Measurement 245 (2025) 116589

Contents lists available at ScienceDirect

Measurement

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

# Feature learning for bearing prognostics: A comprehensive review of machine/deep learning methods, challenges, and opportunities

Ahmed Ayman [a,*], Ahmed Onsy [a], Omneya Attallah [b,c], Hadley Brooks [a], Iman Morsi [b]

[a] School of Engineering and Computing, University of Central Lancashire, Preston PR1 2HE, UK
[b] College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, 21937, Egypt
[c] Wearables, Biosensing, and Biosignal Processing Laboratory, Arab Academy for Science, Technology and Maritime Transport, Alexandria 21937, Egypt

ARTICLE INFO

*Keywords:*
Bearing prognostics
Condition monitoring
Deep learning
Feature learning
Machine learning
Spatiotemporal representation

ABSTRACT

Mechanical bearings are common elements in a wide range of applications, such as wind turbines and manufacturing. Therefore, bearing prognostics are crucial to preventing catastrophic failures and machinery breakdowns. In this context, extracting the influential features is often the most challenging task in the prognosis process. This complexity arises because of the non-linear and non-stationary nature of the acquired vibration signals. Therefore, this paper offers an extensive examination of state-of-the-art feature-learning methods. Initially, the paper introduces a taxonomy of feature learning methods, encompassing both shallow and deep learning approaches. The paper also discusses methods of feature-learning under imbalanced data samples and different operational settings. Furthermore, the paper details the experimental setups of commonly used benchmark datasets to assist scholars and practitioners in understanding the subject area. Finally, the study discusses the challenges associated with calculating bearings' RUL and suggests potential areas for further research.

## 1. Introduction

Bearings are among the most common components in rotating machinery such as gearboxes, wind turbines, and vehicles [1]. A bearing is a mechanical component that supports the load and enables smooth movement between two parts, like the shaft and the housing, with minimal resistance [2]. Bearing degradation, which necessitates timely maintenance, is one of the most frequent causes of operational disruptions in rotating systems [3]. Statistical data shows that bearings account for approximately 45–55 % of operational disruptions in rotating machinery, in contrast to 41 % for motor faults and 10 % for rotor faults [4]. As a result, bearings have been a focus of research in the machinery prognostics field compared to other components or systems [5–7].

Standards such as ISO (ISO 281, ISO 76), and ABMA Standard 9 all try to estimate the remaining useful life (RUL) of types of bearings such as rolling-element bearings (REB) and ball bearings (BEB). The RUL is defined as the time remaining before maintenance should be performed [8]. ISO 281 outlines a systematic approach for determining the remaining operational lifespan of bearings at a 10 % failure rate. According to this standard, a bearing is considered defective if any part of it, such as the outer race, inner race, rollers, balls, or cage, exhibits signs of fatigue [9]. The standard determines the rating life of bearings by considering two types of loads: radial load and axial load, using equations (1–3), where:

$$L_{10} = \left(\frac{C_r}{P_r}\right)^e \tag{1}$$

$$L_{10} = \left(\frac{C_a}{P_a}\right)^e \tag{2}$$

$$L_{h10} = \left(\frac{10^6}{60{*}n}\right){*}L_{10} \tag{3}$$

Where $L_{h10}$ stands for the rating life in hours, given e = 3 if the bearing is a ball bearing or e = $\frac{10}{3}$ for roller bearings. The variable $L_{10}$ represents the fundamental rating life, which denotes the number of revolutions a component is expected to last with a 10 % likelihood of failure, n is measured in millions as it stands for the number of revolutions. Finally, $\frac{C_a}{P_a}$ is the case if the load is axial, whereas $\frac{C_r}{P_r}$ if the load is radial.

From the given formulas, it is clear that the life rating is based on some stationary conditions. Meanwhile, in real-world industries, it is

* Corresponding author.
*E-mail address:* aaabdel-aal@uclan.ac.uk (A. Ayman).

https://doi.org/10.1016/j.measurement.2024.116589
Received 20 July 2024; Received in revised form 19 December 2024; Accepted 23 December 2024
Available online 26 December 2024

common for systems to exhibit varying functioning conditions as a result of challenging operational situations. For instance, wind turbines can be installed in deserts, seas, or elevated areas, leading to variations in wind speed, humidity, and temperature across different sites. Additionally, factors such as lubrication, contamination, and temperature conditions, which are not accounted for, directly influence the RUL of the bearing and consequently impact the performance of the machinery. Therefore, it is essential to establish a dependable estimation of the RUL of bearings considering operational conditions.

Essentially, there are two main methods for bearing RUL predictions: a model-based method and a data-driven method. A model-based method attempts to mathematically describe the degradation pattern. Kalman filters [10], particle filters [11–13] and similar techniques are used to develop a mathematical representation of the degradation pattern [14,15]. As a result, these methods rely heavily on manual work and linear degradation patterns to establish dependable frameworks. Therefore, successful maintenance and accurate RUL calculations rely on expert experience and understanding of various failure modes. Yet, in complex and modern industries, achieving this, is difficult because the degradation pattern relies on various non-linear factors.

Conversely, data-driven approaches seek to estimate the optimal timing for bearing maintenance by considering its expected operational lifespan, while minimizing reliance on human intervention. Data-driven methods can be classified into two main categories: machine learning (ML) and deep learning (DL). ML techniques require more data preparation stages before modelling, unlike DL techniques. This is because ML techniques may not be able to identify non-linear relationships or efficiently process raw non-stationary and non-linear signal characteristics associated with a target bearing. On the contrary, DL techniques aim to automatically extract the latent characteristics within the obtained signals without manual involvement. However, recent research has demonstrated that fusion models, which integrate data processing stages with DL modelling, can accelerate the convergence of DL models and enhance the robustness of the regression model. Fig. 1 illustrates the different stages of model-based and data-driven methods, including ML approaches, DL approaches, and fusion approaches. Meanwhile, Fig. 2 names some of the common techniques of each of the model-based and data-driven methods, including ML and DL approaches.

The study of vibrations has demonstrated its effectiveness in accurately assessing the state of bearings [16–19]. Yet, the estimation of bearing lifespan is considered challenging and open area of research due to several reasons. First, the vibration signals from bearings exhibit a low signal-to-noise ratio (SNR) and non-constant behaviour [20]. In normal operations, a bearing is influenced by other components within the same system. This interference affects the target bearing and needs to be removed to effectively conduct predictive maintenance on the bearing. Therefore, it is crucial to recognise essential bearing features and eliminate any potential noise assumptions. Second is the lack of availability of different failure modes. This is because it is expensive and time-consuming to run industrial equipment until it fails to gather data. Additionally, safety risks may arise in this situation [21,22]. Third, the limited reliability of data-driven prediction methods in the state-of-the-art, which may be attributed to the variations in the distributions of training and testing data across different operational settings. Having consistent distributions of training and testing data suggests that each operational setting should have its own training model, which can be challenging to achieve.

Given that a prognostics and health management (PHM) approach has four stages: data acquisition, data preparation, feature extraction, and RUL calculation, the feature extraction stage is considered the most crucial. Robust techniques for feature learning can enhance the precision of the prognostic approach while also resolving the aforementioned concerns. On the other hand, the poorer the representation, the lower the accuracy of the estimated life. Therefore, the purpose of this study is to review the feature learning methods introduced and examine the progress made in addressing each of the three prognostic difficulties identified. Several review studies have been conducted on bearing prognostics. The authors in [23] conducted a study on the applications of predictive maintenance in different fields, including bearings, aircrafts, and batteries. In [2], Jammu et al. focused on data acquisition methods for bearing prognostics, including vibration and acoustic emission [24]. However, Mbagaya et al. [25], briefly discussed some of the common statistical approaches for bearing prognostics. Authors in [26] summarised the theoretical background of the common data-driven techniques for PHM applications. Chen et al. [27] discussed the challenges of cross-domain prognostics. Kordestani et al. [8] conducted a study to review the methods of failure prognosis, which is the process of predicting the fault state of a malfunctioning component.

In contrast to others, this study focuses on feature-learning methods for bearing prognostics. The study aims to review and summarise the state-of-the-art feature learning methods. The paper provides a review of the common traditional feature learning methods, as well as the methods that leverage the capabilities of deep neural networks (NN) to learn non-linear and non-stationary trends of bearing signals.

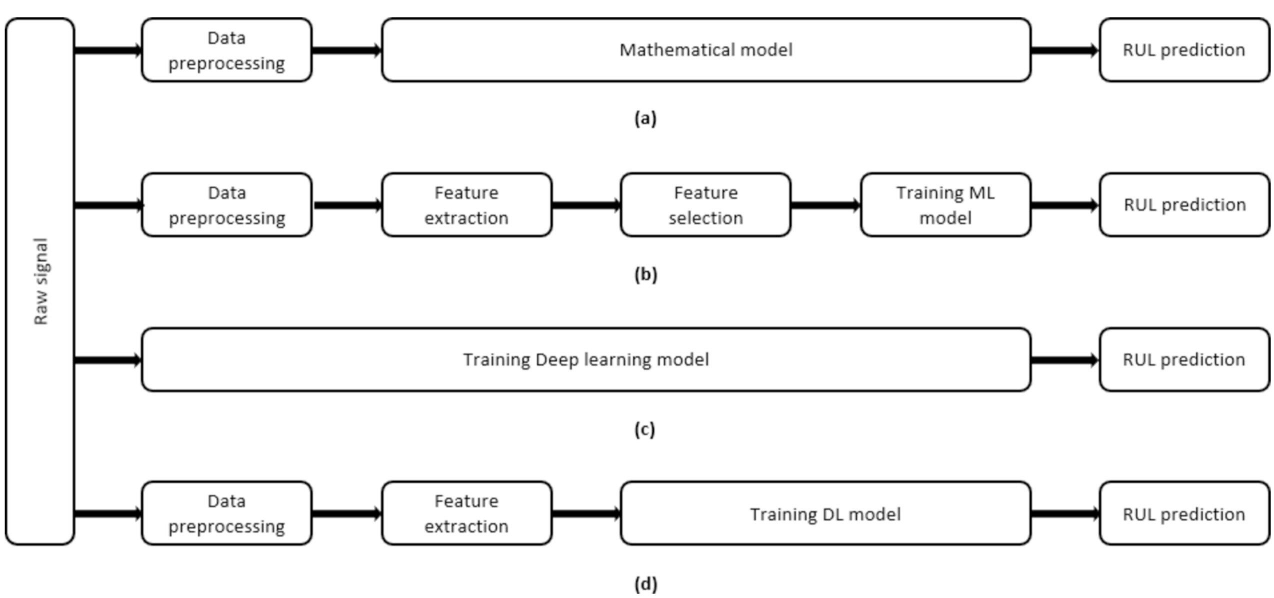


**Fig. 1.** Bearing prognostics frameworks; (a) physical methods, (b) ML methods, (c) DL methods, and (d) fusion methods.

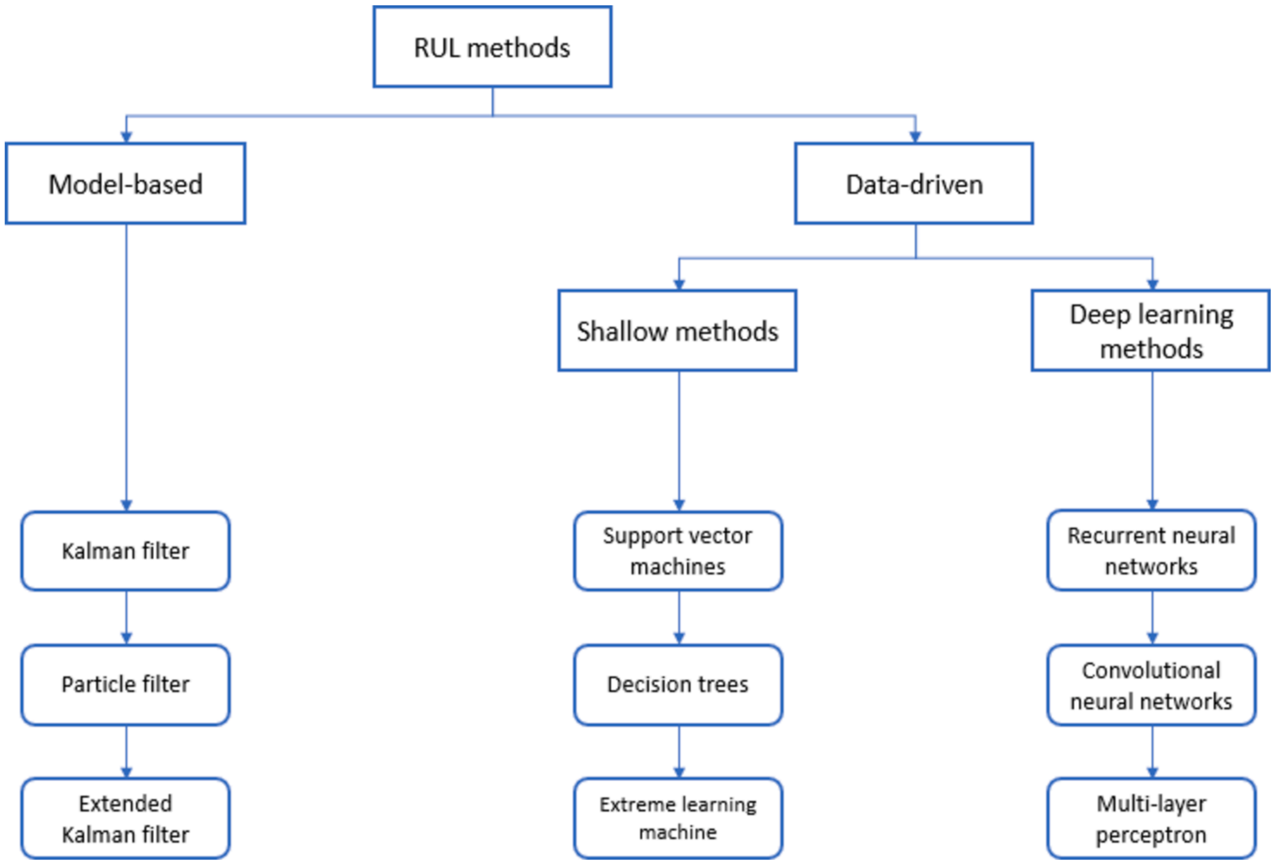


**Fig. 2.** Common model-based and data-driven techniques.

Furthermore, the study reviews and discusses feature learning methods designed to address the scarcity of fault-bearing samples, particularly those with an imbalanced data sample ratio. Additionally, the authors categorize and provide a comprehensive analysis of feature learning approaches under different data distributions between the training and testing data, which result from different operational settings. Furthermore, the authors discuss fusion models for RUL predictions, as they have proven sufficient in recent years. Afterwards, the study discusses the physical setups and data acquisition methods of the common experimental approaches in the state-of-the-art. Finally, the paper outlines current gaps and proposes potential future directions to benefit practitioners and emerging researchers in the field.

The contribution of this study can be listed as follows:

1. The research presents a novel scheme of classification for feature learning techniques in bearing prognostics. This taxonomy includes both shallow machine learning methods and deep learning methods that are based on temporal learning, spatial representation, and spatiotemporal representation.
2. The authors further analysed and presented feature learning methods that address data imbalance challenges by categorizing them into oversampling methods and downsampling methods.
3. The authors discussed recent advancements in overcoming challenges related to feature-invariant learning in different operational scenarios, which involve inconsistent data distributions during training and testing.
4. The paper also explores different methods for predicting the RUL of bearings, categorized as either shallow or advanced approaches depending on the feature learning phase.
5. The paper presents and discusses the experimental setups of the common bearing datasets in the literature.
6. The paper addresses current challenges in the state-of-the-art and highlights future directions.

The rest of the paper is organised as follows: Section 2 discusses the taxonomy of feature learning methods, including those that address the challenges of data imbalances and cross-domain learning. Section 3 reviews different studies developed for bearing RUL prediction. Section 4 presents the common publicly available bearing datasets. Lastly, Section 5 provides future directions for the coming years in the subject field. Finally, section 6 concludes the study.

## 2. Feature learning methods

This section provides a taxonomy of common existing feature learning methods in the state-of-the-art. At first, shallow feature-learning methods are classified into signal decomposition approaches and traditional feature extraction approaches. Signal decomposition techniques consist of empirical mode decomposition (EMD), variational mode decomposition (VMD), and wavelet transform (WT). While Fast Fourier transforms (FFT), short-time Fourier transforms (STFT), statistical analysis, linear discriminant analysis (LDA), and envelope analysis (EA) are among the traditional feature extraction methods that have been discussed. In the second subsection, feature learning methods based on DL techniques are presented and categorised based on the type of learned features, such as temporal learning, spatial representation, and spatiotemporal learning.

The third subsection analyses feature learning methods for imbalanced data samples. They are named downsampling and oversampling methods. Random downsampling, enhanced downsampling, and a one-category learning strategy are all examples of downsampling. Techniques for oversampling encompass geometric and adversarial approaches.

Lastly, the fourth subsection reviews feature learning methods for cross-domain learning. It includes studies on transfer learning, approaches for metric discrepancy, adversarial methods, and few-shot learning methods.

### 2.1. Shallow feature learning methods

This subsection discusses shallow feature learning methods, such as signal decomposition and traditional feature extraction techniques. Although such methods rely on human intervention and domain knowledge, researchers have recently been using them as an introductory stage in hybrid prognostic approaches, aiming to eliminate noise and improve learning speed.

#### 2.1.1. Empirical mode decomposition

Due to the harsh operating environments, the SNR of the acquired signal is significantly low [1]. The EMD [28] is a method that uses mathematical techniques to break down a signal into intrinsic mode functions (IMFs). The monotonicity of the final extracted IMF determines the stopping criterion for this decomposition process. Subsequently, by selecting the related IMFs from the original signal and discarding the irrelevant ones, the noise depicted in the signal can be considered to have been removed. In contrast to other techniques, EMD maintains non-stationary and non-linear relationships in the data. Chen et al. [29] applied EMD to eliminate noise from the collected acceleration data of a motor bearing. They decompose the original signal into fourteen IMFs before training a deep neural network for regression. Liu et al. [30] also adopt a similar approach, using EMD to enhance the backpropagation weight adjustments of a deep neural network that processes signals from multiple sensors. Additionally, another employability of EMD is to reduce fluctuation and extract signal content, as in [31].

The selection criteria for IMFs can be conducted using mathematical expressions like correlation coefficient, covariance [32], or cosine similarity [33]. In [32], the correlation criterion was used to select decomposed signals. Guo et al. [33] introduce a novel selection criterion based on cosine similarity. Although EMD can achieve good results, it may sometimes lead to the loss of hidden information at the edge of a signal during the decomposition process. Therefore, summing up all IMFs may not retain typical input signals [34]. In order to address the limitations of mode mixing in EMD, researchers introduced a noise assistant variant known as ensemble EMD (EEMD) [35]. EEMD is applied in [36] to decompose the vibration signal of a bearing; afterwards, statistical analysis is applied to the selected IMFs to form high-dimensional feature vectors for deep learning network modelling. Guo et al. [37] introduced a bearing prognostic approach based on relevance vector machines (RVM). However, they employ EEMD for feature extraction and noise elimination to reduce the uncertainty of RVM in long-term prognostics. In [38], the decomposed IMFs of EEMD are also utilised for a bearing prognostic study.

However, the added white noise in EEMD greatly affects the computation time. Fortunately, fast EEMD (FEEMD) [39] has been introduced to address this issue. The method has proven efficient in the literature, as in [40,41]. Moreover, Jiang et al. [42] proved the efficiency of FEEMD in extracting fault samples of a REB, while a hybrid denoising autoencoder and contractive autoencoder are then configured to select the higher-order features.

#### 2.1.2. Variational mode decomposition

VMD is a non-recursive signal decomposition technique widely employed in the literature to address the challenges of mode aliasing and noise sensitivity inherent in EMD. VMD seeks to identify a group of modes and their corresponding centre frequencies that can accurately replicate the input signal. After being processed into a baseband signal, every mode in VMD should appear smooth. VMD is considered an extension of the traditional Wiener filter that operates across several adaptable frequency ranges [43]. In [44], in order to detect a fault in its early stages, an approximate entropy VMD approach is developed. At first, the signal is decomposed into IMFs by VMD, and then approximate entropy is calculated for each IMF. An eigenvector is then created that contains fault information. Song et al. [45] used VMD to conduct a fault prognostic study to monitor the development of an element bearing's future degradation trend based on its long-range dependence characteristics. VMD is employed in [46] to detect the initial flaws in the captured vibration signal. VMD splits the original vibration signal into multiple IMFs. Dispersion entropy is computed for each mode, and principal component analysis (PCA) is performed on the two modes with significant variance. The initial step in the PCA learning process is then regarded as the beginning of a degradation output. In, [47], Liu et al., utilised VMD to extract relevant features and eliminate noise from the original signal of a rotating bearing. The extracted modes then undergo a hybrid feature selection methodology based on monotonicity and correlation, which feeds a neural network for lifespan calculation. Han et al. [48] developed an approach based on VMD to extract robust information about an element bearing. VMD has proven to preserve non-Gaussian and non-stationary characteristics during the feature extraction process.

#### 2.1.3. Wavelets transform

Another common signal decomposition technique that has been commonly utilized in the literature is the wavelets transform (WT). The fundamental premise of the wavelet transform involves hierarchically dividing a signal into a series of frequency channels that possess identical bandwidths on a logarithmic scale [49]. The wavelet transform effectively captures both temporal and frequency information in a signal and adeptly addresses signal denoising issues [50,51]. It can be classified into continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The latter is a mathematical method that uses displacement and translation values to create a set of wavelets that are all different from each other [52]. The initial signal, denoted as $x(t)$, undergoes filtration through two distinct filters, specifically a low-pass filter (LPF) and a high-pass filter (HPF). The signal is converted into two distinct components: low-frequency approximations (CA) and high-frequency detail components. For instance, Rathore et al. [52] employed DWT to remove noise from the collected signal before inputting it into a 1D neural network model for higher-order feature-learning. While [53], employed the DWT to differentiate between features in good condition and those in a degraded state.

The literature frequently uses the CWT type to transform a 1-dimensional (1D) signal into two or three dimensions for training neural network algorithms such as convolutional networks [54]. Yoo et al. [55] utilized CWT to extract spatial features and transform the bearing 1D signal array into 2-D dimensions for health index creation using convolutional networks. Similarly, [56,57] employ CWT to transform the acceleration signal into a 2D array, enabling simultaneous diagnostics and prognostics. Furthermore, wavelets can be employed to differentiate features with distinct characteristics, such as identifying spalls on bearing components [58]. Additional studies will be presented within the ensemble methods part, as the employability of WT has recently been introduced as a preliminary stage for training DL rather than being solely applied for regression.

#### 2.1.4. Fast Fourier transform

The FFT is typically utilized in bearing prognostics to get the frequency spectrum of the vibration signal. Specific harmonic component fluctuations in the frequency spectrum can identify a specific problem and serve as a fault signature for bearing durability [59]. Extracted features using FFT can then be input into either a simple prediction model [60] or a deep neural network (NN) model [61]. Ding et al. [62] developed a prognostic approach based on FFT. At first, FFT is applied to map the original signal to a frequency spectrum, and then root mean square (RMS) is applied as a degradation index. Finally, deep convolutional NN (DCNN) is used to estimate the RUL. In [63], a fusion feature-learning approach based on FFT and CWT is introduced. At first, FFT is employed for domain translation, followed by CWT, to transform a 1D signal into 2D vectors for grayscale convolutional neural networks (CNN) modelling. A similar approach based on FFT employability is conducted in [64]. However, a different CNN structure named Le-Net 5 [65] is used, along with a dropout layer designed to prevent overfitting.

Zhou et al. [66] used FFT to understand the variance in learned features from multimode operations affected by load and environment changes. In [67], using FFT before a deep belief network (DBN) improved the performance of the PHM model. This demonstrates the efficiency of learning the frequency distribution of the monitored signal, especially when the ML model struggles to independently learn temporal information. In [68], FFT is used to differentiate the frequency bands of a non-stationary signal that overlap. In [69], FFT is used to obtain Fourier coefficients, which are subsequently inputted into a convolutional network. Finally, in [70], FFT is used to extract useful information before inputting it into unsupervised modelling, reducing the need for manual intervention in extracting relevant features from the vibration signal. Further studies employing FFT will be explored in various sections of this paper, highlighting FFT's crucial role in enabling reliable prognostic processes.

#### 2.1.5. Short-time Fourier transform

STFT is a common technique used for analysing nonstationary signals [19]. The main idea of STFT is to split a bearing's nonstationary vibration signal into small intervals, then apply a Fourier transform to each interval to observe the evolution of a fault early on for a reliable prognostic approach. This method can avoid the loss of fault information over time, such as in FFT [70]. Additionally, once the signal is mapped to the time–frequency domain, further analysis can be performed.

STFT is set up in [71] to create an image representation showing signal frequencies over time before applying CNN for condition monitoring. Zhou et al. [72] applied STFT to the bearing signal so that CNN could learn the progression of fault information and calculate the RUL accordingly. In [73], STFT is used before a stacked sparse autoencoder to enhance the understanding of signal patterns in a condition monitoring approach. Additionally, [74] introduces the use of STFT for tracking specific parts of signals. The study aims to use STFT to understand temporal relationships, while CNN is used to identify degradation in the vibration signal. Similarly, Li et al. [75] used STFT to identify the fault information of an operational bearing early on and estimate its lifespan. Finally, [60] introduces an ensemble feature learning model based on STFT and PCA. The selected features are then fed to a linear regression model for lifespan estimation.

#### 2.1.6. Statistical analysis

Statistical features are the process by which the online signal statistics are monitored using the extracted features. If the observed characteristics deviate significantly from their expected values, the monitored machine would be considered faulty, and an estimate of its lifespan could be made. This technique is commonly used to monitor operating conditions based on extracted features. However, it is considered a manual approach because it relies on human intervention to determine significant features based on the monitored target's characteristics. This is further compounded by the requirement to manually establish a threshold for identifying faults. The analysis can be performed in three domains: time, frequency, or time–frequency. To analyse the frequency domain, the original signal must be transformed into a frequency-based representation. Techniques like the FFT and the STFT can be used to convert the signal into the frequency domain [59,72,76], as discussed in the previous subsection.

Examples of time-domain features include mean, standard deviation, variance, and kurtosis. On the other hand, examples of frequency-domain features are the maximum power spectrum, skewness of the power spectrum, and relative spectral peak per band. To provide a clearer explanation, Table 1 explains the mathematical representation of commonly used features in the literature.

For $k = 1, 2, .., n$, $x(k)$ represents the time domain signal series, and n represents the number of samples. Similarly, $s(i)$ is the frequency domain series for $i = 1, 2, .., n$, and $I_h$ and $I_f$ are features of the same nature in which $h$ and $f$ are features computed from the healthy and faulty states, respectively. $\mu$ is a vector representing the centroid of the acquired signal in the n-dimensional space. While dv $d_b$, $d_p$, and $N_b$ are the ball diameter, pitch ball diameter, and number of balls, respectively. $f_r$ is the rotational frequency. $b$ and $a$ are two positive integers, where $a < b$. $\beta$ is the ball contact angle with the races. Finally, $C$ is the covariance matrix.

As shown in Fig. 1.b, analysing statistical features is an initial step in shallow learning models before modelling [26,77]. On the other hand, in DL methods, the use of statistical feature extraction aims to reduce the computational requirements and remove interference [75,78–80]. For instance, in [78], Wang et al. applied statistical analysis to the 1D vibration signal before applying CWT to convert the 1D array signal into 3D images so that the spatial characteristics of the image could be represented. Similarly, after configuring STFT to translate the time domain signal into the frequency domain, [75] applied frequency domain analysis. Degradation thresholds are then set for a condition monitoring approach. In [79], feature analysis is applied, but a feature ranking criteria is followed using monotonicity and correlation to help a recurrent neural network (RNN) extract the temporal features and achieve better results. Lastly, in [80], time–frequency features were extracted before being represented on a graph for further graph knowledge learning.

In real industries where data sizes are huge, statistical representation can be extensively extracted and then followed by feature selection to simplify the computational cost. Traditional feature selection approaches can be classified into two categories: filter-based methods and wrapper-based methods [47,48]. Lately, neural network algorithms such as autoencoders (AE) can be adopted for the same approach without human intervention. Leveraging the capacity of ML, Patil et al. [81] employed a decision tree (DT) framework to select the most influential time domain features. In [82], PCA is used. However, the PCA approach achieved slightly higher results compared to the feature selection approach done using DTs. The authors in [83] and [84] applied the same methods for predicting bearing RUL. The framework involves applying wavelet decomposition, followed by time–frequency representation, and training support vector machines (SVM). In contrast, Benkedjouh et al. [84] used isometric feature mapping reduction (IFMR) to reduce dimensions and speed up computation. In [85], the authors extracted a set of statistical features and compared the efficiency of SVM and DT under similar conditions, showing that DT can yield better results.

It is worth mentioning that for statistical analysis to be effective, segmenting the signal into uniform windows of a predetermined duration to enable the extraction of distinct values is deemed essential, often called the "windowing" technique. There is always a trade-off between the quality of the extracted features and the window length, as a fixed window size may lead to the loss of non-stationarity characteristics. One possible solution to this issue is to reduce the window length; however, this approach requires a considerable amount of computation [26].

#### 2.1.7. Linear discriminant analysis (LDA)

Different from the previous feature learning methods that directly extract an observed characteristic, LDA aims to construct a feature map of a target signal. The algorithm can then identify unnecessary features and classify samples to distinguish between faulty and normal ones, enhancing the accuracy of regression models afterwards. In their study, Harmouche et al. [86] used LDA to distinguish the severity of faults on different parts of a ball-bearing: the outer-race, inner-race, and balls. In the same study, LDA, PCA, and SVM for feature learning were evaluated, and the researchers concluded that LDA outperformed the other

**Table 1**
Formulas of common statistical features in the literature.

| Domain | Feature | Formula |
|---|---|---|
| Time-domain | Mean value ($I_{mv}$) | $I_{mv} = \sum_{k=1}^{n}\frac{k_i}{n}$ |
| | Max ($I_{max}$) | $I_{max} = max\,(x_{k=1,2,\ldots,n}(\mathrm{k}))$ |
| | Min ($I_{min}$) | $I_{min} = min\,(x_{k=1,2,\ldots,n}(\mathrm{k}))$ |
| | Root mean square ($I_{rms}$) | $I_{rms} = \sqrt{\sum_{k=1}^{n}\frac{{x_k}^2}{N}}$ |
| | Variance ($I_{vr}$) | $I_{vr} = \frac{\sum_{k=1}^{n}(x_k - I_{mv})^2}{n-1}$ |
| | Standard Deviation ($I_{sd}$) | $I_{sd} = \sqrt{I_{vr}}$ |
| | peak-to-peak ($I_{p2p}$) | $I_{p2p} = I_{max}\text{-}I_{min})$ |
| | Absolute mean | $I_{amean} = \sum_{k=1}^{n}\frac{\lvert x_k\rvert}{n}$ |
| | Wave Factor ($I_{wf}$) | $I_{wf} = n.\,I_{rms}/\sum_{k=1}^{n}x_k$ |
| | Root mean squared error ($I_{rmse}$) | $I_{rmse} = \sqrt{\sum_{k=1}^{n}\frac{(x_k - I_{mv})^2}{n}}$ |
| | Peak value | $I_p = 1/2[I_{max} - I_{min}]$ |
| | Peak to RMS | $I_{p2rms} = \lvert I_{max}\rvert/I_{rms}$ |
| | Impulse factor | $I_{impulse} = I_p/I_{amean}$ |
| | Skewness factor ($I_{sf}$) | $I_{sf} = \frac{\sum_{k=1}^{n}(x_k - I_{mv})^3}{n.{I_{sd}}^3}$ |
| | Kurtosis factor ($I_{kf}$) | $I_{kf} = \frac{\sum_{k=1}^{n}(x_k - I_{mv})^4}{n.{I_{sd}}^4} - 3$ |
| | Crest factor ($I_{cf}$) | $I_{cf} = I_{max}/I_{rms}$ |
| | Absolute max ($I_{\lvert max\rvert}$) | $I_{\lvert max\rvert} = max\,\lvert x_{k=1,2,\ldots,n}(\mathrm{k})\rvert$ |
| | Median ($I_{median}$) | $I_{median} = \mathrm{median}\,(x_{k=1,2,\ldots,n}(\mathrm{k}))$ |
| | Mode ($I_{mod}$) | $I_{mod} = \mathrm{mode}\,(x_{k=1,2,\ldots,n}(\mathrm{k}))$ |
| | Mean absolute deviation ($I_{mad}$) | $I_{mad} = \mathrm{mad}\,(x_{k=1,2,\ldots,n}(\mathrm{k}))$ |
| | Harmonic mean ($I_{hmean}$) | $I_{hmean} = \mathrm{n}/\sum_{k=1}^{n}\frac{1}{x_k}$ |
| | Percentiles ($I_{perc}$) | $I_{perc} = perc\,(x_{k=1,2,\ldots,n}(\mathrm{k}))$ |
| | Interquartile range ($I_{IQR}$) | $I_{IQR} = IQR\,(x_{k=1,2,\ldots,n}(\mathrm{k}))$ |
| | Energy quantification related ($I_{\sigma^2}$) | $I_{\sigma^2} = \frac{1}{n}\sum_{k=1}^{n}(x(k) - I_{mv})^2$ |
| | Variance coefficient ($I_{vc}$) | $I_{vc} = I_{mv}/I_{\sigma}$ |
| | Skewness coefficient ($I_{sc}$) | $I_{sc} = \frac{\frac{1}{n}\sum_{k=1}^{n}(x(k))^3}{(I_{\sigma})^3}$ |
| | Kurtosis coefficient ($I_{kc}$) | $I_{kc} = \frac{\frac{1}{n}\sum_{k=1}^{n}(x(k))^4}{(I_{\sigma})^4}$ |
| | Fisher criterion ($I_F$) | $I_F = \frac{(I_{mv} - I_{mv_h})^2}{(I_{sd})^2 + (I_{sd_h})^2}$ |
| | Euclidian distance ($I_{ED}$) | $I_{ED} = \sqrt{\sum_{k=1}^{n}(I_h(k) - I_f(k))^2}$ |
| | Sum square error distance ($I_{ssed}$) | $I_{ssed} = \lVert I_h(k) - I_f(k)\rVert^2$ |
| | Mahalanobis distance ($I_{MD}$) | $I_{MD} = \sqrt{(x_k - \mu)^T C^{-1}(k - \mu)}$ |
| | Manhattan distance ($I_{ManD}$) | $I_{ManD} = \sum_{k=1}^{n}\lvert I_h(k) - I_f(k)\rvert$ |
| | Median error distance ($I_{Me}$) | $I_{Me} = \mathrm{argmin}\sum_{k=1}^{n}\lVert I_h(k) - I_f(k)\rVert_2$ |
| Frequency-domain | Outer-race fault (ORF) frequency ($I_{orf}$) | $I_{orf} = \frac{N_b}{2}f_r(1-(d_b cos\beta/d_p))$ |
| | Inner-race fault (IRF) frequency ($I_{irf}$) | $I_{irf} = \frac{N_b}{2}f_r(1 + (d_b cos\beta/d_p))$ |
| | Roller (ball) fault (BBF) frequency ($I_{bbf}$) | $I_{bbf} = \frac{d_p}{d_b}f_r(1-(d_b cos\beta/d_p)^2)$ |
| | RMS frequency of the 1st harmonic ($I_{rms-1sth}$) | $I_{rms-1sth} = \sqrt{\frac{1}{b_1 - a_1}\sum_{k=a_1}^{b_1}{f_k}^2}$ |
| | RMS frequency of the 2nd harmonic ($I_{rms-2ndh}$) | $I_{rms-2ndh} = \sqrt{\frac{1}{b_2 - a_2}\sum_{k=a_2}^{b_2}{f_k}^2}$ |
| | RMS frequency of the 3rd harmonic ($I_{rms-3rdh}$) | $I_{rms-3rdh} = \sqrt{\frac{1}{b_3 - a_3}\sum_{k=a_3}^{b_3}{f_k}^2}$ |
| | Mean power of spectrum ($F_{mv}$) | $F_{mv} = \sum_{i=1}^{n}\frac{s_i}{n}$ |
| | Maximum of power spectrum ($F_{max}$) | $F_{max} = max\,(s_i)$ |
| | Root mean square of power spectrum ($F_{rms}$) | $F_{rms} = \sqrt{\sum_{i=1}^{n}\frac{{s_i}^2}{n}}$ |
| | Variance of power spectrum ($F_{vr}$) | $F_{vr} = \frac{\sum_{i=1}^{n}(s_i - F_{mv})^2}{n-1}$ |
| | Standard deviation of power spectrum ($F_{sd}$) | $F_{sd} = \sqrt{F_{vr}}$ |
| | Skewness of power spectrum ($F_{sf}$) | $F_{sf} = \frac{\sum_{i=1}^{n}(s_i - F_{mv})^3}{n.{F_{sd}}^{3/2}}$ |
| | Kurtosis of power spectrum ($F_{kf}$) | $F_{kf} = \frac{\sum_{i=1}^{n}(s_i - F_{mv})^3}{n.{F_{sd}}^2}$ |
| | Relative spectral peak per band ($F_{rs}$) | $F_{rs} = \frac{F_{max}}{F_{mv}}$ |

techniques. M. Zhao et al. [87] demonstrated that LDA can eliminate noise while maintaining the important local structure of bearing samples, which is essential for PHM. Furthermore, authors in [88,89] utilised LDA as a denoising method for non-stationary vibration signals from bearings. Compared to PCA and marginal Fisher analysis, LDA demonstrated superior performance [89]. Ciabattoni et al. [90] presented a variant of LDA called delta-LDA, which addresses certain issues by utilising covariance matrices. This variant utilises covariance matrices to resolve the problem of a between-class scatter matrix trace approaching zero. This method produced outstanding results and was validated using electrical motor bearings.

### 2.1.8. *Envelope analysis*

Envelope analysis (EA) is an example of a shallow feature learning technique commonly used in bearing diagnostics and prognostics. In PHM applications, EA primarily extracts the wide variation of acquired signal amplitudes, such as vibration signals. The technique demodulates the accelerometer signal through band-pass filtering. According to Antoni et al. [91], EA methods can effectively find and boost fault-related high-frequency components. This can help in the early detection of bearing faults like cracking and spalling. Swalhi et al. [92] further enhanced the performance of EA for bearing early fault detection by integrating spectral kurtosis. Another approach was proposed in [93], which suggests the integration of wavelet packet transform. The authors demonstrated the study's effectiveness compared to a traditional STFT kurtogram. Qin et al. [94] proposed a novel SVD approach in order to facilitate the process of selecting the appropriate scale for the wavelet transform integrated with EA. The study has proven reliable in detecting a weak signature of mechanical impulses. In order to identify the influential sub-band signal in EA, Kang et al. [95] proposed a Gaussian mixture model-based residual component-to-defect component ratio, which serves as an effective metric for assessing the severity of defects. While to distinguish between healthy and faulty features, [96,97] feed the extracted features from EA to SVM. In a similar vein, Wein et al. [98] proposed a 2D feature vector based on EA as an input to K-nearest neighbour (KNN) for rolling bearing condition monitoring. It can be concluded that EA enhances the effectiveness of extracting fault-related features, thereby improving the early prognostics process.

Table 2 presents an overview of the techniques discussed, along with a compilation of studies that have implemented these methodologies.

## 2.2. *Deep feature learning methods*

Feature learning methods using deep learning algorithms, are capable of capturing important features of an operational bearing automatically and without manual intervention. This is because DL networks have the capacity to learn complex relationships among patterns extracted from the raw signal using non-linear functions like hyperbolic functions, rectified linear units, the sigmoid function, and SoftMax [99]. In the context of bearing prognostics, methods can be classified as temporal feature learning, spatial feature learning, and spatiotemporal feature learning.

**Table 2**
Summary of discussed feature learning techniques.

| Technique | Advantages | Disadvantages | Studies |
|---|---|---|---|
| EMD | offers adaptive and data-driven decomposition of nonlinear and nonstationary signals | Sensitivity to signal noise and mode mixing. | [28–42] |
| VMD | Provides a robust and non-recursive method for signal decomposition which would overcome the mode mixing | Dependence on predefined parameters such as number of modes which limits its adaptability | [43–48] |
| WT | Capturing time and frequency information simultaneously, making it effective for examining non-stationary signals. | Sensitivity to the choice of mother wavelet which can limit the adaptability in different operational environments | [49–58] |
| FFT | Enabling rapid conversion of time-domain signals into the frequency domain | Cannot effectively analyse non-stationary signals | [59–70] |
| STFT | Effective in analysing non-stationary signals by capturing frequency content in localized time windows | Fixed time–frequency resolution which can limits its adaptability to signal with varying features | [19,71–75] |
| Statistical analysis | Effective in detecting variation of signal trends | Depends on pre-defined thresholds and may cause the fading of non-stationary characteristics of the captured signals | [26,59,72,75–80] |
| LDA | Ability to maximize the class separability of the normal and fault data features | Sensitivity to outliers | [86–90] |
| EA | Enhance the understanding of complex non-stationary signals | May lead to the masking of important high frequency characteristics leading to the loss of critical features | [91–98] |

### 2.2.1. *Temporal learning*

Temporal learning uses DL algorithms to detect variations in signal patterns over time, eliminating the need for manually established thresholds that are typically used in conventional statistical techniques. Additionally, it models the relationships among sequential input features to automatically calculate the bearing RUL at each time step [100]. Several deep learning techniques have been developed to learn the dependencies of features over time. RNN, long-short-term memory (LSTM), and gated recurrent units (GRU) are examples of these techniques.

RNN is widely recognized for its success in learning time sequences, making it a prominent model in this context. However, because of the long backpropagation training that is required to learn a whole sequence of bearing life cycles, training an RNN can lead to vanishing or exploding gradients [101]. Other variants of RNN, such as LSTM and GRU, have avoided the vanishing gradient problem by introducing a gating mechanism in which the input gate and forget gate in LSTM can only consider important temporal information while discarding unimportant features [101]. Meanwhile, the number of gates in GRU is minimised to reduce complexity and, at the same time, preserve long-sequence learning [102]. Fig. 3 illustrates the internal structure of RNN, LSTM, and GRU. Bi-directional LSTM (Bi-LSTM) and bi-directional GRU (Bi-GRU) are variations of LSTM and GRU, incorporating bidirectional learning. It aims to learn the sequence of features in the forward and backward directions in order to provide more precise learning [103–105]. Table 3 summarises the state-of-the-art variants of common temporal networks.

To the best of the authors' knowledge, RNN was first introduced for bearing prognostics in [106]. The novel approach aims to predict the RUL using RNN and extended KF (EKF). The RNN is employed to extract the hidden temporal features, while the EKF is used to map the extracted features to the remaining life values. Guo et al. [79] employed RNN to automatically detect the failure threshold and, hence, estimate the RUL based on the severity of the failure. Mao et al. [107] used LSTM to extract temporal information for bearing RUL estimation. At first, the Hilbert-Huang transform is applied to the raw data, and then LSTM is employed for both feature learning and RUL calculation. Chen et al. [29] conducted a comparison between extracting the relevant features using LSTM and SVM after applying signal decomposition using EMD. It is stated that LSTM performed with a higher prediction accuracy of 36 %.

**Table 3**
Variant of recurrent neural networks.

| Variant | Description |
|---|---|
| RNN | • Can make use of the sequential data<br>• Capable of remembering short-term information |
| LSTM | • Resolve the gradient vanishing and exploding problem due to the gating mechanism.<br>• Capable of retaining both short- and long-term knowledge |
| GRU | • Similar to LSTM, however, it is considered a faster learner as it depends on only two gates.<br>• Capable of capturing the intrinsic relationship for long-term forecasting. |
| Bi-LSTM, Bi-GRU | • The learning process should be in both forward and reverse directions.<br>• Suitable for moderate forecasting |

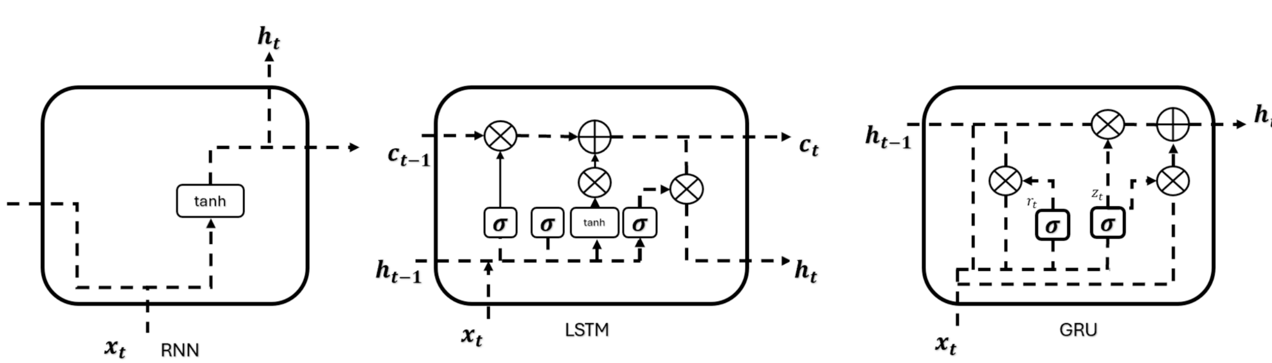


**Fig. 3.** Shallow rnn structure.

Zhan et al. [32] applied the same model configuration without applying EMD and did a comparison between RNN, LSTM, and Bi-LSTM on the same bearing dataset. LSTM has demonstrated markedly better performance than RNN, while only marginally surpassing Bi-LSTM, highlighting the effectiveness of bidirectional approaches. However, when the signal is decomposed using EMD and a set of statistical features is extracted, Bi-LSTM improves prediction accuracy by nearly 50 %. A fusion-temporal learning approach is conducted in [108]. At first, statistical features are represented to reduce computational costs and eliminate interference. Subsequently, GRU is used for temporal learning to improve prediction accuracy. However, in [109], a bi-GRU network is configured to rely on the CNN network instead of statistical representation. Another GRU-based approach is introduced in [110]. An attention mechanism is applied to GRU so that the features selected can then be utilised for life calculations with less computational complexity. Finally, in an industrial bearing prognostics approach, Zhou et al. [111] used features extracted by GRU to measure differences in probability distributions across various operational conditions.

*2.2.2. Spatial features learning methods*

In addition to temporal methods, spatial representation has demonstrated significant efficiency in bearing prognostics. Within the framework of spatial learning, the prognostic process typically transforms the input vibration signal into the spatial domain to capture the key spatial characteristics at each time step. [112]. Deep convolutional networks are widely used for this purpose. CNNs have had tremendous success in computer vision applications [113,114]. It extracts representative features from grid structures using convolutional, pooling, and fully connected layers. Fig. 4 explains the structure of CNN.

In addition to CNNs, a recent method for bearing prognostics is based on graph representations [80]. The method represents features in a graph structure and learns by using deep neural networks on graphs. This neural network type is known as a graph neural network (GNN) and has recently been introduced for bearing prognostics. One significant advantage of GNN over traditional spatial networks is its capability to handle high-dimensional data without needing to calculate the Euclidean distance between features in the feature space, which has significantly improved machinery prognostic results. This subsection provides an illustration of both methods.

*2.2.2.1. Spatial learning via CNN approaches.* The main goal of the convolution layer is to create feature maps by applying convolutional operations to the input with filters. The main aim of the pooling layer is to reduce the input size, which helps compress data through downsampling. The fully connected layer, also known as a fully connected forward network, serves the primary purpose of extracting deeper characteristics from the data. Typically, it is linked to the implementation of multiple convolutional and pooling layers. Fig. 4 is a shallow explanation of a native CNN structure. Various CNN network topologies have been developed in the last decade. Some of them aim for better learning performance in terms of computational complexity and speed, while others are mainly focused on feature manipulation [114–117]. Table 4 states some of the common variants of typical CNN networks that are widely used for bearing prognostics.

In [118], a pioneering study introduces CNN for PHM. The authors demonstrated that they could use CNN for spatial feature extraction to estimate the RUL of bearings. The authors proved that extracting spatial features using CNN can be employed for estimating the RUL of bearings. Ding et al. [62] employed CNN but removed the pooling layer. A typical topology of a CNN algorithm in this study consists of three convolutional layers followed by two fully connected layers. Zhou et al. [72] applied STFT instead to ensure effective learning of the fault characteristics. The authors represented the time–frequency features after applying STFT, so the extracted features are then fed to the CNN network, facilitating the learning process of CNN. Meanwhile, Wang et al. [64] conducted another approach. The authors employed CWT to convert the 1D vibration signal into 2D images. Then the LeNet-5 convolutional network topology, which is introduced in [65], is applied to bearing RUL prediction.

To diagnose and predict a bearing's RUL, [119] introduced a new CNN architecture that shares learning weights in fully connected layers for both tasks. The methodology has markedly enhanced the error

**Table 4**
Common variants of CNN for bearing prognostics.

| Variant | Description |
|---|---|
| Traditional CNN | • It consists of convolutional, pooling, and dense layers.<br>• Effective in catching the most prominent signal patterns |
| Multi-layered CNN | • Multiple CNN architectures are layered one upon another.<br>• The output of one CNN becomes the input of subsequent CNNs.<br>• Effective at dealing with raw signals, as instead of reliance on feature extractors |
| Multi-scale CNN (MSCNN) | • It contains two primary sequential stages: local and full convolutions.<br>• In local convolution, features are retrieved at each layer<br>• In the entire convolution layer, all retrieved features are concatenated to produce the final result.<br>• Effective preservation of the prediction's numerous layers of abstraction |

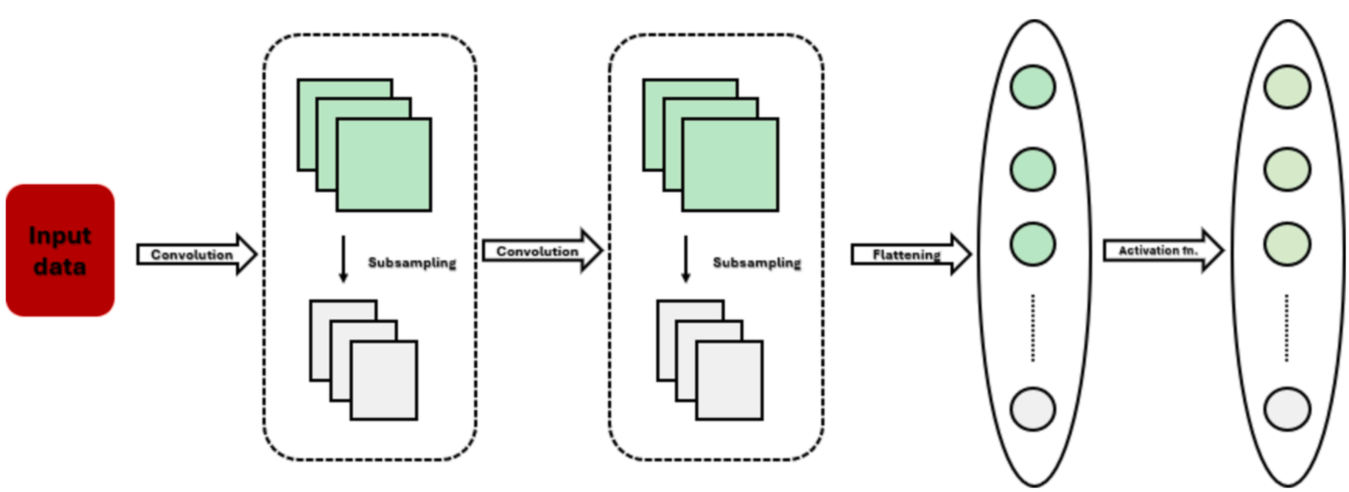


**Fig. 4.** CNN structure.

compared to utilizing the CNN approach for each task separately. MSCNN, initially introduced in [120], leverages both global and local information within a CNN network. MSCNN is also considered in [75]; however, STFT is applied to extract time–frequency representations. The hamming window function of 20-ms window length is configured for the statistical feature extraction process with 50 % window overlap between each of the two consecutive windows. However, [121] suggests an alternate implementation of MSCNN. Instead of extracting the features from local and global convolutional layers with fixed-size kernels, different kernel sizes are applied to extract the hidden patterns from different receptive fields. The results obtained in this study show a relatively higher performance compared to the original architecture introduced in [120]. Finally, and benefiting from the residual networks (ResNet) approach, which is mainly dependent on forming bypass links between different convolutional layers to ensure effective learning of features in deeper layers [115], Wang et al. [122] used separable CNN with residual connections between layers to prevent the loss of important features that could lead to inaccurate predictions.

*2.2.2.2. Spatial learning via GNN approaches.* PHM applications have recently incorporated GNN. In PHM applications, graph learning aims to achieve spatial feature learning and causal learning of manufacturing systems. Once the acquired signal is represented on a graph, spatial learning can be achieved by applying convolutions to graph structures [123]. One benefit of using convolutions on graph structures is that the method is not reliant on the distances between extracted features in the hidden space. CNNs are primarily created for grid-like structures like images and are commonly used in computer vision applications. That is why it may lead to model uncertainty when applied for applications such as bearing prognostics, as in that case, the high-level feature representation does not need to be attained. In addition, the number of papers that have adopted this approach is still limited. Using GNN in PHM applications involves representing the obtained sequence of signals on a graph [124]. A typical graph structure consists of three main elements: vertex, edge, and feature set. So, it can be denoted as $G = (V, E, X)$ [125].

To the best of the author's knowledge, a pioneering study of applying graph learning to calculate the RUL of an element bearing was first introduced in [80]. The researchers studied how spatial learning affects the accuracy of predicting the RUL of bearings without considering Euclidean distance. The results are significantly higher than those obtained using traditional CNNs. Manipulating the design of a GNN is straightforward. For example, we can represent each sensor in the system as a graph node, each containing a feature set of its signal values. Also, each degradation stage can be considered a node in the graph, with the characteristics of each node serving as the feature set. The formulation of the graph edge can be done through similarity metrics based on cosine similarity, correlation, or any mathematical formula according to the prognostic task that can ensure the relationship between a feature set and another.

In [126], a multi-scale convolutional architecture is applied to graph representation to better incorporate spatial learning in the regression process. At first, the spectral energies of the entire collected signal are extracted through a sliding window mechanism. Second, PCA is applied so that the variance between the first extracted feature and the rest of the features over time can be calculated and the fault occurrence can be determined. Finally, the graph is structured in such a way that each feature extracted is considered a graph node, and PCA latent space is used to build the relationship between the nodes. When two nodes make significant contributions to PCA, the edges that depict the relationship between these nodes exhibit a considerable relationship. The structure of the graph is considered changeable through time; thus, when convolution is applied to each of the constructed graphs, it is considered spatiotemporal learning. Further studies focusing on analysing spatiotemporal features through graph learning will be explored in the following subsection.

#### 2.2.3. Spatiotemporal Learning

Despite the continuous development of spatial learning techniques, ignoring temporal features can result in inaccurate condition monitoring (CdM) and unreliable predictions of bearing lifespans [127,128]. Zhao et al. [129] aimed to study the impact of extracting both temporal and spatial features on bearing prognostics. The authors introduced a 1D-CNN network named temporal convolutional network (TCN), where the processing is done on a 1D network instead of 2D or 3D networks, focusing on the time array for spatial learning through convolution. Notably, the certainty of the achieved results increased. The TCN approach is compared to traditional CNN models, and the results achieved by TCN are 10–20 % higher compared to conventional CNN structures. Cao et al. [130] applied TCN for the same purpose, but a self-attention mechanism was adopted to obtain the contribution degree of each of the extracted features. The study results outperformed the native TCN approach significantly compared to TCN without feature selection.

Moreover, authors in [56] studied the impact of converting the 1D signal to a 2D image, thus applying 2D CNN instead of 1D TCN. The transformation methodology in the state-of-the-art for such conversion commonly depends on wavelet transform structures such as Morlet wavelets. The study found that combining spatial features extracted with 2D-CNN with temporal features performs better than extracting spatiotemporal features simultaneously with 1D-CNN methods. Following the same approach, Wang et al. [78] mapped the 1D acceleration signal into 3D images using the Morlet wavelet transform and applied an overlapped windowing strategy. The spatiotemporal features are then extracted using 3D CNN, and Gaussian process regression (GPR) is employed for calculating the bearing lifespan. Interestingly, this approach achieved even better results compared to other spatiotemporal approaches. Wang et al. used a similar approach in another study [112], focusing on the computational performance of spatiotemporal learning. To address this challenge, the author used a technique called non-negative matrix factorization (NMF) for reducing the dimensions of latent space.

Jiang et al. [131] separated the raw bearing signal into various channels. Each channel underwent CNN processing independently before combining the spatial features with the temporal ones extracted using attention-LSTM. This approach aims to link the spatial features of each channel in the sequence with the extracted temporal features to enhance the understanding of spatiotemporal representation. Aside from the multiple channels concept applied in [131,132] applied recurrent convolutional networks for remaining life calculations. However, variational inference is used to measure the uncertainty of the model. In the same study [132], the authors departed from the usual method of connecting CNN sequentially to a temporal algorithm. Instead, they applied a convolutional operation separately to the input-to-state and state-to-state parts of an LSTM.

Moreover, GNN has been recently employed for spatiotemporal learning [133,134]. In [135], spatial features via graph learning are adopted, and then temporal convolutional regression (TCR) based on LSTM and graph convolutional network (GCN) is implemented for RUL prediction. While in [136], a Bi-LSTM is adopted to extract important features, followed by employing GCN to calculate the bearing RUL. Wei et al. [137] developed a self-attention spectral graph network for bearing PHM. The goal is to help the regression model learn from graphs independently during training, leading to significant improvements in prediction accuracy and reliability. Liang et al. [138] developed a graph network based on transformers. The primary goal is to understand how different sensors are related in space using the graph structure and to learn about time dependencies using the transformer network. While, in [134], an attention layer is added to fuse the learned spatial and time dependency information of the extracted data. Wang et al. [139] proposed a strategy to select the relevant sensor data to form neighbouring nodes of a graph-structured network for spatiotemporal learning. That strategy is based on the Pearson correlation coefficient calculated between different sensory readings, and the higher the correlated data, the more adjacent these sensors—graph nodes—are. On the other hand, to overcome the problem of graphically representing a univariate vibration signal on a graph, a ChebGCN-BiLSTM network is developed in [140,141]. The path graph at the node level is created to depict the connections between time-discrete signals. Edges indicate the ordered sequence, while nodes represent the signals themselves. It's important to note that Bi-LSTM is used to understand temporal dependencies in addition to the spatial ones, which are captured by the GCN network.

Cao et al. [142] introduced a graph convolutional structure known as a complex picture in a picture (PIP) to offer various levels of fault severity analysis. This approach differs from connecting graph nodes based on clustering or similarity methods like cosine similarity. The PIP approach uses the path graph as a node within a larger graph and systematically constructs a new embedded graph. Researchers in [143] explored the correlation between extracted features at various time points. To get that correlation at different stages of a rolling bearing's degradation, an LSTM-based Siamese network is created to classify the different stages of degradation. Next, a GCN based on attention mechanisms is put in place to make the prediction more reliable by using the correlation of information gathered at different time points. In a similar approach, a one-dimensional GCN is built to establish a connection between the reference data and the current working conditions and to accurately predict the RUL [144]. Considering the spatial dependencies of the acquired raw signal and temporal correlation, the short-time Fourier transform is utilised to extract node attributes, while dynamic edge connections are formed based on node importance weights in [145]. This is achieved by utilising the GCN to identify spatial relationships within the input graphs and employing a Bi-LSTM network to record overall temporal connections. Finally, a graph readout layer based on autoencoders is created to capture and transmit the key graph features. The authors [146] came up with a way to avoid the problems that come with combining graph structures, which can make it hard to see how the observed signal changes over time and create features that aren't very useful. They suggested using a residual graph connection and a self-attention GCN to pick out the most important features.

The deep GNN model also has a problem called "over-smoothing," which makes it hard to represent the difference between nodes in the network [147]. Consequently, the classification and prediction of nodes in the network will be inaccurate and ineffective. To overcome this, the authors proposed a multi-scale graph convolution network. First, they configure a sliding window to identify the fault occurrence time, after which the spatiotemporal graph network learns the dependencies and estimates the RUL [126]. In [143], an LSTM network is serially connected to a GCN for spatiotemporal learning. However, on the graph structure, a self-attention mechanism is implemented to select the most influential features and overcome the smoothing challenge in RUL prediction. In [141], dilation convolution is used to combine spatial and temporal data, preventing long sensor signals from over smoothing. While a GRU-GCN is introduced in [148], a skipping connection to extract the temporal features and overcome the challenge of signal over-smoothing is proposed.

Finally, Mylonas et al. [135] did a comparison study of a spatio-temporal network based on LSTM with a temporal convolutional head and a graph network based on convolutional structure. The results showed that the graph network structure did better than the non-graph network structure because it could learn features of times that were not equally distant.

### 2.3. Feature learning under imbalanced data samples

The difficulty of acquiring fault data samples imposes a challenge for a reliable training process and, hence, accurate estimation results. In real-world industrial settings, mechanical systems typically operate under normal conditions for the majority of their operational lifespan, with occurrences of fault states being relatively rare. When the system enters a short-duration defective state, gathering degradation indications is expensive and difficult. Hence, imbalanced datasets are common [149].

An imbalanced dataset is a set of samples with a disproportionate number of representative samples from one class. This proportion defines the "imbalance ratio," which is crucial and might be a challenge in a regression problem. Due to an imbalanced dataset, training a DL model to estimate RUL is difficult. Hence, the imbalanced dataset problem limits RUL estimation methodologies. In order to mitigate this limitation, numerous sampling methodologies have been extensively documented in scholarly literature to artificially equate the ratio between different classes of training samples, including under-sampling and over-sampling techniques.

#### 2.3.1. Downsampling methods

Under-sampling is an approach that reduces the number of instances belonging to the majority class in order to align the imbalance ratio with the smaller number of samples in the minority class. There are three main techniques of downsampling in the PHM literature, which can be summarised as:

1- **Random downsampling**: random downsampling is the simplest downsampling technique that implies discarding random samples from the majority class to alleviate the features of the imbalanced class. However, discarding important samples during random downsampling can directly affect the predicted RUL [150].
2- **Enhanced downsampling:** opposite to random downsampling, enhanced downsampling aims to select the discarded or retained samples to achieve a balanced class of data. This can be accomplished through three methods: clustering methods, sample density methods, or dynamic methods. In clustering methods, distances calculated between samples, such as the Euclidean distance, play a significant role. Samples that are closer to the class centre are considered more important in clustering methods. Accordingly, latent features that are distorted or on the edges of clusters are discarded [151]. Similarly, in terms of sample density, samples in less dense regions of latent space are removed while those in denser regions are kept [152]. While the dynamic process selects the samples for retention based on their influence on RUL calculations. So, the probability of losing an important, sparse feature is reduced [153].
3- **One-category learning strategy:** This method aims to select the relative samples based on similarity [153]. This can be achieved by applying statistical formulas to majority-class samples or through similarity-based ML approaches such as SVM. Moreover, utilising autoencoders (AE), variational autoencoders (VAE), and generative adversarial networks (GAN) can help identify important samples based on their data distribution while disregarding others [154].

Despite improvements in downsampling techniques to refine feature selection, downsampling can lead to the omission of crucial features, affecting the accuracy of RUL predictions and reducing robustness. In particular, bearing defects can take various forms and may not conform to a specific pattern of similarity for selection, particularly in challenging working conditions.

#### 2.3.2. Oversampling methods

Oversampling methods involve randomly adding samples from the minority class to balance the dataset. Oversampling methods can be classified into classical and adversarial methods. Random oversampling, windowing overlapping, Synthetic minority oversampling technique (SMOTE), and adaptive synthetic sampling approach for imbalanced learning (ADASYN) are all methods used to address imbalanced features of a given dataset.

##### 2.3.2.1. Geometric oversampling methods

In random oversampling, latent features and clustering centres are randomly added to the classified minority class samples [155]. Windowing overlap oversampling involves overlapping signal lengths between consecutive windows for data augmentation [156]. This can be achieved by varying the lengths of windows of fault samples relative to the lengths of normal samples. Yang et al. [157] introduced a variable-scale windowing approach based on AEs. Leveraging the advantages of sparse autoencoders (SAE) and denoising autoencoders (DAE), extracted features using signal windows overlapping can then be employed to solve the skewness of data. Such an approach can be effective, although in large machinery datasets, it may require extensive computational complexity and storage and may also lead to discarding the non-stationary characteristics of the acquired bearing signal [26].

On the other hand, SMOTE [158] is an oversampling approach that generates synthetic samples to oversample the minority class. Each new synthetic sample in SMOTE is created along the line connecting a chosen minority class sample and its nearest neighbour. SMOTE is extensively used in PHM research [159–162]. The steps of generating samples using SMOTE can be simplified as follows:

- Select KNN samples of the sample $x_i$ from the same class.
- Randomly select n samples from the KNN samples.
- For each chosen neighbour, a new sample is constructed according to:

$$x_{new} = x_i + rand(0,1)^{*}x_i - x_n$$

SMOTE is usually applied after a signal denoising stage to ensure the quality of the newly generated feature. For instance, Ruifeng et al. [163] applied SMOTE to address the challenge of data imbalance, following a noise elimination approach based on extended learning machine (ELM). In [164], a similar approach is conducted, where the relevance vector machine is used for feature denoising instead. Additionally, the authors in [154] introduced an ensemble learning approach based on Naive Bayes SVM and KNN to classify data samples before applying SMOTE. Interestingly, Yang et al. [165] conducted a comparison study between SMOTE and random oversampling, showing that the quality of features generated using SMOTE outperformed random oversampling-generated features.

A drawback of traditional SMOTE is that if noisy data and its neighbour are chosen as a cluster, oversampling may occur within the wrong class region. That's why different approaches are introduced in this context. For instance, Zhang et al. [166] proposed a learning strategy based on Euclidean distance calculations between minority class samples to avoid oversampling noisy features. Fan et al. [167] relied on the calculation of Euclidean distances between minority-class samples and the centre of the cluster. In a different approach, Zhu et al. [168] proposed the Mahalanobis distance as a metric of qualification. This approach showed superior performance in preserving the data distribution because the Mahalanobis distance is less influenced by data magnitude, unlike the Euclidean metric. Another approach is to divide the minority samples into different subspaces, and then oversampling can be performed for each subspace accordingly. K-means clustering [166], hierarchal clustering [169], density-based clustering, and fuzzy c-means [167] can all be adequate in such cases.

ADASYN [170] is another effective strategy for handling imbalanced data and has proven superior for bearing PHM [171]. The authors in this study demonstrated that ADASYN can better avoid the risk of synthesising noisy features compared to SMOTE. The experiment is done on bearings under different operational conditions. It uses a weighted distribution for minority classes. It is worth mentioning that most complex systems operate under diverse conditions, leading to multiple fault modes. Thus, ADASYN requires fine configurations to capture the complexity of numerous failure situations and avoid contributing to undesired noise [172]. So, despite the effectiveness of oversampling methods compared to under sampling methods, they may lead to the generation of noisy synthetic samples with disregard for the data distribution.

*2.3.2.2. Adversarial oversampling methods.* On the other hand, adversarial oversampling methods aim to preserve the feature distribution during the sample generation process. In this context, commonly, two techniques are used for data generation: VAE [173] and GAN [174].

GAN [174] is composed of a composite network structure of generative and discriminative networks. The model trains alternatingly until it reaches Nash equilibrium, at which point the discriminator is unable to distinguish between real data and generated samples. Lee et al. [175] developed a GAN model based on multi-layer perceptron (MLP). Initially, EMD is applied to denoise the vibration signal and then the selected IMFs are fed to the generative-discriminative MLP network for oversampling. Another generative-discriminative approach based on CNN was developed in [176]. The newly generated features proved to mimic actual faults, such as roller faults and inner race faults of an element bearing. In [177], researchers introduce a generative approach based on TCN to retain spatiotemporal features. Another approach based on statistical analysis of the extracted features is conducted in [178]. The time-domain statistical features are extracted then fed to the discriminator network to improve the quality of the features that GAN generates. However, instead of time–frequency representation, Yang et al. [179] implemented a feature mapping approach based on stacked AE to enhance the quality of the newly generated samples. Moreover, [180] configured an autoencoder network to measure the similarity between artificially generated samples by the generator and the original samples for further improvements.

Despite preserving the distribution, the basic GAN had poor sample quality, mode collapse, and unstable training. Numerous GAN variants, like the Wasserstein GAN (WGAN), have been introduced to improve training stability by comparing real and generated sample distributions using Wasserstein distance instead of KL-divergence in the traditional GANs. But to keep the Wasserstein distance reasonable, the WGAN has to cut the weights in the discriminative network to a certain range [−c, c] after backpropagation. Phan et al. [181] developed a WGAN using a 2D representation of signals obtained from bearing acoustic emissions to retain spatial features in the generation process. In [182], Cabrera et al. proposed a novel WGAN approach based on similarity and sparsity centroids. The aim is to efficiently identify the relative-generated samples of the minority class before feature generation. A spatiotemporal WGAN, designed for retrieving both spatial and temporal features, is introduced in [183], aiming to generate samples of minority classes while preserving the spatiotemporal characteristics.

Despite the superiority of WGAN compared to GAN in terms of learning convergence, gradient vanishing or explosion is common, specifically in the backpropagation process. The authors of [184] suggested a gradient penalty to control the discriminative network's gradient. Pu et al. [185] developed a gradient penalty WGAN (WGAN-GP) to overcome the deficiencies of the WGAN model. The authors also looked at the differences in how accurate minor samples generated using GAN were compared to those generated with traditional oversampling methods like SMOTE and random down sampling. In this context, WGAN-GP demonstrated a slightly higher accuracy performance. Another approach is introduced by Behera et al. [186] as a conditional GAN (CGAN). It is designed to learn specific classes instead of any minority class to avoid capturing irrelevant features. This helps eliminate the probability of generating samples of the majority class [187]. For a similar approach, frequency-domain features and features of the normalisation spectrum are fed to CGAN for minor fault mode feature generation. Generated samples are then sent to stacked autoencoders for denoising before being fed to a discriminator network [188].

VAE consists of three main parts: the encoding network, the sampling component, and the decoding network. The encoding network converts the samples into a latent variable $z$, characterised by a mean vector $\mu$ and a standard deviation vector $\sigma$. The decoding network duplicates the input by utilising the latent variable. During training, the VAE ensures the latent variable aligns with a normal distribution. After obtaining a novel latent variable $z^\wedge$ from the normal distribution, decoding is used to create a new sample. The Conditional Variational Autoencoder (CVAE) is an extended version of the VAE that incorporates label information [189]. By influencing the mean vector, one can generate corresponding categories while generating samples. Although VAE and CVAE have been effective in image recognition, their application in imbalanced learning for bearing prognosis is still limited [190]. Zhao et al. [191] and Karamti et al. [192] used a VAE to create artificial features that enhanced the accuracy of life prediction with imbalanced data. Xie et al. [193] used PCA to reduce feature space dimensions before using CVAE to generate artificial features representing different failure modes. This method was chosen to speed up the convergence process. Finally, Yang et al. [194] presented a novel technique that merges VAE and GAN to generate a variety of samples. They used a conditional VAE as the generator network in their ensemble approach. Table 5 lists the common techniques utilized for imbalanced data problems.

### 2.4. Feature learning under varying working conditions

Another challenge of data-driven approaches for bearing prognostics is their requirement for (1) sufficient labelled data, namely, historical data; and (2) training and testing data following a similar distribution, which are hard to fulfil in bearing prognostics. That is because industrial rolling bearings operate under complex, frequent, and inconsistent working conditions, making it difficult to preserve a common data distribution from one environment to another. In this context, several methods have been conducted and validated in the literature to address this gap. The methods include transfer learning (TL), distribution metrics analysis, adversarial learning, and few-shot learning. All these methods are discussed in this subsection.

#### 2.4.1. Transfer learning

TL is a technique that transfers knowledge from one domain to another and has been widely adopted in fields such as text classification [195], computer vision [196], and natural language processing [197]. TL's role in RUL prediction is to extract domain-invariant feature representations and transfer degradation knowledge acquired under different working conditions. It relies on transferring the learned parameters from one domain to another. This method is based on the underlying assumption that the model architectures of both the source and target domains should have specific shared parameters [198]. In this context, the tuning strategy involves transferring the learned model parameters, such as weights, from the source domain to the target domain [199]. The process consists of two separate stages: (1) the initial training of a deep neural network in the source domain, where the optimal model parameters are saved, and (2) the preservation and transfer of the model parameters from the hidden layers in the source domain, which can be utilised as the initialization parameters for the model in the target domain. Moreover, the target data can randomly initialise the parameters of the final layer in the target domain, enabling further refinement of the model parameters [200].

In a study by Zhang et al. [201], a TL framework was developed using Bi-LSTM. The purpose of using Bi-LSTM in this case is to capture the hidden temporal patterns in the target domain data in one direction. Meanwhile, the reverse direction of the LSTM architecture was designed specifically for predicting RUL. In their study, Sun et al. [202] present an alternative approach that utilises SAE. The study uses SAE to map the weights and features of the encoder layer from the source to the target domain. Furthermore, Zhang et al. [200] used a TL methodology to transfer spatial characteristics to the target domain by converting the vibration data of rolling bearings into RGB images. Simultaneously, the LSTM network is employed to learn the temporal patterns and estimate the RUL respectively. Huang et al. [203] approach utilises depth-wise separable convolution and Bi-LSTM. After training, the model parameters of the networks in the source domain are shared, and a dense layer is added to allow for separate weight updates.

#### 2.4.2. Metric discrepancy evaluation

Distribution metric evaluation methods aim to minimise feature distribution differences between the source and target domains. This can be determined and minimised through the utilisation of various discrepancy metrics such as maximum mean discrepancy (MMD) [204], Jensen–Shannon (JS) divergence [205], multi-kernel MMD (MK-MMD), Euclidean distance [206], and Kullback-Leibler (KL) divergence [207]. Afterwards, reliable life prediction values are obtained by leveraging the shared feature spaces.

Chang et al. [208] developed a MK-MMD approach to predict the RUL of bearings under different data distributions between the source and target domains. Initially, TCN was used to extract hidden spatio-temporal features. Hence, the MK-MMD method is utilised to minimise the distribution discrepancy of the spatiotemporal features between the source and target domains. Rathore et al. [209] introduced another MK-MMD approach based on Bi-LSTM algorithm. However, Zhang et al. [210] incorporated spatiotemporal characteristics into the learning procedures by first employing TCN with self-attention, followed by MK-MMD. Another cross-domain adaptation method based on deep transferable metric learning is proposed in [211] for predicting RULs, using TCN to extract domain-invariant representations and enhance transformation invariance. Mao et al. [212] proposed another metric discrepancy approach based on transfer component analysis (TCA). Initially, the vibration signal of the bearing is subjected to transformation using the Hilbert-Huang method. The utilisation of Pearson's correlation coefficient is employed to determine the relevant degradation state features. Then, TCA is employed to map the extracted features from the two domains. This mapping is then incorporated into the least-squares SVM (LSSVM) algorithm to predict the lifespan. Instead, a MMD approach is introduced in [213]. At first, SAE is utilised for the purpose of signal denoising. The MMD method is then employed to quantify the variation in distributions and obtain invariant features. Finally, a Bi-LSTM network is configured for the task of regression. Replacing the Bi-LSTM, the configuration of Bi-GRU is introduced in [214], with the same incorporation of MMD. In [215], the authors introduced a novel metric-learning strategy that incorporates a clustering mechanism and a temporal learning network to measure similarity between two domains.

One of the primary benefits of employing such methodologies is their capacity to be applied without the need for supervised (labelled) data. In addition, it can also be of use in different working conditions, but not in complex and harsh working environments where negative learning might occur. Accordingly, adversarial learning approaches can be beneficial for that purpose.

**Table 5**
Summary of the imbalanced data techniques.

| Method | Technique | Studies |
|---|---|---|
| Downsampling | Random downsampling | [150] |
| | Enhanced downsampling | [151,152] |
| | One-category learning strategy | [153,154] |
| Oversampling | Random oversampling | [155–157] |
| | SMOTE | [158–167] |
| | ADASYN | [170–172] |
| | Adversarial techniques | [174–194] |

### 2.4.3. Deep adversarial learning

Domain-adversarial training of neural networks, introduced in [216], has been widely used in various fields. However, unlike methods like GAN and VAE that create new feature samples to address data imbalances, in this case, adversarial approaches focus on learning invariance between two distributions.

The authors of [217] propose a GAN-based adversarial model for cross-domain PHM learning. At first, the authors developed a CNN network to extract the representative spatial features. Subsequently, they introduced a GAN architecture consisting of three distinct discriminators. The main task of the discriminator is to extract invariant features, carry out online monitoring and predict RUL. The authors of [218] propose an adversarial method focusing on invariant spatial feature learning. Initially, the CWT is used to transform the raw signal into 3D images. Subsequently, a CNN with self-attention is employed to facilitate the extraction of the influential features. Zou et al. [219] developed an adversarial approach based on CNN and autoencoders. Afterwards, the KL divergence is incorporated to quantitatively measure and minimise the discrepancy between the source and target domains. Finally, the extracted features are then employed for lifespan calculation using Bi-LSTM to encounter temporal dependencies.

The authors in [220] present a novel integration between metric evaluation and adversarial approaches. The research proposed the utilisation of Mk-MMD and adversarial learning in a parallel structure for a prognostic task. The Mk-MMD method is employed to evaluate the disparity between distributions, while adversarial learning is utilised to acquire invariant features. In the same way, the authors in [221] and [222] employed a methodology that combined adversarial techniques and discrepancy metrics, utilising a CNN for the purpose of bearing prognostics. This method was used to address the difficulties arising from different working conditions and incomplete run-to-failure (RTF) target data. A cross-operating condition degradation knowledge learning method is proposed in [223], constructing a shared latent feature space across different operating conditions. This method effectively extracts bearing degradation features and weakens cross-domain feature differences. To get domain-invariant features, the method uses time–frequency representation samples and joint dictionary matrix factorization. Another multi-source adversarial online regression (MAOR) method is proposed in [224] to address the limitation of the health prognosis of target data under unknown conditions. It uses pseudo-domain extension, domain-level adaptation, and feature-level adaptation to build robust domain-invariant features. An offline-online prediction framework is developed to predict online target data streams and update the online model. Finally, Adversarial methods are capable of learning dynamically without the need for manual settings. However, combining them with another feature learning method may result in a more robust approach.

### 2.4.4. Advanced approaches

Few-shot learning is another approach that aims to overcome the challenge of poor learning for limited samples in a given task. It aims to address the challenge of prognosis with rare and sparse training data in the target domain and under different working conditions and different fault modes in the source and target domains. The proposed approach can improve estimation results in the source domain while conducting generalisations to ensure a reliable prognosis. Cheng et al. [225] propose a PHM approach based on a multimodal few-shot learning method (MMFSL). The multimodal method uses time-series data for temporal representation and images for spatial learning. Leveraging the scarce fault modes of the multimodal representation, the few-shot strategy proved to efficiently provide a reliable PHM. Another few-shot PHM learning method is introduced in [226]. It incorporates spatial features derived from 2D images to provide cross-domain CdM. Another approach is introduced in [227], which suggests a feature disentanglement and restoration (FDR) few-shot method for initial fault observation, aiming to deal with the target of early prognosis. The method processes vibration signals, pulls out general features that are relevant to the task, puts together task-specific features, and sorts nonlinear relationships between features for a PHM task. Jang et al. [228] introduce a novel health representation learning method based on a Siamese network that prevents overfitting by incorporating a multitask learning scheme. The method uses the learned embedding space, enabling robust RUL prediction.

Meta-learning is another approach that has recently been studied in PHM. Meta-learning aims to use knowledge from previous tasks to help with new tasks. Dividing a large dataset into small-sample tasks allows the model to adapt rapidly to the current tasks. The authors of [229] propose a *meta*-network framework for RUL prediction, addressing distribution discrepancies in transient working conditions. An iterative *meta*-network pruning algorithm is introduced that calculates the *meta*-gradients of convolutional kernels, deleting unimportant connections. Yang et al. [230] present a novel MetaDESK model for RUL prediction with limited data, based on *meta*-learning and a deep sparse kernel network. The model uses a Gaussian process, variational inference, and KL-divergence of sparse approximation to estimate latent variables and reduce overfitting. It also combines common knowledge and task-specific information to improve decision-making. Sun et al. [231] proposed a lightweight Bi-LSTM based on automated model pruning for bearing RUL prediction. The method uses a learning technique that rewards correct actions to find and remove unnecessary elements, eliminating the need for manual search and choosing the best pruning structures. The lightweight Bi-LSTM accurately forecasts bearing wear patterns using an RMS value as a health indicator. It achieves a 36 % reduction in model size and a 3 % enhancement in prediction accuracy over the initial Bi-LSTM. In the same context, [232] introduced *meta*-learning for prognostics with limited data and variable working conditions. It extracts degradation indicators that remain consistent across different domains from time-frequency images and time-series data. For this, two adaptation networks are configured, named *meta*-CNN and meta GRU, for handling spatial and temporal features, respectively. A *meta*-learning algorithm, which is not dependent on specific models, is used in [233] to adjust model parameters swiftly based on test samples. It also recommends artificial task sets resembling *meta*-RUL tasks, enhancing the method's applicability across various scenarios. In essence, the issue of bearing prognosis in a dynamic working environment remains a research question and is deemed essential for real-world industries.

## 3. Prediction task using hybrid approaches

After reviewing recent advancements and methods of effective feature learning for bearing prognosis, this section aims to provide a comprehensive review of end-to-end prognostic methods, including ensemble approaches, rather than focusing solely on the feature learning stage. A prediction task is always subsequent to the feature-learning stage, which involves modelling the learned features and mapping each of the extracted features to a corresponding lifespan estimation value. Prognostic approaches can be categorized into shallow and neural network methods, similar to feature-learning methods, based on the method used for feature learning and regression tasks. This subsection aims to provide an end-to-end examination of bearing prognostic studies instead of focusing on feature learning.

For shallow methods, support vector regression (SVR) [83], decision trees (DTs) [85], random forests (RF) [82], XG-boost [234], and ELM are considered common shallow bearing regression techniques in the state-of-the-art. The SVR tries to figure out how an input sample affects the output, which in this case is the predicted RUL value. It accomplishes this by assuming that the joint distribution between the input vibration sample and the predicted lifespan is unknown. SVR draws an insensitive tube [235], so the penalty on samples inside this region of the insensitive tube does not apply. Otherwise, a penalty function applies to samples far away from real values. The support vectors are then employed to measure the degradation severity of an input sample and calculate the corresponding estimate of remaining life [236].

The authors in [82–85,237–239] applied SVR to bearing RUL prediction. In [83] and [84], SVR is applied subsequent to wavelet decomposition and time–frequency representation. However, Benkedjouh et al. [84] utilised isometric feature mapping reduction (IFMR) for dimensionality reduction and faster computation. In [237,239], signal decomposition using EMD is applied prior to SVR modelling. The aim is to eliminate noise samples and extract relevant features before training SVR for better regression results. However, Cao et al. [239] calculated RMS and kurtosis for more reliable supporting vector construction. Additionally, Saidi et al. [238] calculated spectral kurtosis for a wind turbine high-speed shaft bearing. Afterwards, time domain features are calculated and then selected using monotonicity and trend ability before training SVR. At last, SVR is of great use in PHM and is memory efficient, so it can be adopted for online modelling [236].

DT is a supervised tree-life model that aims to perform regression or class classification through its hierarchical structure. A typical DT model consists of a root node, branches, and leaf nodes representing a response to a regression task. The path from the root node to the leaf nodes through interval nodes identifies the machinery state according to the objective task. RF is a variant developed from DT and is widely adopted for bearing prognostics. Instead of having one tree that represents the whole vibration dataset, RF aims to build several trees and then provide the mean prediction of independent trees [240]. This can be effective for noise elimination that may be encountered using traditional DT. Ren et al. [241] applied DT as an approach to avoid overfitting in bearing prognostics. Singh et al. [240] conducted a comparative study to validate the performance of RF and ordinary DT in bearing RUL prediction as a result of wear phenomena. At first, time domain features are extracted, and then the Pearson correlation methodology is applied to select the relevant features. In the regression task, the authors demonstrated that RF outperformed DT in terms of root mean square error quantification. In [242], the authors conducted an approach to study the impact of temporal feature extraction on the bearing RUL prediction task. In this context, RF has proven to be an effective modelling approach that can incorporate the influential temporal features in the regression task with noise elimination. In [82], an SVR-RF approach is studied. At first, statistical features are extracted, followed by PCA for feature ranking. Afterwards, the performance of SVR and RF is studied solely. The pattern learned by each of them is then fed to the Weibull Hazard Function for calculating the corresponding RUL. Notably, the RF outperformed the SVR.

ELM is a shallow learning algorithm that aims to speed up the process of traditional feed-forward neural networks. Unlike neural networks, ELM does not depend on gradient descent, which requires extensive learning iterations for weights and parameters and may cause overfitting or fall into local extrema values [243,244]. Thus, it can be deployed for online monitoring. ELM consists of a single hidden NN layer in which the number of hidden nodes in this layer is chosen randomly, and a linear or nonlinear input–output relation can be drawn based on the selected activation function. In bearing prognostics, ELM is commonly employed as a regression mechanism subsequent to a feature learning stage. For instance, the authors in [245] extracted the standardized RMS [246] for fitting a bearing health index, and then a moving average filter for pattern smoothing was developed. Subsequently, ELM is employed for RUL prediction. Similarly, [246,247] adopted an ELM prognostic approach based on relative RMS (RRMS) [248]. RRMS is able to record degradation at an early stage due to its sensitivity compared to RMS. Thus, once degradation is observed, the ELM can estimate the corresponding RUL. However, in [247], Pearson-correlation coefficient combined entropy weight methods are developed to assure degradation and eliminate anomaly readings before training ELM. A variant of ELM, called kernel ELM, is adopted in [249]. Inspired by the kernel SVM strategy, a kernel ELM is proposed in [244], which has demonstrated superiority compared to ELM [249]. Table 6 presents the techniques and shallow regression models adopted in some of the discussed frameworks.

**Table 6**
Comparative analysis of different ml approaches.

| Study | Feature extraction | ML method |
|---|---|---|
| [83] | WT + time–frequency analysis | SVM |
| [84] | wavelets + Statistical analysis on the time-domain + IFMR | SVM |
| [60] | time–frequency analysis + PCA + Linear Discriminant Analysis | Linear Regression |
| [237] | EMD | SVM |
| [250] | Spectral Kurtosis | SVM |
| [239] | EMD + extraction of RMS and Kurtosis on the time-domain | SVM |
| [85] | Hamiltonian harmonic oscillation | SVM<br>KNN<br>DT |
| [81] | Statistical analysis on the time-domain + feature ranking using DT | RF<br>Gradient Boosting |
| [82] | Statistical analysis on the time-domain + PCA | SVM<br>RF |

On the other hand, CNN, RNN, Neuro-fuzzy systems, and their variants are considered common neural network prognostic techniques. Initially, the RBM was employed for feature extraction in [251]. The loss function was modified by incorporating a slope component, which serves to incentivise the features to acquire knowledge about the trend inherent in the degradation pattern. Subsequently, the acquired features were inputted into a self-organizing map in order to predict the RUL. CWT was employed in [55] to convert unprocessed vibration inputs into time–frequency image characteristics. The construction of Le-Net-5 CNN network, aimed to extract spatial information from images and subsequently map them to the bearing's health score. The initial extraction of tri-domain characteristics from the raw vibration signal was conducted in [79]. The degradation-sensitive characteristics were further selected using a feature selection method that relied on correlation and monotonicity. Subsequently, an LSTM model was developed to associate the chosen characteristics with the bearing's health score.

Wang et al. [64] utilised the LeNET-5 architecture to forecast RUL by extracting spatial features. In that study, the FFT was utilized to extract the relative features, and subsequently, the CWT was employed to convert the 1D signal into 2D pictures. Another prognostic strategy utilizing CNN is presented in reference [120]. However, this approach incorporates features retrieved from various levels of the convolutional network for RUL prediction rather than solely relying on the final dense layer. An alternative CNN methodology is employed in [118]. The usefulness of a unique feature learning method that relies on the spectrum main energy vector utilising FFT has been demonstrated. In [75], spatial features were used to calculate RUL. However, the initial step involved applying feature denoising using STFT. Ren et al. [252] utilised a deep neural network to estimate health index and longevity by inputting a fusion matrix consisting of time-domain information and frequency-domain features. In [253], another deep NN was set up. However, the features were first chosen using autoencoders. An estimation of temporal life prediction is performed in reference [241]. Initially, RBM was employed to compress vast time–frequency information, enabling GRU to eliminate duplicate features during the regression phase. In reference [212], a spatiotemporal prediction model is shown. The utilization of CNN was employed to extract spatial characteristics subsequent to their selection by SVD. These extracted features were subsequently inputted into an LSTM network for the purpose of predicting RUL.

Additionally, the investigation of deep structure networks and fuzzy inference systems is also a significant area of focus in the field of prognostic studies. The concept of fuzzy sets was initially introduced by L. A. Zadeh in 1965 in [254], with regards to the utilisation of FIS and PNN. Mamdani (1975) and Sugeno (1985) introduced two distinct categories of fuzzy inference systems (FIS). The ANFIS combines the principles of artificial neural networks (ANNs) and fuzzy logic, enabling the utilization of the advantages of both in a unified approach [255]. In [256], ANFIS is utilised to combine and extract the relative characteristics of a compact latent space for data arrays from multiple sensors. The authors in [257] employed a CdM methodology utilizing ANFIS. The ANFIS method is utilized to execute IF-THEN rules on extracted features through the EEMD algorithm, which is subsequently regarded as a regressor algorithm for RUL prediction. The authors of [258] have developed a unique ANFIS that incorporates a particle swarm structure for power coefficient prediction. Ultimately, the integration of ANFIS and RNN is employed to forecast the forthcoming value associated with the health index [259]. Table 7 summarises some of the frameworks that utilised DL regression models.

## 4. Benchmark bearing datasets

To help researchers validate their studies, several labs introduced RTF datasets for research purposes. This section presents five common experiments in the literature and the main characteristics of each dataset.

### 4.1. FEMTO dataset

This dataset was submitted at the IEEE International Conference on PHM 2012 for the prognostic challenge and was provided by the Franche-Comté Electronics Mechanics Thermal Science and Optics–Sciences and Technologies institute [265]. FEMTO dataset comprises RTF analysis for 17 rolling element bearings. To accelerate the degradation of the bearings within hours, bearings were exposed to a radial force that surpassed their maximum dynamic load. The experimental setup is shown in Fig. 5. Throughout testing, the bearing's speed remained constant. Bearing temperatures and vibrations were recorded using two accelerometers and a thermocouple for accurate data collection. When the vibration signal surpasses 20 g, it indicates that the bearing has reached the end of its useful life. The experiment captures data at a rate of 25,600 samples per second to ensure detailed measurements. Each sample has a period of 0.1 s; therefore, each sample has a total of 2560 points. The data is recorded in intervals of 10 s to capture specific information during the experiment. The test is stopped to prevent damage when the signal's amplitude surpasses a predefined threshold. The dataset consists of 17 RTF datasets that represent bearings in three different operating scenarios, as outlined in Table 8. The dataset has commonly been adopted in the literature for prognosis studies as in [61,75,266,267].

**Table 7**
Summary of some DL approaches.

| Study | FE | Model |
|---|---|---|
| [29] | EMD | SVM Vs LSTM |
| [260] | Time and frequency domain features | Multi-Layer Perceptron (MLP) |
| [51] | WPT + statistical features: mean, variance, kurtosis, RMS, crest factor | MLP |
| [79] | Time and frequency domain features | RNN |
| [253] | Time and frequency domain features + AEs | MLP |
| [261] | EEMD | CNN |
| [262] | CNN | Bi-LSTM |
| [122] | ResNET | FC layer |
| [263] | Hilbert-Huang Transform | CNN |
| [128] | CNN | LSTM |
| [55] | CWT | CNN |
| [264] | Self-organizing maps | backpropagation |
| [112] | CWT | 3DNN |
| [78] | Time and frequency domain features + wavelets + CNN | GPR |
| [75] | Time and frequency domain features | MSCNN |
| [72] | STFT | CNN |
| [120] | Time and frequency domain features | MSCNN |
| [124] | GCN | GRU |
| [80] | Time and frequency domain features | GCN |

### 4.2. XJTU-SY dataset

The XJTU-SY rolling dataset is another experiment that has been used for bearing prognostic studies [268]. Fifteen different bearing datasets from a testing platform were recorded using vibration sensors under various operating conditions. Similar to the previous experiment, radial force and motor speed are the main factors for simulating different operating conditions. Two PCB 352C33 accelerometers are utilized to capture vibration signals at a sampling rate of 25.6 kHz in the horizontal and vertical directions, respectively, to monitor the vibrations. Each sample consists of 32,768 data points collected at a 1-minute sampling interval. Table 9 details the specific working conditions of the tested bearings. Fig. 6 illustrates the experimental setup of the dataset. The XJTU-SY dataset has mainly been used for bearing prognostics studies [61,120,267,269].

### 4.3. IMS dataset

The IMS bearing dataset [94] is the third benchmark dataset that has been widely used in research. This dataset comprises three experimental datasets that are obtained from a test rig. On a shaft, four bearings were placed for each experiment. The rotational speed was maintained at 2,000 r/min with the aid of an AC motor attached to the shaft. A force mechanism applied a 6,000-pound radial load to the shaft and bearings. A sample of 20,480 vibration data points was acquired every 10 min. The data was collected at a sample rate equal to 20 KHz. The test ended when debris stuck to the magnetic plug reached a specific amount, causing an electrical switch to activate. Fig. 7 displays the setup of the experiment.

For the first experiment, two accelerometer sensors were placed on each bearing. This made eight data channels. For the other two experiments, only one accelerometer was mounted on each bearing. At the end of the first experiment, bearing 3 had damage to the inner race, and bearing 4 had issues with the roller elements. At the end of the second and third experiments, the outer races of bearing 1 and bearing 3 experienced structural damage, respectively. The dataset has been used for condition monitoring and PHM studies such as in [62,270–273].

### 4.4. CRWU dataset

Data acquired from Case Western Reserve University (CRWU) has considered standard in the PHM community [274]. It comprises a 2-horsepower reliance electric motor that drives a shaft, on which a torque transducer and encoder are affixed. Torque is applied to the shaft using a dynamometer and an electronic control system. Fig. 8 illustrates the experimental setup of the test rig, depicting the arrangement of the equipment used in the experiment.

Electro discharge machining was employed to introduce defects to the drive and fan-end bearings of the motor, specifically the SKF deep-groove ball bearings: 6205-2RS JEM and 6203-2RS JEM, with diameters ranging from 0.007 to 0.028 in.. The defects were intentionally introduced on the rotating components and on the inner and outer rings. Subsequently, each defective bearing was replaced individually on the experimental setup, which was then run at a constant speed with motor loads varying from 0 to 3 horsepower (equivalent to motor speeds around 1797 to 1720 rpm). Table 10 displays the pertinent information on bearing specifications and rates of faults. Acceleration was recorded vertically on the housing of the drive-end bearing throughout each test. In some tests, acceleration was also measured vertically on the fan-end

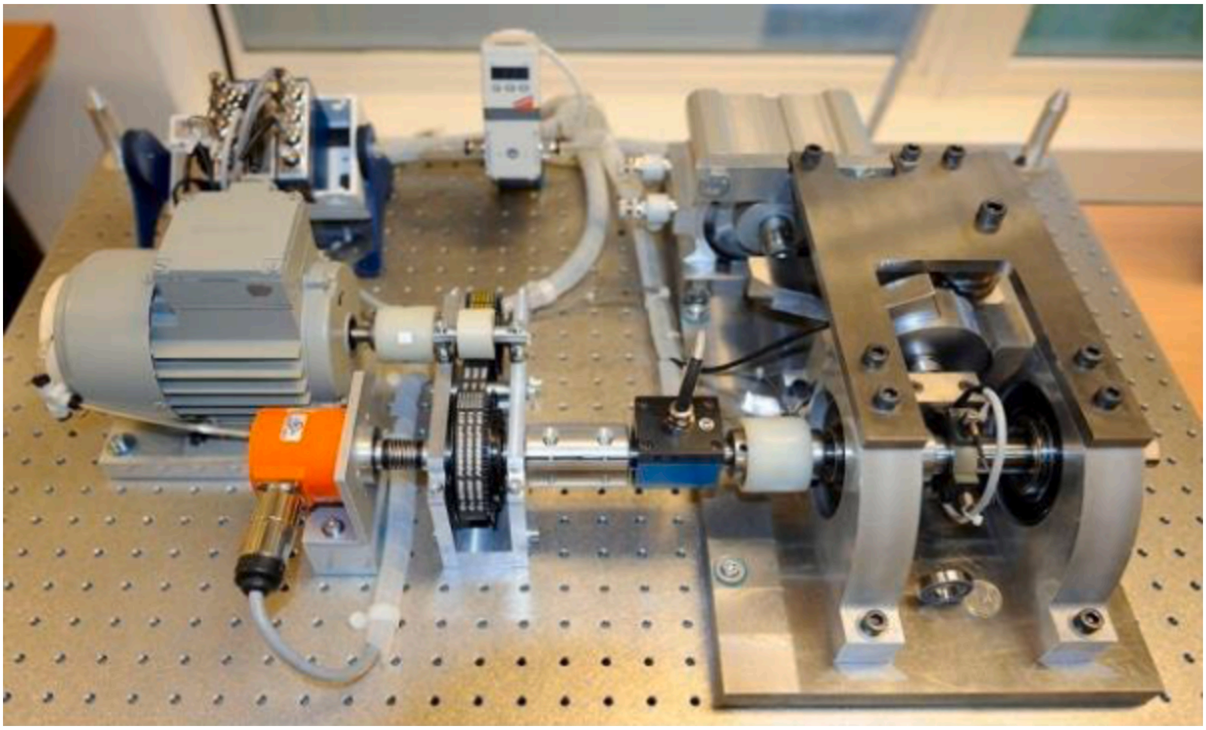

**Fig. 5.** Laboratory setup of FEMTO Dataset.

**Table 8**
Characteristics of the FEMTO dataset.

| Characteristics | 1 | 2 | 3 |
|---|---|---|---|
| Load (N) | 4000 | 4200 | 5000 |
| Speed (rpm) | 1800 | 1650 | 1500 |
| Training | B-1.1<br>B-1.2 | B-2.1<br>B-2.2 | B-3.1<br>B-3.2 |
| Validation/ Testing | B-1.3<br>B-1.4<br>B-1.5<br>B-1.6<br>B-1.7 | B-2.3<br>B-2.4<br>B-2.5<br>B-2.6<br>B-2.7 | B-3.3 |

**Table 9**
XJTU-SY bearing sets summary and working conditions.

| Characteristics | First working conditions | Second working conditions | Third working conditions |
|---|---|---|---|
| Load (N) | 12 K | 11 k | 10 k |
| Speed (rpm) | 2100 | 2250 | 2400 |
| Bearing sets | B-1.1<br>B-1.2<br>B-1.3<br>B-1.4<br>B-1.5 | B-2.1<br>B-2.2<br>B-2.3<br>B-2.4<br>B-2.5 | B-3.1<br>B-3.2<br>B-3.3<br>B-3.4<br>B-3.5 |

bearing housing and, on the motor, –supporting base plate. The tests were conducted using sampling rates of 12 kHz and 48 kHz. The dataset has been extensively employed in studies on bearing diagnostics [21,275,276].

BPFI and BPFO are the ball pass frequencies of the inner and outer races, respectively. FTF represents the fundamental train frequency (cage speed) and BSF is the ball spin frequency.

### 4.5. Paderborn university dataset

The Paderborn University dataset is a publicly accessible repository for bearing data [277]. The data collection setup includes a test motor, measuring shaft, bearing module, and load motor, shown in Fig. 9. The collection includes both synchronous vibration and motor current measurements. The device uses one accelerometer, two current sensors, and one thermocouple. The vibration signals are captured at a sample frequency of 64 kHz. A total of 32 bearings were used in the experiments, including 6 fully functional bearings and 26 damaged bearings. Out of the damaged bearings, 12 were intentionally damaged, while the others were naturally damaged due to accelerated tests. This dataset has been used in bearing diagnostics research [278–280].

Finally, Table 11 aims to summarize the reviewed datasets based on the adopted sensor type and sample frequency of each dataset.

## 5. Challenges and future directions

Despite the great advancements that have been achieved in bearing prognostics, there are some challenges and future trends that can lead to better results and applicability in real industries. These can be sorted as:

1- **Multi-modal learning:** Experimental setups in the literature for bearing prognostics are based solely on vibration signals. However, an accelerometer sensor, which detects vibration signals, may not capture all factors contributing to bearing wear, such as temperature and lubrication [19,281]. Furthermore, the sensor is unable to accurately detect the initial and early-stage defects that may develop in bearing components during operation [9]. Given this challenge, thermal imaging methods have been evaluated and shown to be effective in diagnostic procedures. Specifically, thermal imaging was used to categorize bearing failures by recording temperature readings. Choudhry et al. [282] correctly classified six different faults based on thermal imaging using Le-Net CNN [18]. The authors in [283] demonstrated a successful diagnostic-bearing approach using RF. However, a study in [284] proved that SVM can perform better compared to RF on the same dataset. A novel method in [285]

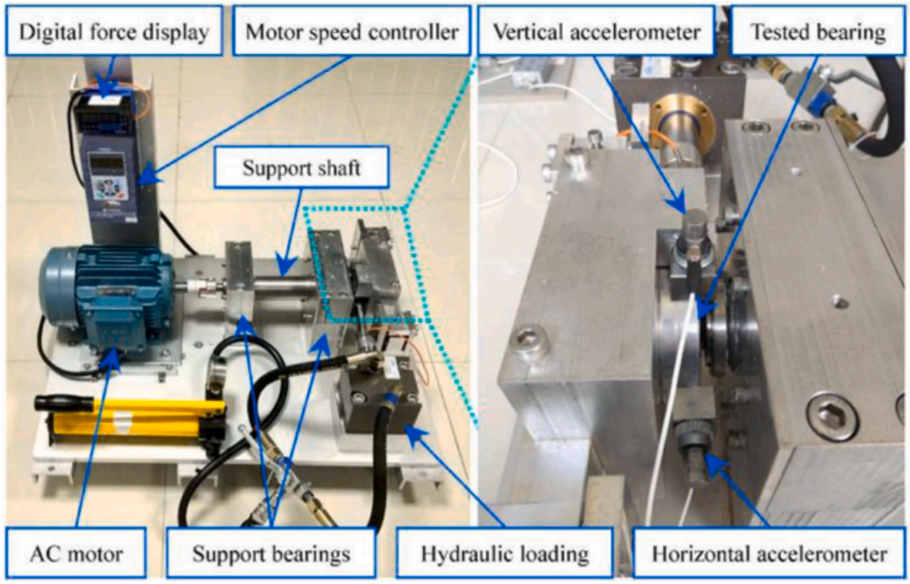


**Fig. 6.** Laboratory setup of xjtu-sy datset.

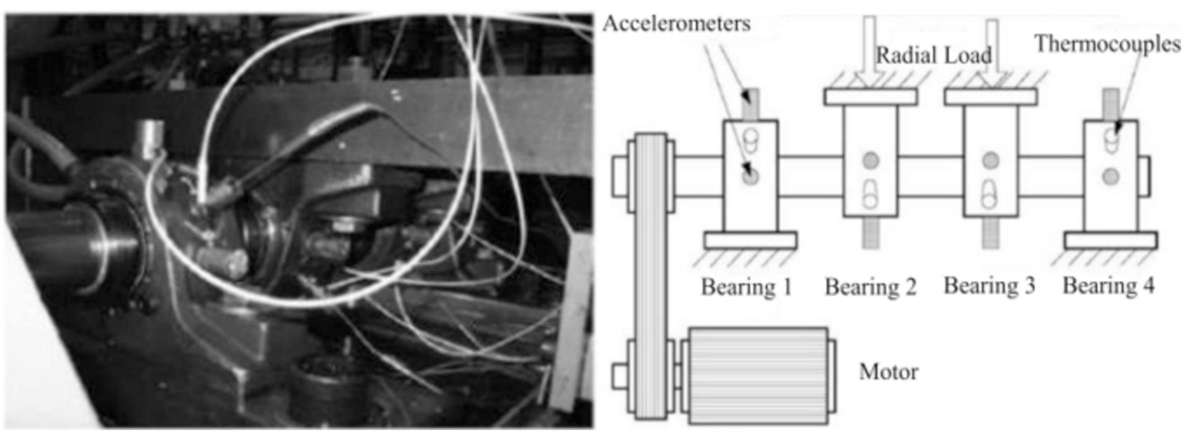


**Fig. 7.** Laboratory setup of ims dataset.

introduced a novel approach that combines vibration signals with thermal images. This method was 100 % accurate at finding the fault point of a bearing that was starting to wear out early on. So, the ability to add more instruments, either using imaging or traditional sensors, to measure more meaningful characteristics of bearings while they are running is an area that needs to be looked into for a predictive approach.

2- **Management of time complexity:** Now that data-driven methods are being used in the field of predictive maintenance (PdM), the proposed algorithms are getting more complicated. Accordingly, the encounter with time and computational complexity is becoming important and should be considered. Real-world applications often involve massive amounts of data, emphasising the significance of considering time and computational aspects in the learning process. However, there may be a trade-off between the performance of complex algorithms and the training time required. Therefore, the interest in compressed models for real-time monitoring is crucial due to the importance of available computing resources in industry.

3- **System prognostics rather than component prognostics:** Most prognostic methods in the literature are conducted for specific components such as bearings [135], gearboxes [286], and turbofans [103]. This is because there aren't any benchmark datasets available for research on real systems. This may lead to the inapplicability of such approaches in real and complex industries. Thus, availing data from real manufacturing systems for research and considering the possibility of developing models that consider the condition monitoring of several components at a time with condition reasoning is a potential aspect of research to mimic real-world industries.

4- **Measurement of uncertainty:** Many methods primarily concentrate on forecasting the average values of RUL, yet it is crucial to consider model accuracy for making informed decisions in practical scenarios. In this context, precision is important in order to make better-informed decisions. So, it would be a point of interest to conduct more studies to measure uncertainty and study the progression of inferences such as those conducted using Bayesian models [108,287,288].

## 6. Conclusion

This paper provides a thorough review of recent advancements in feature learning methods for a successful bearing prognostic approach. The authors classified the methods into shallow and deep learning approaches. Further, the paper provides a new taxonomy for feature learning methods based on temporal learning, spatial learning, and spatiotemporal learning. The paper then addressed feature-learning challenges under an imbalanced ratio of data samples and classified

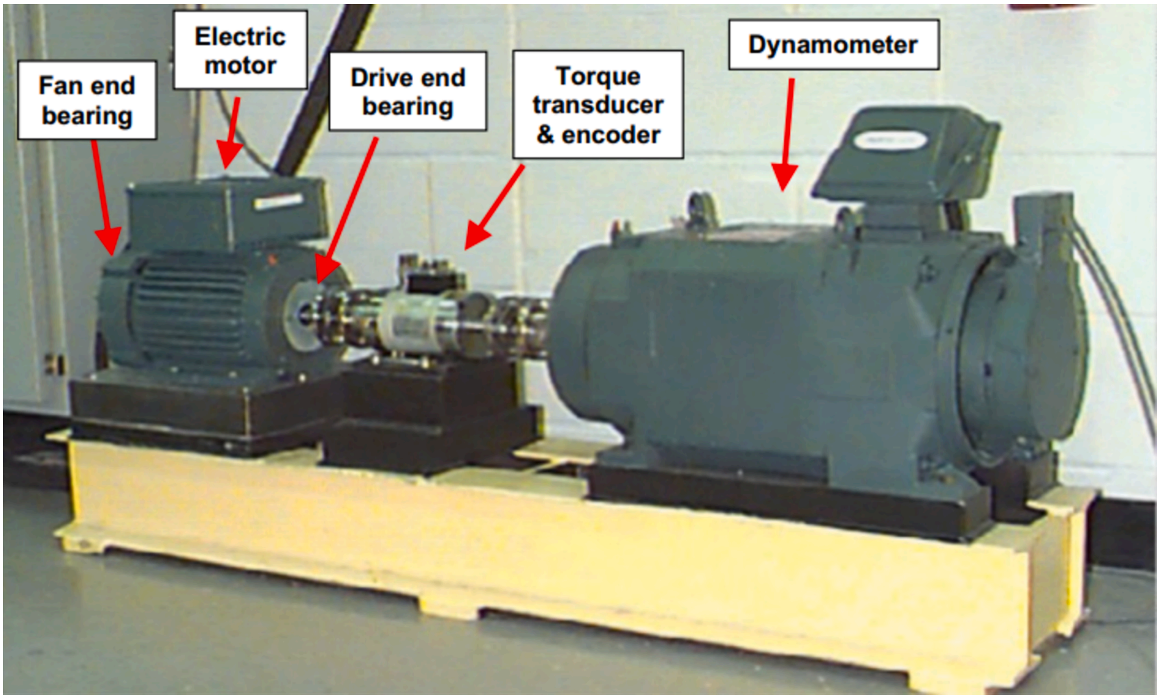


**Fig. 8.** Laboratory setup of crwu dataset.

**Table 10**
Bearings specifications of CRWU dataset.

| Position on rig | Model no. | Fault frequencies (multiple of shaft speed) | | | |
|---|---|---|---|---|---|
| | | BPFI | BPFO | FTF | BSF |
| Drive end | SKF-6205-2RS JEM | 5.415 | 3.585 | 0.3983 | 2.357 |
| Fan end | | 4.947 | 3.053 | 0.3816 | 1.994 |

the literature methods into downsampling and oversampling methods, including the merits and drawbacks of each technique. Additionally, the paper explained the challenges and methods used for feature representation under different operational conditions. Moreover, the paper presents recent ensemble regression techniques for reliable bearing prognostics. It can be concluded that recent research has demonstrated that feature learning is the most crucial phase in bearing prognostics. Additionally, each of the common data-driven techniques has an effect on the type of learned characteristic in the latent space. The paper then outlined the benchmark datasets frequently used in studies on bearing prognostics. Finally, this paper provided insights into the challenges and future directions in this field, aiming to assist new researchers and practitioners in identifying opportunities for future research. Overall, this research provides insights into the study of bearing prognostic challenges from a data-driven methods perspective and lays the groundwork for future investigations.

## CRediT authorship contribution statement

**Ahmed Ayman:** Writing – original draft, Methodology. **Ahmed Onsy:** Writing – original draft, Methodology. **Omneya Attallah:** Writing – review & editing, Supervision. **Hadley Brooks:** Supervision. **Iman Morsi:** Supervision.

## Declaration of competing interest

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

**Table 11**
Summary of the discussed datasets.

| Dataset | Sensors | Sample Frequency |
|---|---|---|
| FEMTO | Accelerometer and thermocouple | 25.6KHz |
| XJTU-SY | Accelerometer | 25.6KHz |
| IMS | Accelerometer | 20KHz |
| CWRU | Accelerometer | 12 and 48KHz |
| Paderborn | Accelerometer and thermocouple | 64KHz |

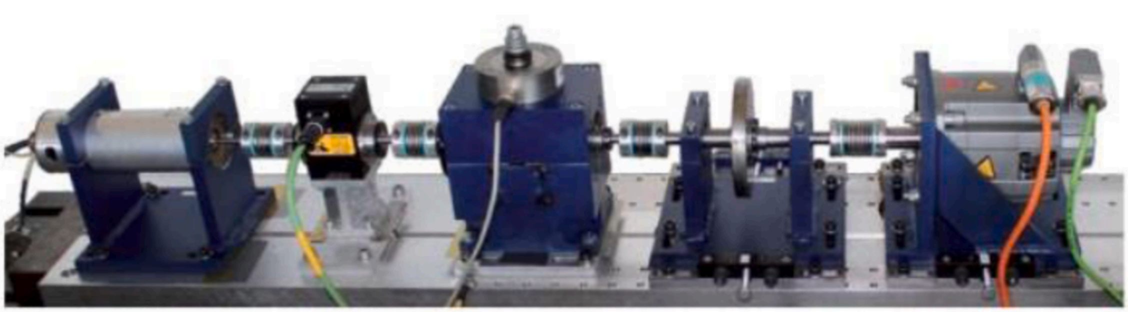

**Fig. 9.** Laboratory setup of Paderborn dataset.

## Data availability

No data was used for the research described in the article.

## References


[1] W. Qiao, D. Lu, A survey on wind turbine condition monitoring and fault diagnosis—Part I: components and subsystems, IEEE Trans. Ind. Electron. 62 (10) (2015) 6536–6545.
[2] N.S. Jammu, P.K. Kankar, A review on prognosis of rolling element bearings, Int. J. Eng. Sci. Technol. 3 (10) (2011) 7497–7503.
[3] A. Rai, S.H. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, Tribol. Int. 96 (2016) 289–306.
[4] R.N. Bell, D.W. McWilliams, P. O'donnell, C. Singh, S.J. Wells, Report of large motor reliability survey of industrial and commercial installations. I, IEEE Trans. Ind. Appl. 21 (4) (1985) 853–864.
[5] M. Hamadache, J.H. Jung, J. Park, B.D. Youn, A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: shallow and deep learning, JMST Adv. 1 (2019) 125–151.
[6] T. Han, J. Pang, A.C.C. Tan, Remaining useful life prediction of bearing based on stacked autoencoder and recurrent neural network, J. Manuf. Syst. 61 (2021) 576–591.
[7] D. Wang, K.-L. Tsui, Q. Miao, Prognostics and health management: a review of vibration based bearing and gear health indicators, IEEE Access 6 (2017) 665–676.
[8] M. Kordestani, M. Saif, M.E. Orchard, R. Razavi-Far, K. Khorasani, Failure prognosis and applications—a survey of recent literature, IEEE Trans. Reliab. 70 (2) (2019) 728–748.
[9] A. Abufroukh, A. Onsy, and I. Sherrington, "Intelligent Management of Wind Turbine Bearings," 2016.
[10] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.
[11] L. Saidi, J. Ben Ali, E. Bechhoefer, M. Benbouzid, Particle filter-based prognostic approach for high-speed shaft bearing wind turbine progressive degradations, in: IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2017, pp. 8099–8104.
[12] Y. Qian, R. Yan, Remaining useful life prediction of rolling bearings using an enhanced particle filter, IEEE Trans. Instrum. Meas. 64 (10) (2015) 2696–2707.
[13] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, N. Zerhouni, Particle filter-based prognostics: review, discussion and perspectives, Mech. Syst. Sig. Process. 72 (2016) 2–31.
[14] F. Shen, J. Xu, C. Sun, X. Chen, R. Yan, Transfer between multiple working conditions: a new TCCHC-based exponential semi-deterministic extended Kalman filter for bearing remaining useful life prediction, Measurement 142 (2019) 148–162.
[15] L. Cui, X. Wang, Y. Xu, H. Jiang, J. Zhou, A novel switching unscented Kalman filter method for remaining useful life prediction of rolling bearing, Measurement 135 (2019) 678–684.
[16] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T.A. Tameghe, G. Ekemb, Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges, Energies (Basel) 7 (4) (2014) 2595–2630.
[17] E. Artigao, S. Koukoura, A. Honrubia-Escribano, J. Carroll, A. McDonald, E. Gómez-Lázaro, Current signature and vibration analyses to diagnose an in-service wind turbine drive train, Energies (Basel) 11 (4) (2018) 960.
[18] B. Corley, S. Koukoura, J. Carroll, A. McDonald, Combination of thermal modelling and machine learning approaches for fault detection in wind turbine gearboxes, Energies (Basel) 14 (5) (2021) 1375.
[19] W. Qiao, D. Lu, A survey on wind turbine condition monitoring and fault diagnosis—Part II: signals and signal processing methods, IEEE Trans. Ind. Electron. 62 (10) (2015) 6546–6557.
[20] M.S. Kan, A.C.C. Tan, J. Mathew, A review on prognostic techniques for non-stationary and non-linear rotating systems, Mech. Syst. Sig. Process. 62 (2015) 1–20.
[21] X. Li, Q. Ding, J.-Q. Sun, Remaining useful life estimation in prognostics using deep convolution neural networks, Reliab. Eng. Syst. Saf. 172 (2018) 1–11.
[22] D. Kim, S. Lee, D. Kim, An applicable predictive maintenance framework for the absence of run-to-failure data, Appl. Sci. 11 (11) (2021) 5180.
[23] T. Zonta, C.A. Da Costa, R. da Rosa Righi, M.J. de Lima, E.S. da Trindade, G.P. Li, Predictive maintenance in the Industry 4.0: a systematic literature review, Comput. Ind. Eng. 150 (2020) 106889.
[24] A. Onsy, A new acoustic emission wireless monitoring system; an experimental validation of bearing condition monitoring, SAE Int. J. Aerosp. 6 (2013) 534–539, 2013–01–2221.
[25] L.L. Mbagaya J.K. Kimotho J.G. Njiri A review on prognosis of rolling element bearings operated under non-stationary conditions In: Proceedings of the Sustainable Research and Innovation Conference 2022 141 147.
[26] Z. Zhao, J. Wu, T. Li, C. Sun, R. Yan, X. Chen, Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: a review, Chinese J. Mech. Eng. 34 (1) (2021) 1–29.
[27] J. Chen, R. Huang, Z. Chen, W. Mao, W. Li, Transfer learning algorithms for bearing remaining useful life prediction: a comprehensive review from an industrial application perspective, Mech. Syst. Sig. Process. 193 (2023) 110239.
[28] N. E. Huang et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, In: Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, 454(1971), pp. 903–995, 1998.
[29] Z. Chen, Y. Liu, S. Liu, Mechanical state prediction based on LSTM neural netwok, in: IEEE, 2017, 2017, pp. 3876–3881.
[30] M. Liu, X. Yao, J. Zhang, W. Chen, X. Jing, K. Wang, Multi-sensor data fusion for remaining useful life prediction of machining tools by IABC-BPNN in dry milling operations, Sensors 20 (17) (2020) 4657.
[31] H. Gu, J. Zhao, X. Zhang, Hybrid methodology of degradation feature extraction for bearing prognostics, Eksploatacja i Niezawodność 15 (2) (2013) 195–201.
[32] Y. Zhan, S. Sun, X. Li, F. Wang, Combined remaining life prediction of multiple bearings based on EEMD-BILSTM, Symmetry (Basel) 14 (2) (2022) 251.
[33] R. Guo, Y. Wang, H. Zhang, G. Zhang, Remaining useful life prediction for rolling bearings using EMD-RISI-LSTM, IEEE Trans. Instrum. Meas. 70 (2021) 1–12.
[34] A.A. Tabrizi, L. Garibaldi, A. Fasana, S. Marchesiello, Performance improvement of ensemble empirical mode decomposition for roller bearings damage detection, Shock Vib. (2015) 2015.
[35] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adaptive Data Anal. 1 (01) (2009) 1–41.
[36] V.H. Nguyen, J.S. Cheng, Y. Yu, V.T. Thai, An architecture of deep learning network based on ensemble empirical mode decomposition in precise identification of bearing vibration signal, J. Mech. Sci. Technol. 33 (2019) 41–50.
[37] R. Guo, Y. Wang, Remaining useful life prognostics for the rolling bearing based on a hybrid data-driven method, Proc. Inst. Mech. Eng., Part I: J. Syst. Control Eng. 235 (4) (2021) 517–531.
[38] A. Rai, S.H. Upadhyay, An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen-Rényi divergence, Appl. Soft Comput. 71 (2018) 36–50.
[39] Y.-H. Wang, C.-H. Yeh, H.-W.-V. Young, K. Hu, M.-T. Lo, On the computational complexity of the empirical mode decomposition algorithm, Physica A 400 (2014) 159–167.
[40] H. Liu, H. Tian, Y. Li, Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, Wavelet Packet-MLP and Wavelet Packet-ANFIS for wind speed predictions, Energy Convers. Manag. 89 (2015) 1–11.
[41] W. Sun, M. Liu, Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China, Energy Convers. Manag. 114 (2016) 197–208.
[42] W. Jiang, J. Zhou, H. Liu, Y. Shan, A multi-step progressive fault diagnosis method for rolling element bearing based on energy entropy theory and hybrid ensemble auto-encoder, ISA Trans. 87 (2019) 235–250.
[43] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2013) 531–544.
[44] K. Li, L. Su, J. Wu, H. Wang, P. Chen, A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine, Appl. Sci. 7 (10) (2017) 1004.
[45] W. Song, H. Liu, E. Zio, Long-range dependence and heavy tail characteristics for remaining useful life prediction in rolling bearing degradation, App. Math. Model. 102 (2022) 268–284.
[46] C. Yi, S. Li, T. Huang, H. Xiao, Y. Jiang, On a prediction method for remaining useful life of rolling bearings via VMD-based dispersion entropy and GAN, IEEE Sens. J. (2023).
[47] C. Liu, L. Zhang, J. Niu, R. Yao, C. Wu, Prognostics for rotating machinery using variational mode decomposition and long short-term memory network, in: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, 2019, pp. 7–12.
[48] T. Han, D. Jiang, Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier, Shock Vib. 2016 (2016).
[49] Q. Huang, D. Jiang, L. Hong, Y. Ding, Application of wavelet neural networks on vibration fault diagnosis for wind turbine gearbox, in: Springer, 2008, 2008, pp. 313–320.
[50] G.F. Bin, J.J. Gao, X.J. Li, B.S. Dhillon, Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network, Mech. Syst. Sig. Process. 27 (2012) 696–711.
[51] A. Rohani Bastami, A. Aasi, H.A. Arghand, Estimation of remaining useful life of rolling element bearings using wavelet packet decomposition and artificial neural network, Iranian J. Sci. Technol., Trans. Electr. Eng. 43 (2019) 233–245.
[52] M.S. Rathore, S.P. Harsha, An attention-based stacked BiLSTM framework for predicting remaining useful life of rolling bearings, Appl. Soft Comput. 131 (2022) 109765.
[53] K. Bhavsar, V. Vakharia, R. Chaudhari, J. Vora, D.Y. Pimenov, K. Giasin, A comparative study to predict bearing degradation using discrete wavelet transform (DWT), tabular generative adversarial networks (TGAN) and machine learning models, Machines 10 (3) (2022) 176.
[54] A. Eltotongy, M.I. Awad, S.A. Maged, A. Onsy, Fault detection and classification of machinery bearing under variable operating conditions based on wavelet transform and cnn, in: 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE, 2021, pp. 117–123.
[55] Y. Yoo, J.-G. Baek, A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network, Appl. Sci. 8 (7) (2018) 1102.
[56] B. Zhao, Q. Yuan, A novel deep learning scheme for multi-condition remaining useful life prediction of rolling element bearings, J. Manuf. Syst. 61 (2021) 450–460.
[57] C. Chen, T. Wang, Y. Liu, L. Cheng, J. Qin, Spatial attention-based convolutional transformer for bearing remaining useful life prediction, Meas. Sci. Technol. 33 (11) (2022) 114001.
[58] P.K. Kankar, S.C. Sharma, S.P. Harsha, Rolling element bearing fault diagnosis using wavelet transform, Neurocomputing 74 (10) (2011) 1638–1645.

[59] S.E. Pandarakone, M. Masuko, Y. Mizuno, H. Nakamura, Deep neural network based bearing fault diagnosis of induction motor using fast Fourier transform analysis, in: 2018 IEEE energy conversion congress and exposition (ECCE), IEEE, 2018, pp. 3214–3221.
[60] M. Zhao, B. Tang, Q. Tan, Bearing remaining useful life estimation based on time–frequency representation and supervised dimensionality reduction, Measurement 86 (2016) 41–55.
[61] C.-G. Huang, H.-Z. Huang, Y.-F. Li, W. Peng, A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing, J. Manuf. Syst. 61 (2021) 757–772.
[62] H. Ding, L. Yang, Z. Cheng, Z. Yang, A remaining useful life prediction method for bearing based on deep neural networks, Measurement 172 (2021) 108878.
[63] A. Majali, A. Mulay, V. Iyengar, A. Nayak, P. Singru, Fault identification and remaining useful life prediction of bearings using Poincare maps, fast Fourier transform and convolutional neural networks, Math. Models Eng. 8 (1) (2022) 1–14.
[64] Q. Wang, B. Zhao, H. Ma, J. Chang, G. Mao, A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion, J. Mech. Sci. Technol. 33 (2019) 2561–2571.
[65] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[66] F. Zhou, Y. Gao, C. Wen, A novel multimode fault classification method based on deep learning, J. Control Sci. Eng. 2017 (2017).
[67] S.-Y. Shao, W.-J. Sun, R.-Q. Yan, P. Wang, R.X. Gao, A deep learning approach for fault diagnosis of induction motors in manufacturing, Chinese J. Mech. Eng. 30 (2017) 1347–1356.
[68] Y. Liu, L. Duan, Z. Yuan, N. Wang, J. Zhao, An intelligent fault diagnosis method for reciprocating compressors based on LMD and SDAE, Sensors 19 (5) (2019) 1041.
[69] 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. Sig. Process. 72 (2016) 303–315.
[70] X. Zhao, J. Wu, Y. Zhang, Y. Shi, L. Wang, Fault Diagnosis of Motor in Frequency Domain Signal by Stacked De-noising Auto-encoder, Comput., Mater. Continua 57 (2) (2018).
[71] L.-H. Wang, X.-P. Zhao, J.-X. Wu, Y.-Y. Xie, Y.-H. Zhang, Motor fault diagnosis based on short-time Fourier transform and convolutional neural network, Chinese J. Mech. Eng. 30 (2017) 1357–1368.
[72] S. Zhou, M. Xiao, P. Bartos, M. Filip, G. Geng, Remaining useful life prediction and fault diagnosis of rolling bearings based on short-time fourier transform and convolutional neural network, Shock Vib. 2020 (2020) 1–14.
[73] M. Ragulskis, L. Chen, G. Song, and A. El Sinawi, "Intelligent fault diagnosis based on vibration signal analysis," 2017, *Hindawi*.
[74] S. Dong, G. Wen, Z. Zhang, Y. Yuan, J. Luo, Rolling bearing incipient degradation monitoring and performance assessment based on signal component tracking, IEEE Access 7 (2019) 45983–45993.
[75] X. Li, W. Zhang, Q. Ding, Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction, Reliab. Eng. Syst. Saf. 182 (2019) 208–218.
[76] M. Jalayer, C. Orsenigo, C. Vercellis, Fault detection and diagnosis for rotating machinery: a model based on convolutional LSTM, Fast Fourier and continuous wavelet transforms, Comput. Ind. 125 (2021) 103378.
[77] A. Ayman, O. Attalah, H. Shaban, Smart system for recognizing daily human activities based on wrist IMU sensors, in: 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), IEEE, 2020, pp. 1–6.
[78] X. Wang, et al., Deep spatiotemporal convolutional-neural-network-based remaining useful life estimation of bearings, Chinese J. Mech. Eng. 34 (1) (2021) 1–15.
[79] 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 (2017) 98–109.
[80] P. Li, X. Liu, Y. Yang, Remaining useful life prognostics of bearings based on a novel spatial graph-temporal convolution network, Sensors 21 (12) (2021) 4217.
[81] S. Patil, A. Patil, V. Handikherkar, S. Desai, V.M. Phalle, F.S. Kazi, Remaining useful life (RUL) prediction of rolling element bearing using random forest and gradient boosting technique. ASME international mechanical engineering congress and exposition, American Society of Mechanical Engineers, 2018 p. V013T05A019.
[82] A. Tayade, S. Patil, V. Phalle, F. Kazi, S. Powar, Remaining useful life (RUL) prediction of bearing by using regression model and principal component analysis (PCA) technique, Vibroengineering Procedia 23 (2019) 30–36.
[83] F. Sloukia, M. el Aroussi, H. Medromi, M. Wahbi, Bearings prognostic using mixture of gaussians hidden markov model and support vector machine, in: IEEE, 2013, pp. 1–4.
[84] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Remaining useful life estimation based on nonlinear feature reduction and support vector regression, Eng. Appl. Artif. Intel. 26 (7) (2013) 1751–1760.
[85] V.M. Nistane, S.P. Harsha, Prognosis of degradation progress of ball bearings using supervised machine learning, Proc. Inst. Mech. Eng., Part K: J. Multi-Body Dyn. 232 (2) (2018) 183–198.
[86] 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, IEEE, 2014, pp. 182–187.
[87] 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 (7) (2014) 3391–3401.
[88] T. Liu, J. Chen, X.N. Zhou, W.B. Xiao, Bearing performance degradation assessment using linear discriminant analysis and coupled HMM. Journal of Physics: Conference Series, IOP Publishing, 2012.
[89] X. Jin, M. Zhao, T.W.S. Chow, M. Pecht, Motor bearing fault diagnosis using trace ratio linear discriminant analysis, IEEE Trans. Ind. Electron. 61 (5) (2013) 2441–2451.
[90] L. Ciabattoni, 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.
[91] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, Cyclostationary modelling of rotating machine vibration signals, Mech. Syst. Sig. Process. 18 (6) (2004) 1285–1314.
[92] N. Sawalhi, R.B. Randall, H. Endo, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, Mech. Syst. Sig. Process. 21 (6) (2007) 2616–2633.
[93] Y. Lei, J. Lin, Z. He, Y. Zi, Application of an improved kurtogram method for fault diagnosis of rolling element bearings, Mech. Syst. Sig. Process. 25 (5) (2011) 1738–1749.
[94] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, J. Sound Vib. 289 (4–5) (2006) 1066–1090.
[95] 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. Ind. Electron. 62 (12) (2015) 7749–7761.
[96] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, Automatic bearing fault diagnosis based on one-class $\nu$-SVM, Comput. Ind. Eng. 64 (1) (2013) 357–365.
[97] L. Guo, J. Chen, X. Li, Rolling bearing fault classification based on envelope spectrum and support vector machine, J. Vib. Control 15 (9) (2009) 1349–1363.
[98] Z. Wei, J. Gao, X. Zhong, Z. Jiang, B. Ma, Incipient fault diagnosis of rolling element bearing based on wavelet packet transform and energy operator, WSEAS Trans. Syst. 10 (3) (2011) 81–90.
[99] S. Sharma, S. Sharma, A. Athaiya, Activation functions in neural networks, Towards Data Sci 6 (12) (2017) 310–316.
[100] S. Zhao, Y. Zhang, S. Wang, B. Zhou, C. Cheng, A recurrent neural network approach for remaining useful life prediction utilizing a novel trend features construction method, Measurement 146 (2019) 279–288.
[101] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[102] K. Cho, et al., Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).
[103] J. Wang, G. Wen, S. Yang, Y. Liu, Remaining useful life estimation in prognostics using deep bidirectional lstm neural network, in: IEEE, 2018, pp. 1037–1042.
[104] A. Elsheikh, S. Yacout, M.-S. Ouali, Bidirectional handshaking LSTM for remaining useful life prediction, Neurocomputing 323 (2019) 148–156.
[105] J. Wang, J. Yan, C. Li, R.X. Gao, R. Zhao, Deep heterogeneous GRU model for predictive analytics in smart manufacturing: Application to tool wear prediction, Comput. Ind. 111 (2019) 1–14.
[106] F.O. Heimes, Recurrent neural networks for remaining useful life estimation, in: IEEE, 2008, pp. 1–6.
[107] W. Mao, J. He, J. Tang, Y. Li, Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network, Adv. Mech. Eng. 10 (12) (2018), 1687814018817184.
[108] Z. Que, X. Jin, Z. Xu, Remaining useful life prediction for bearings based on a gated recurrent unit, IEEE Trans. Instrum. Meas. 70 (2021) 1–11.
[109] K.G. Eknath, G. Diwakar, Prediction of remaining useful life of rolling bearing using hybrid DCNN-BiGRU model, J. Vib. Eng. Technol. 11 (3) (2023) 997–1010.
[110] Y. Chen, G. Peng, Z. Zhu, S. Li, A novel deep learning method based on attention mechanism for bearing remaining useful life prediction, Appl. Soft Comput. 86 (2020) 105919.
[111] J. Zhou Y. Qin J. Luo T. Zhu Remaining useful life prediction by distribution contact ratio health indicator and consolidated memory GRU, IEEE Trans Industr Inform 2022.
[112] X. Wang, T. Wang, A. Ming, W. Zhang, A. Li, F. Chu, Spatiotemporal non-negative projected convolutional network with bidirectional NMF and 3DCNN for remaining useful life estimation of bearings, Neurocomputing 450 (2021) 294–310.
[113] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
[114] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv Preprint arXiv:1409.1556 (2014).
[115] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[116] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122 (2015).
[117] S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, Adv Neural Inf Process Syst 30 (2017).
[118] L. Ren, Y. Sun, H. Wang, L. Zhang, Prediction of bearing remaining useful life with deep convolution neural network, IEEE Access 6 (2018) 13041–13049.

[119] R. Liu, B. Yang, A.G. Hauptmann, Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network, IEEE Trans Industr Inform 16 (1) (2019) 87–96.
[120] J. Zhu, N. Chen, W. Peng, Estimation of bearing remaining useful life based on multiscale convolutional neural network, IEEE Trans. Ind. Electron. 66 (4) (2018) 3208–3216.
[121] R. Wang, R. Shi, X. Hu, C. Shen, Remaining useful life prediction of rolling bearings based on multiscale convolutional neural network with integrated dilated convolution blocks, Shock Vib. 2021 (2021) 1–11.
[122] B. Wang, Y. Lei, N. Li, T. Yan, Deep separable convolutional network for remaining useful life prediction of machinery, Mech. Syst. Sig. Process. 134 (2019) 106330.
[123] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, K. Weinberger, Simplifying graph convolutional networks, in: International conference on machine learning, PMLR, 2019, pp. 6861–6871.
[124] X. Yang, Y. Zheng, Y. Zhang, D.-S.-H. Wong, W. Yang, Bearing remaining useful life prediction based on regression shapalet and graph neural network, IEEE Trans. Instrum. Meas. 71 (2022) 1–12.
[125] F. Xia, et al., Graph learning: a survey, IEEE Trans. Artif. Intell. 2 (2) (2021) 109–127.
[126] X. Yang, X. Li, Y. Zheng, Y. Zhang, D.-S.-H. Wong, Bearing remaining useful life prediction using spatial-temporal multiscale graph convolutional neural network, Meas. Sci. Technol. 34 (8) (2023) 085009.
[127] M. Ma, Z. Mao, Deep-convolution-based LSTM network for remaining useful life prediction, IEEE Trans Industr Inform 17 (3) (2020) 1658–1667.
[128] A.Z. Hinchi, M. Tkiouat, Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network, Procedia Comput. Sci. 127 (2018) 123–132.
[129] R. Zhao, R. Yan, J. Wang, K. Mao, Learning to monitor machine health with convolutional bi-directional LSTM networks, Sensors 17 (2) (2017) 273.
[130] Y. Cao, Y. Ding, M. Jia, R. Tian, A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings, Reliab. Eng. Syst. Saf. 215 (2021) 107813.
[131] J.-R. Jiang, J.-E. Lee, Y.-M. Zeng, Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life, Sensors 20 (1) (2019) 166.
[132] B. Wang, Y. Lei, T. Yan, N. Li, L. Guo, Recurrent convolutional neural network: a new framework for remaining useful life prediction of machinery, Neurocomputing 379 (2020) 117–129.
[133] T. Li, Z. Zhao, C. Sun, R. Yan, X. Chen, Hierarchical attention graph convolutional network to fuse multi-sensor signals for remaining useful life prediction, Reliab. Eng. Syst. Saf. 215 (2021) 107878.
[134] Z. Kong, X. Jin, Z. Xu, B. Zhang, Spatio-temporal fusion attention: a novel approach for remaining useful life prediction based on graph neural network, IEEE Trans. Instrum. Meas. 71 (2022) 1–12.
[135] C. Mylonas, E. Chatzi, Remaining useful life estimation for engineered systems operating under uncertainty with causal GraphNets, Sensors 21 (19) (2021) 6325.
[136] G. Gao, B. Zhao, X. Chang, and X. Ming, “Remaining useful life prediction of bearings based on GCN and Bi-LSTM,” 2022.
[137] Y. Wei, D. Wu, Remaining useful life prediction of bearings with attention-awared graph convolutional network, Adv. Eng. Inf. 58 (2023) 102143.
[138] P. Liang, Y. Li, B. Wang, X. Yuan, L. Zhang, Remaining useful life prediction via a deep adaptive transformer framework enhanced by graph attention network, Int. J. Fatigue 174 (2023) 107722.
[139] M. Wang, Y. Li, Y. Zhang, L. Jia, Spatio-temporal graph convolutional neural network for remaining useful life estimation of aircraft engines, Aerosp. Syst. 4 (2021) 29–36.
[140] C. Yang, J. Liu, K. Zhou, X. Jiang, M.-F. Ge, Y. Liu, A node-level PathGraph-based bearing remaining useful life prediction method, IEEE Trans. Instrum. Meas. 71 (2022) 1–10.
[141] Y. Zhang, Y. Li, Y. Wang, Y. Yang, X. Wei, Adaptive spatio-temporal graph information fusion for remaining useful life prediction, IEEE Sens. J. 22 (4) (2021) 3334–3347.
[142] Y. Cao, J. Zhuang, M. Jia, X. Zhao, X. Yan, Z. Liu, Picture-in-picture strategy based complex graph neural network for remaining useful life prediction of rotating machinery, IEEE Trans. Instrum. Meas. (2023).
[143] Y. Wei, D. Wu, J. Terpenny, Bearing remaining useful life prediction using self-adaptive graph convolutional networks with self-attention mechanism, Mech. Syst. Sig. Process. 188 (2023) 110010.
[144] A. Kumar, C. Parkash, H. Tang, J. Xiang, Intelligent framework for degradation monitoring, defect identification and estimation of remaining useful life (RUL) of bearing, Adv. Eng. Inf. 58 (2023) 102206.
[145] X. Zeng, et al., Remaining useful life prediction for rotating machinery based on dynamic graph and spatial–temporal network, Meas. Sci. Technol. 34 (3) (2022) 035102.
[146] Y. Wei, D. Wu, J. Terpenny, Remaining useful life prediction using graph convolutional attention networks with temporal convolution-aware nested residual connections, Reliab. Eng. Syst. Saf. 242 (2024) 109776.
[147] W. Huang, Y. Rong, T. Xu, F. Sun, J. Huang, Tackling over-smoothing for general graph convolutional networks, arXiv preprint arXiv:2008.09864 (2020).
[148] L. Wang, H. Cao, H. Xu, H. Liu, A gated graph convolutional network with multi-sensor signals for remaining useful life prediction, Knowl. Based Syst. 252 (2022) 109340.
[149] K.F. Goebel, Management of uncertainty in sensor validation, sensor fusion, and diagnosis of mechanical systems using soft computing techniques, University of California, Berkeley, 1996.
[150] T.M. Khoshgoftaar, K. Gao, Feature selection with imbalanced data for software defect prediction, in: 2009 International Conference on Machine Learning and Applications, IEEE, 2009, pp. 235–240.
[151] E.F. Swana, W. Doorsamy, P. Bokoro, Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset, Sensors 22 (9) (2022) 3246.
[152] R. Li, W. Li, Y. Sun, Y. Wu, Resampling algorithm for imbalanced data based on their neighbor relationship, Chinese J. Eng. 43 (6) (2021) 862–869.
[153] Z. Ren, T. Lin, K. Feng, Y. Zhu, Z. Liu, K. Yan, A systematic review on imbalanced learning methods in intelligent fault diagnosis, IEEE Trans. Instrum. Meas. (2023).
[154] X. Zhang, P. Wu, J. He, S. Lou, J. Gao, A gan based fault detection of wind turbines gearbox, in: 2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), IEEE, 2020, pp. 271–275.
[155] S. Wang, L.L. Minku, X. Yao, Online class imbalance learning and its applications in fault detection, Int. J. Comput. Intell. Appl. 12 (04) (2013) 1340001.
[156] A. Ayman, O. Attalah, H. Shaban, An efficient human activity recognition framework based on wearable imu wrist sensors, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2019, pp. 1–5.
[157] J. Yang, G. Xie, Y. Yang, An improved ensemble fusion autoencoder model for fault diagnosis from imbalanced and incomplete data, Control Eng. Pract. 98 (2020) 104358.
[158] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[159] Y. Wang, K. Zhu, M. Sun, Y. Deng, An ensemble learning approach for fault diagnosis in self-organizing heterogeneous networks, IEEE Access 7 (2019) 125662–125675.
[160] A. Buabeng, A. Simons, N.K. Frempong, Y.Y. Ziggah, A novel hybrid predictive maintenance model based on clustering, smote and multi-layer perceptron neural network optimised with grey wolf algorithm, SN Appl. Sci. 3 (5) (2021) 593.
[161] H. Li, R. Fan, Q. Shi, Generative oversampling and deep forest based minority-class sensitive fault diagnosis approach, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2020, pp. 3629–3636.
[162] L. Yang, Z. Li, Y. Xiao, K. Feng, Evaluation method of rotating machinery health state based on TPE-XGBoost, in: 2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing), IEEE, 2021, pp. 1–6.
[163] L.I. Ruifeng, X.U. Aiqiang, S.U.N. Weichao, W.U. Yangyong, Preprocessing method based on sample resampling for imbalanced data of electronic circuits, Syst. Eng. Electron. 42 (11) (2020).
[164] Y.G. Xu, W.K. Deng, B. Song, X.Y. Deng, F. Luo, Pre-processing of imbalanced samples and the effective contribution in fault diagnosis in wastewater treatment plants, J. Hydroinf. 19 (2) (2017) 251–260.
[165] N.-C. Yang, H. Ismail, Robust intelligent learning algorithm using random forest and modified-independent component analysis for PV fault detection: In case of imbalanced data, IEEE Access 10 (2022) 41119–41130.
[166] Y. Zhang, X. Li, L. Gao, L. Wang, L. Wen, Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning, J. Manuf. Syst. 48 (2018) 34–50.
[167] M. Fan, W. Yan, Z. Ji, Fault diagnosis for bearings of unbalanced data based on feature generation, J. Syst. Simul. 32 (12) (2020) 2438–2448.
[168] Q.-X. Zhu, N. Zhang, Y.-L. He, Y. Xu, Novel Imbalanced Fault Diagnosis Method Based on CSMOTE Integrated with LSDA and LightGBM for Industrial Process, in: 2022 Decision and Information Technologies (CoDIT), IEEE, 2022, pp. 326–331.
[169] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, D. Huang, New imbalanced fault diagnosis framework based on Cluster-MWMOTE and MFO-optimized LS-SVM using limited and complex bearing data, Eng. Appl. Artif. Intel. 96 (2020) 103966.
[170] S. Barua, M.M. Islam, X. Yao, K. Murase, MWMOTE–majority weighted minority oversampling technique for imbalanced data set learning, IEEE Trans. Knowl. Data Eng. 26 (2) (2012) 405–425.
[171] Y. Han, B. Li, Y. Huang, L. Li, K. Yan, Imbalanced fault classification of rolling bearing based on an improved oversampling method, J. Braz. Soc. Mech. Sci. Eng. 45 (4) (2023) 223.
[172] S. Shao, P. Wang, R. Yan, Generative adversarial networks for data augmentation in machine fault diagnosis, Comput. Ind. 106 (2019) 85–93.
[173] D.P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).
[174] I. Goodfellow, et al., Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.
[175] Y.O. Lee, J. Jo, J. Hwang, Application of deep neural network and generative adversarial network to industrial maintenance: A case study of induction motor fault detection, in: 2017 IEEE international conference on big data (big data), IEEE, 2017, pp. 3248–3253.
[176] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, X. Li, Machinery fault diagnosis with imbalanced data using deep generative adversarial networks, Measurement 152 (2020) 107377.
[177] Q. Liu, G. Ma, C. Cheng, Generative adversarial network based multi-class imbalanced fault diagnosis of rolling bearing, in: 2019 4th International Conference on System Reliability and Safety (ICSRS), IEEE, 2019, pp. 318–324.
[178] Y. Xie, T. Zhang, Imbalanced learning for fault diagnosis problem of rotating machinery based on generative adversarial networks, in: IEEE, 2018, pp. 6017–6022.
[179] S. Yang, F. Zhou, D. Chen, C. Wen, Deep learning fault diagnosis method based on feature generative adversarial networks for unbalanced data, in: 2019 CAA

Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), IEEE, 2019, pp. 465–470.
[180] B. Zhao, Q. Yuan, Improved generative adversarial network for vibration-based fault diagnosis with imbalanced data, Measurement 169 (2021) 108522.
[181] M.T. Pham, J.-M. Kim, C.H. Kim, Rolling bearing fault diagnosis based on improved GAN and 2-D representation of acoustic emission signals, IEEE Access 10 (2022) 78056–78069.
[182] D. Cabrera, et al., Generative adversarial networks selection approach for extremely imbalanced fault diagnosis of reciprocating machinery, IEEE Access 7 (2019) 70643–70653.
[183] Y. Li, W. Zou, L. Jiang, Fault diagnosis of rotating machinery based on combination of Wasserstein generative adversarial networks and long short term memory fully convolutional network, Measurement 191 (2022) 110826.
[184] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein gans, Adv. Neural. Inf. Process Syst. 30 (2017).
[185] Z. Pu, D. Cabrera, R.-V. Sánchez, M. Cerrada, C. Li, J. Valente de Oliveira, Exploiting generative adversarial networks as an oversampling method for fault diagnosis of an industrial robotic manipulator, Appl. Sci. 10 (21) (2020) 7712.
[186] S. Behera, R. Misra, Generative adversarial networks based remaining useful life estimation for IIoT, Comput. Electr. Eng. 92 (2021) 107195.
[187] L. Fetty, et al., Investigating conditional GAN performance with different generator architectures, an ensemble model, and different MR scanners for MR-sCT conversion, Phys. Med. Biol. 65 (10) (2020) 105004.
[188] J. Wang, B. Han, H. Bao, M. Wang, Z. Chu, Y. Shen, Data augment method for machine fault diagnosis using conditional generative adversarial networks, Proc. Inst. Mech. Eng., Part D: J. Automobile Eng. 234 (12) (2020) 2719–2727.
[189] K. Sohn, H. Lee, X. Yan, Learning structured output representation using deep conditional generative models, Adv. Neural. Inf. Process Syst. 28 (2015).
[190] R. Zemouri, et al., Recent research and applications in variational autoencoders for industrial prognosis and health management: A survey, in: 2022 Prognostics and Health Management Conference (PHM-2022 London), IEEE, 2022, pp. 193–203.
[191] D. Zhao, et al., Enhanced data-driven fault diagnosis for machines with small and unbalanced data based on variational auto-encoder, Meas. Sci. Technol. 31 (3) (2019) 035004.
[192] H. Karamti, M.M.A. Lashin, F.M. Alrowais, A.M. Mahmoud, A new deep stacked architecture for multi-fault machinery identification with imbalanced samples, IEEE Access 9 (2021) 58838–58851.
[193] H. Xie Y. Lei M. Yan Y. Yu X. Xia X. Mao A universal data augmentation approach for fault localization In: Proceedings of the 44th International Conference on Software Engineering 2022 48 60.
[194] J. Yang, Y. Li, C. Yang, T. Gao, Fault diagnosis method of analog circuit based on enhanced boundary equilibrium generative adversarial networks, in: 2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing), IEEE, 2021, pp. 1–7.
[195] C.B. Do, A.Y. Ng, Transfer learning for text classification, Adv. Neural. Inf. Process Syst. 18 (2005).
[196] A. Brodzicki, M. Piekarski, D. Kucharski, J. Jaworek-Korjakowska, M. Gorgon, Transfer learning methods as a new approach in computer vision tasks with small datasets, Found. Comput. Decision Sci. 45 (3) (2020) 179–193.
[197] J. Wang, Y. Chen, Transfer Learning for Natural Language Processing, in: Introduction to Transfer Learning Algorithms and Practice, Springer, 2022, pp. 275–279.
[198] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[199] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu, A survey on deep transfer learning, in: Artificial Neural Networks and Machine Learning–ICANN 27th International Conference on Artificial Neural Networks Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27, Springer, 2018, pp. 270–279.
[200] H. Zhang, Q. Zhang, S. Shao, T. Niu, X. Yang, H. Ding, Sequential network with residual neural network for rotatory machine remaining useful life prediction using deep transfer learning, Shock Vib. 2020 (2020) 1–16.
[201] A. Zhang, et al., Transfer learning with deep recurrent neural networks for remaining useful life estimation, Appl. Sci. 8 (12) (2018) 2416.
[202] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, X. Chen, Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing, IEEE Trans. Industr. Inform 15 (4) (2018) 2416–2425.
[203] G. Huang, Y. Zhang, J. Ou, Transfer remaining useful life estimation of bearing using depth-wise separable convolution recurrent network, Measurement 176 (2021) 109090.
[204] B.K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, G.R.G. Lanckriet, Hilbert space embeddings and metrics on probability measures, J. Mach. Learn. Res. 11 (2010) 1517–1561.
[205] D. Virosztek, The metric property of the quantum Jensen-Shannon divergence, Adv Math (n y) 380 (2021) 107595.
[206] Y. Che, Z. Deng, X. Lin, L. Hu, X. Hu, Predictive battery health management with transfer learning and online model correction, IEEE Trans. Veh. Technol. 70 (2) (2021) 1269–1277.
[207] T.-K. Nguyen, Z. Ahmad, J.-M. Kim, A deep-learning-based health indicator constructor using Kullback–Leibler divergence for predicting the remaining useful life of concrete structures, Sensors 22 (10) (2022) 3687.
[208] H. Cheng, X. Kong, G. Chen, Q. Wang, R. Wang, Transferable convolutional neural network based remaining useful life prediction of bearing under multiple failure behaviors, Measurement 168 (2021) 108286.
[209] M.S. Rathore, S.P. Harsha, Rolling bearing prognostic analysis for domain adaptation under different operating conditions, Eng. Fail. Anal. 139 (2022) 106414.
[210] J. Zhuang, M. Jia, Y. Ding, P. Ding, Temporal convolution-based transferable cross-domain adaptation approach for remaining useful life estimation under variable failure behaviors, Reliab. Eng. Syst. Saf. 216 (2021) 107946.
[211] J. Zhuang, M. Jia, Y. Ding, P. Ding, Remaining useful life estimation under variable failure behaviors via transferable metric learning, in: 2021 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence (ICSMD), IEEE, 2021, pp. 1–6.
[212] W. Mao, J. He, M.J. Zuo, Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning, IEEE Trans. Instrum. Meas. 69 (4) (2019) 1594–1608.
[213] J. Xu, M. Fang, W. Zhao, Y. Fan, X. Ding, Deep Transfer Learning Remaining Useful Life Prediction of Different Bearings, in: 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1–8.
[214] Y. Cao, M. Jia, P. Ding, Y. Ding, Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network, Measurement 178 (2021) 109287.
[215] W. Mao, J. He, B. Sun, L. Wang, Prediction of bearings remaining useful life across working conditions based on transfer learning and time series clustering, IEEE Access 9 (2021) 135285–135303.
[216] Y. Ganin, et al., Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (1) (2016) 2030–2096.
[217] X. Li, W. Zhang, H. Ma, Z. Luo, X. Li, Data alignments in machinery remaining useful life prediction using deep adversarial neural networks, Knowl. Based Syst. 197 (2020) 105843.
[218] J. Zhuang, M. Jia, X. Zhao, An adversarial transfer network with supervised metric for remaining useful life prediction of rolling bearing under multiple working conditions, Reliab. Eng. Syst. Saf. 225 (2022) 108599.
[219] Y. Zou, Z. Li, Y. Liu, S. Zhao, Y. Liu, G. Ding, A method for predicting the remaining useful life of rolling bearings under different working conditions based on multi-domain adversarial networks, Measurement 188 (2022) 110393.
[220] S. Fu, Y. Zhang, L. Lin, M. Zhao, S. Zhong, Deep residual LSTM with domain-invariance for remaining useful life prediction across domains, Reliab. Eng. Syst. Saf. 216 (2021) 108012.
[221] M. Miao, J. Yu, Z. Zhao, A sparse domain adaption network for remaining useful life prediction of rolling bearings under different working conditions, Reliab. Eng. Syst. Saf. 219 (2022) 108259.
[222] H. Cheng, X. Kong, Q. Wang, H. Ma, S. Yang, G. Chen, Deep transfer learning based on dynamic domain adaptation for remaining useful life prediction under different working conditions, J. Intell. Manuf. 34 (2) (2023) 587–613.
[223] X. Wang, T. Wang, A. Ming, W. Zhang, A. Li, F. Chu, Cross-operating condition degradation knowledge learning for remaining useful life estimation of bearings, IEEE Trans. Instrum. Meas. 70 (2021) 1–11.
[224] J. Zhuang, Y. Cao, M. Jia, X. Zhao, Q. Peng, Remaining useful life prediction of bearings using multi-source adversarial online regression under online unknown conditions, Expert Syst. Appl. 227 (2023) 120276.
[225] L. Cheng, Z. An, Y. Guo, M. Ren, Z. Yang, S. McLoone, MMFSL: a novel multi-modal few-shot learning framework for fault diagnosis of industrial bearings, IEEE Trans. Instrum. Meas. (2023).
[226] T. Wang, C. Chen, X. Dong, H. Liu, A novel method of production line bearing fault diagnosis based on 2D image and cross-domain few-shot learning, Appl. Sci. 13 (3) (2023) 1809.
[227] S. Wang, D. Wang, D. Kong, W. Li, J. Wang, H. Wang, Few-shot multiscene fault diagnosis of rolling bearing under compound variable working conditions, IET Control Theory Appl. 16 (14) (2022) 1405–1416.
[228] J. Jang, C.O. Kim, Siamese network-based health representation learning and robust reference-based remaining useful life prediction, IEEE Trans Industr Inform 18 (8) (2021) 5264–5274.
[229] T. Pan, S. Zhang, F. Li, J. Chen, A. Li, A meta network pruning framework for remaining useful life prediction of rocket engine bearings with temporal distribution discrepancy, Mech. Syst. Sig. Process. 195 (2023) 110271.
[230] J. Yang, X. Wang, Z. Luo, Few-shot remaining useful life prediction based on meta-learning with deep sparse kernel network, Inf Sci (N Y) 653 (2024) 119795.
[231] J. Sun, X. Zhang, J. Wang, Lightweight bidirectional long short-term memory based on automated model pruning with application to bearing remaining useful life prediction, Eng. Appl. Artif. Intel. 118 (2023) 105662.
[232] P. Ding, M. Jia, X. Zhao, Meta deep learning based rotating machinery health prognostics toward few-shot prognostics, Appl. Soft Comput. 104 (2021) 107211.
[233] Y. Mo, L. Li, B. Huang, X. Li, Few-shot RUL estimation based on model-agnostic meta-learning, J. Intell. Manuf. 34 (5) (2023) 2359–2372.
[234] C. Guo, Z. Xu, Q. Yao, Information Fusion and XGBoost Algorithm Used for Bearing Remaining Useful Life Prediction, in: 2021 China Automation Congress (CAC), IEEE, 2021, pp. 1689–1693.
[235] P.J.G. Nieto, E. García-Gonzalo, F.S. Lasheras, F.J. de Cos Juez, Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability, Reliab. Eng. Syst. Saf. 138 (2015) 219–231.
[236] Y. Wen, M.F. Rahman, H. Xu, T.-L.-B. Tseng, Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective, Measurement 187 (2022) 110276.
[237] C.B.S. Maior, M. das Chagas Moura, I.D. Lins, E.L. Droguett, H.H.L. Diniz, Remaining useful life estimation by empirical mode decomposition and support vector machine, IEEE Lat. Am. Trans. 14 (11) (2016) 4603–4610.

[238] L. Saidi, J. Ben Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR, Appl. Acoust. 120 (2017) 1–8.
[239] L. Cao, Z. Qian, Y. Pei, Remaining useful life prediction of wind turbine generator bearing based on EMD with an indicator, in: 2018 Prognostics and System Health Management Conference (PHM-Chongqing), IEEE, 2018, pp. 375–379.
[240] S. Singh, T. Agarwal, G. Kumar, O.M. Yadav, Predicting the remaining useful life of ball bearing under dynamic loading using supervised learning, in: 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, 2019, pp. 1119–1123.
[241] L. Ren, X. Cheng, X. Wang, J. Cui, L. Zhang, Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction, Futur. Gener. Comput. Syst. 94 (2019) 601–609.
[242] C. Bienefeld, E. Kirchner, A. Vogt, M. Kacmar, On the importance of temporal information for remaining useful life prediction of rolling bearings using a random forest regressor, Lubricants 10 (4) (2022) 67.
[243] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
[244] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, "Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst., Man, Cybernetics Part B (Cybernetics) 42 (2) (2011) 513–529.
[245] J. Lee, et al., Remaining useful life estimation for ball bearings using feature engineering and extreme learning machine, IFAC-PapersOnLine 55 (7) (2022) 198–203.
[246] Z. Pan, Z. Meng, Z. Chen, W. Gao, Y. Shi, A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings, Mech. Syst. Sig. Process. 144 (2020) 106899.
[247] C. Wu, H. Sun, Z. Zhang, Stages prediction of the remaining useful life of rolling bearing based on regularized extreme learning machine, Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. 235 (22) (2021) 6599–6610.
[248] M. Elforjani, S. Shanbr, Prognosis of bearing acoustic emission signals using supervised machine learning, IEEE Trans. Ind. Electron. 65 (7) (2017) 5864–5871.
[249] H. Zhao, H. Liu, Y. Jin, X. Dang, W. Deng, Feature extraction for data-driven remaining useful life prediction of rolling bearings, IEEE Trans. Instrum. Meas. 70 (2021) 1–10.
[250] L. Saidi, J. Ben Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR, Appl. Acoust. 120 (2017) 1–8.
[251] L. Liao, W. Jin, R. Pavel, Enhanced restricted Boltzmann machine with prognosability regularization for prognostics and health assessment, IEEE Trans. Ind. Electron. 63 (11) (2016) 7076–7083.
[252] L. Ren, J. Cui, Y. Sun, X. Cheng, Multi-bearing remaining useful life collaborative prediction: a deep learning approach, J. Manuf. Syst. 43 (2017) 248–256.
[253] L. Ren, Y. Sun, J. Cui, L. Zhang, Bearing remaining useful life prediction based on deep autoencoder and deep neural networks, J. Manuf. Syst. 48 (2018) 71–77.
[254] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[255] F. Cheng, L. Qu, W. Qiao, A case-based data-driven prediction framework for machine fault prognostics, in: 2015 IEEE Energy Conversion Congress and Exposition (ECCE), IEEE, 2015, pp. 3957–3963.
[256] J. Wu, Y. Su, Y. Cheng, X. Shao, C. Deng, C. Liu, Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system, Appl. Soft Comput. 68 (2018) 13–23.
[257] R.R. Mutra, D. Mallikarjuna Reddy, J. Srinivas, D. Sachin, K. Babu Rao, Signal-based parameter and fault identification in roller bearings using adaptive neuro-fuzzy inference systems, J. Braz. Soc. Mech. Sci. Eng. 45 (1) (2023) 45.
[258] A. Khosravi, R.N.N. Koury, L. Machado, J.J.G. Pabon, Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system, Sustain. Energy Technol. Assess. 25 (2018) 146–160.
[259] F. Cheng, L. Qu, W. Qiao, Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis, IEEE Trans. Sustain. Energy 9 (1) (2017) 157–167.
[260] W. Teng, X. Zhang, Y. Liu, A. Kusiak, Z. Ma, Prognosis of the remaining useful life of bearings in a wind turbine gearbox, Energies (Basel) 10 (1) (2016) 32.
[261] S.K. Singh, S. Kumar, J.P. Dwivedi, Compound fault prediction of rolling bearing using multimedia data, Multimed. Tools Appl. 76 (2017) 18771–18788.
[262] Y. Cheng, K. Hu, J. Wu, H. Zhu, X. Shao, A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings, Adv. Eng. Inf. 48 (2021) 101247.
[263] C. Cheng, et al., A deep learning-based remaining useful life prediction approach for bearings, IEEE/ASME Trans. Mechatron. 25 (3) (2020) 1243–1254.
[264] R. Huang, L. Xi, X. Li, C.R. Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, Mech. Syst. Sig. Process. 21 (1) (2007) 193–207.
[265] P. Nectoux *et al.*, PRONOSTIA: An experimental platform for bearings accelerated degradation tests, In: IEEE International Conference on Prognostics and Health Management, PHM'12., IEEE Catalog Number: CPF12PHM-CDR, 2012, pp. 1–8.
[266] X. Xu, X. Li, W. Ming, M. Chen, A novel multi-scale CNN and attention mechanism method with multi-sensor signal for remaining useful life prediction, Comput. Ind. Eng. 169 (2022) 108204.
[267] C.-G. Huang, H.-Z. Huang, Y.-F. Li, A bidirectional LSTM prognostics method under multiple operational conditions, IEEE Trans. Ind. Electron. 66 (11) (2019) 8792–8802.
[268] B. Wang, Y. Lei, N. Li, N. Li, A hybrid prognostics approach for estimating remaining useful life of rolling element bearings, IEEE Trans. Reliab. 69 (1) (2018) 401–412.
[269] H. Wei, Q. Zhang, Y. Gu, Remaining useful life prediction of bearings based on self-attention mechanism, multi-scale dilated causal convolution, and temporal convolution network, Meas. Sci. Technol. 34 (4) (2023) 045107.
[270] B.P. Duong, et al., A reliable health indicator for fault prognosis of bearings, Sensors 18 (11) (2018) 3740.
[271] 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 (3) (2017) 549.
[272] W. Gousseau, J. Antoni, F. Girardin, J. Griffaton, Analysis of the rolling element bearing data set of the center for intelligent maintenance systems of the university of Cincinnati, CM2016 (2016).
[273] M. Yan, X. Wang, B. Wang, M. Chang, I. Muhammad, Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model, ISA Trans. 98 (2020) 471–482.
[274] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study, Mech. Syst. Sig. Process. 64 (2015) 100–131.
[275] O. Janssens, et al., Convolutional neural network based fault detection for rotating machinery, J. Sound Vib. 377 (2016) 331–345.
[276] L. Eren, Bearing fault detection by one-dimensional convolutional neural networks, Math. Probl. Eng. 2017 (1) (2017) 8617315.
[277] Y. Deng, Paderborn bearing dataset and PHM2009 gearbox dataset.
[278] V. Pandhare, J. Singh, J. Lee, Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features, in: 2019 Prognostics and System Health Management Conference (PHM-Paris), IEEE, 2019, pp. 320–326.
[279] Y. Chen, G. Peng, C. Xie, W. Zhang, C. Li, S. Liu, ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis, Neurocomputing 294 (2018) 61–71.
[280] Z. Zhu, G. Peng, Y. Chen, H. Gao, A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, Neurocomputing 323 (2019) 62–75.
[281] M. Benbouzid, T. Berghout, N. Sarma, S. Djurović, Y. Wu, X. Ma, Intelligent condition monitoring of wind power systems: state of the art review, Energies (Basel) 14 (18) (2021) 5967.
[282] A. Choudhary, T. Mian, S. Fatima, Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images, Measurement 176 (2021) 109196.
[283] O. Janssens, et al., Thermal image based fault diagnosis for rotating machinery, Infrared Phys. Technol. 73 (2015) 78–87.
[284] A. Choudhary, D. Goyal, S.S. Letha, Infrared thermography-based fault diagnosis of induction motor bearings using machine learning, IEEE Sens. J. 21 (2) (2020) 1727–1734.
[285] O. Janssens, M. Loccufier, S. van Hoecke, Thermal imaging and vibration-based multisensor fault detection for rotating machinery, IEEE Trans Industr Inform 15 (1) (2018) 434–444.
[286] J. Carroll, S. Koukoura, A. McDonald, A. Charalambous, S. Weiss, S. McArthur, Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques, Wind Energy 22 (3) (2019) 360–375.
[287] A. Mosallam, K. Medjaher, N. Zerhouni, Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction, J. Intell. Manuf. 27 (2016) 1037–1048.
[288] T. Gao, Y. Li, X. Huang, C. Wang, Data-driven method for predicting remaining useful life of bearing based on Bayesian theory, Sensors 21 (1) (2020) 182.