions and classify their different faults. Xie and Zhang [178] used two methods, echo state network (ESN) and recurrent multilayer perceptron (RMLP), which are functionals of RNN, for vibration-based REB fault prognosis. The two methods used were able to predict the REB health condition in a relatively short time and with only limited data available, contrary to the autoregressive

moving average (ARMA) and SVM methods. More recently, the RNN was used as the main tool for REB prognosis [179], [180], where in [180] it was applied in the time and the frequency domains; test results showed that the RNN can be used to do fault prognosis in general, and especially for bearing health conditions. These prior studies showed promising results regarding the ability of RNNs to predict the RUL, is an important factor for decision-making to alleviate emergency situations. Thus, the use of an RNN for REB fault prognosis is worth further in-depth study.

4.3 RBM-based DNN approaches for REB PHM

Deep neural networks (DNNs) belong to the category of artificial NNs, but they are generally superior since they are known to have strong power for learning representation. A DNN that builds an architecture using a deep learning technique, which is a layer-by-layer learning technique, has the ability to deal with the issue of a local-optimal to train the parameters of the network [111]. A deep DNN structure can be built either by the restricted Boltzmann machine (RBM) or by the autoencoder (AE) technique. In the next two subsections, research on RBM-based DNN and the AE-based DNN for REB PHM is reviewed, respectively. Thus, a brief description of RBM is presented first, with variant models that used it as the basic learning module, i.e., deep-belief networks (DBN) and the deep Boltzmann machine (DBM). Then, a comprehensive review of existing studies examining RBM-based DNN for REB PHM is presented.

An RBM is a network of symmetrically coupled stochastic binary units composed of a visible layer with visible nodes, v_i , and a hidden layer with hidden nodes, h_j . To build the RBM, there must be a symmetric connection between the visible and the hidden units and no connection among the same layer, as can be seen in Fig. 12a. Further details can be found in [181] and in [21].

Stacking multiple RBMs, the DBN is constructed, as can be seen in Fig. 12b. Thus, the DBN is a NN of multiple layers that has stochastic latent variables (hidden units) and a generative graphical model [182]. The DBN has two steps of training, first an unsupervised layer-wise pre-training (RBM 1, RBM 2, and RBM 3 in Fig. 12b), and then a supervised fine-tuning (fully connected (FC) layer in Fig. 12b). In pre-training, each hidden layer serves as the visible layer for the next layer.

In contrast to a DBN, the DBM is built by grouping hidden units into a hierarchy of layers instead of a single one. Thus, the DBM is simply a deep structured RBM, where any adjacent layers can be connected, but non-adjacent layers cannot be connected. In addition, no connection is permitted within units of the same layer. The DBM adopts learning a complex, fully connected Boltzmann machine, in which each layer captures complicated, higher order correlations

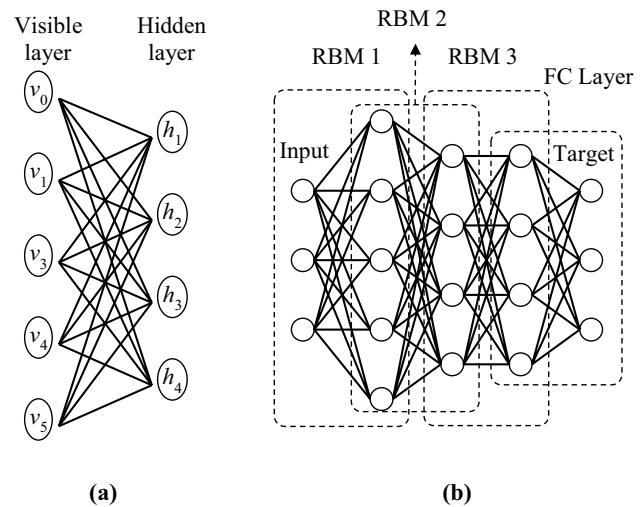


Fig. 12 Frameworks of **a** RBM architecture, and **b** RBM-based DBN structure

between the activities of hidden features in the layer below [183].

First, the RBM-based DNN structure that uses RBM as the basic learning module, i.e., the deep-belief network (DBN), was employed as a bearing condition monitoring tool to overcome the presence of noise and transient impacts in the acquired vibration signals in [184]. Another research paper that uses an optimization DBN method to deal with REB fault diagnosis was achieved by Shao et al. [185]. Different research works have been performed by combining the DBN with other techniques to improve the REB detection, diagnosis, and prognosis capability. Wang et al. [186] proposed a bearing fault diagnosis method based on the Hilbert envelope spectrum and a DBN. Getting the right parameters of the DBN is crucial, however, it can be time-consuming due to the training process. Thus, a research study in [187] was proposed to deal with this issue and to avoid both the overfitting and the under-fitting problems. An assessment of the bearing degradation based on the Weibull distribution and a DBN was investigated by Ma et al. [188]. Bearing fault diagnosis based on a DBN and multi-sensor information fusion techniques was carried out based on use of multi-vibration signals to adaptively fuse multi-feature data and identify various bearing faults [189]. Yin et al. [190] developed a combined machine health assessment model based on an Isomap and a DBN, which effectively evaluated the degradation of the bearing health conditions, since it was found to be more sensitive to the incipient faults. A two-layer hierarchical diagnosis network (HDN) [191] that deals with REB diagnosis in two stages was carried out using a wavelet packet energy feature. The bearing fault types were identified by the first layer, then their severity ranking was recognized in the second layer. Finally, the HDN was

compared to two similar networks constructed by SVM, and to a backpropagation neuron network (BPNN); according to the experimental results, it could deal with the presence of noises and disturbances that gave rise to the overlapping problem among the different fault classes and was more reliable for precise, multi-stage diagnosis.

One critical challenge for performing prognosis of bearings in the era of the IoT and 4th Industrial Revolution is to automatically process massive amounts of data and accurately predict the RUL of bearings. Recently, a study of Deutsch et al. [192] addressed the limitations of SL-based REB prognostics, and presented a new method that integrates a DBN and a particle filter for RUL prediction of hybrid ceramic bearings; the study then compared the results with DBN and particle filter-based approaches. The validation and comparison results showed promising RUL prediction performance of the integrated method. Early bearing fault diagnosis using effective feature selection methods was proposed by Devendiran et al. [193]; these researchers used a DBN as one of the neural network classification algorithms. In contrast to the conventional fault diagnosis and classification methods, which usually do not consider the temporal coherence of time series data, Zhang et al. [194] proposed a REB FDD model based on a DBN. It can directly recognize raw time series sensor data without feature selection and signal processing. It also takes advantage of the temporal coherence of the data, thus, expertise in feature selection and signal processing is not required.

In the current year, 2018, three papers have already been published. All of them used the DBN-based deep learning method to deal with REB PHM. In [195], in contrast to the shallow learning methods, which require establishing explicit model equations and much prior knowledge (and therefore are limited in the age of big data as explained in the previous section), this paper presented a deep learning-based approach for RUL prediction of rotating components with big data. The developed deep learning-based approach was a DBN-feedforward neural network (DBN-FNN) algorithm that takes advantage of the self-taught, feature-learning capability of the DBN and the predicting power of the FNN; together, these strategies overcome the above-mentioned limitations. A novel convolutional deep-belief network (CDBN) was proposed for REB PHM in [196]. First, an autoencoder was used to compress data and reduce the dimension. Second, a novel CDBN was constructed with Gaussian visible units to learn the representative features. Finally, the exponential moving average (EMA) was considered to improve the performance of the constructed deep model. Another study was performed by Oh et al. [197] where the researchers developed a DBN-based deep learning method with vibration images as the inputs. The developed method was found scalable, due to the fact that the

vibration imaging approach devised incorporates data from systems with various scales, such as small testbeds and real field-deployed systems. Further, the method was proposed for unsupervised feature engineering. The proposed DBN-based deep learning algorithm was pre-trained for high-level feature extraction, where a large amount of field data without any label can be incorporated since pre-training can be achieved in an unsupervised manner. Then, the pre-trained DBN was fine-tuned by combining it with a multilayer perceptron (MLP), leading to a fault classifier. The pre-trained DBN could also be used as a fault cluster by combining it with a self-organizing map (SOM).

Second, an RBM-based DNN structure that stacks multiple RBMs, i.e., deep Boltzmann machine (DBM), was investigated and used for REB condition monitoring in [198]. In the study, several time, frequency, and time–frequency domain features were extracted from an acquired data set with seven fault patterns to assess the performance of the proposed DBM for REB fault diagnosis. The seven parameters were used as the input parameters of the DBM model. Their results clarify the accuracy and reliability of the DBM model. An enhanced RBM was considered with prognosability regularization for prognostics and health assessment of the REBs. The proposed DBM method was benchmarked with deep structure of the regular RBM algorithm and the PCA [199]. A scoring method based on the benchmarking score was used to evaluate each PHM method in its ability to predict the RUL. He et al. [200] proposed a novel bearing diagnosis method based on the Gaussian restricted Boltzmann machine (Gaussian RBM) algorithm using vibration signal data. The envelope spectrums were used directly as the feature vectors to represent the fault types of the bearing and then classified using the proposed Gaussian RBM algorithm.

The deep-statistical feature learning (DSFL) of the machinery condition health monitoring can be constructed by Gaussian-Bernoulli deep Boltzmann machine (GDBM) [201], where in the GDBM, each neuron in the intermediate layers is connected to both top-down and bottom-up information, unlike in other RBM-based deep models, such as the DBN and the deep autoencoder. For deep learning of statistical features with unknown value boundaries, real-value Gaussian-Bernoulli restricted Boltzmann machines (GRBMs) were stacked to develop the GDBM-based DSFL method in [202] and were applied for both bearing and gearbox systems. Deutsch and He [203] dealt with bearing RUL prediction with big data based on a deep learning technique that used the DBM, which predicts L steps ahead in the future, to predict the RUL by predicting the RMS values and the time of the bearing's failure.

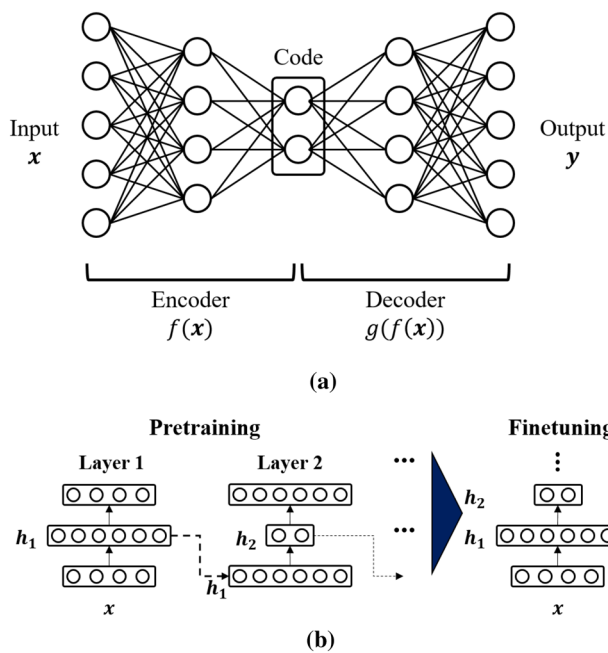


Fig. 13 Frameworks of **a** AE architecture, and **b** the AE-based DNN structure

4.4 AE-based DNN approaches for REB PHM

The autoencoder [204] belongs to the unsupervised machine learning structure; it is a feedforward NN that consists of two phases, the encoder phase and the decoder phase. Its main characteristic is that it tries to predict the output equal to the input. A common architecture of AE is shown in Fig. 13a and the AE-based DNN structure is shown in Fig. 13b. In the encoder phase, several features will be extracted from the input vector to form the hidden layer via a nonlinear mapping, using a weight matrix W_1 . In the decoder phase, the output vector will be predicted to reconstruct the original input vector in a similar way by employing a weight matrix W_2 . It should be noted that the hidden layer could have a smaller dimension m than the input layer dimension n (i.e., $m < n$) so the AE will be forced to learn a compressed form of the input, which will try to find the correlation among the input vectors. Also, it can have a bigger dimension $m > n$ and the AE still can have an interesting structure by imposing other constraints on the network.

The AE-based trained layers can be stacked into a new network, which is an AE-based DNN. By training and stacking various layers of AEs, diverse structures of AE-based DNN can be generated to extract features that present health states of various engineered systems as shown in Fig. 13b. Moreover, a deeper layered structure than the one in Fig. 13a is another widely used form of AE-based DNN, which can discover highly representative features from extremely complex signals.

The AE was applied to rolling element bearing fault diagnosis by W. Lu et al. [205] to extract the features from the raw signal and guarantee sensitivity to every interested fault category to avoid incomplete diagnosis results and the appearance of unknown-category faults. Six classes of the different bearing faults were considered to evaluate the proposed method after data preprocessing using the FFT to generate a 600 points length. Therefore, the built DNN was a two-layer structure with 800 and 400 neurons in the hidden layers, which were constructed by an AE. The authors concluded that the constructed DNN-based AE could extract useful features and further studies should be carried out to better classify the fault categories with high accuracy and to handle the unknown-category fault cases. F. Jia et al., [206] stated that the AE-based DNN method could overcome the two issues hindering ANN-based intelligent fault diagnosis of rotating machinery—1) the need for prior expertise and knowledge to manually extract fault features, and 2) the limitation in learning the complex nonlinear relationships in fault diagnosis. Thus, F. Jia et al., [206] suggested a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, based on AE with deep architectures, instead of shallow ones. Another study [207], took advantage of the learning capability of the AE and combined it with the digital wavelet frame (DWF) and nonlinear soft threshold method to de-noise the fault vibration signal first. The study applied a stacked autoencoder (SAE) to extract the features, which were the inputs to a BP network classifier. A method was employed to deal with extracting features from the stationary and nonlinear characteristics of bearing deep learning stacked de-noising autoencoder (SDAE) [208] non-vibration signals. The deep learning SDAE, combined with dropout, was found to be useful to learn good representations of those features and improve fault pattern classification robustness.

The AE-based DNN was also used to classify REB fault classes under two and three hidden layers in an unsupervised manner using the encoder part of the AE [209]. L. Guo et al., [210] suggested a new multi-feature extraction and a nonlinear dimension reduction algorithm based on deep learning for bearing condition recognition. Different time domain, frequency domain, and time–frequency domain features were calculated and then their dimension was reduced. Finally, the different bearing faults were classified using a top-layer classifier of AE-based DNN outputs.

S. Tao et al. [211] combined the SAE with the softmax regression method, which is a classification method that generalizes logistic regression to multiclass problems, for examining the bearing fault diagnosis problem. Their results showed that combining SAE with the softmax regression method had a strong robustness and eliminated the impact of noises remarkably. H. Liu et al., [212], for the first time, combined the short-time Fourier transform (STFT) and the

stacked sparse autoencoder (SSAE) with the softmax regression to automatically extract the features from the sound signals and classify the different REB fault modes, respectively. The effectiveness of the proposed STFT-SSAE method was investigated and compared to empirical mode decomposition (EMD), Teager energy operator (TEO), and SSAE to evaluate its performance in deploying the PCA technique for dimensionality reduction. Taking the advantage of the high training speed of the ELM method and the AE extraction capability, another deep learning algorithm named the “AE-ELM-based diagnosis method” was proposed by Mao et al. [213] to build a universal extraction and a fast-trained method to deal with the different REB diagnosis issues. High-level features, which were extracted in the frequency domain and the Wavelet packet transform domain, were extracted in [214] using a DNN. First, the weights of the DNN were initialized using Stacked Denoising Sparse Autoencoders (SDSAE), then those weights were finetuned based on the softmax regression and the centering towards the median. These high-level features were then classified using the SVM and the random forest technique simultaneously. In [215], the SDAE was also applied to denoise random noises in the raw signals and to represent fault features in fault pattern diagnosis for both bearing rolling faults and gearbox faults. The SDAE was trained in a greedy, layer-wise fashion. The proposed method was compared to the DBN algorithm in a highly noisy environment and the results showed its superiority for fault diagnosis.

More recent works, just accepted or lately published in last/current year, 2017/2018, have been conducted to investigate the AE method or its varieties for use in REB fault diagnosis. Chen and Li [216] proposed a multi-sensor feature fusion for bearing fault diagnosis using sparse auto encoder and a DBN. Lu et al. [217] solved the health state identification problem in REB fault diagnosis using a SDAE method. Another deep learning method named “automated AE correlation-based (AEC)” was developed by Hasani et al. [218]; it was used for health monitoring and for prognostics of machine bearings. A hybrid feature pool method was proposed in [219] that was combined with SAE-based DNNs to perform effective diagnosis of REB faults of multiple severities. The authors found that the hybrid feature pool could extract more discriminating information from the raw vibration signals to overcome the non-stationary behavior of the signals caused by multiple crack sizes; the proposed method outperforms the SVM and the BPNN. In [220], a locality preserving projection (LPP) was adopted to fuse the deep features, and thus to build a new deep AE method constructed with a denoising autoencoder (DAE) and a contractive autoencoder (CAE) for the enhancement of feature-learning ability with the goal of diagnosing REB faults. A hybrid deep model consisting of a multi-channel CNN followed by a stack of denoising autoencoders (MCNN-SDAE)

was developed by A. Shaheryar et al., [221] for fault identification in rotary machines. In the study, these researchers explored the MCNN for unsupervised feature learning on vibration signals and SDAE for extracting vibration features that are robust and invariant to the noises in vibration signals.

Two papers, [222] and [223], that proposed intelligent bearing fault diagnosis methods were just published in the current year, 2018. [222] was built by combining compressed data acquisition and SSAE-based deep learning; in [223], the authors explored, highly compressed measurements of REB vibration signals under different operating conditions, taking advantage of the compressed sensing (CS) method [225] that allows sampling the signals below the Nyquist frequency. A summary of the classification of the DL-based REB PHM methods and a list of all reviewed DL-based REB PHM papers with the principle, the pros and cons of these methods, and their applications, is presented in Table 4.

5 Summary and concluding remarks

This paper has presented a comprehensive review and summary of recent techniques aimed at REB fault detection, diagnosis, prognosis, and their applications. Thus, this paper attempts to elegantly represent the widespread, contemporary REB PHM techniques by considering two main categories, shallow learning algorithms and deep learning methods. First, the different bearing failure modes were briefly described, focusing on fatigue, wear, plastic deformation, corrosion, electrical erosion, and fracture & cracking modes. Then, the different health features (indexes, criteria), which are used by these contemporary REB PHM techniques, were thoroughly described (with their physical meaning where applicable—some of the features do not have any) to provide an overall background for researchers, system engineers, and experts—in the general PHM field and in the specific REB PHM field—to select and adopt the best fit for their specific applications.

Several SL-based algorithms were found and were applied to REB PHM systems; some originated from artificial neural networks (NN), some did not. Thus, three categories were proposed in this paper: (1) Statistical approaches, which were divided to LDA-based REB PHM, SVM-based REB PHM, KNN-based REB PHM, ELM-based REB PHM, and other statistical algorithms for REB PHM; (2) NN approaches; and (3) combined methods. Further, DL-based REB PHM techniques were also reviewed and classified into four groups in this paper, as follows: (1) (Deep) CNN methods; (2) (Deep) RNN methods; (3) RBM-based DNN methods—subdivided into DBN-based REB PHM and DBM-based REB PHM; and (4) AE-based DNN methods. Furthermore, the principle, the

Table 4 Summary of the reviewed DL-based REB PHM methods

DL-based REB PHM algorithms	Principle	Pros	Cons	Application	References
(Deep) CNN approach	CNN is simply a feedforward NN of multiple layers, which assumes inputs as images	Less complexity in terms of the required number of neurons compared to the artificial NN Many open networks are available: GoogLeNet, AlexNet, VGG, and Clarifai [108], [153]. Can handle nonlinear data and noisy signals	High network complexity (i.e., many layers) is needed to model high hierarchical training data High computational cost Weak generalization ability	FDD Fault prognosis None	[161–172]
(Deep) RNN approach	RNN is a deep NN structure that applies the same weights recurrently over structured inputs	Powerful in analyzing sequential information A very deep network in which the current output depends on all the past data Work well with short-term information	Can frequently face a vanishing gradient problem during back-propagation for model training Can face difficulties when dealing with long-term information Huge amounts of data are needed for training	FDD Fault prognosis	[176], [178] [177], [179], [180], [181]
<i>RBM-based DNN approaches</i>					
DBN-based	The DBN is a NN of multiple layers, which has stochastic latent variables (hidden units) and a generative graphical model; it is constructed by stacking multiple RBMs	Can deal with the issue of local-optimal to fix the parameters of the network when benefiting from a regularization method such as dropout, L2 regularization, etc Has strong power of representation Does not need much prior knowledge or much expert knowledge Can handle nonlinear data Considers the temporary coherence of time series data	Significant computations are needed, especially in the training procedure that requires initialization and sampling Time-consuming due to the optimization process	FDD Fault prognosis	[185–188], [190], [192], [193], [195], [196], [198], [199] [189], [191], [194], [197]
DBM-based	DBM is constructed by grouping hidden units into a hierarchy of layers instead of a single one. The DBM thus is simply a deep structured RBMs	Can handle nonlinear data Robust when dealing with obscured data Top-down feedbacks are integrated	Time complexity for the inference is higher than that of a DBN [212] Cannot handle big data well, especially during the optimization of the network parameters	FDD Fault prognosis	[200], [202], [204] [201], [205]
<i>AE-based approaches and its variants</i>					
	AE is a feedforward NN with an unsupervised machine learning structure that aims at predicting accurately the output. Further, it reduces the data dimensionality during the encoder phase	Can extract features from raw signal and guarantee sensitivity to each considered fault, thus avoiding incomplete diagnosis results and the appearance of the unknown-category faults An unsupervised learning technique Does not need much prior knowledge or much expert knowledge Can handle directly complex nonlinear data	During the propagation, errors can appear Requires a pre-training phase The possibility of the sparse representation	FDD Fault prognosis	[158], [207], [208–219], [221–225] [220]

pros and cons of these SL-based REB PHM, and DL-based REB PHM methods, and their advancements and applications were reviewed and summarized.

From this survey study, several key points can be concluded, including:

- Although both SL-based and DL-based REB PHM techniques achieve good results in detecting and diagnosing different REB faults (sometimes they achieve perfect results with 100% accuracy), they are still not adopted in industry due to a lack of studies that consider how these contemporary techniques (i.e., SL-based and DL-based) will be applied in practice. Thus, it will be very interesting if academics and industrial experts work together to adopt and study these strategies. To consider different scales, different fault modes (i.e., a single failure mode as well as compound failures), and different bearing types, such as journal bearings and magnetic bearings that are becoming more incorporated in real-world applications nowadays should be studied. Furthermore, it is recommended that companies and industry experts to share their data healthy, faulty, and run-to-failure data—with academics, who usually use only data collected from their in-lab test bench; this shared data would help to achieve better advancements not only in research, but also for practical industry use.
- It is well known in the PHM field that if an accurate enough reference model exists, using a model-based technique for detecting, diagnosing, and predicting the faults is the best choice. Thus, incorporating dynamic models of REBs could improve the accuracy of the REB PHM methods. Further, since fault data are very rare and hard to get from modern engineered systems, researchers can benefit from the recently developed generative adversarial networks (GAN) technique for generating faulty data.
- Although there have been significant advancements in the development of both the SL-based REB PHM techniques and the DL-based REB PHM techniques, there is still no formula or law that exists to select the optimal values of the network geometry or hyper-parameters (e.g., number of layers) to achieve the best results in detecting, diagnosing the bearing faults, and (ultimately) predicting health conditions. Thus, providing a standardized platform or at least a streamline of how deep those algorithms should be, with consideration of the fact that most companies lack software, modeling, and expertise to understand deeply those algorithms and to interpret their results, will enable integration of these contemporary techniques into real-world applications.
- Finally, nearly all existing REB PHM, whether based on shallow learning or deep learning techniques, have targeted only the REB fault detection and diagnosis (condition monitoring) problem. Very few studies were found

that deal with the REB prognosis with the aim of predicting the remaining useful lifetime (RUL) with the goal of providing a better condition-based maintenance (CBM) strategy. Strategies that begin to enable improved CBM will be of great interest to the rolling element bearing PHM field in particular, and to PHM for any modern engineered system in general, especially in the forthcoming years in the age of IoT and big data.

Acknowledgements This research was supported by Korea Electric Power Corporation (R17TH02), the Basic Research Lab Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) (No. 2018R1A4A1059976), and a grant from the Institute of Advanced Machinery and Design at Seoul National University (SNU-IAMD).

References

1. V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fiedjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015)
2. V. Venkatasubramanian, R. Rengaswamy, S.N. Kavuri, K. Yin, A review of process fault detection and diagnosis Part III: process history based methods. *Comput. Chem. Eng.* **27**(3), 327–346 (2003)
3. S.X. Ding, P. Zhang, T. Jeinsch, E.L. Ding, P. Engel, W. Gui, A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis. Preprints of the 18th IFAC World Congress Milano (Italy) (2011), pp. 12380–12388
4. S. Yin, S.X. Ding, X. Xie, H. Luo, A review on basic data-driven approaches for industrial process monitoring. *IEEE Trans. on Industrial Electronics* **61**(11), 6418–6428 (2014)
5. C. Hu, B.D. Youn, P. Wang, Time-dependent reliability analysis in operation: prognostics and health management. in *Engineering Design Under Uncertainty and Health Prognostics*. Springer Series in Reliability Engineering (Springer, Cham, 2019), pp. 233–301
6. C. Hu, B.D. Youn, P. Wang, Case studies: prognostics and health management (PHM). in *Engineering Design under Uncertainty and Health Prognostics*. Springer Series in Reliability Engineering (Springer, Cham, pp. 303–342), 2019
7. I. Shin, J. Lee, J.Y. Lee, K. Jung, D. Kwon, B.D. Youn, A framework for prognostics and health management applications toward smart manufacturing systems. *International Journal of Precision Engineering and Manufacturing-Green Technology* **5**, 519–538 (2018)
8. C. Hu, B.D. Youn, P. Wang, J.T. Yoon, Ensemble of Data-Driven Prognostic Algorithms for Robust Prediction of Remaining Useful Life. *Reliability Engineering and System Safety* **103**, 120–135 (2012)
9. G. Niu, J. Jiang, B.D. Youn, M. Pecht, Autonomous health management for PMSM rail vehicles through demagnetization monitoring and prognosis control. *ISA Trans.* **72**, 245–255 (2018)
10. C. Hu, B.D. Youn, P. Wang, *Engineering Design Under Uncertainty and Health Prognostics* (Springer, Cham, 2018). ISBN 978-3-319-92572-1
11. A. Rai, S.H. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **96**, 289–306 (2016)

12. S. Choi, B. Akin, M. Rahimian, H. Toliyat, Performance-oriented electric motors diagnostics in modern energy conversion systems. *IEEE Trans. Ind. Elect.* **59**(2), 1266–1277 (2012)
13. C. Lanham, Understanding the tests that are recommended for electric motor predictive maintenance. *Baker Instrument Company* (2002)
14. S. Nandi, H. Toliyat, X. Li, Condition monitoring and fault diagnosis of electrical motors- A review. *IEEE Trans. Energy Convers.* **20**(4), 719–729 (2005)
15. IEEE Motor Reliability Working Group, Report of large motor reliability survey of industrial and commercial installations. *IEEE Trans. Industrial Appl.* **21**(4), 853–872 (1985)
16. W. Zhou, T.G. Habetler, R.G. Harley, Bearing condition monitoring methods for electrical machines: a general review. *Proc. IEEE SPEEDAM*, 6–8 (2007)
17. M. Hamadache, D. Lee, K.C. Veluvolu, Rotor speed-based bearing fault diagnosis (RSB-BFD) under variable speed and constant load. *IEEE Trans. Ind. Electro.* **62**(10), 6486–6495 (2015)
18. S.X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms and tools* (Springer, Germany, 2008)
19. S. Schuet, D. Timuçin, K. Wheeler, Physics-based precursors or wiring diagnostics for shielded-twisted-pair cable. *IEEE Trans. on Instrum. and Measurement* **64**(2), 378–391 (2015)
20. J. Liu, W. Luo, X. Yang, L. Wu, Robust model-based fault diagnosis for PEM fuel cell air-feed system. *IEEE Trans. on Industrial Electronics* **63**(5), 3261–3270 (2016)
21. R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring: a survey. *Journal of Latex Class Files* **14**(8), 1–13 (2015)
22. G. Zurita, V. Sánchez, D. Cabrera, A review of vibration machine diagnostics by using artificial intelligence methods. *Investigación & Desarrollo* **1**(16), 102–114 (2016)
23. J. Wang, Y. Ma, L. Zhang, R.X. Gao, D. Wu, Deep learning for smart manufacturing: methods and applications. *J. Manuf. Syst.* **13**. Available online Jan. 2018 (**In Press**)
24. B. Sung, J. Lee, Reliability improvement of machine tool changing servo motor. *Journal of International Council on Electrical Engineering* **1**(1), 28–32 (2011)
25. J. Slavic, A. Brkovic, M. Boltezar, Typical bearing-fault rating using force measurements-application to real data. *J. Vib. Control* **17**(14), 2164–2174 (2011)
26. Emerson, Bearing failure analysis, ebook. <http://www.emerson-bearing.com/bearing-failure-modes> (2017)
27. ISO 15243 Rolling bearings: damage and failures—terms, characteristics and causes (2004)
28. P.P. Kharche, S.V. Kshirsagar, Review of fault detection in rolling element bearing. *Int. J. Innov Res Adv Eng (IJIRAE)* **1**(5), 169–174 (2014)
29. S. Devendiran, K. Manivannan, S.C. Kamani, R. Refai, An early bearing fault diagnosis using effective feature selection methods and data mining techniques. *Int. J. Eng. Technol. (IJET)* **7**(2), 583–598 (2015)
30. L.S. Dhamande, M.B. Chaudhari, Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. *Measurement* **125**, 63–77 (2018)
31. L. Gelman, T.H. Patel, B. Murray, A. Thomson, Rolling bearing diagnosis based on the higher order spectra. *Int. J. Prog. Health Manag.* **022** (2013) (ISSN 2153-2648)
32. M. Hamadache, D. Lee, Principal component analysis based signal-to-noise ratio improvement for inchoate faulty signals: application to ball bearing fault detection. *Int. J. Control Autom. Syst.* **15**(2), 506–517 (2017)
33. J. Lin, Q. Chen, Fault diagnosis of rolling bearings based on multifractal detrended fluctuation analysis and Mahalanobis distance criterion. *Mech. Syst. Signal Proc.* **38**(2), 515–533 (2013)
34. P.H. Nguyen, J.M. Kim, Multifault diagnosis of rolling element bearings using a wavelet kurtogram and vector median-based feature analysis. *Shock Vib.* **215**, 14 (2015). **Article ID 320508**
35. W. Li, M. Qiu, Z. Zhu, B. Wu, G. Zhou, Bearing fault diagnosis based on spectrum images of vibration signals. *Meas. Sci. Technol.* **27**(035005), 10 (2016)
36. B. Attaran, A. Ghanbarzadeh, Bearing Fault Detection Based on Maximum Likelihood Estimation and Optimized ANN Using the Bees Algorithm. *Journal of Applied and Computational Mechanics* **1**(1), 35–43 (2015)
37. M. Hamadache, Rotor speed based bearing fault diagnosis using absolute value PCA, *PhD Thesis, School of electronics Engineering, Kyungpook National University*, (2015), pp. 50–54
38. G. Georgoulas, G. Nikolakopoulos, Bearing fault detection and diagnosis by fusing vibration data. in *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, IECON 2016—42nd Annual Conference of the IEEE*, (2016), pp. 6955–6960
39. J. Harmouche, C. Delpha, D. Diallo, Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals. *IEEE Trans. Energy Convers.* **30**(1), 376–383 (2015)
40. J. Park, M. Hamadache, J.M. Ha, Y. Kim, K. Na, B.D. Youn, A positive energy residual (per) based planetary gear fault detection method under variable speed conditions. *Mechanical Systems and Signal Processing* **117**, 347–360 (2019)
41. J.M. Ha, J. Park, K. Na, Y. Kim, B.D. Youn, Toothwise fault identification for a planetary gearbox based on a health data map. *IEEE Trans. on Ind. Electronics* **65**(7), 5903–5912 (2018)
42. J.H. Jung, B.C. Jeon, B.D. Youn, M. Kim, D. Kim, Y. Kim, Omnidirectional regeneration (ODR) of proximity sensor signals for robust diagnosis of journal bearing systems. *Mechanical Systems and Signal Processing* **90**, 189–207 (2017)
43. C. Hu, P. Wang, B.D. Youn, W. Lee, J.T. Yoon, Copula-based statistical health grade system against mechanical faults of power transformers. *IEEE Trans. Power Deliv.* **27**(4), 1809–1819 (2012)
44. B.D. Youn, B.C. Jeon, J.H. Jung, Apparatus and method for diagnosing rotor shaft. *US Patent App.* 15/239,987, June 2017
45. J.M. Ha, H. Oh, J. Park, B.D. Youn, Classification of operating conditions of wind turbines for a class-wise condition monitoring strategy. *Renewable Energy* **103**, 594–605 (2017)
46. W. Zhou, T.G. Habetler, R.G. Harley, Bearing fault detection via stator current noise cancellation and statistical control. *IEEE Trans. Ind. Elect.* **55**(12), 4260–4269 (2008)
47. H. Zoubek, S. Villwock, M. Pacas, Frequency response analysis for rolling-bearing damage diagnosis. *IEEE Trans. on Ind. Elect.* **55**(12), 4270–4276 (2008)
48. M. Kang, J. Kim, L.M. Wills, J.M. Kim, Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis. *IEEE Trans. on Ind. Elect.* **62**(12), 7749–7761 (2015)
49. F. Zhang, T. Zhang, H. Yu, A Novel rolling bearing fault diagnosis method. in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (2016) pp. 1148–1152
50. D. Rossetti, Y. Zhang, S. Squartini, S. Collura, Classification of bearing faults through time-frequency analysis and image processing. in *2016 17th International Conference on Mechatronics-Mechatronika (ME)*
51. Z. Huo, Y. Zhang, P. Francq, L. Shu, J. Huang, Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures. *IEEE Access* **5**, 19442–19456 (2017)
52. A.A. Krishnamurthy, M.N. Belur, D. Chakraborty, Comparison of various linear discriminant analysis techniques for fault diagnosis of Re-usable Launch Vehicle. in *2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, Orlando, FL, USA, pp. 3050–3055 (2011)

53. J. Harmouche, C. Delpha, D. Diallo, Linear discriminant analysis for the discrimination of faults in bearing balls by using spectral features. in *2014 First International Conference on Green Energy ICGE 2014*, (2014), pp. 182–187
54. T. Liu, J. Chen, X.N. Zhou, W.B. Xiao, Bearing performance degradation assessment using linear discriminant analysis and coupled HMM. *J. Phys: Conf. Ser.* **364**(012028), 12 (2012)
55. M. Zhao, X. Jin, Z. Zhang, B. Li, Fault diagnosis of rolling element bearings via discriminative subspace learning: visualization and classification. *Expert Syst. Appl.* **41**, 3391–3401 (2014)
56. X. Jin, M. Zhao, T.W.S. Chow, M. Pecht, Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Tran. on Ind. Electr.* **61**(5), 2441–2451 (2014)
57. L. Ciabattini, G. Cimini, F. Ferracuti, A. Freddi, G. Ippoliti, A. Monteriu, A novel LDA-based approach for motor bearing fault detection. in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN) (IEEE, 2015)*, pp. 771–776
58. C.P. Mbo'o, K. Hameyer, Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection. *IEEE Trans. Ind. Appl.* **52**(5), 3861–3868 (2016)
59. W. Yan, H. Shao, Application of support vector machine nonlinear classifier to fault diagnoses. in *Proceedings of the 4th World Congress on Intelligent Control and Automation, 2002*, vol. 4, (IEEE, 2002), pp. 2697–2700
60. V. Sugumaran, K.I. Ramachandran, Effect of number of features on classification of roller bearing faults using SVM and PSVM. *Expert Syst. Appl.* **38**(4), 4088–4096 (2011)
61. K.C. Gryllias, I. Ioannis, A. Antoniadis, A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Eng. Appl. Artif. Intell.* **25**(2), 326–344 (2012)
62. D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, Automatic bearing fault diagnosis based on one-class ν -SVM. *Comput. Ind. Eng.* **64**(1), 357–365 (2013)
63. G. Wang, Y. He, K. He, Multi-layer kernel learning method faced on roller bearing fault diagnosis. *J. Softw.* **7**(7), 1531–1538 (2012)
64. X.M. Liu, J.W. Yin, Z.L. Feng, J. Dong, Incremental manifold learning via tangent space alignment. in *Artificial Neural Networks in Pattern Recognition*, (Ulm, Germany, 2006), pp. 107–121
65. X. Li, A. Zheng, X. Zhang, C. Li, L. Zhang, Rolling element bearing fault detection using support vector machine with improved ant colony optimization. *Measurement* **46**(8), 2726–2734 (2013)
66. D. Hwang, Y. Youn, J. Sun, K. Choi, J. Lee, Y. Kim, Support vector machine based bearing fault diagnosis for induction motors using vibration signals. *J. Electr. Eng. Technol.* **10**, 30–40 (2015)
67. R. Liua, B. Yang, X. Zhang, S. Wang, X. Chen, Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *Mech. Syst. Signal Proc.* **75**, 345–370 (2016)
68. Y. Li, M. Xu, Y. Wei, W. Huang, A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* **77**, 80–94 (2016)
69. M.M. Manjurul Islam, J. Kim, S.A. Khan, J. Kima, Reliable bearing fault diagnosis using Bayesian inference-based multi-class support vector machines. *J. Acoust. Soc. Am.* **141**(2), 7 (2017)
70. N. Zhang, L. Wu, J. Yang, Y. Guan, Naive bayes bearing fault diagnosis based on enhanced independence of data. *Sensors* **18**(463), 17 (2018)
71. M. Tabaszewski, Optimization of a nearest neighbors classifier for diagnosis of condition of rolling bearings. *Diagnostyka* **15**(1), 37–42 (2014)
72. M. Tabaszewski, Fault detection and diagnosis in low speed rolling element bearings Part II The use of nearest neighbour classification. *Mech. Syst. Signal Proc.* **6**(4), 309–316 (1992)
73. Y. Lei, Z. He, Y. Zi, A combination of WKNN to fault diagnosis of rolling element bearings. *J. Vib. Acoust.* **131**, 6 (2009)
74. A.B. Andre, E. Beltrame, J. Wainer, A combination of support vector machine and k-nearest neighbors for machine fault detection. *Applied Artificial Intelligence: An Int. J.* **27**(1), 36–49 (2013)
75. Q. Wang, Y. Liu, X. He, S. Liu, J. Liu, Fault diagnosis of bearing based on KPCA and KNN method. *Advanced Materials Research* **986–987**, 1491–1496 (2014)
76. S. Dong, X. Xu, R. Chen, Application of fuzzy C-means method and classification model of optimized K-nearest neighbor for fault diagnosis of bearing. *J. Braz. Soc. Mech. Sci. Eng.* **38**(8), 2255–2263 (2016)
77. P. Baraldi, F. Cannarile, F.D. Maio, E. Zio, Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. *Eng. App. of Artificial Intell.* **56**, 1–13 (2016)
78. R.K. Sharma, V. Sugumaran, H. Kumar, M. Amarnath, Condition monitoring of roller bearing by k-star classifier and k-nearest neighborhood classifier using sound signal. *SDHM Structural Durability and Health Monitoring* **12**(1), 1–16 (2017)
79. S. Dong, T. Luo, L. Zhong, L. Chen, X. Xu, Fault diagnosis of bearing based on the kernel principal component analysis and optimized k-nearest neighbour model. *J. Low Freq. Noise Vib. Active Control* **36**(4), 354–365 (2017)
80. G. Huang, Q. Zhu, C. Siew, Extreme learning machine: theory and applications. *Neurocomputing* **70**, 489–501 (2006)
81. R. Razavi-Far, M. Saif, Ensemble of extreme learning machines for diagnosing bearing defects in non-stationary environments under class imbalance condition. in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (2016)
82. G. Ditzler, R. Polikar, N. Chawla, An incremental learning algorithm for non-stationary environments and class imbalance. in *International Conference on Pattern Recognition (ICPR)* (2010), pp. 2997–3000
83. G. Ditzler, R. Polikar, Incremental learning of concept drift from streaming imbalanced data. *IEEE Trans. Knowl. Data Eng.* **25**(10), 2283–2301 (2013)
84. W. Mao, L. He, Y. Yan, J. Wang, Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine. *Mech. Syst. Signal Proc.* **83**, 450–473 (2017)
85. V. Sugumaran, K.I. Ramachandran, Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. *Mech. Syst. Signal Proc.* **21**(5), 2237–2247 (2007)
86. V. Sugumarana, K.I. Ramachandran, Fault diagnosis of roller bearing using fuzzy classifier and histogram features with focus on automatic rule learning. *Expert Syst. Appl.* **38**(5), 4901–4907 (2011)
87. J. Yu, Local and nonlocal preserving projection for bearing defect classification and performance assessment. *IEEE Trans. Ind. Electr.* **59**(5), 2363–2376 (2012)
88. P.K. Kankar, S.C. Sharma, S.P. Harsha, Rolling element bearing fault diagnosis using wavelet transform. *Neurocomputing* **74**(5), 1638–1645 (2011)
89. S. Cao, X. Ma, Y. Zhang, L. Luo, F. Yi, A fault diagnosis method based on semisupervised fuzzy c-means cluster analysis. *Inter. J. on Cyber. & Informatics (IJCI)* **4**(2), 281–289 (2015)
90. T. Han, D. Jiang, Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier. *Shock Vib* **216**, 11 (2016). **Article ID 5132046**
91. Y. Mohsenzadeh, H. Sheikhzadeh, A.M. Reza, N. Bathaee, M.M. Kalayeh, The relevance sample-feature machine: a sparse bayesian learning approach to joint feature-sample selection. *IEEE Trans. Cybern.* **43**(6), 2241–2254 (2013)

92. P.K. Wong, J. Zhong, Z. Yang, C.M. Vong, A new framework for intelligent simultaneous-fault diagnosis of rotating machinery using pairwise-coupled sparse Bayesian extreme learning committee machine. *Arch. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* 1989–1996 **203–210**, 16 (2016)
93. F. Shen, C. Chen, R. Yan, R.X. Gao, Bearing fault diagnosis based on SVD feature extraction and transfer learning classification. in *Prognostics and System Health Management Conference (PHM) (IEEE, 2015)*, pp. 1–6
94. S.J. Pan, Q. Yang, A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010)
95. B. Lei, L.Y. Soon, E.L. Tan, Robust SVD-based audio watermarking scheme with differential evolution optimization. *IEEE Trans. Audio Speech Lang. Process.* **21**(1), 2368–2378 (2013)
96. H.S. Seung, D.L. Daniel, The manifold ways of perception. *Science* **290**, 2268–2269 (2000)
97. S. Kadoury, M.D. Levine, Face detection in gray scale images using locally linear embeddings. *Comput. Vis. Image Underst.* **105**, 1–20 (2007)
98. X. Liu, D. Tosun, M.W. Weiner, N. Schuff, Locally linear embedding (LLE) for MRI based Alzheimer’s disease classification. *Neuroimage* **83**, 148–157 (2013)
99. K. Kima, J. Lee, Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognit.* **47**, 758–768 (2014)
100. J.H. Yang, J.W. Xu, D.B. Yang, Noise reduction method for non-linear time series based on principal manifold learning and its application to fault diagnosis. *Chin. J. Mech. Eng.* **42**, 154–158 (2006)
101. X. Wang, Y. Zheng, Z. Zhao, J. Wang, Bearing fault diagnosis based on statistical locally linear embedding. *Sensors* **15**, 16225–16247 (2015)
102. S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000)
103. Y. Wang, G. Xu, L. Liang, K. Jiang, Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **54–55**, 259–276 (2015)
104. A. Hertzmann, *Introduction to Bayesian Learning, Course Notes* (University of Toronto, Ontario, 2004)
105. M.R.G. Meireles, P.E.M. Almeida, M.G. Simões, A comprehensive review for industrial applicability of artificial neural networks. *IEEE Trans. Ind. Electr.* **50**(3), 585–601 (2003)
106. N. Qian, On the momentum term in gradient descent learning algorithms. *Neural Netw.* **12**(1), 145–151 (1999)
107. K.F. Al-Raheem, W. Abdul-Karem, Rolling bearing fault diagnostics using artificial neural networks based on Laplace wavelet analysis. *Int. J. Eng. Sci. Technol.* **2**(6), 278–290 (2010)
108. M. Nielsen, Chapter 6, *Neural Networks and Deep Learning* (2015)
109. A.T. Vemuri, M.M. Polycarpou, Neural-network-based robust fault diagnosis in robotic systems. *IEEE Trans. Neural Netw.* **8**(6), 1410–1420 (1997)
110. V.N. Ghate, S.V. Dudul, Cascade neural-network-based fault classifier for three-phase induction motor. *IEEE Trans. Ind. Electr.* **58**(5), 1555–1563 (2011)
111. S.S. Moosavi, A. Djerdir, Y. Ait-Amirat, D.A. Khaburi, A. N’Diaye, Artificial neural network-based fault diagnosis in the AC–DC converter of the power supply of series hybrid electric vehicle. *IET Electr. Syst. Transp.* **6**(2), 96–106 (2016)
112. B. Li, M. Chow, Y. Tipsuwan, J.C. Hung, Neural-network-based motor rolling bearing fault diagnosis. *IEEE Trans. Ind. Electr.* **47**(5), 1060–1069 (2000)
113. B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Sys. and Sig. Proc.* **17**(2), 317–328 (2003)
114. D.H. Pandya, S.H. Upadhyay, S.P. Harsha, “ANN based fault diagnosis of rolling element bearing using time-frequency domain feature,” *Int. J. Eng. Science and Technology (IJEST)* **4**(06), 2878–2886 (2012)
115. B. Samanta, K.R. Al-Balushi, S.A. Al-Araimi, Bearing fault detection using artificial neural networks and genetic algorithm. *J. on Applied Sig. Processing* **2004**(3), 366–377 (2004)
116. H. Yang, J. Mathew, L. Ma, V. Kosse, Matching pursuit feature based neural network pattern recognition of ball bearing faults. in *International Conference of Maintenance Societies (Australia, 2004)*, pp. 25–28
117. N. Gebraeel, M. Lawley, R. Liu, V. Parmeshwaran, Residual life predictions from vibration-based degradation signals: a neural network approach. *IEEE Trans. Ind. Electr.* **51**(3), 694–700 (2004)
118. V. Hariharan, P.S.S. Srinivasan, New approach of classification of rolling element bearing fault using artificial neural network. *J. Mech. Eng.* **40**(2), 119–130 (2009)
119. M. Delgado, G. Cirrincione, A.G. Espinosa, J.A. Ortega, H. Henao, Bearing faults detection by a novel condition monitoring scheme based on statistical-time features and neural networks. *IEEE Trans. Ind. Electr.* **60**(8), 3398–3407 (2013)
120. M. Unal, M. DEMetgul, M. Onat, H. Kucuk, Fault diagnosis of rolling bearing based on feature extraction and neural network algorithm. *Recent Adv. Telecom Signal Syst* 179–185 (2013)
121. S.S. Refaat, H. Abu-Rub, M.S. Saad, E.M. Aboul-Zahab, A. Iqbal, ANN-based for detection, diagnosis the bearing fault for three phase induction motors using current signal. in *2013 IEEE International Conference on Industrial Technology (ICIT)*, (2013), pp. 253–258
122. J.P. Patela, S.H. Upadhyayb, Comparison between artificial neural network and support vector method for a fault diagnostics in rolling element bearings. *Proc. Eng. 12th Int. Conf. Vib. Probl. ICOVP2015* **144**, 390–397 (2016)
123. D.K. Gaud, P. Jayaswal, Effects of artificial neural network parameters on rolling element bearing fault diagnosis. *Int. J. Curr Eng. Sci. Res.* **3**(1), 55–60 (2016)
124. N. Zhao, H. Zheng, L. Yang, Z. Wang, A fault diagnosis approach for rolling element bearing based on S-transform and artificial neural network. in *Proceedings of ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition GT2017*, USA, (2017)
125. R.G. Stockwell, L. Mansinha, R.P. Lowe, Localization of the complex spectrum: the S-transform. *IEEE Trans. Signal Process.* **44**(4), 998–1001 (1996)
126. J.B. Ali, L. Saidi, A. Mouelhi, B. Chebel-Morello, F. Fnaiech, Linear feature selection and classification using PNN and SFAM neural networks for an early online diagnosis of bearing naturally progressing degradations. *Eng. Appl. Artif. Intell.* **42**, 67–81 (2015)
127. A.A. Jaber, R. Bicker, Fault diagnosis of industrial robot bearings based on discrete wavelet transform and artificial neural network. *Int. J. Progn. Health Manag.* **017**, 13 (2016). **ISSN 2153-2648**
128. J. Zheng, H. Pan, J. Cheng, Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **85**, 746–759 (2017)
129. D. Yao, J. Yang, Y. Bai, X. Cheng, Railway rolling bearing fault diagnosis based on multi-scale intrinsic mode function permutation entropy and extreme learning machine classifier. *Adv. Mech. Eng.* **8**(10), 1–9 (2016)
130. Q. Tong, J. Cao, B. Han, X. Zhang, Z. Nie, J. Wang, Y. Lin, W. Zhang, A fault diagnosis approach for rolling element bearings based on RSGWPT-LCD bilayer screening and extreme learning machine. *IEEE Access* **5**, 5515–5530 (2017)

131. M. Liang, D. Su, D. Hu, M. Ge, A novel faults diagnosis method for rolling element bearings based on ELCD and extreme learning machine. *Shock Vib.* **218**, 10 (2018). **Article ID 1891453**
132. L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mech. Syst. Signal Process.* **16**(2–3), 373–390 (2002)
133. P. Jayaswal, S.N. Verma, A.K. Wadhvani, Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis. *J. Vib. Control* **17**(8), 1131–1148 (2011)
134. H.M. Ertunc, H. Ocak, C. Aliustaoglu, ANN- and ANFIS-based multi-staged decision algorithm for the detection and diagnosis of bearing faults. *Neural Comput. Appl.* **22**(1), S435–S446 (2013)
135. B.A. Paya, I.I. Esat, Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a pre-processor. *Mech. Syst. Signal Process.* **11**(5), 751–765 (1997)
136. Y. Yu, Y. Dejie, C. Junsheng, A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J. Sound Vib.* **294**, 269–277 (2006)
137. K.F. Al-Raheem, A. Roy, K.P. Ramachandran, D.K. Harrison, S. Grainger, Application of the laplace-wavelet combined with ANN for rolling bearing fault diagnosis. *J. Vib. Acoust.* **130**, 9 (2008)
138. Y. Hwang, K. Jen, Y. Shen, Application of cepstrum and neural network to bearing fault detection. *J. Mech. Sci. Technol.* **23**, 2730–2737 (2009)
139. K. Al-Raheem, Wavelet analysis and neural networks for bearing fault diagnosis. *Advances in Wavelet Theory and Their Applications in Eng., Physics and Technology*, (2012), pp. 313–352
140. J.B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, F. Fnaiech, Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mech. Syst. Signal. Process.* **56–57**, 150–172 (2015)
141. J.B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, F. Fnaiech, Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **89**, 16–27 (2015)
142. R. Dubey, D. Agrawal, Bearing fault classification using ANN-based Hilbert footprint analysis. *IET Sci. Meas. Technol.* **9**(8), 1016–1022 (2015)
143. Q. Hu, Z. He, Z. Zhang, Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mech. Syst. Signal Proc. ess.* **21**, 688–705 (2007)
144. J. Yang, Y. Zhang, Y. Zhu, Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mech. Syst. Signal Process.* **21**, 2012–2024 (2007)
145. L. Guo, J. Chen, X. Li, Rolling bearing fault classification based on envelope spectrum and support vector machine. *J. Vib. Control* **15**(9), 1349–1363 (2009)
146. P. Konar, P. Chattopadhyay, Bearing fault detection of induction motor using wavelet and support vector machines (SVMs). *Appl. Soft Comput.* **11**, 4203–4211 (2011)
147. S. Wu, P. Wu, C. Wu, J. Ding, C. Wang, Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **14**, 1343–1356 (2012)
148. Z. Liu, H. Cao, X. Chen, Z. He, Z. Shen, Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing* **99**, 399–410 (2013)
149. X. Zhang, Y. Liang, J. Zhou, Y. Zang, A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement* **69**, 164–179 (2015)
150. L. Saidi, J.B. Ali, F. Fnaiech, Application of higher order spectral features and support vector machines for bearing faults classification. *ISA Trans.* **54**, 193–206 (2015)
151. Y. Li, M. Xu, H. Zhao, W. Huang, Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis. *Mech. Mach. Theory* **98**, 114–132 (2016)
152. J. Tian, C. Morillo, M.H. Azarian, M. Pecht, Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with k-nearest neighbor distance analysis. *IEEE Trans. Ind. Electr.* **63**, 3 (2016)
153. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**(7553), 436–444 (2015)
154. I. Goodfellow, Y. Bengio, A. Courville, Deep learning. MIT Press, www.deeplearningbook.org (2016)
155. W. Yan, L. Yu, On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach. in *Annual Conference of The Prognostics and Health Management Society 2015*, vol. 6, 2015
156. H. Dong, L. Yang, H. Li, Small fault diagnosis of front-end speed controlled wind generator based on deep learning. *WSEAS Trans. Circ. Syst.* **15**, 64–72 (2016)
157. F. Lv, C. Wen, Z. Bao, M. Liu, Fault diagnosis based on deep learning. *2016 American Control Conference (ACC)*. Boston Marriott Copley Place, Boston, MA, USA, July 6–8, (2016)
158. H. Liu, C. Liu, Y. Huang, Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mech. Syst. Signal Process.* **25**(2), 558–574 (2011)
159. N.K. Verma, V.K. Gupta, M. Sharma, R.K. Sevakula, Intelligent condition based monitoring of rotating machines using sparse autoencoders. in *Proceedings of IEEE Conference on Prognostics and Health Management*, (Gaithersburg, 2013) pp. 1–7, June 24–27
160. S. Min, B. Lee, S. Yoo, Deep learning in bioinformatics. *Briefings Bioinf* **18**, 851–869 (2017)
161. D. Lee, V. Siu, R. Cruz, C. Yetman, Convolutional neural net and bearing fault analysis. in *Proceedings of the International Conference on Data Mining (DMIN'16)*, (2016), pp. 194–200
162. X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolutional neural network and its application to bearing fault diagnosis. *Measurement* **93**, 490–502 (2016)
163. X. Ding, Q. He, Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Trans. Inst. Meas* **66**(8), 1926–1935 (2017)
164. O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, Convolutional neural network based fault detection for rotating machinery. *J. Sound Vib.* **377**, 331–345 (2016)
165. W. You, C. Shen, X. Guo, Z. Zhu, Bearing fault diagnosis using convolution neural network and support vector regression. in *2017 International Conference on Mech. Engineering and Cont. Automation*, (2017), pp. 6–11
166. W. You, C. Shen, X. Guo, X. Jiang, J. Shi, Z. Zhu, A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery. *Adv. Mech. Eng.* **9**(6), 1–17 (2017)
167. W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **17**(425), 1–21 (2017)
168. W. Fuan, J. Hongkai, S. Haidong, D. Wenjing, W. Shuaipeng, An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Meas. Sci. Technol.* **28**(9), 1–25 (2017)
169. S. Li, G. Liu, X. Tang, J. Lu, J. Hu, An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis. *Sensors* **17**(1729), 1–19 (2017)
170. Y. Xie, T. Zhang, Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition. *Shock Vib.* **2017**, 12 (2017)

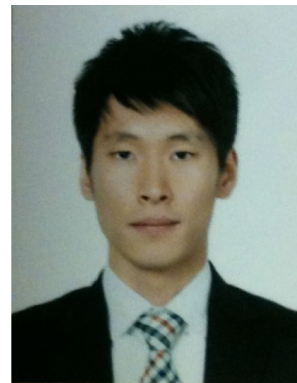
171. C. Lu, Z. Wang, B. Zhou, Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Adv. Eng. Inf.* **32**, 139–151 (2017)
172. R. Socher, C.C. Lin, A.Y. Ng, C.D. Manning, Parsing natural scenes and natural language with recursive neural networks. *The 28th International Conference on Machine Learning (ICML 2011)*, (2011)
173. K. Cho, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation. in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Doha, Qatar, October 25–29, 2014), pp. 1724–1734
174. R. Dey, F.M. Salem, Gate-variants of gated recurrent unit (GRU) neural networks. in *IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017 (2017), pp. 5
175. W. Abed, S. Sharma, R. Sutton, A. Motwani, A robust bearing fault detection and diagnosis technique for brushless DC motors under non-stationary operating conditions. *J. Control Autom. Electr. Syst.* **26**, 14 (2015)
176. A. Malhi, R. Yan, R.X. Gao, Prognosis of defect propagation based on recurrent neural networks. in *IEEE Transaction on Instrumentation and Measurement*, (vol. 60, no. 3, March 2011)
177. S. Sharma, W. Abed, R. Sutton, B. Subudhi, Corrosion fault diagnosis of rolling element bearing under constant and variable load and speed conditions. *IFAC-PapersOnLine* **48–30**, 049–054 (2015)
178. Y. Xie, T. Zhang, The application of echo state network and recurrent multilayer perceptron in rotating machinery fault prognosis. in *Proceedings of 2016 IEEE Chinese Guidance, Navigation and Control Conference*, (China, 2016), pp. 2286–2291
179. L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **240**, 98–109 (2017)
180. Q. Cui, Z. Li, J. Yang, B. Liang, Rolling bearing fault prognosis using recurrent neural network. in *29th Chinese Control And Decision Conference (CCDC)*, (2017), pp. 1196–1201
181. G.E. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets. *Neural Comput.* **18**, 16 (2006)
182. G. Hinton, Deep belief nets. *Encycl. Mach. Learn.* **4**, 5947 (2010)
183. R. Salakhutdinov, G. Hinton, Deep boltzmann machines. in *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics 2009*, (Florida, USA, vol. 5 of JMLR: W&CP 2009), p. 5
184. T. Jie, L. Yi-Lun, Y. Da-Lian, T. Fang, L. Chi, Fault diagnosis of rolling bearing using deep belief networks. in *International Symposium on Material, Energy and Environment Engineering*, (2015), pp. 566–569
185. H. Shao, H. Jiang, X. Zhang, M. Niu, Rolling bearing fault diagnosis using an optimization deep belief network. *Meas. Sci. Technol.* **26**(115002), 17 (2015)
186. X. Wang, Y. Li, T. Rui, H. Zhu, J. Fei, Bearing fault diagnosis method based on Hilbert envelope spectrum and deep belief network. *J. Vibroeng.* **17**(3), 1295–1308 (2015)
187. R. Zhang, L. Wu, X. Fu, B. Yao, Classification of bearing data based on deep belief networks. in *Prognostics and System Health Management Conference (PHM-Chengdu)*, (2016), pp. 1–6
188. M. Ma, X. Chen, S. Wang, Y. Liu, W. Li, Bearing degradation assessment based on weibull distribution and deep belief network. in *2016 Internatinal Symposium on Flexible Automat.*, (Ohio, U.S.A., 2016), pp. 1–4
189. Y. Liu, D. Yang, Bearing fault diagnosis based on deep belief network and multisensor information fusion. *Shock Vib.* **216**, 9 (2016). (Article ID 9306205)
190. A. Yin, J. Lu, Z. Dai, J. Li, Q. Ouyang, Isomap and deep belief network-based machine health health combined assessment model. *Strojniški vestnik J. Mech. Eng.* **62**(12), 740–750 (2016)
191. M. Gan, C. Wang, C. Zhu, Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mech. Syst. Signal Process.* **72–73**, 92–104 (2016)
192. J. Deutsch, M. He, D. He, Remaining useful life prediction of hybrid ceramic bearings using an integrated deep learning and particle filter approach. *Appl. Sci.* **7**(649), 17 (2017)
193. S. Devendiran, K. Manivannan, S.C. Kamani, R. Refai, An early bearing fault diagnosis using effective feature selection methods and data mining techniques. *Int. J. Eng. Technol. (IJET)* **7**(2), 583–598 (2015)
194. R. Zhang, Z. Peng, L. Wu, B. Yao, Y. Guan, Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence. *Sensors* **17**(549), 17 (2017)
195. J. Deutsch, D. He, Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Trans. Syst. Man Cybern. Syst.* **48**(1), 11–20 (2018)
196. H. Shao, H. Jiang, H. Zhang, T. Liang, Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network. *IEEE Trans. Ind. Electr.* **65**(3), 2727–2736 (2018)
197. H. Oh, J.H. Jung, B.C. Jeon, B.D. Youn, Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis. *IEEE Trans. Ind. Electr.* **65**(4), 3539–3549 (2018)
198. S. Deng, Z. Cheng, C. Li, X. Yao, Z. Chen, R.V. Sanchez, Rolling bearing fault diagnosis based on deep boltzmann machines. in *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, (2016), pp. 19–21
199. L. Liao, W. Jin, R. Pavel, Enhanced restricted boltzmann machine with prognosability regularization for prognostics and health assessment. *IEEE Trans. Ind. Electr.* **63**(11), 7076–7083 (2016)
200. X. He, D. Wang, Y. Li, C. Zhou, A novel bearing fault diagnosis method based on gaussian restricted boltzmann machine. *Math. Probl. Eng.* **216**, 8 (2016). Article ID 2957083
201. K.H. Cho, A. Ilin, T. Raiko, Improved learning of gaussian-bernoulli restricted boltzmann machines. in *Artificial Neural Networks and Machine Learning—ICANN 2011*, (Springer Berlin Heidelberg: Berlin, Germany, vol. 6791), pp. 10–17
202. C. Li, R. Sánchez, G. Zurita, M. Cerrada, D. Cabrera, Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. *Sensors* **16**(895), 19 (2016)
203. J. Deutsch, D. He, Using deep learning based approaches for bearing remaining useful life prediction. in *Annual Conference of the Prognostics and Health Management Society*, (2016)
204. G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006)
205. W. Lu, X. Wang, C. Yang, T. Zhang, A novel feature extraction method using deep neural network for rolling bearing fault diagnosis. in *The 27th Chinese Control and Decision Conference (2015 CCDC)*. (IEEE, 2015) pp. 2427–2431
206. F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **72**, 303–315 (2016)
207. T. Junbo, L. Weining, A. Juneng, W. Xueqian, Fault diagnosis method study in roller bearing based on wavelet transform and stacked autoencoder. in *The 27th Chinese Control and Decision Conference (2015 CCDC)* (IEEE, 2015), pp. 4608–4613
208. W. Zhao, C. Lu, J. Ma, Z. Wang, A deep learning method using SDA combined with dropout for bearing fault diagnosis. *Vibroeng. Proc.* **5**(151), 156 (2015)
209. H. O. A. Ahmed, M. L. Dennis Wong, and A. K. Nandi, Effects of deep neural network parameters on classification of the bearing faults. in *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, (2016), pp. 6329–6334
210. L. Guo, H. Gao, H. Huang, X. He, S. Li, Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring. *Shock Vib* **216**, 10 (2016). (Article ID 4632562)

211. S. Tao, T. Zhang, J. Yang, X. Wang, W. Lu, Bearing fault diagnosis method based on stacked autoencoder and softmax regression. in *Control Conference (CCC)*, 2015 34th Chinese. (IEEE, 2015), pp. 6331–6335
212. H. Liu, L. Li, J. Ma, Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock Vib.* **2016**, 12 (2016). (Article ID 6127479)
213. W. Mao, J. He, Y. Li, Y. Yan, Bearing fault diagnosis with autoencoder extreme learning machine: a comparative study. *Proc. Mech. E Part C J. Mech. Eng. Sci.* **231**, 1560–1578 (2016)
214. R. Thirukovalluru, S. Dixit, R. K. Sevakula, N. K. Verma, and A. Salour, Generating feature sets for fault diagnosis using denoising stacked autoencoder. in *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)* (IEEE, 2016), pp. 1–7, 2016
215. X. Guo, C. Shen, L. Chen, Deep fault recognizer: an integrated model to denoise and extract features for fault diagnosis in rotating machinery. *Appl. Sci.* **7**(41), 1–17 (2017)
216. Z. Chen, W. Li, Multi-sensor feature fusion for bearing fault diagnosis using sparse auto encoder and deep belief network. *IEEE Trans. Instr. Meas.* **66**(7), 1693–1702 (2017)
217. C. Lu, Z. Wang, W. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process.* **130**, 377–388 (2017)
218. R. M. Hasani, G. Wang, R. Grosu, An automated autoencoder correlation-based health-monitoring and prognostic method for machine bearings. [arXiv:1703.06272v1](https://arxiv.org/abs/1703.06272v1) [cs.LG], 2017
219. M. Sohaib, C. Kim, J. Kim, A hybrid feature model and deep-learning-based bearing fault diagnosis. *Sensors* **17**(2876), 1–16 (2017)
220. H. Shao, H. Jiang, F. Wang, H. Zhao, An enhancement deep feature fusion method for rotating machinery fault diagnosis. *Knowl Based Syst* **119**, 200–220 (2017)
221. A. Shaheryar, X. Yin, W.Y. Ramay, Deep-learning framework: an application for fault identification in rotary machines. *Int. J. Comput. Appl.* (0975–8887) **167**(4), 37–45 (2017)
222. J. Sun, C. Yan, J. Wen, Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Trans. Instr. Meas.* **67**(1), 185–195 (2018)
223. H.O.A. Ahmed, M.L.D. Wong, A.K. Nandi, Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features. *Mech. Syst. Signal Process.* **99**, 459–477 (2018)
224. D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G. Yang, Deep learning for health informatics. *IEEE J. Biomed. Health Inform.* **21**(1), 4–21 (2017)
225. D.L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006)



Moussa Hamadache received his Engineer of State degree in Automation of Industrial Processes (summa cum laude) in 2008 from the University of M'hamed-Bougara Boumerdes, Algeria, and the Ph.D. degree in Electrical and Electronic Engineering in 2015 from the Kyungpook National University, Republic of Korea. He has worked as a postdoctoral research associate at the same University. He also worked as a visiting researcher at the Seoul National University (SNU),

Republic of Korea. In addition, he worked as a principal researcher (PI) in OnePredict Inc., a SNU startup, from December 2016 to May 2017. Plus, he worked as a postdoctoral research fellow at the University of Ferrara, Italy. He is currently a research fellow in Mechatronic Systems for Railways at the University of Birmingham, Department of Electronic, Electrical and Systems Engineering, Birmingham Center for Railway Research and Education (BCRRE). His scientific activity is mainly in the field of advanced system safety engineering with reference to condition monitoring, diagnostics, and experimental vibration measurements, prognostic and health management (PHM). He is the editorial board member of the automation, control and intelligent systems, science publishing group journal. He is also a reviewer for several important journals and international conferences in the field of system safety engineering and vibration. He has received several awards including three best student and best representation awards at several international conferences.



Joon Ha Jung received his B.S. degree in mechanical and aerospace engineering from Seoul National University (SNU), Seoul, Republic of Korea, in 2012. He received Ph.D. degree in mechanical engineering from SNU in 2019. His research area includes prognostics and health management for machine tools, and machine learning-based diagnostics for rotating machineries. He currently works as a senior researcher in Korea Institute of Machine and Materials.



Jungho Park received his B.S. degree in mechanical and aerospace engineering from Seoul National University, Seoul, South Korea in 2012, where he is currently working toward the Ph.D. degree. His research interest focuses on prognostics and health management for rotating machinery. Mr. Park is the recipient of the PHM Society Data Challenge Competition Winner in 2014 and 2015, and PHMAP 2017 Data Challenge.



Byeng D. Youn is a Full Professor in the Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul, South Korea. Dr. Youn received the Ph.D. degree in mechanical engineering from the University of Iowa, Iowa City, IA, USA, in 2001. Dr. Youn was the recipient of ASME IDETC Best Paper Awards (2001 and 2008), the ISSMO/Springer Prize for a Young Scientist (2005), the IEEE PHM Competition Winner (2014), the PHM Society Data Challenge Competition Winner (2014, 2015, 2017), etc.