




Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals

Bayu Adhi Tama¹ · Malinda Vania^{2,3} · Seungchul Lee^{4,5} · Sunghoon Lim^{2,3} 

© The Author(s) 2022

Abstract

Vibration measurement and monitoring are essential in a wide variety of applications. Vibration measurements are critical for diagnosing industrial machinery malfunctions because they provide information about the condition of the rotating equipment. Vibration analysis is considered the most effective method for predictive maintenance because it is used to troubleshoot instantaneous faults as well as periodic maintenance. Numerous studies conducted in this vein have been published in a variety of outlets. This review documents data-driven and recently published deep learning techniques for vibration-based condition monitoring. Numerous studies were obtained from two reputable indexing databases, Web of Science and Scopus. Following a thorough review, 59 studies were selected for synthesis. The selected studies are then systematically discussed to provide researchers with an in-depth view of deep learning-based fault diagnosis methods based on vibration signals. Additionally, a few remarks regarding future research directions are made, including graph-based neural networks, physics-informed ML, and a transformer convolutional network-based fault diagnosis method.

Keywords Deep learning · Vibration signal · Fault diagnosis · Vibration sensor · Condition monitoring

✉ Seungchul Lee
seunglee@postech.ac.kr

✉ Sunghoon Lim
sunghoonlim@unist.ac.kr

¹ Department of Information Systems, University of Maryland, Baltimore, MD 21250, USA

² Department of Industrial Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

³ Industry Intelligentization Institute, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea

⁴ Department of Mechanical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea

⁵ Institute for Convergence Research and Education in Advanced Technology, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

1 Introduction

Condition monitoring has become an essential aspect of successful industrial and manufacturing systems. Various condition monitoring methods, such as acoustic and acoustic emission-based techniques (Govekar et al. 2000; Potočník et al. 2007) have been prevalent. However, among others, vibration analysis (Randall 2011; Ruiz-Cárcel et al. 2016) is deemed the most effective method for predictive maintenance since it can be used for troubleshooting instantaneous faults as well as periodic maintenance. As a rule of thumb, an early fault diagnosis is very crucial in modern industry since it is strongly correlated with the reduction of maintenance costs, the increase of the machine's operating time, and the prevention of huge economic losses and devastating accidents (Lei 2016; Mohanty 2014). For machines that work under different speed conditions, the non-stationary vibration signals can be preprocessed by using the speed signals so that the preprocessed signals can be readily employed for diagnosis. Therefore, it implies that the variation in the speed signals is very critical for effective fault diagnosis (Rao and Zuo 2018).

Obtaining rotating speed signals of rotating machines is a challenging task. To date, there exist two main approaches to obtaining the rotating speed: (i) placing extra speed sensors to estimate speed immediately (Sarma et al. 2008), and (ii) collecting speed from vibration signals using vibration level sensors (e.g., accelerometers) (Peeters et al. 2019). Most state-of-the-art fault diagnosis techniques have been focused on the latter, as no additional speed sensors are needed, and generally satisfactory results can be obtained. Different kinds of mechanical components, such as gears and bearings, can be examined and analyzed by measuring vibration levels from sensors. In contrast, engineers often find the former challenging since placing extra speed sensors is almost impractical due to the limited operating conditions. Moreover, it would not be uneconomical concerning the cost of machine condition monitoring and maintenance, if even possible. Analyzing vibration signals involves a process of acquiring raw signals data from the device and specifying crucial properties or features that are heavily dependent on defects. Over the last few years, a variety of techniques for obtaining the speed signal from vibration measurements have been studied, with the vast majority relying on the time-frequency (TF) representation technique (Dziedziech et al. 2018; Iatsenko et al. 2016; Khan et al. 2017; Schmidt et al. 2018). The typical examples of TF representations are short-time Fourier transform (STFT) and wavelet transform. TF-based techniques have been widely used and are deemed effective for speed extraction. Other methods include band-pass filtering and phase demodulation (Urbanek et al. 2011, 2013), Teager Kaiser energy operator (Randall and Smith 2016), and Kalman filter (Cardona-Morales et al. 2014).

There have been a number of speed estimators; however, some improvements are still required to be addressed. The key challenges for speed estimators are operating changes that affect the harmonic structure and sudden speed fluctuations (Peeters et al. 2019). Moreover, there exists a need to manually select the proper model parameters to analyze the vibration signals. This circumstance will make the process labor-intensive since a huge amount of data with high dimensional features is increasing progressively. In condition monitoring, the most common way is to inspect each speed sensor measurement and set a minimum and maximum threshold value. The machine is healthy if the value is within the range, otherwise it is unhealthy. This static limit measurement might lead to a complex piece of equipment that cannot be reliably judged.

Machine learning (ML) techniques have been utilized for the health status diagnosis of rotating machinery. Fault diagnosis technologies are a mechanical extension of pattern

recognition theory that aims to solve the problem of state classification in engineering systems and operating equipment (Xu et al. 2020a; Zhang et al. 2021b). According to Xu et al. (2020a), fault detection, problem isolation, and fault identification are the three tasks that fault diagnosis systems must perform. Fault diagnosis systems can be used to assess the state of a machine while it is in use and detect any faults in real-time. Informative feature representations are typically extracted using advanced signal processing techniques, allowing the accurate discrimination of faults with respect to their location (Kateris et al. 2014). This feature extraction is combined with classification methods to construct mathematical representations of the relationships between all machine parameters. Classification algorithms take into consideration empirical data that has been measured. The algorithms learn such relationships from this data and classify new measured data to support an early diagnosis system for rotating machinery. Classification techniques are comprised of various statistical methods, neural networks (Kateris et al. 2014), and fuzzy logics (Subbaraj and Kanapiran 2014), and, in the past few years, deep learning (DL) (Jia et al. 2016).

1.1 Research aims and contributions

In the past few years, DL (LeCun et al. 2015), as a sub-field of ML, has become a leading intelligent technique applied in wide-array applications. DL-based techniques such as deep neural network (DNN) have been developed that can, automatically and without human intervention, extract deep patterns from huge raw data (Jia et al. 2016). This paper focuses on documenting an overview of DL techniques for condition monitoring using vibration data. Some recently published studies are discussed in this paper, providing researchers and practitioners an insight into the current status of DL-based condition monitoring approaches. This review article is an extension of a previously published paper in (Tama et al. 2020) and has two main contributions:

- (i) To provide documentation on how DL techniques have been employed and are presently addressed for vibration signal-based machine fault diagnosis;
- (ii) To present a benchmark on which DL models from the selected studies have performed better.

1.2 Research questions

Following the recommendation of (Kitchenham et al. 2015), we determine research questions (RQs) based on the problems being addressed and the goals of this study as follows.

- (a) RQ₁: What are the distribution per year and publication venues of prior works related to the applications of DL for vibration-based fault diagnosis techniques?
- (b) RQ₂: What types of DL techniques have been developed the most?
- (c) RQ₃: What are the future research directions concerning the challenges and limitations of the current studies that remain?

RQ₁ is addressed in Sects. 4.1 and 4.2, while Sect. 4.3 addresses RQ₂. Section 5 is devoted to answer RQ₃.

1.3 Structure

The rest of the paper is broken down as follows. Section 2 conveys the basic concept of vibration signal analysis and DL, followed by Sect. 3 that presents the methodology of conducting the review. The results of this review are systematically classified and discussed in Sect. 4, while several directions for future research are provided in Sect. 5. Finally, Sect. 6 anticipates the threat to validity, while Sect. 7 concludes this study.

2 Fundamental concepts

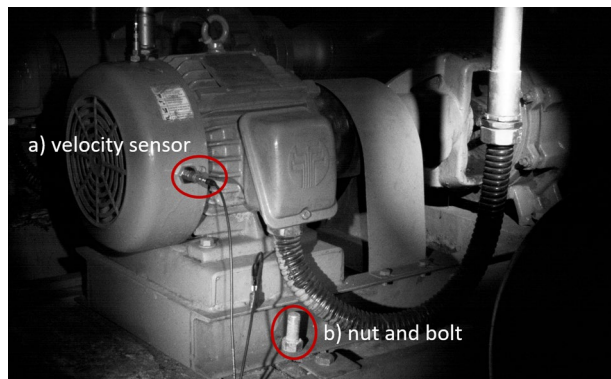
This section discusses the fundamental of fault diagnosis approach by using vibration signal analysis and an overview of DL-based techniques (e.g., DNNs).

2.1 Vibration signal analysis for rotating machinery

Besides lubricant analysis, vibration analysis and monitoring of rotating equipment are the key activities in modern industries (Randall 2011). Vibration measurement is typically performed online, providing a real-time source of diagnostic information on the machine's current condition. This information is combined with other parameters to help the diagnostic interpretation of vibration data. A common vibration-based condition monitoring system is composed of several accelerometers attached to each bearing of the machine. An extra sensor is also needed to allow the correlated analysis of signals from different sensors. Figure 1 shows some sensors installed on a laboratory-scale water pump while a nut under the looseness is typically used for developing mechanical signatures that can be related to the fault. Note that examples of faults resulting in high vibration levels in rotating machinery include misalignment of bearings, damaged bearings and gears, bent shafts, and resonance.

Other components that are necessarily important in the system include on-site signal conditioning modules, data acquisition, and a processing sub-system. This system enables maintenance staff or site engineers to perform real-time frequency analysis of machine vibration, leading to a reasonably effective diagnostic assessment of the machine's dynamic circumstances. Approaches that can considerably enhance the diagnostic performance of vibration monitoring are (i) an in-depth analysis of the

Fig. 1 Several speed sensors are attached on a laboratory-scale water pump (a) and a nut under the looseness for developing mechanical signature (b) (Sun et al. 2020)



huge amount of historical information of vibration signals from each sensor; and (ii) a time-correlated analysis of the signals from different sensors. Both are capable of early detection of possible damage, and improved diagnostic assurance and accuracy can be achieved.

Most machine components bring about particular vibration signals, allowing them to be discriminated against by others, including distinguishing between faulty and healthy conditions. A fundamental understanding of vibration signals will diagnose faults and control machinery vibration. In order to acquire relevant information from signals, they need to be analyzed. Signals substantially carry features (e.g., information), which can be constant or vary for a specified period. Furthermore, they can be broken down into two fundamental types: (i) stationary, which means that variances in their statistical features do not exist, and (ii) non-stationary, where the time domain statistical features change with time (Mohanty 2014; Randall 2011).

Stationary signals can be analyzed in either the time domain or the frequency domain. Raw signals obtained from rotating machinery are typically in the time domain, representing a collection of time-indexed data points (e.g., acceleration, velocity, and background noise) obtained over a period of time. Time-domain fault diagnosis might be divided into two main categories (Ahmed and Nandi 2020):

(a) Visual inspection

It is performed by comparing a measured vibration signal to a formerly measured vibration signal of a machine in a normal condition. However, this type of inspection is deemed to be unreliable for the condition monitoring of rotating machines for several reasons: (i) the difference between the waveform signals is not always clearly visible, (ii) a large collection of vibration signals along with some background noises is dealt with, and (iii) early detection of faults is much desired, thus making such a manual inspection unfeasible to perform.

(b) Feature-based inspection

It is done by calculating particular features of the raw vibration signals. Various techniques include statistical functions (e.g., root mean square amplitude, peak amplitude, mean amplitude, variance, and standard deviation) and advanced techniques (e.g., time-synchronous averaging, autoregressive moving average, and filter-based method).

Machinery condition monitoring commonly uses frequency (e.g., spectral) domain analysis since it provides a mechanical signature of the machine as well as the ability to discover information based on frequency characteristics. A possible change in the dynamics due to a fault or a defect in the machine can be considered as a corresponding change in the amplitude of the signals. As mentioned earlier, a rotating machine generates various vibration signals in the time domain. Hence, to get the vibration signals in the frequency domain, those signals are often performed using Fourier analysis, which can be categorized into three types: Fourier series, continuous Fourier transform, and discrete Fourier transform.

It is generally known that transforming a signal into the frequency domain relies upon the assumption that a frequency component does not change over time (e.g., the signal is stationary). Hence, more specifically, the Fourier transform in the frequency domain cannot serve time distribution information. Rotating machinery typically generates stationary signals, but it would be possible to have non-stationary signals due to a speed swing (e.g., speed fluctuates over time) (Brandt 2011). The time-frequency

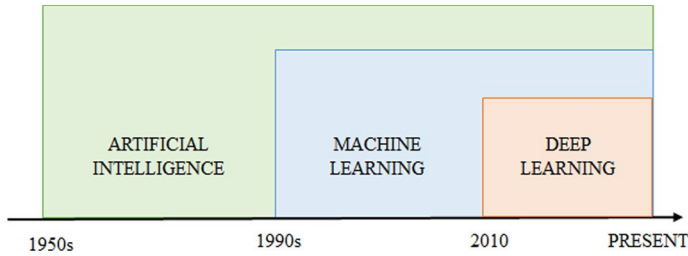


Fig. 2 The relationship and a time frame between AI, ML, and DL

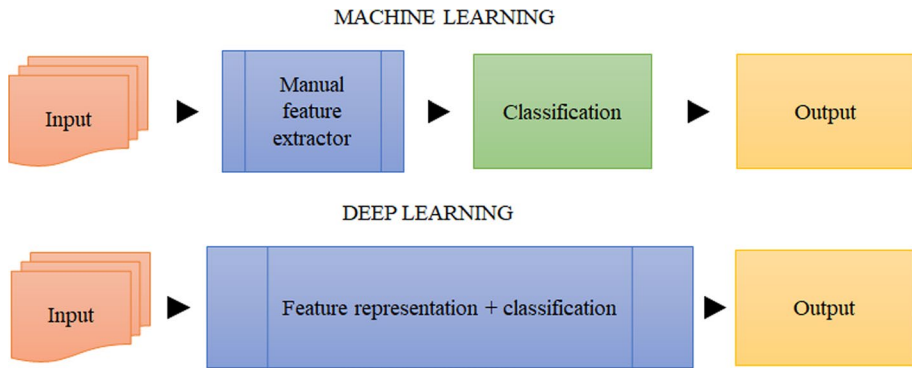


Fig. 3 Difference between ML and DL

domain has been utilized thus far for non-stationary signals, which are prevalent when machine faults happen. Several time-frequency signal analysis methods have been established, such as wavelet transform, Hilbert-Huang transform, and short-time Fourier transform.

2.2 Deep learning

The birth of DNN, hereafter referred to as DL, in 2010 marked a significant breakthrough in the learning paradigm after being previously dominated by ML techniques since the 1990s. More importantly, DNN and ML form artificial intelligence (AI), which is a term that refers to the ability of machines to mimic human intelligence without explicit programming (Simon 1969). Figure 2 shows the relationship and a time frame between AI, ML, and DL.

DL uses feature learning rather than handcrafted feature engineering to perform the task (LeCun et al. 2015). It possesses unique properties over shallow ML algorithms with respect to feature learning, model construction, and model training (see Fig. 3). In addition, DL comprises tens or even hundreds of layer representations that are mainly learned through models, called neural networks. A neural network is a learning algorithm in which a number of neurons and activation functions are utilized to produce a nonlinear transformation of the input samples.

A ML algorithm learns the features of input data X with the help of a feature extractor. The extracted features ϕ are used to train a classifier $\mathcal{F}_\theta(\cdot)$ that outputs predictions Y . In contrast, a DL algorithm requires no handcrafted feature representation. The input data X is transformed using \mathcal{T}_ϕ , where ϕ is made up of several *learnable* parameters of the transformation. The transformation provides a new representation of $\mathcal{T}_\phi(X)$ that is further used for classification $\mathcal{F}_\phi(\cdot)$ to produce final output Y .

DL techniques have been predominantly used in the fault diagnosis of rotating machinery. They can be trained either in a supervised or unsupervised manner or in other learning paradigms, e.g., reinforcement learning (Salimans and Kingma 2016). The readers are encouraged to see a good overview of DL available in the current literature (Pouyanfar et al. 2018; Schmidhuber 2015). Several well-known DL architectures are convolutional neural networks (CNNs), deep belief networks (DBNs), recurrent neural networks (RNNs), generative neural networks, and graph neural network (GNN). We briefly discuss these numerous architectures in the following section.

2.2.1 Convolutional neural network (CNN)

CNN, a type of neural network that specializes in processing data with a grid-like topology, was first introduced in the LeNet-5 architecture (LeCun et al. 1995) and is referred to as convolutional 2D. CNN is typically utilized on image data. Because the kernel moves along two dimensions on the data, CNN is known as a two-dimensional CNN. Consequently, this neural network is distinct from the one-dimensional convolutional neural networks (1D CNN). In 1D CNN, the kernel slides along a single dimension that is typically used for time series data (Kiranyaz et al. 2021). The primary distinction between 1D and 2D CNNs is that both the kernels and the feature maps are represented by 1D arrays rather than 2D matrices (see Fig. 4 for typical CNN architecture).

In addition, the computational difficulty of 1D and 2D convolutions differs significantly. For example, an image with $M \times M$ dimensions convolved with a $T \times T$ kernel will incur

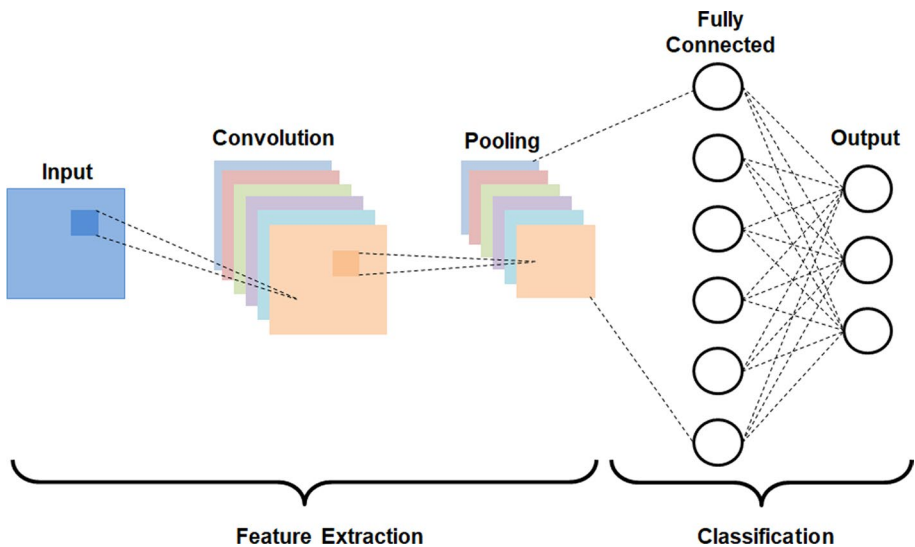


Fig. 4 A typical architecture of CNN (LeCun et al. 1995)

a computational cost of $\mathcal{O}(M^2T^2)$, whereas a comparable 1D convolution with the same dimensions (M and T) will incur a computational cost of $\mathcal{O}(MT)$. This indicates that the computational complexity of a 1D CNN is considerably lower than a 2D CNN under comparable circumstances (i.e., same design, network, and hyperparameters).

Furthermore, in contrast to typical neural networks, a CNN is made up of convolutional layers followed by pooling (e.g., subsampling) layers and fully connected layers (e.g., the final stage layers) (LeCun et al. 1995) (see Fig. 4). Each layer has its own customizable parameter set and does something different to the input data. These differences are shown in Table 1. A CNN's main building block is the convolution layer, which is responsible for the majority of the network's computational load. This layer does a dot product between two matrices, one of which is a set of *learnable* parameters known as a kernel, and the other is the restricted section of the receptive field. The kernel is a fraction of the size of an input image. Suppose we have an input image of size $m \times n$ and let $q \times r$ be the size of t kernels of the convolutional layer. The kernel slides over the image's height and width during the forward pass, providing an image representation of that receptive region. These create an activation map or feature map, a two-dimensional image representation that shows the kernel's response at each spatial point in the image. A stride S denotes the kernel's sliding size with the amount of padding P . Note that the size of the feature map is less than the size of the input image (see Fig. 5). Formally, the output of the convolutional layer is defined as a set of t feature maps of size $((m - q + 2P)/S) + 1 \times ((n - r + 2P)/S) + 1$.

Pixels in the convolutional layer are obtained by registering a weight w_{ij} to each pixel of the input image, calculating the weighted sums, and extracting related features of the image. Next, a bias ϵ is added to the weighted sums, and they are forwarded to a non-linear activation function φ , which can be either $\tanh()$ or $\text{sigm}()$. The output of the activation function y_r^s of a given feature map r in convolutional layer s is defined as:

$$y_r^s = \varphi \left[\epsilon_r^s + \sum_{i \in M_r^s} y_i^{s-1} \otimes w_{ir}^s \right] \quad (1)$$

where φ is non-linear activation function and ϵ_r^s is the bias for the s -th layer. M_r^s and \otimes denote the selected feature map i in the $(s - 1)$ -th layer and a convolutional operator, respectively.

The procedure is continue with the pooling layer. Pooling layers are similar to convolutional layers, but they execute a specific function, such as max pooling, which takes the largest value in a certain filter region, or average pooling, which takes the average value in a particular filter region. These are commonly employed to lower the network's dimensionality. Activation map as the output f_h^r after reduction-sampling the feature map r into a feature map h in a layer s is calculated as:

$$f_h^s = \delta(g_r^s, N^s) \quad (2)$$

where δ is the reduction-sampling function by a factor of N^s , while g_r^s is the convoluted feature map to be reduced. At the final layer, the network is characterized by the classification layer. The output of the classification layer o is calculated as:

$$o = \xi(b_o + W_z) \quad (3)$$

where b_o is the bias of the output layer, W is weight matrix between the second-to-final layer and the output layer, and z is the concatenated feature maps of the second-to-final layer.

Table 1 Comparison of different layers in CNN

Layers	Action	Parameters
Convolution layer	Apply filters to extract features	Number of kernels
	Filters are composed of small kernels, learned	Size of kernels (height and width)
	One bias per filter	Activation function
	Apply activation function on every value of feature map	Stride
Pooling layer		Padding
	Reduce dimensionality	Regularization type and value
	Extract maximum or average value of a region	Stride
	Sliding window approach	Size of window
Fully connected layer	Aggregate information from final feature maps	Number of nodes
	Generate final classification	Activation function, depending on role of layer. If aggregating info, use ReLU. If producing final classification, use Softmax

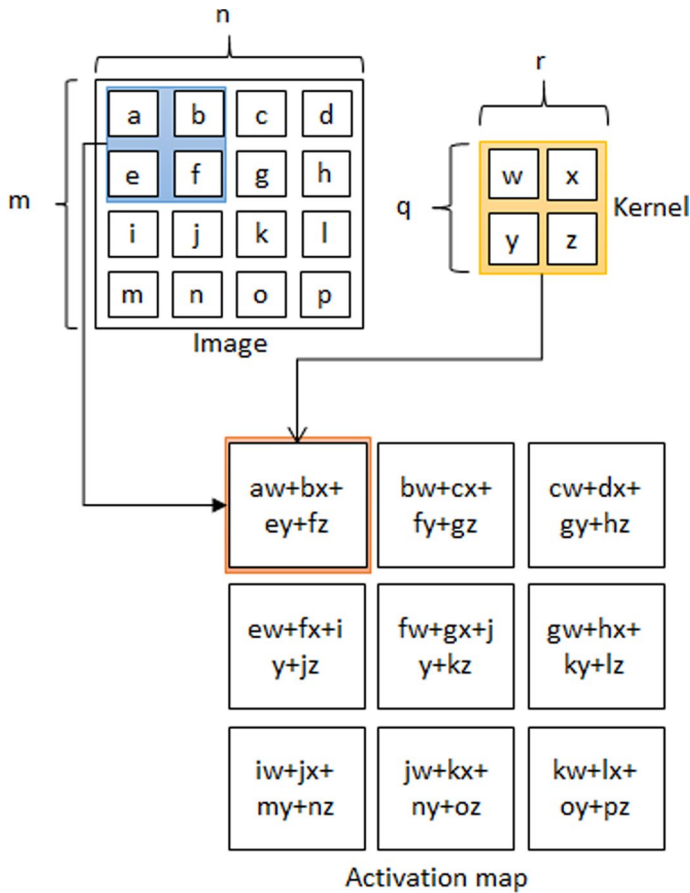


Fig. 5 Convolutional process in a CNN's convolution layer

The limitations of the traditional CNN have led to the emergence of a new type of CNN known as Capsule Network (CapsNet). The CapsNet architecture is comprised of an encoder and a decoder (Sabour et al. 2017), each with three layers. A convolutional layer, a PrimaryCaps layer, and a DigitCaps layer make up an encoder. A decoder has three fully connected layers (see Fig. 6). The Capsule Network and CNN operate similarly. Convolutional neural networks (CNNs) use max-pooling and scalar-output feature detectors. However, in the capsule network, routing by agreement replaces max-pooling and CNNs' scalar-output feature detectors are replaced with vector-output capsules. Therefore, capsules networks outputs are vectors that have a direction. Sabour et al. (2017) emphasize the preservation of specific information concerning an object's location and posture across the network. This process, known as equivariance, is a crucial characteristic of CapsNet. Sabour et al. (2017) develop the process by employing CNN-like higher-level capsules that cover larger areas of the image. However, unlike max-pooling, CapsNet does not discard information regarding the precise location of an entity within the region. CapsNet has four main steps:

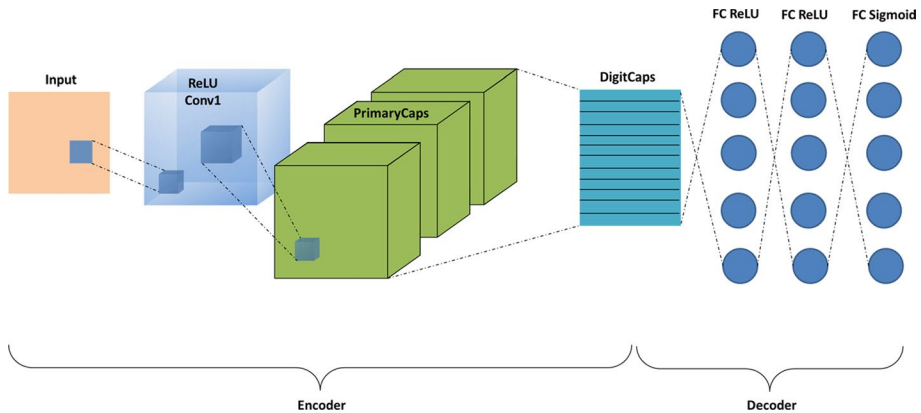


Fig. 6 A typical architecture of a Capsule Network (CapsNet) (Sabour et al. 2017)

- (i) Matrix multiplication of input vectors with weight matrices: This operation, performed on the input layer, converts the input image to vector values. These vectors are initially multiplied by the weight matrices. As mentioned previously, the weight matrix represents spatial relationships.
- (ii) Weighting the input vectors: The input weights determine the direction that the current capsules should move in the next layer. It complements the component of the dynamic routing algorithm described below.
- (iii) Dynamic routing algorithm: The dynamic routing technique, which follows the aims of CapsNet, facilitates communication between the PrimaryCaps and DigitCaps layers. The PrimaryCaps layer (i.e., lower capsule) sends its input to the DigitCaps layer (i.e., higher-level capsules) that “agrees” with its input during this routing process. This means that lower-level attributes will only be transferred to higher-level layers whose contents match.
- (iv) Applying a non-linear function to condense the information: CapsNet utilizes the squashing (i.e., nonlinear) function to compress information to a length between 0 and 1 as opposed to the ReLU function used by conventional CNNs. Long vectors are compressed to a length just under one, while short vectors are compressed to nearly zero. Therefore, the length of a capsule’s output vector reflects the likelihood that a particular object is present in the current input.

It is worth mentioning that CapsNet appears promising as a superior alternative to CNN, but it is still in its infancy. Therefore, researchers must continue to investigate and develop this network in order to enhance its performance and application-use potential.

2.2.2 Deep belief network (DBN)

DBN is a hybrid probabilistic generative model. It can be viewed as a composition of several Restricted Boltzmann Machines (RBMs) (Hinton 2009; Hinton et al. 2006). Figure 7 shows a feedforward architecture of DBN with three hidden layers. Let $\{(x^1, y^1), (x^2, y^2), \dots, (x^n, y^n)\}$ be a training dataset with N instances. An input vector \vec{x} comes across the nodes in the network, where DBN provides initial weights \vec{w} . The objective is to minimize a cost function C :

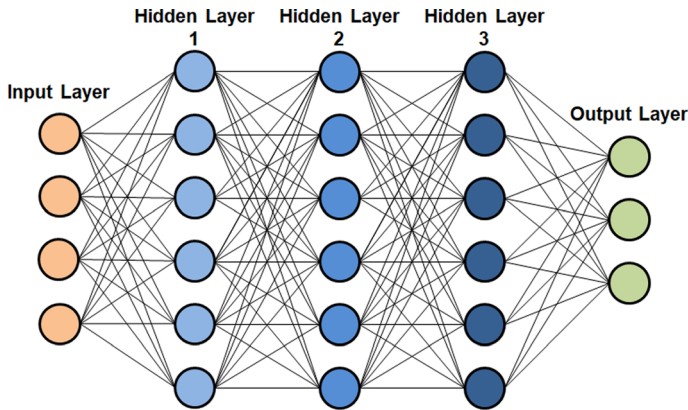


Fig. 7 A typical architecture of a DBN (Hinton 2009; Hinton et al. 2006)

$$C(\vec{w}, \vec{x}, y) = \frac{1}{2} \|h_w(\vec{x}) - y\|^2 \quad (4)$$

where $h_w(\vec{x})$ is a hypothesis function, providing an predicted output. The overall cost is calculated as:

$$C(\vec{w}) = \frac{1}{N} \sum_n C(\vec{w}, \vec{x}^n, y^n) + \frac{\lambda}{2} \sum_k^K \sum_i^{L_m} \sum_j^{L_{m+1}} (w_{ij}^k)^2 \quad (5)$$

where K denotes the depth of network, L_m is the number of nodes in the m -th layer, while $w_{ij}^k \in \vec{w}$ is the weight of edges between a node i in the layer $k - 1$ and a node j in the layer k . The parameter set \vec{w}^* is calculated to minimize the overall cost function as follows.

$$\vec{w}^* = \underset{w}{\operatorname{argmin}} C(\vec{w}) \quad (6)$$

A backpropagation algorithm is commonly used to get the weight vectors \vec{w} , which are updated from the top layer to the bottom layer by solving this equation:

$$w_{ij}^k = w_{ij}^{k-1} + \xi \frac{\partial}{\partial w_{ij}^k} C(\vec{w}) \quad (7)$$

where ξ specifies the adaptation parameter.

2.2.3 Recurrent neural network (RNN)

RNN is a well-known algorithm in DL that is prevalently used in natural language processing and speech processing (Cho et al. 2014). It employs the sequential information in the network, thus allowing an extraction of useful knowledge from the embedded structure in a data sequence. A long short-term memory (LSTM) network is an improved version of RNN. A typical LSTM network is shown in Fig. 8. The LSTM module is denoted by blue circles in the figure.

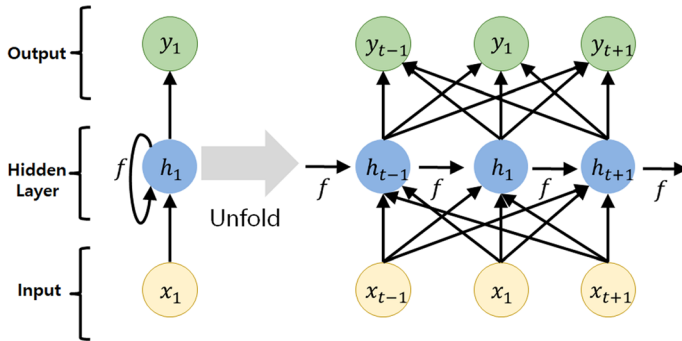


Fig. 8 A typical architecture of a long short-term memory network (Cho et al. 2014)

The fundamental concept of LSTM is forward propagation. Considering Fig. 8 as an example, a conventional LSTM consists of the input vector, hidden state, and output vector at time t is \vec{x}_t , \vec{h}_t , \vec{y}_t , respectively. Next, the hidden state at time t is given by a function f of the input vector at time t and the hidden vector at time $(t - 1)$ is:

$$\vec{h}_t = f(\vec{h}_{t-1}, \vec{x}_t) \quad (8)$$

A separate function g is used to learn the output probabilities from the hidden states. Therefore, the output probabilities can be defined as follows.

$$\vec{y}_t = g(\vec{h}_t) \quad (9)$$

More concretely, let W_{xh} be the $p \times d$ input-hidden matrix, W_{hh} be the $p \times p$ hidden-hidden matrix, and W_{hy} be the $d \times p$ hidden-output matrix, then the outputs of Eq. 9 can be rewritten as:

$$\begin{aligned} \vec{h}_t &= \tanh(W_{xh}\vec{x}_t + W_{hh}\vec{h}_{t-1}) \\ \vec{y}_t &= W_{hy}\vec{h}_t \end{aligned} \quad (10)$$

2.2.4 Generative neural network

Other DL algorithms that are prevalently used as generative neural networks are variational autoencoders (VAEs) (Kingma and Welling 2013) and generative adversarial networks (GANs) (Goodfellow et al. 2014). Generative neural networks work with two neural network models simultaneously. The objective of a generative neural network is to generate synthetic objects that are too realistic for the trained classifier to distinguish whether a particular object belongs to the original dataset or not.

VAE is an autoencoder whose training is regularized in order to avoid overfitting and ensure that the latent space has good properties that enable a generative process. VAE consists of both an encoder and a decoder and is trained to minimize the reconstruction error between the encoded-decoded data and the original data. First, the input is encoded as the distribution over the latent space and a point from the latent space is sampled from that distribution. Then, the sampled point is decoded and the reconstruction error is backpropagated. VAE uses KL-divergence as its loss function to measure the difference between the

generated distribution and the original distribution of the data. Suppose that a distribution z is known and generating the observation x from this distribution, which means calculating the probability using Bayes' theorem as Eq. 11, is necessary. However, due to the probability $p(x)$, which is calculated as Eq. 12 with the integral form, calculating $p(z|x)$ is difficult.

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} \quad (11)$$

$$p(x) = \int p(x|z)p(z)dz \quad (12)$$

Instead of directly calculating $p(z|x)$, $p(z|x)$ can be estimated as $q(z|x)$ through minimizing the KL-divergence loss as Eq. 13. Equation 13 can be converted to the maximization problem as Eq. 14, where $E_{q(z|x)} \log p(x|z)$ is the reconstruction likelihood and $KL(q(z|x)||p(z))$ indicates how the estimated distribution q is similar to the prior distribution p . Finally, the total loss for VAE can be expressed as Eq. 15 and is backpropagated to the network (Kingma and Welling 2013).

$$\min KL(q(z|x)||p(z|x)) \quad (13)$$

$$\max E_{q(z|x)} \log p(x|z) - KL(q(z|x)||p(z)) \quad (14)$$

$$Loss = L(x, \hat{x}) + \sum_i KL(q_i(z|x)||p(z)) \quad (15)$$

GAN is another kind of generative neural networks, which learns deep representations without extensively annotated training data and can be used for both supervised and unsupervised learning. GAN can be described as a two-player game between the "generator" and the "discriminator". On the one hand, the generator artificially generates samples that are intended to come from the same distribution of the real data and is trained to deceive the discriminator. On the other hand, the discriminator classifies samples into two classes (i.e., real and fake) and is trained using the traditional classifiers not to be deceived by the generator (Figs. 9 and 10).

GAN is typically implemented by multi-layer neural networks with convolutional and/or fully connected layers. A prior on input noise variables is represented as $p_z(z)$ to learn the generator's distribution p_g over the generated data x . Mapping to a data space is illustrated as $G(z; \theta_g)$ and $D(x; \theta_d)$, where G is a differentiable function, which is represented by a multi-layer perceptron with θ_g , and $D(x)$ indicates the probability that x came from the real data rather than the generator's distribution (i.e., p_g). D is trained to maximize the probability of assigning the correct labels to both the samples from the real data and the samples from G , while G is simultaneously trained to minimize $\log(1 - D(G(z)))$. The process above can be illustrated as follows, which is a minimax game with value function $V(G, D)$ (Goodfellow 2016; Goodfellow et al. 2014).

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{x \sim p_z(z)} [\log(1 - D(G(z)))] \quad (16)$$

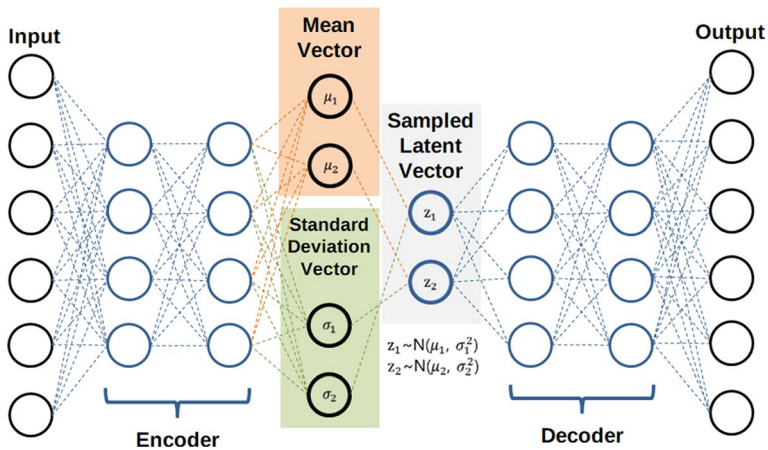


Fig. 9 A basic architecture of a variational autoencoder (VAE) (Kingma and Welling 2013)

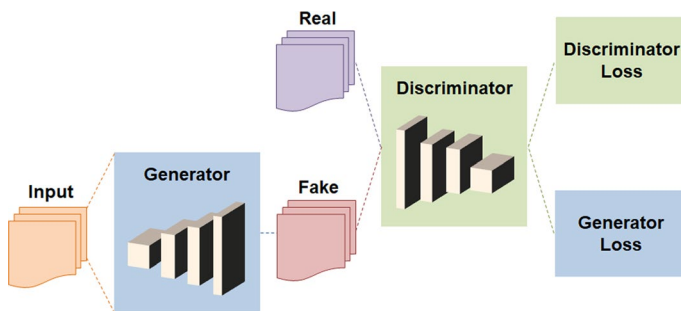


Fig. 10 A typical architecture of GAN (Goodfellow et al. 2014)

2.2.5 Graph neural network (GNN)

Scarselli et al. (2009) first proposed GNNs, which aim to use graph theory to develop a neural network that could model graph data. In response to the success of CNNs, RNNs, and autoencoders in DL, new methodologies and descriptions have been extended to complex graph data, resulting in graph convolutional neural networks (GCNs), graph recurrent neural networks (GRNNs), and graph auto-encoders (GAEs). These GNNs have been applied successfully in a wide number of domains. However, there is a dearth of applications for using GNNs to diagnose faults in rotating machinery via vibration data. Figure 11 illustrates a typical GNN architecture for fault diagnosis.

A simple graph can be represented as:

$$G = G(V, E) \quad (17)$$

where V and E are the sets of nodes and edges, respectively. Let $v_i \in V$ be a node and $e_{ij} = (v_i, v_j) \in E$ denote an edge between v_i and v_j . Then, the neighborhood of a node v can be defined as $M(v) = \{u \in V | (v, u) \in E\}$. Usually, a graph can be described by an adjacency matrix $A \in R^{M \times M}$ where M is the number of nodes, that is, $M = |V|$. In particular, $A_{ij} = 1$ if $\{v_i, v_j\} \in E$ and $i \neq j$; $A_{ij} = 0$, otherwise. In a directed graph, A_{ij} represents an

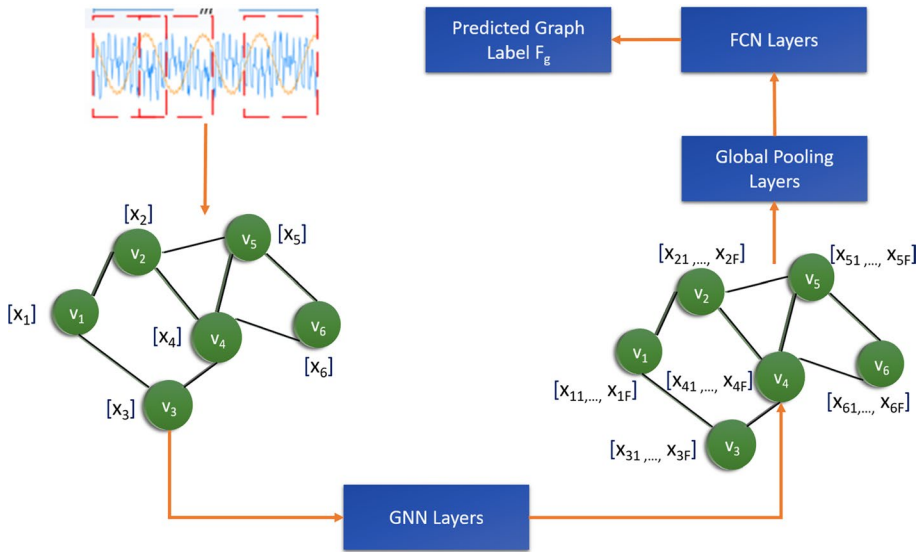


Fig. 11 Typical GNNs architecture for fault diagnosis (Chen et al. 2020)

edge pointing from v_i to v_j while, in an undirected graph, A_{ij} denotes an edge connection between nodes v_i and v_j .

In practical applications, a graph has nodes known as features that represent as feature matrix $X \in R^{M \times c}$ where c is the dimension of a node feature matrix. A degree matrix $DM \in R^{M \times M}$ is a diagonal matrix, which can be obtained as $DM_{ii} = \sum_{j=1}^N A_{ij}$. The graph can be represented as the Laplacian matrix LM . The relationship between the degree matrix, the adjacency matrix, and the Laplacian matrix is illustrated in Fig. 12. The Laplacian matrix is defined as:

$$LM = DM - A \quad (18)$$

GNN receives two piece of information as the input, the feature matrix X and the adjacency matrix A . The output of GNNs is obtained by the general forward propagation that is defined as :

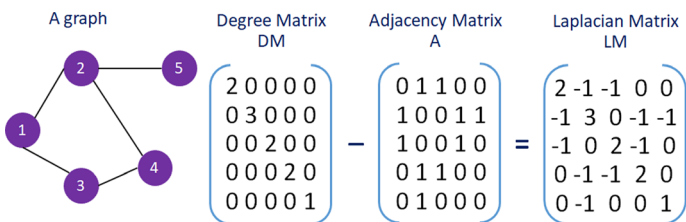


Fig. 12 Illustration of the relationship between the degree matrix, the adjacency matrix, and the Laplacian matrix

$$\mathbf{Z}_G = \text{softmax}(G_k(\dots, G_2(\mathbf{A}, G_1(\mathbf{A}, \mathbf{X})))) \quad (19)$$

where G denotes an operation of GNN, and k is the number of layers in the architecture.

Like images, industrial networks often represent a structure. For example, a station voltage in an electric power system may be independent of a remote load but be highly correlated to a nearby generator. Unfortunately, CNNs perform well only for regular data structures like images and cannot achieve the same performance for domains with an irregular structure like industrial networks. The data's geometrical structure in the graph structure can provide additional information, as they encompass the nodes' values and their relationships. As a result, the graph structure can convey more information than regular data structures. Recently, GCNs have been introduced to generalize CNN to some irregular or, more generally, non-Euclidean domains (Bruna et al. 2013). Input, convolution, coarsening, and pooling are the four fundamental phases of a graph convolutional network. More precisely, four fundamental phases can be described as follows.

- (i) Graphs may express the geometry and structure of data as input to a neural network and therefore provide more information than other general data formats.
- (ii) Fast-localized filters based on graph spectral theory and Chebyshev expansion are utilized to conduct a convolution operator on a graph.
- (iii) A graph coarsening process is conducted to group similar vertices together.
- (iv) For graph pooling, the vertices are rearranged by first generating a balanced binary tree then applying regular 1D pooling.

The GCN's output can be derived based on Eq. 19 and following the suggestion by Kipf and Welling (2017), therefore it is defined as:

$$\mathbf{Z} = \sigma\left(\left(\tilde{\mathbf{D}}\tilde{\mathbf{M}}^{-0.5}\tilde{\mathbf{A}}\tilde{\mathbf{D}}\tilde{\mathbf{M}}^{-0.5}\right)\mathbf{X}\boldsymbol{\Theta}\right) \quad (20)$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_M$, \mathbf{I}_M is the identity matrix of order M , and $\tilde{\mathbf{D}}\tilde{\mathbf{M}}$ is a diagonal matrix, whose diagonal elements are $\tilde{\mathbf{D}}\tilde{\mathbf{M}}_{ii} = \sum_{j=1}^M \tilde{\mathbf{A}}_{ij}$. σ is an activation function, e.g., the ReLU function. $\boldsymbol{\Theta} \in \mathcal{R}^{c \times c'}$ is the parameter matrix of the network to be learned, c' represents the dimension of the output, and $\mathbf{Z} \in \mathcal{R}^{M \times c'}$ is the output matrix.

The cross-entropy loss function that is generally used to train the parameters is defined as:

$$Loss = - \sum_{i=1}^{M_r} \sum_{j=1}^P \mathbf{Z}_{ij} \log \mathbf{Y}_{ij} \quad (21)$$

here M_r represents the number of training points, P is the number of categories, \mathbf{Z} is obtained from Eq. 20, and \mathbf{Y} is the corresponding label of the training set. It should be noted that in a classification problem, the dimension of the output equals the number of categories, i.e., $c' = P$ (Kipf and Welling 2017). Figure 13 shows a graph convolutional network-based fault diagnosis method using vibration signals. The framework is comprised of several steps, including a graph construction of the vibration signals, followed by applying a certain activation function like the softmax function as the classifier for mechanical fault diagnosis.

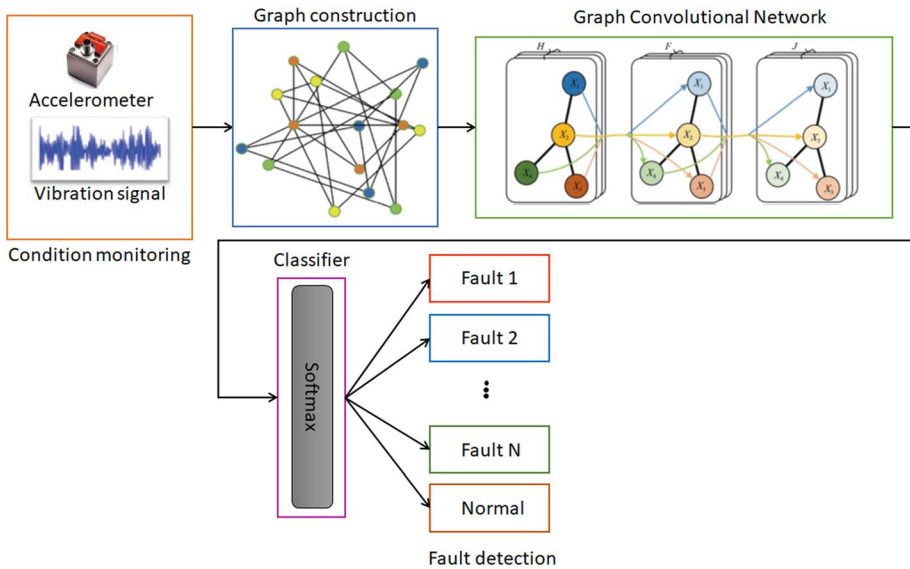


Fig. 13 Graph convolutional network-based approach for fault diagnosis (Liu et al. 2021)

3 Methodology

This section discusses the methodology used for conducting the systematic review. We follow the PRISMA guideline for conducting this systematic review (Liberati et al. 2009; Moher et al. 2009).

3.1 Search procedure

Considering that DL techniques were first introduced in 2010, we investigate primary publications that appeared over the last 6 years, from January 2016 to January 2022. An automatic search was utilized for hunting relevant publications as completely as possible. More precisely, we searched two well-known indexing databases, such as Scopus and Web of Science, to incorporate all prior works comprehensively. The two indexing databases normally provide indexing services for journals published by IEEE, ACM, Springer, Elsevier, John Wiley & Sons, and other publishers. We limited our search to these databases because: (i) the databases are supposed to archive high-quality papers; and (ii) to avoid multiple searches, which are typically available at each publication outlet. Some search keywords were derived from index terms found in some papers. The search keywords were combined using Boolean operators, i.e., AND and OR, resulting in the following combination of the keywords.

(Condition monitoring OR vibration signals OR vibration analysis OR rotating machinery OR fault detection OR fault diagnosis OR rolling bearings OR health monitoring OR planetary gearbox)
AND
(Deep learning OR convolutional neural network OR autoencoder OR recurrent neural network OR deep belief network OR feature learning OR graph neural network)

3.2 Study inclusion and exclusion procedure

In order to obtain the paper as relevant as possible, the search results were refined using a set of inclusion and exclusion criteria as shown in Table 2. All studies considered in this review must be papers that were published in journals because journal articles represent the highest level of research outcomes and have been thoroughly peer-reviewed (Nord and Nord 1995). Furthermore, the papers must be written in English and discuss DL and vibration signals for rotating machinery. The steps involved in the study inclusion and exclusion, including the number of relevant studies at each step, are presented in Fig. 14.

4 Result and discussion

This section is devoted to answer RQ₁ and RQ₂. It is comprised of the discussion about the research trend, publication venues, and the categorization of the selected studies w.r.t DL techniques.

4.1 Research trend

According to the trend (see Fig. 15), the number of studies has been varying over time. There have been a surge studies between 2017 and 2021, while in contrast, there exist limited studies in 2016. Moreover, there has been at least five studies have been identified discussing the application of DL techniques in fault diagnosis using vibration signals.

4.2 Publication venue

This section presents the dissemination of the selected studies by considering the publication outlets. Table 3 enumerates all publication outlets, where the selected studies were published based on the journal title, number of studies, publisher, and the relative percentage to the total number of selected studies. The selected studies appeared in 27 kinds of journals, where the Measurement journal (10 articles) takes the predominant publications in this field. It is followed by IEEE Transactions on Instrumentation and Measurement that shares 7 articles. Next, IEEE Transactions on Industrial Electronics and Mechanical Systems and Signal Processing share 5 articles each, while other well-known journals are IEEE Access (3 articles) and IEEE Transactions on Industrial Informatics (3 articles). The other outlets that share the same number of publications (2 articles) are Advances in Mechanical Engineering, Computers in Industry, International Journal of Advanced Manufacturing Technology, ISA Transactions, and Neurocomputing. Lastly, the rest outlets had published one article of the total selected studies.

4.3 Categorization of selected studies w.r.t. deep learning methods

As previously noted, condition monitoring is an integral component of condition-based maintenance. Maintenance is determined in accordance with the machine's health, which can be identified through the condition monitoring system. The essential purpose of condition monitoring is to avert a catastrophic machine failure that might trigger a chain reaction

Table 2 Set of inclusion and exclusion criteria to get the relevant papers

Type	Inclusion	Exclusion
Type of publication	Journal	Conference proceedings, PhD thesis, or gray literature, i.e., working papers, editorial comments, technical reports
Language	English publications	Non-English publications
Scope	DL for machine fault diagnosis using vibration signals	Publications that fall outside the broader field of vibration signal analysis, DL, and machine fault diagnosis

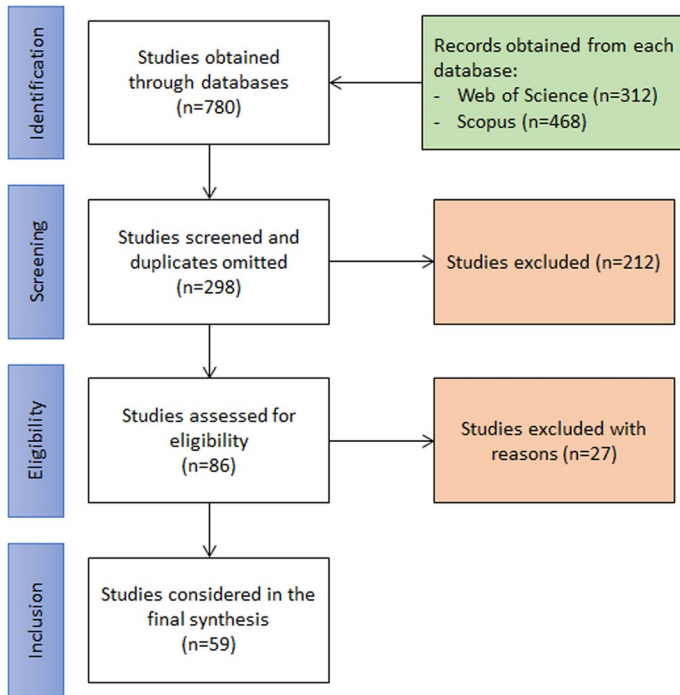


Fig. 14 Procedure in study inclusion and exclusion by following PRISMA framework (Liberati et al. 2009; Moher et al. 2009)

Fig. 15 Proportion of the selected studies w.r.t. year of publication

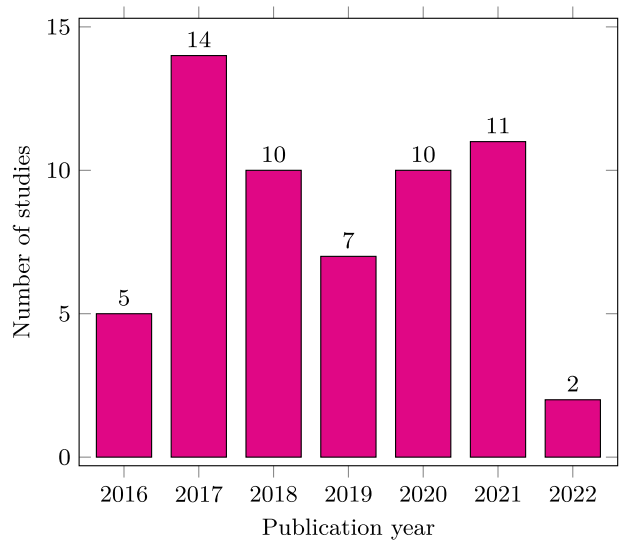
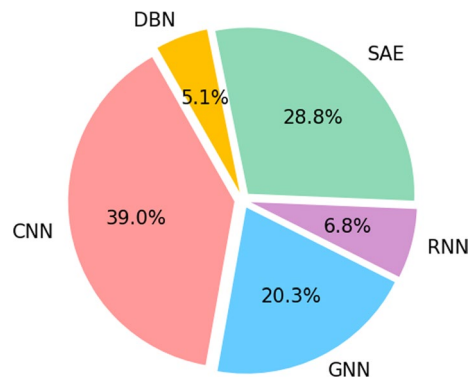


Table 3 Dissemination of the selected studies w.r.t. publication outlets (in alphabetically order)

No.	Title	Publisher	#	Proportion
1.	Advanced Engineering Informatics	Elsevier	1	1.69%
2.	Advances in Mechanical Engineering	SAGE	2	3.39%
3.	Applied Acoustics	Elsevier	1	1.69%
4.	Applied Intelligence	Springer	1	1.69%
5.	Chinese Journal of Aeronautics	Elsevier	1	1.69%
6.	Cognitive Systems Research	Elsevier	1	1.69%
7.	Computers in Industry	Elsevier	2	3.39%
8.	Frontiers of Mechanical Engineering	Springer	1	1.69%
9.	IEEE Access	IEEE	3	5.08%
10.	IEEE Transactions on Cybernetics	IEEE	1	1.69%
11.	IEEE Transactions on Industrial Electronics	IEEE	5	8.47%
12.	IEEE Transactions on Industrial Informatics	IEEE	3	5.08%
13.	IEEE Transactions on Instrumentation and Measurement	IEEE	7	11.86%
14.	International Journal of Advanced Manufacturing Technology	Springer	2	3.39%
15.	ISA Transactions	Elsevier	2	3.39%
16.	Journal of Intelligent Manufacturing	Springer	1	1.69%
17.	Journal of Manufacturing Systems	Springer	1	1.69%
18.	Journal of Mechanical Engineering Science	SAGE	1	1.69%
19.	Journal of Risk and Reliability	SAGE	1	1.69%
20.	Journal of Sound and Vibration	Elsevier	1	1.69%
21.	Knowledge-Based Systems	Elsevier	1	1.69%
22.	Measurement	Elsevier	10	16.95
23.	Measurement Science and Technology	IOP	1	1.69%
24.	Mechanical Systems and Signal Processing	Elsevier	5	8.47%
25.	Neurocomputing	Elsevier	2	3.39%
26.	Reliability Engineering & System Safety	Elsevier	1	1.69%
27.	Signal Processing	Elsevier	1	1.69%

Fig. 16 Proportion of the selected studies w.r.t. DL techniques

of more damage, breakdown, production failure, and rising maintenance costs. In recent years, there has been an increase in the quantity of work applying DL to condition monitoring. This section discusses state-of-the-art approaches for fault diagnosis based on DL and vibration analysis. Figure 16 illustrates the contribution of each DL technique to the overall number of studies chosen. The vast majority of research uses CNNs, followed by stacked autoencoders (SAEs), accounting for 39.0% and 28.8% of the total studies, respectively. GNNs account for 20.3% of the total research, while RNNs and DBNs account for 6.8% and 5.1%, respectively.

4.3.1 Convolutional neural network (CNN)

The results of selected studies utilizing CNNs for vibration-based fault diagnosis are summarized (in chronological order) in Table 4. A novel hierarchical adaptive CNN is studied in (Guo et al. 2016). By adopting a CNN architecture provided in (Slavkovikj et al. 2015), Janssens et al. (2016) propose a feature learning model for conditional monitoring. A novel contribution to the spindle bearing fault diagnosis is proposed (Ding and He 2017). The proposed approach is based on energy-fluctuated multiscale feature mining using the wavelet packet energy image and CNN. An effective and reliable bearing fault diagnosis approach based on CNN is investigated in (Lu et al. 2017a). The CNN model can reduce learning computation requirements in the temporal dimension. It is worth mentioning that a dislocated layer can be added to the original CNN. The authors in (Liu et al. 2017) propose a method called dislocated time series CNN to develop a new diagnosis framework based on the characteristics of industrial vibration signals. The authors of (Jing et al. 2017) develop CNN to learn features directly from the frequency data of vibration signals. The effectiveness of the proposed CNN is evaluated on the PHM 2009 gearbox challenge data and planetary gearbox test rig. A new approach for fault types and severity of the gearbox health monitoring system is presented (Wang et al. 2017). A deep CNN learns the underlying features in the time-frequency domain and performs classification. Experiments show the efficiency and effectiveness of the proposed model with a classification accuracy of more than 99.5%.

A combination of CNN and support vector machine (SVM) for machinery fault diagnosis techniques is proposed (You et al. 2017). CNN promotes feature extraction capability, while SVM is used for multi-class classification. The proposed model is evaluated using the real acoustic signals from locomotive bearings and vibration signals measured from the automobile transmission gearbox. A new CNN based on LeNet-5 for fault diagnosis method is detailed in (Wen et al. 2018). By converting signals into two-dimensional images, the proposed method can extract the features and wipe out the ones with manual design features. A novel method suitable for mechanical data analysis, called LiftingNet, is proposed (Pan et al. 2018). Two different motor-bearing datasets validate the effectiveness of LiftingNet. A learning paradigm called transfer learning based on CNN is employed in (Cao et al. 2018). In this research, transfer learning is applied to gearbox fault diagnosis, while vibration responses are obtained using accelerometers during gearbox operation. A study undertaken in (Chen et al. 2019a) employs a data fusion technique to determine the health condition of a machine. The classification is then performed using a deep convolutional neural network (DCNN). The findings indicate that DCNNs outperform SVMs and backpropagation neural networks.

A multi-sensor data fusion and bottleneck layer optimization of CNN are discussed in (Wang et al. 2019a). The proposed approach has been validated on a wind-powered test

Table 4 Summary of CNNs-based fault diagnosis methods using vibration signals

Ref.	Element(s)	Pre-processing approach(es)	Number of convolutional layer(s)	Number of subsampling layer(s)	Dataset	Validation method(s)	Testing accuracy (%)	Average computational time (s)
Guo et al. (2016)	Rolling bearing	–	3	3	CWRU	10cv and holdout	10cv: 98.1; holdout: 99.7	–
Janssens et al. (2016)	Bearing	Discrete Fourier transform	3	–	Private	LOCV	93.61	–
Ding and He (2017)	Spindle bearing	Wavelet packet transform	3	2	CWRU	Hold-out	98.8	–
Lu et al. (2017a)	Rolling bearing	–	2	2	CWRU and Private	10cv	CWRU: 92.60%±1.17; Private: 95.78%±0.66	CWRU: 217±14; private: 40±5
Liu et al. (2017)	Induction motor	–	4	2	Private	Holdout	96.32	–
Jing et al. (2017)	Gearbox	FFT	1	1	PHM 2009 and Private	Holdout	PHM 2009: 99.33±0.50; Private: 98.67±0.52	–
Wang et al. (2017)	Gearbox	Wavelet transform	2	2	Private	Holdout	99.58	–
You et al. (2017)	Locomotive bearing, automobile transmission gearbox	–	3	3	Private	Holdout	99.67	–
Wen et al. (2018)	Motor bearing, centrifugal pump, hydraulic pump	Time-domain to images	4	4	CWRU, centrifugal pump, and hydraulic pump (Lu et al. (2017b))	Holdout	CWRU: 99.79; centrifugal pump: 99.48; hydraulic pump: 100	–

Table 4 (continued)

Ref.	Element(s)	Pre-processing approach(es)	Number of convolutional layer(s)	Number of subsampling layer(s)	Dataset	Validation method(s)	Testing accuracy (%)	Average computational time (s)
Pan et al. (2018)	Motor bearing	Wavelet transform	6	1	CWRU and Private	10cv	CWRU: 99.6; private: 93.19	CWRU: 0.0538; Private: 0.0508
Cao et al. (2018)	Gearbox	–	5	3	Private	Holdout	53–74.92	–
Chen et al. (2019a)	Gearbox	Data stitching, channel overlay	6	6	CWRU	Holdout	99.22	–
Wang et al. (2019a)	Wind power, centrifugal pump	Signal-to-image conversion	4	3	Private	10cv	Wind power: 99.47; centrifugal pump: 97.32	–
Hoang and Kang (2019)	Rolling bearing	Signal-to-image conversion	2	2	CWRU	Holdout	100	–
Chen et al. (2019b)	Rolling bearing	–	4	–	CWRU	10cv	96.7	–
Khodja et al. (2020)	Rolling bearing	Vibration spectrum imaging, FFT	2	2	CWRU and private	Holdout	CWRU: 94.74; private: 96.92	–
Wu et al. (2021)	Rolling bearing	–	3	3	CWRU and private	Holdout	CWRU: 95.25; private: 88.71	2.719
Liang et al. (2021)	Rolling bearing	CWT	6	2	CWRU	Holdout	93.58	–
Zhu et al. (2020)	Rolling bearing	Teager energy operator	4	2	CWRU	Multiple holdout	96.25±1.47	–
Yu et al. (2020)	Rolling bearing	–	4	4	Private	Holdout	91.00	–
Yang et al. (2020)	Motor, gearbox, and locomotive bearings	–	34	–	CWRU and private	Holdout	CWRU: 88.27±4.27; private: 80.12±3.43	–

Table 4 (continued)

Ref.	Element(s)	Pre-processing approach(es)	Number of convolutional layer(s)	Number of subsampling layer(s)	Dataset	Validation method(s)	Testing accuracy (%)	Average computational time (s)
Wu et al. (2020b)	Planetary gearbox	FFT	6	3	PU (Lessmeier et al. (2016) and private	Holdout	PU: 98.62; private: 89.50	–
Wang et al. (2020)	Rolling bearings	FFT	5	2	CWRU and private	Multiple holdout	CWRU: 96.10±0.08; private: 96.18±0.037	–

10cv tenfold cross validation, *CWRU* Case Western Reserve University, *CWT* continuous wavelet transform, *FFT* fast Fourier transform, *LOCV* leave-one-out cross validation, *PU* Paderbon University

rig and a centrifugal pump test rig. A rolling-element bearing fault diagnosis technique based on CNN's deep structure is carried out in (Hoang and Kang 2019). When using vibration signals directly as input data, the proposed technique has high accuracy and robustness in noisy environments. Another CNN variant named deep capsule network with a stochastic delta rule (DCN-SDR) for rolling bearing fault diagnosis is proposed in (Chen et al. 2019b). DCN-SDR uses a raw temporal signal as its input and achieves very high accuracy under a noisy environment. The authors in (Khodja et al. 2020) use CNN and vibration spectrum imaging for classifying bearing faults. The proposed classifier is evaluated in terms of classification accuracy, variable load and speed testing, generalization, and robustness by adding noise to the collected data. Wu et al. (2021) consider a semi-supervised CNN for the intelligent diagnosis of bearings. A 1-D CNN is applied to learn spatial features and generates class probabilities of unlabeled samples. A pre-fully connected deep CNN approach is presented in (Zhu et al. 2020) to obtain comprehensive bearing vibration characteristics. The front fully connected layer, designed to reduce the signal's complexity, extracts the signal's global characteristics.

Yu et al. (2020) introduce one-dimensional CNN as a feature extractor, while the fully connected layer is adopted as the main element of two-symmetric classifiers. Moreover, the cosine similarity is used to guide the adversarial training process of the two classifiers. Yang et al. (2020) propose a diagnosis model combined with polynomial kernel induced maximum mean discrepancy (PK-MMD) to reuse diagnosis knowledge from one machine to the other. The proposed method uses the ResNet model to extract features from the samples in the source and target domains. Wu et al. (2020b) introduce two transfer situations for rotating machinery intelligent diagnosis and construct seven few-shot transfer learning methods based on a unified 1-D CNN. A novel deep transfer learning based on modified ResNet-50 is proposed by (Wang et al. 2020). The ResNet-50 model extracts low-level features, constructs a multiple scale feature learner, and obtains high-level features as the classifier's input. Finally, to tackle a small dataset, Liang et al. (2021) utilize parallel CNN for bearing fault identification. Two different CNNs are constructed parallel to extract features in the time and frequency domains. The two features are then fused in the merged layer as the inputs of the final classifier.

4.3.2 Deep belief network (DBN)

Yu et al. (2020) introduce one-dimensional CNN as a feature extractor, while the fully connected layer is adopted as the main element of two-symmetric classifiers. Moreover, the cosine similarity is used to guide the adversarial training process of the two classifiers. Yang et al. (2020) propose a diagnosis model combined with polynomial kernel induced maximum mean discrepancy (PK-MMD) to reuse diagnosis knowledge from one machine to the other. The proposed method uses the ResNet model to extract the features from the samples in the source and target domains. Wu et al. (2020b) introduce two transfer situations for rotating machinery intelligent diagnosis and construct seven few-shot transfer learning methods based on a unified 1-D CNN. A novel deep transfer learning based on modified ResNet-50 is proposed by (Wang et al. 2020). The ResNet-50 model is used to extract low-level features, construct a multiple scale feature learner, and obtain high-level features as the classifier's input. Lastly, to tackle a small dataset, Liang et al. (2021) utilize parallel CNN for bearing fault identification. Two different CNNs are constructed parallel to extract features in the time and frequency domains. The two kinds of features are then fused in the merged layer as the inputs of the final classifier.

4.3.3 Recurrent neural network (RNN)

A study presented by (Liu et al. 2018) discusses a fault-bearing diagnosis method based on RNN in the form of an autoencoder. In this technique, the rolling bearings' vibration data for the next period are predicted based on the previous period using a gated recurrent unit (GRU)-based denoising autoencoder. Assessing the performance degradation of bearings using LSTM is proposed in (Zhang et al. 2019). The proposed model can effectively identify degradation states and accurately predict the bearings' remaining useful life. A DL model called many-to-many-to-one bidirectional long short-term memory (MMO-BLSTM) is proposed to automatically extract the rotating speed from vibration signals. The proposed model's performance is validated on an internal combustion engine dataset, a rotor system dataset, and a fixed-shaft gearbox dataset (Rao et al. 2020). Xu et al. (2020b) propose a deep residual network (DRN) and LSTM to process extracted features obtained from multisensory system condition data. The proposed method shows robust and superior performance for the two datasets, including cutting tool monitoring and bearing fault diagnosis.

4.3.4 Stacked autoencoders

Jia et al. (2016) use two main procedures to train DNN such as pre-train DNNs with autoencoder and fine-tune DNNs with a backpropagation algorithm for classification. Similarly, Sun et al. (2016) utilize a sparse autoencoder to improve the robustness of feature representation. Features learned from SAE are further used to train a neural network classifier. An intelligent severity identification of rolling bearing faults can be obtained using a stack of multiple sparse autoencoders with a classifier layer (Chen et al. 2017). A feature learning ability can be improved by combining denoising and contractive autoencoders (Shao et al. 2017a). To further increase the quality of the learnt features, a locality preserving projection is used to fuse the deep features. Likewise, a stacked sparse autoencoder (SAE) is employed for machine fault diagnosis (Qi et al. 2017). Preprocessing is performed on the collected non-stationary and transient signals using ensemble empirical model decomposition and autoregressive models. The study (Mao et al. 2017) proposes an autoencoder-ELM-based approach for identifying bearing defects. Research published in (Lu et al. 2017b) analyzes an effective and reliable approach for identifying the health condition of a machine using a stacked denoising autoencoder. The suggested model is shown to be a promising method for identifying specific health states from data incorporating ambient noise and variations in working conditions (Tables 5 and 6).

Furthermore, a novel deep autoencoder feature learning is proposed for diagnosing rotating machinery fault (Shao et al. 2017b). The proposed method is built based on maximum correntropy to design the new loss function for the improvement of feature learning from the measured vibration signals. A local connection network (LCN) developed using a normalized sparse autoencoder (NSAE), called LCN-NSAE, is suggested for intelligent fault diagnosis (Jia et al. 2018). The proposed technique incorporates feature extraction and fault detection into a general-purpose learning procedure to improve the fault diagnosis system. An effective DNN with an unsupervised feature learning algorithm based on a sparse autoencoder is presented in (Ahmed et al. 2018). The study explores the effects of a sparse autoencoder on the classification performance of highly compressed measurements of bearing vibration signals.

Table 5 Summary of stacked autoencoder-based fault diagnosis methods using vibration signals

Ref.	Element(s)	Number of autoencod- ers	Classifier	Dataset	Validation method(s)	Testing accuracy (%)	Average computational time (s)
Jia et al. (2016)	Rolling bearing, planetary gearbox	–	DNN	CWRU	Multiple holdout	100	–
Sun et al. (2016)	Induction motor	600	DNN	Private	Holdout	97.61	–
Chen et al. (2017)	Rolling bearing	4	Neural network	CWRU	Holdout	97.82	66.78
Shao et al. (2017a)	Rotating machinery	3	Softmax	CWRU and private	Holdout	CWRU: 95.23; private: 95.19	CWRU: –; private: 113.27
Qi et al. (2017)	Rolling bearing, gearbox	3	Softmax	CWRU and private	Holdout	CWRU: 99.85; private: 99.83	–
Mao et al. (2017)	Rolling bearing	–	Extreme learning machine	CWRU	Holdout	100	0.62
Lu et al. (2017b)	Rolling bearing	3	Softmax	CWRU	Holdout	99.83	–
Shao et al. (2017b)	Gearbox	3	Softmax	Private	Multiple holdout	94.05±1.34	37.67
Jia et al. (2018)	Gearbox	–	Softmax	Private	Multiple holdout	99.43±0.105	–
Ahmed et al. (2018)	Rolling bearing	2	Softmax	CWRU and Private	Multiple holdout	CWRU: 100±0.1; private: 99.8±0.1	–
Shao et al. (2018)	Rolling bearing	15	Softmax	CWRU	Multiple holdout	97.18±0.114	589.17
Li et al. (2018)	Rolling bearing	–	Softmax	CWRU	Holdout	98.53	–
Sun et al. (2018)	Rolling bearing	2	DNN	CWRU	Multiple holdout	97.47±0.43	172.5
He and He (2019)	Rolling bearing	–	Softmax	Private	Holdout	99.98	85.23
Li et al. (2020b)	Gearbox	–	Softmax	Private	Holdout	99.14	–
Wu et al. (2020a)	Rotating machinery	–	Linear discriminator	Private	Holdout	90.61	–
Shao et al. (2020)	Rolling bearing	3	Softmax	CWRU and private	Holdout	CWRU: 91.40; private: 89.90	–

CWRU Case Western Reserve University

Table 6 Summary of GNNs-based fault diagnosis methods using vibration signals

Ref.	Element(s)	Number of GNN layer(s)	Number of classifier(s)	Dataset	Validation method(s)	Training samples (%)	Testing accuracy (%)	Average computational time (s)
Li et al. (2020a)	Rolling bearing and Gearbox	3	Softmax	CWRU and SEU	10cv	80	CWRU: 96.2±2.6; SEU: 95±2.2	–
Zhang et al. (2020)	Rolling bearing	4	Softmax	Private	Holdout	80	89.45	–
Chen et al. (2021)	dSPACE Fault Simulation	2	Softmax	Private	8cv	27	97	–
Li et al. (2021a)	Gearbox and Aeroengine Bevel Gear	4	Linear discriminator	Private	Holdout	80	Gearbox:87.2; Aeroengine Bevel Gear:79.81	–
Li et al. (2021b)	Gearbox	2	Linear discriminator	Private	Holdout	80	98.31	508
Liu et al. (2021)	Rolling Bearing and Gearbox	3	DNN	Private	10cv	30	Rolling Bearing:98.27; Gearbox:99.08	Rolling Bearing:42; Gearbox: 41
Yu et al. (2021)	Gearbox	4	Softmax	Private	Holdout	DDS test rig: 71.43; Wind turbine gearbox: 66.67	DDS test rig:98.70; Wind turbine gearbox:93.09	DDS test rig:51.25; Wind turbine gearbox:48.76
Zhang et al. (2021b)	Rolling Bearing	1	Multi layer perceptron	CWRU and SPCP	10cv	83.33	CWRU:99.96; SPCP: 99.99	CWRU:0.99; SPCP: –
Zhao et al. (2021a)	Rolling bearing	2	Fuzzy clustering	Private	10cv	DatasetA:50; DatasetB:40	DatasetA:99.15; DatasetB: 97.23	DatasetA: 128.18; DatasetB: 128.23
Zhao et al. (2021b)	Motor bearing	2	Multi layer perceptron	Private	5cv	40	97.9	297.68
Zhou et al. (2022)	Rotating machinery	3	Softmax	Private	Holdout	40	99.62	–

Table 6 (continued)

Ref.	Element(s)	Number of GNN layer(s)	Number of classifier(s)	Dataset	Validation method(s)	Training samples (%)	Testing accuracy (%)	Average computational time (s)
Yang et al. (2022)	Rotating machinery	2	Softmax	CWRU	10cv	30	99.72	–
10cv tenfold cross validation, 8cv eighth-fold cross validation, 5cv five-fold cross validation, CWRU Case Western Reserve University, SEU Southeast University, SPCP self-priming centrifugal pump, DDS drivetrain diagnostics simulator								

An automatic and accurate identification of rolling bearing faults using an ensemble of deep autoencoders (EDAEs) is studied in (Shao et al. 2018). The proposed method is applied to analyze experimental bearing vibration signals. A new bearing fault diagnosis method based on a fully-connected winner-take-all autoencoder has been developed (Li et al. 2018). A soft voting method is applied to combine the prediction results of signal segments sliced by the slicing window to improve accuracy and stability. A DNN based on stacked sparse autoencoders (SAEs) for bearing fault diagnosis is built (Sun et al. 2018). The proposed method is compared with traditional methods such as a single-hidden-layer backward propagation neural network (BPNN) and SVM. A work in (He and He 2019) utilizes a discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) autoencoder to calculate the spectral average of the vibration signals. The reconstructed vibration signals are then fed into the softmax classifier for bearing fault diagnosis. Li et al. (2020b) introduce a unique augmented deep sparse autoencoder that diagnoses gear pitting faults with few raw vibration data. By validating six types of gear pitting conditions, the proposed method is very accurate and robust. Wu et al. (2020a) propose a fault-attention probabilistic adversarial autoencoder that obtains a low-dimensional manifold embedded in the vibration signal's high-dimensional space. Lastly, Shao et al. (2020) combine particle swarm optimization and a stacked autoencoder for intelligent fault diagnosis techniques on various rotating machines.

4.3.5 Graph neural network (GNN)

Researchers have been paying close attention to GNNs, an emerging specialized graph signal processing algorithm (Zhou et al. 2020). By aggregating information from the node's neighbors at any depth, GNN can more effectively extract and inference data relationships. Using a horizontal visibility graph (HVG) and a GNN, Li et al. (2020a) suggested a new model for bearing faults diagnosis. The HVG algorithm turns a time series sample into a graph with condition-specific topology. Compared to pure numerical information, the proposed method gives additional valuable information for classification. Li et al. (2020a) prove that the GNN model outperforms the RNN model for bearing faults diagnosis.

The graph convolutional network (GCN) is a variety of neural networks that establishes associations between data using an association graph to speed up training and improve model performance (Bruna et al. 2013). Zhou et al. (2022), propose a GCN-based rotating machinery fault diagnosis method that uses multi-sensor data. To diagnose problems in wind turbine gearboxes, Yu et al. (2021) suggested a fast deep graph convolutional network (FDGCN). The proposed FDGCN efficiently and adaptively learns the discriminative fault features from the initial graph input and then uses these learnt features to categorize the associated fault type. Zhang et al. (2020) applied a deep GCN (DGCN) to diagnose acoustic-based faults for roller bearings. Overall, the DGCN method has outperformed existing approaches in terms of classification accuracy, and it can be used to detect various types and degrees of faults in roller bearings.

To handle the unbalanced dataset problem for fault diagnosis, Liu et al. (2021) use an autoencoder-based SuperGraph feature learning method. However, the constructed SuperGraph by Yang et al. (2022) has redundant edges; all labeled signals with the same fault type are interconnected, resulting in excessive computational costs. Li et al. (2021b), on the other hand, used a multireceptive field graph convolutional network (MRF-GCN) to solve the constraints of GCNs for effective intelligent fault diagnosis. Case studies demonstrate that MRF-GCN outperforms other algorithms even when the dataset is unbalanced.

Furthermore, Chen et al. (2021) pre-diagnose faults by using the structural analysis (SA) method and construct it into an association graph. Then, to adjust the influence of measurements and prior knowledge, a weight coefficient is introduced. The suggested method produces superior diagnosis findings than existing methods based on common evaluation indicators. Zhang et al. (2021b) propose using gated recurrent units (GRUs) in their novel FDGRU method for fault diagnosis. FDGRU stabilizes the training process and improves the accuracy of fault diagnosis.

Additionally, in the unsupervised fault diagnosis domain, (Li et al. 2021a) propose using a domain adversarial graph convolutional network (DAGCN) to solve the data discrepancy between source and target domain in unsupervised domain adaptation (UDA) for machinery fault diagnosis. According to experimental data, the suggested DAGCN extracts the transferable domain-invariant and discriminative features for domain adaptation and achieves excellent performance among the six comparison approaches. Zhao et al. (2021a) propose a multiple-order graphical deep ELM (MGDELM) algorithm to synchronously extract local and global structural information from raw data. To handle data limitations, Zhao et al. (2021b) propose a new semisupervised deep convolutional belief network (SSD-CDBN) for motor-bearing fault diagnosis. This method used both labeled and unlabeled data information.

4.4 Datasets

This section explains various publicly available datasets employed to diagnose faults using vibration signals. More specifically, we include datasets well-known in the literature: the Case Western Reserve University (CWRU) bearing dataset, the University of Paderborn (UPB) bearing dataset, the Southeast University (SEU) gearbox dataset, and several other datasets. The following sections provide additional information about the datasets.

(a) Case Western Reserve University (CWRU) Dataset (Loparo 2012)

Widely used in fault diagnosis studies, this benchmark dataset can be obtained via CWRU's public-bearing data center. The test rig is comprised of a two-horsepower induction motor, a torque transducer or encoder, and a dynamometer. The accelerometer near the motor-driven end acquires the vibration signal at a sampling frequency of 12 kHz. Electric discharge machining (EDM) introduces single-point flaws into test bearings, resulting in three severity levels of damage with diameters of 0.007, 0.014, and 0.021 in, respectively. Bearing faults are classified into three types based on their location: inner-race fault, outer-race fault, and ball fault. In addition, a normal condition bearing is evaluated. As a result, these datasets represent ten bearing defect patterns under four different motor loads of 0, 1, 2, and 3 horsepower.

(b) Paderborn University Dataset (Lessmeier et al. 2016)

Paderborn University (PU) provides this experimental bearing dataset for condition monitoring and diagnostics based on motor current and vibration signals. Because of the vast amount of data and variety of bearing testing types, the dataset includes a large mining space test rig for bearing condition monitoring as well as an accelerated lifetime test rig. The bearing condition monitoring test rig is a modular system that includes a load motor, a flywheel, a bearing test module, a torque measurement shaft, and a driving motor. To collect the essential data, it is necessary to reproduce relevant bearing faults. The vibration signal is one measured variable from the housing vibrations in the form of acceleration at the bearing housing in the bearing test stand.

(c) University of Connecticut Dataset (Cao et al. 2018)

A two-stage gearbox with changeable gears was used to capture vibration signals. The gear speed is controlled by a motor. The torque is provided by a magnetic brake, the input voltage of which can be modified. A 32-tooth pinion and an 80-tooth gear are coupled to the first stage input shaft. The second stage is made up of a 48-tooth pinion and a 64-tooth gear. The input shaft speed is determined by a tachometer, and an accelerometer determines the gear vibration signals. A dSPACE system is used to sample the signals at a rate of 20 kHz. There are nine different gear conditions listed on the input shaft, including healthy condition, missing tooth, root fracture, spalling, and chipped tip. Using a non-preprocessing technique, the vibration signals were converted into images in order to reveal the 2D features of the raw data. The final image representation of the signal is a grayscale image 227×227 in size. Using the experimental gearbox, 104 signals were collected for each gear state, with 3600 angle-even samples taken for each signal.

(d) XJTU-SY Bearing Dataset (Wang et al. 2018)

An alternating current (AC) induction motor, a motor speed controller, a support shaft, two support bearings (heavy-duty roller bearings), and a hydraulic loading system are all part of the bearing testbed. This testbed was designed to speed up the deterioration of rolling element bearings under a variety of operating situations. There are three different operating circumstances, and each working state was tested with five bearings. Two accelerometers, one on the horizontal axis and one on the vertical axis, are mounted at 90° on the housing of the tested bearings to gather vibration signals. The sampling frequency is set at 25.6 kHz. Each sample collects 32,768 data points (i.e., 1.28 s), and the sampling time is 1 min.

(e) ABLT-1A Bearing Dataset 6308 (Ding et al. 2022)

At zero load conditions, five different health states are simulated: normal, inner ring fault, outer ring fault, inner and outer ring compound fault, and inner and outer ring compound weak fault. As a result, each health data type was gathered at a rate of 17.5 Hz, and a sampling frequency of 12,800 Hz was intercepted using sample group interception. Two thousand sample groups are intercepted by 1024 signal points treated as a group of samples, yielding a total of $2000 \times 5 = 10,000$ sample points.

(f) ABLT-1A Bearing Dataset 6205 (Ding et al. 2022)

At zero load conditions, seven different health states are simulated: normal, outer ring fault, inner and outer ring compound fault, inner ring fault, inner and outer ring weak compound fault, and inner and outer ring weak fault. As a result, each health data type was gathered at a rate of 17.5 Hz, and a sampling frequency of 12,800 Hz was intercepted using sample group interception. Two thousand sample groups are intercepted by 1024 signal points treated as a group of samples, yielding a total of $2000 \times 7 = 14,000$ sample points.

5 Future research directions

This current study has reviewed the application of DL algorithms for fault diagnosis methods using vibration signals. Even though present-day developments in DL have been eye-opening, several challenges and limitations remain, some of which are mentioned as follows. Noise signals containing working environment fluctuations might be obtained from real-world industrial applications. Hence, it is still challenging to determine whether the

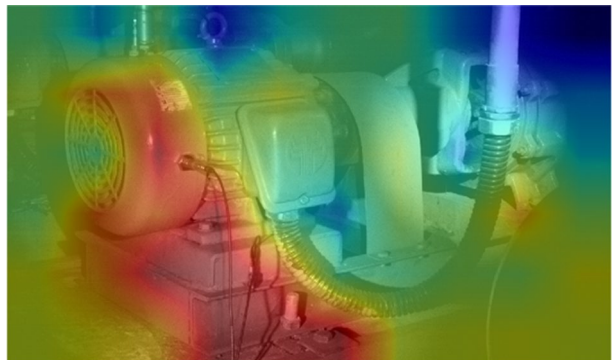
current DL algorithms can perform well under such noisy data. Like DL techniques applied in other application domains, an imbalanced dataset might be crucial. The vast majority of the machine health data obtained would be under normal or healthy circumstances, which directly affects the training process of fault diagnosis models.

Additionally, the interpretability of DL algorithms for fault diagnosis utilizing vibration signals is still underexplored. The internal mechanism of DL algorithms has not been divulged due to a lack of interpretability and explainability. To address this issue, some solutions have been developed, such as feature visualization (Matthew and Fergus 2014) and coarse localization maps (e.g., class activation maps (CAMs)) (Selvaraju et al. 2017; Zhou et al. 2016). The former focuses on visualizing feature maps in which the first layers define local patterns such as edges, circles, and so on, while the subsequent layers blend them into more meaningful structures. Similarly, the latter generates CAMs that identify the important portions of a picture used in prediction. By using CAMs, it will be feasible to gain a better grasp of the image's problematic regions. It is also worth noting that the interpretability of CAMs will aid us in selecting the critical component for detecting the most vibrating regime without the need for vibration sensors. An example of CAMs utilized for defect detection in a laboratory-scale water pump is shown in Fig. 17.

Other research utilizes SincNet to address interpretability (Abid et al. 2020) from a frequency domain perspective. SincNet uses parameterized Sinc functions to assist the first layer in identifying more meaningful filters (Ravanelli and Bengio 2018). Temporal representations of the Sinc filter kernels are expressed in both frequency and time domains with distinct low and high cut-off frequencies. This approach makes the network focus on high-level parameters, which must have a clear physical meaning. Based on layerwise relevance propagation (LRP) (Bach et al. 2015), Grezmaek et al. (2019) introduce an interpretable CNN for gearbox fault diagnosis. Utilizing LRP, the network can explain the reasoning behind the results by indicating which time-frequency points in the spectra images contribute most to the fault (Grezmaek et al. 2019). Li et al. (2022) propose WaveletKernelNet (WKN) and introduce an interpretable convolution kernel that extracts significant fault information components in vibration signals. A continuous wavelet convolution (CWConv) layer is used to replace the standard CNN's initial convolution layer in order to uncover kernels with a specific physical meaning (Li et al. 2022).

Furthermore, while the DL algorithm for fault diagnosis is a solely data-driven algorithm that may well suit observations, predictions may be inconsistent or unreasonable due to extrapolation or observational biases, resulting in poor generalization

Fig. 17 An example of CAM used for identifying the most vibrating regime of a laboratory-scaled water pump (Sun et al. 2020). The red area indicates the most vibrating regime



performance. As a result, there is an urgent need to incorporate fundamental physical principles and domain knowledge into the DL algorithm in order to govern physical norms, which could provide informative priors or substantial theoretical limitations and inductive biases on top of observational biases. Therefore, physics-informed learning, an act of leveraging prior knowledge derived from the observational, empirical, physical, or mathematical understanding of the environment to improve a learning algorithm's performance, is required (Karniadakis et al. 2021). This recent development is known as physics-informed neural networks (PINNs) (Raissi et al. 2019). DL based on physics is still in its early research stages and requires careful configuration for each challenge. One key goal is to develop PINNs for uncertainty quantification to create a robust and reliable forecast. The basic idea is to employ neural networks as universal approximators of the intended solution, then constrain the training process by defining the loss function based on domain knowledge (Raissi et al. 2019). By using automatic differentiation to integrate partial differential equations (PDEs) into the loss function of a neural network, physics-informed neural networks (PINNs) incorporate information from both measurements and PDEs.

Finally, another consideration for further direction is the use of the self-attention mechanism. Transformer (Vaswani et al. 2017), the most recent development for this solution, has outperformed previous techniques, such as CNNs and RNNs, in various tasks, including natural language processing (NLP), point cloud, and picture analysis (Pei et al. 2021). Although the transformer is a revolutionary design for DL model architectures, it has been underexplored in the realm of fault diagnosis until Pei et al. (2021) introduced the transformer convolution network (TCN) for rotating machinery fault diagnosis. The TCN uses a multi-layered attention stacking strategy for remarkable parallel and learning abilities. In addition, the TCN demonstrates remarkable stability and resilience against noise due to the stability of the automatic feature extraction by the transformer with the attention mechanism. However, some modest modifications are required to make the TCN completely appropriate for fault diagnosis utilizing vibration signals. Moreover, there are several limitations that must be addressed in order to improve TCN performance. Including scenarios with extremely limited fault samples, its effectiveness in real-world industrial applications has not been thoroughly verified. The transformer encoder's computational complexity is another unresolved research gap that requires further investigation (Pei et al. 2021).

Most of these studies concentrate on automatic fault diagnosis of the bearing fault, the most prevalent issue in industrial and manufacturing systems. However, machine failures reveal a reaction chain between cause and defect. The rotating machinery works involve rotational motion such as gears, rotors and shafts, rolling element bearings, flexible couplings, and electrical machines (Gu et al. 2018; Guan 2017; Guo et al. 2021; Gupta and Pradhan 2017; Peng et al. 2022; Wang et al. 2019b; Zhang et al. 2021a; Zhu et al. 2021). Due to the complex structure and interaction of multiple components in rotating machinery, there are additional faults that will need to be diagnosed in the future in order to improve results:

- (i) Coupling faults.
- (ii) Pitting of race and ball/roller (i.e., rolling element faults).
- (iii) Ball screw faults.
- (iv) Rotors and shaft faults (e.g., misalignment, unbalance, and loose components).
- (v) Electrical machine faults.

6 Threat to validity

Some of the factors considered as potential threats to this study are summarily discussed here. Some search keywords were drawn from keywords found in some papers to reduce the risk of excluding relevant studies. The search was restricted to two primary indexing databases to minimize bias and miss relevant papers. Other researchers' biases can be unavoidable, such as inaccurate data extraction and mapping procedure. A consensus mechanism was utilized to select all the extracted studies to minimize this issue. The final chosen studies and classification were made by consensus among the four authors.

7 Conclusion

DL techniques have solved complex problems across various applications and research concerning the fault diagnosis of rotating machinery. This paper has documented an overview of DL algorithms for condition monitoring using vibration signals. Through this study, existing studies were reviewed and classified according to the DL architectures. Moreover, some current research challenges were reported, providing researchers insights into the latest DL techniques for fault diagnosis. More importantly, a 'white-box' model for fault diagnosis will be sought after as it helps the prediction be understandable to humans. To conclude, this section sums up the answers to the RQs mentioned earlier.

- (a) RQ₁: What are the distribution per year and publication venues of prior works related to the applications of DL for vibration-based fault diagnosis techniques? This study has observed DL for fault diagnosis using vibration signals either in supervised, unsupervised, or semi-supervised learning mechanisms. This study reveals that there has been a considerable increase between 2017 and 2021.
- (b) RQ₂: What types of DL techniques are developed the most? Supervised learning task based on CNN was still the major player in this domain, which shared more than one-third of the total studies. Other significant DL techniques that have been developed include stack autoencoder and graph-based neural networks.
- (c) RQ₃: What are the future research directions concerning the challenges and limitations of the current studies that remain? DL models are black-box; therefore, the model's interpretability and explainability would be very promising tools in the future. In addition, physics-informed ML and the self-attention mechanism remain underexplored in the current literature.

Acknowledgements This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIT) (No. 2020R1A2C1009744), in part by the Institute of Civil Military Technology Cooperation funded by the Defense Acquisition Program Administration and Ministry of Trade, Industry and Energy of the Korean government under Grant No. 19-CM-GU-01, and in part by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) Grant funded by the Korean Government [Ministry of Trade, Industry, and Energy (MOTIE)] under Grant 20206610100290. This work was partly supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (No. 2021R1F1A1046416). This work was supported by the AI Collaboration Project Fund (1.220083) of UNIST (Ulsan National Institute of Science and Technology).

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abid FB, Sallem M, Braham A (2020) Robust interpretable deep learning for intelligent fault diagnosis of induction motors. *IEEE Trans Instrum Meas* 69(6):3506–3515. <https://doi.org/10.1109/TIM.2019.2932162>
- Ahmed H, Nandi AK (2020) Condition monitoring with vibration signals: compressive sampling and learning algorithms for rotating machines. Wiley, Hoboken
- Ahmed HOA, Wong MLD, Nandi AK (2018) Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features. *Mech Syst Signal Process* 99:459–477
- Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W (2015) On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* 10(7):e0130140
- Brandt A (2011) Noise and vibration analysis: signal analysis and experimental procedures. Wiley, Hoboken
- Bruna J, Zaremba W, Szlam A, LeCun Y (2013) Spectral networks and locally connected networks on graphs. arXiv preprint [arXiv:1312.6203](https://arxiv.org/abs/1312.6203)
- Cao P, Zhang S, Tang J (2018) Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access* 6:26241–26253
- Cardona-Morales O, Avendaño L, Castellanos-Dominguez G (2014) Nonlinear model for condition monitoring of non-stationary vibration signals in ship driveline application. *Mech Syst Signal Process* 44(1–2):134–148
- Chen R, Chen S, He M, He D, Tang B (2017) Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion. *Proc Inst Mech Eng O* 231(6):666–679. <https://doi.org/10.1177/1748006X17726452>
- Chen H, Hu N, Cheng Z, Zhang L, Zhang Y (2019a) A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. *Measurement* 146:268–278. <https://doi.org/10.1016/j.measurement.2019.04.093>
- Chen T, Wang Z, Yang X, Jiang K (2019b) A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals. *Measurement* 148:106857. <https://doi.org/10.1016/j.measurement.2019.106857>
- Chen K, Hu J, Zhang Y, Yu Z, He J (2020) Fault location in power distribution systems via deep graph convolutional networks. *IEEE J Sel Areas Commun* 38(1):119–131
- Chen Z, Xu J, Peng T, Yang C (2021) Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge. *IEEE Trans Cybern*. <https://doi.org/10.1109/TCYB.2021.3059002>
- Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint [arXiv:1406.1078](https://arxiv.org/abs/1406.1078)
- Ding X, He Q (2017) Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Trans Instrum Meas* 66(8):1926–1935
- Ding Y, Jia M, Miao Q, Cao Y (2022) A novel time-frequency transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings. *Mech Syst Signal Process* 168:108616
- Dziedziech K, Jablonski A, Dworakowski Z (2018) A novel method for speed recovery from vibration signal under highly non-stationary conditions. *Measurement* 128:13–22
- Goodfellow I (2016) Nips 2016 tutorial: generative adversarial networks. arXiv preprint [arXiv:1701.00160](https://arxiv.org/abs/1701.00160)

- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: *Advances in neural information processing systems*, pp 2672–2680
- Govekar E, Gradišek J, Grabec I (2000) Analysis of acoustic emission signals and monitoring of machining processes. *Ultrasonics* 38(1–8):598–603
- Grezmak J, Wang P, Sun C, Gao RX (2019) Explainable convolutional neural network for gearbox fault diagnosis. *Procedia CIRP*. 80:476–481. <https://doi.org/10.1016/j.procir.2018.12.008>. (26th CIRP Conference on Life Cycle Engineering (LCE) Purdue University, West Lafayette, IN, USA May 7–9, 201)
- Gu FC, Bian JY, Hsu CL, Chen HC, Lu SD (2018) Rotor fault identification of induction motor based on discrete fractional fourier transform. In: *2018 international symposium on computer, consumer and control (IS3C)*, pp 205–208. <https://doi.org/10.1109/IS3C.2018.00059>
- Guan Z (2017) Vibration analysis of shaft misalignment and diagnosis method of structure faults for rotating machinery. *Int J Perform Eng* 13:337
- Guo X, Chen L, Shen C (2016) Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* 93:490–502. <https://doi.org/10.1016/j.measurement.2016.07.054>
- Guo S, Yang T, Hua H, Cao J (2021) Coupling fault diagnosis of wind turbine gearbox based on multitask parallel convolutional neural networks with overall information. *Renew Energy* 178:639–650. <https://doi.org/10.1016/j.renene.2021.06.088>
- Gupta P, Pradhan M (2017) Fault detection analysis in rolling element bearing: a review. *Mater Today: Proc* 4(2, Part A):2085–2094. <https://doi.org/10.1016/j.matpr.2017.02.054>
- He M, He D (2019) A new hybrid deep signal processing approach for bearing fault diagnosis using vibration signals. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2018.12.088>
- Hinton GE (2009) Deep belief networks. *Scholarpedia* 4(5):5947
- Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
- Hoang DT, Kang HJ (2019) Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognit Syst Res* 53:42–50. <https://doi.org/10.1016/j.cogsys.2018.03.002>
- Iatsenko D, McClintock PV, Stefanovska A (2016) Extraction of instantaneous frequencies from ridges in time-frequency representations of signals. *Signal Process* 125:290–303
- Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccuifer M, Verstockt S, Van de Walle R, Van Hoecke S (2016) Convolutional neural network based fault detection for rotating machinery. *J Sound Vib* 377:331–345. <https://doi.org/10.1016/j.jsv.2016.05.027>
- Jia F, Lei Y, Lin J, Zhou X, Lu N (2016) Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech Syst Signal Process* 72:303–315
- Jia F, Lei Y, Guo L, Lin J, Xing S (2018) A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing* 272:619–628. <https://doi.org/10.1016/j.neucom.2017.07.032>
- Jing L, Zhao M, Li P, Xu X (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* 111:1–10. <https://doi.org/10.1016/j.measurement.2017.07.017>
- Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L (2021) Physics-informed machine learning. *Nat Rev Phys* 3(6):422–440
- Kateris D, Moshou D, Pantazi XE, Gravalos I, Sawalhi N, Loutridis S (2014) A machine learning approach for the condition monitoring of rotating machinery. *J Mech Sci Technol* 28(1):61–71
- Khan NA, Jönsson P, Sandsten M (2017) Performance comparison of time-frequency distributions for estimation of instantaneous frequency of heart rate variability signals. *Appl Sci* 7(3):221
- Khodja AY, Guersi N, Saadi MN, Boutasseta N (2020) Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. *Int J Adv Manuf Technol* 106(5–6):1737–1751. <https://doi.org/10.1007/s00170-019-04726-7>
- Kingma DP, Welling M (2013) Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*
- Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: *5th international conference on learning representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings, OpenReview.net*
- Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ (2021) 1d convolutional neural networks and applications: a survey. *Mech Syst Signal Process* 151:107398. <https://doi.org/10.1016/j.ymssp.2020.107398>
- Kitchenham BA, Budgen D, Brereton P (2015) *Evidence-based software engineering and systematic reviews*, vol 4. CRC Press, Boca Raton

- LeCun Y, Bengio Y et al (1995) Convolutional networks for images, speech, and time series. *Handb Brain Theory Neural Netw* 3361(10):1995
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
- Lei Y (2016) Intelligent fault diagnosis and remaining useful life prediction of rotating machinery. Butterworth-Heinemann, Oxford
- Lessmeier C, Kimotho JK, Zimmer D, Sextro W (2016) Condition monitoring of bearing damage in electro-mechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification. In: PHM Society European conference, vol 3
- Li C, Zhang W, Peng G, Liu S (2018) Bearing fault diagnosis using fully-connected winner-take-all autoencoder. *IEEE Access* 6:6103–6115
- Li C, Mo L, Yan R (2020a) Rolling bearing fault diagnosis based on horizontal visibility graph and graph neural networks. In: International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), 2020, pp 275–279. <https://doi.org/10.1109/ICSMD50554.2020.9261687>
- Li X, Li J, Qu Y, He D (2020b) Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning. *Chin J Aeronaut* 33(2):418–426. <https://doi.org/10.1016/j.cja.2019.04.018>
- Li T, Zhao Z, Sun C, Yan R, Chen X (2021a) Domain adversarial graph convolutional network for fault diagnosis under variable working conditions. *IEEE Trans Instrum Meas* 70:1–10
- Li T, Zhao Z, Sun C, Yan R, Chen X (2021b) Multireceptive field graph convolutional networks for machine fault diagnosis. *IEEE Trans Ind Electron* 68(12):12739–12749
- Li T, Zhao Z, Sun C, Cheng L, Chen X, Yan R, Gao RX (2022) Waveletkernelnet: an interpretable deep neural network for industrial intelligent diagnosis. *IEEE Trans Syst Man Cybern: Syst* 52(4):2302–2312. <https://doi.org/10.1109/TSMC.2020.3048950>
- Liang MX, Cao P, Tang J (2021) Rolling bearing fault diagnosis based on feature fusion with parallel convolutional neural network. *Int J Adv Manuf Technol* 112(3):819–831
- Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis J et al (2009) The prisma statement for reporting systematic and meta-analyses of studies that evaluate interventions: explanation and elaboration. *PLoS Med* 6(7):1–28
- Liu R, Meng G, Yang B, Sun C, Chen X (2017) Dislocated time series convolutional neural architecture: an intelligent fault diagnosis approach for electric machine. *IEEE Trans Ind Inform* 13(3):1310–1320
- Liu H, Zhou J, Zheng Y, Jiang W, Zhang Y (2018) Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans* 77:167–178. <https://doi.org/10.1016/j.isatra.2018.04.005>
- Liu J, Zhou K, Yang C, Lu G (2021) Imbalanced fault diagnosis of rotating machinery using autoencoder-based SuperGraph feature learning. *Front Mech Eng* 16:829–839
- Loparo K (2012) Case western reserve university bearing data center. Bearings vibration data sets, Case Western Reserve University pp 22–28
- Lu C, Wang Z, Zhou B (2017a) Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Adv Eng Inform* 32:139–151. <https://doi.org/10.1016/j.aei.2017.02.005>
- Lu C, Wang ZY, Qin WL, Ma J (2017b) Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process* 130:377–388. <https://doi.org/10.1016/j.sigpro.2016.07.028>
- Mao W, He J, Li Y, Yan Y (2017) Bearing fault diagnosis with auto-encoder extreme learning machine: a comparative study. *Proc Inst Mech Eng C* 231(8):1560–1578. <https://doi.org/10.1177/0954406216675896>
- Matthew D, Fergus R (2014) Visualizing and understanding convolutional neural networks. In: Proceedings of the 13th European Conference Computer Vision and Pattern Recognition, Zurich, Switzerland, pp 6–12
- Mohanty AR (2014) Machinery condition monitoring: principles and practices. CRC Press, Boca Raton
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P et al (2009) Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *PLoS Med* 6(7):e1000097
- Nord JH, Nord GD (1995) Mis research: journal status assessment and analysis. *Inf Manage* 29(1):29–42
- Pan J, Zi Y, Chen J, Zhou Z, Wang B (2018) LiftingNet: a novel deep learning network with layer-wise feature learning from noisy mechanical data for fault classification. *IEEE Trans Ind Electron* 65(6):4973–4982
- Peeters C, Leclère Q, Antoni J, Lindahl P, Donnal J, Leeb S, Helsen J (2019) Review and comparison of tacholeless instantaneous speed estimation methods on experimental vibration data. *Mech Syst Signal Process* 129:407–436

- Pei X, Zheng X, Wu J (2021) Rotating machinery fault diagnosis through a transformer convolution network subjected to transfer learning. *IEEE Trans Instrum Meas* 70:1–11
- Peng B, Xia H, Lv X, Annor-Nyarko M, Zhu S, Liu Y, Zhang J (2022) An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network. *Appl Intell* 52(3):3051–3065. <https://doi.org/10.1007/s10489-021-02555-4>
- Potočník P, Govekar E, Grabec I (2007) Acoustic and acoustic emission based condition monitoring of production processes. In: *Proceedings of the second world congress on asset management and the fourth international conference on condition monitoring*, pp 11–14
- Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu ML, Chen SC, Iyengar S (2018) A survey on deep learning: algorithms, techniques, and applications. *ACM Comput Surv (CSUR)* 51(5):1–36
- Qi Y, Shen C, Wang D, Shi J, Jiang X, Zhu Z (2017) Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. *IEEE Access* 5:15066–15079
- Raissi M, Perdikaris P, Karniadakis G (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686–707
- Randall RB (2011) *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. Wiley, Hoboken
- Randall R, Smith W (2016) Use of the teager kaiser energy operator to estimate machine speed. In: *Paper to be presented at PHM Europe conference, Bilbao, Spain*, pp 5–8
- Rao M, Zuo MJ (2018) A new strategy for rotating machinery fault diagnosis under varying speed conditions based on deep neural networks and order tracking. In: *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, IEEE, pp 1214–1218
- Rao M, Li Q, Wei D, Zuo MJ (2020) A deep bi-directional long short-term memory model for automatic rotating speed extraction from raw vibration signals. *Measurement* 158:107719. <https://doi.org/10.1016/j.measurement.2020.107719>
- Ravanelli M, Bengio Y (2018) Speech and speaker recognition from raw waveform with sincnet. <https://doi.org/10.48550/ARXIV.1812.05920>, URL <https://arxiv.org/abs/1812.05920>
- Ruiz-Cárcel C, Jaramillo VH, Mba D, Ottewill JR, Cao Y (2016) Combination of process and vibration data for improved condition monitoring of industrial systems working under variable operating conditions. *Mech Syst Signal Process* 66:699–714
- Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. In: *Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems*, vol 30. Curran Associates Inc, Red Hook
- Salimans T, Kingma DP (2016) Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In: *Advances in neural information processing systems*, pp 901–909
- Sarma S, Agrawal V, Udupa S, Parameswaran K (2008) Instantaneous angular position and speed measurement using a dsp based resolver-to-digital converter. *Measurement* 41(7):788–796
- Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G (2009) The graph neural network model. *IEEE Trans Neural Netw* 20(1):61–80
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- Schmidt S, Heyns PS, De Villiers JP (2018) A tacholeless order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process. *Mech Syst Signal Process* 100:630–646
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*, pp 618–626
- Shao H, Jiang H, Wang F, Zhao H (2017a) An enhancement deep feature fusion method for rotating machinery fault diagnosis. *Knowl-Based Syst* 119:200–220. <https://doi.org/10.1016/j.knosys.2016.12.012>
- Shao H, Jiang H, Zhao H, Wang F (2017b) A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mech Syst Signal Process* 95:187–204. <https://doi.org/10.1016/j.ymssp.2017.03.034>
- Shao H, Jiang H, Lin Y, Li X (2018) A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mech Syst Signal Process* 102:278–297. <https://doi.org/10.1016/j.ymssp.2017.09.026>
- Shao HD, Ding ZY, Cheng JS, Jiang HK (2020) Intelligent fault diagnosis among different rotating machines using novel stacked transfer auto-encoder optimized by PSO. *ISA Trans* 105:308–319
- Simon HA (1969) *The sciences of the artificial*. MIT Press, Cambridge
- Slavkovikj V, Verstockt S, De Neve W, Van Hoecke S, Van de Walle R (2015) Hyperspectral image classification with convolutional neural networks. In: *Proceedings of the 23rd ACM international conference on Multimedia*, pp 1159–1162

- Subbaraj P, Kannapiran B (2014) Fault detection and diagnosis of pneumatic valve using adaptive neuro-fuzzy inference system approach. *Appl Soft Comput* 19:362–371
- Sun W, Shao S, Zhao R, Yan R, Zhang X, Chen X (2016) A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement* 89:171–178. <https://doi.org/10.1016/j.measurement.2016.04.007>
- Sun J, Yan C, Wen J (2018) Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Trans Instrum Meas* 67(1):185–195
- Sun KH, Huh H, Tama BA, Lee SY, Jung JH, Lee S (2020) Vision-based fault diagnostics using explainable deep learning with class activation maps. *IEEE Access* 8:129169–129179
- Tama BA, Lee SY, Lee S (2020) An overview of deep learning techniques for fault detection using vibration signal. In: INTER-NOISE and NOISE-CON congress and conference proceedings, institute of noise control engineering, pp 5701–5706
- Urbanek J, Barszcz T, Sawalhi N, Randall RB (2011) Comparison of amplitude-based and phase-based method for speed tracking in application to wind turbines. *Metrol Meas Syst* 18(2):295–303
- Urbanek J, Barszcz T, Antoni J (2013) A two-step procedure for estimation of instantaneous rotational speed with large fluctuations. *Mech Syst Signal Process* 38(1):96–102
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Lu, Polosukhin I (2017) Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) *Advances in neural information processing systems*, vol 30. Curran Associates Inc., Red Hook
- Wang P, Ananya Yan R, Gao RX (2017) Virtualization and deep recognition for system fault classification. *J Manuf Syst* 44:310–316
- Wang B, Lei Y, Li N, Li N (2018) A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Trans Reliab* 69(1):401–412
- Wang H, Li S, Song L, Cui L (2019a) A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals. *Comput Ind* 105:182–190
- Wang J, Sánchez J, Iturrioz J, Ayesta I (2019b) Artificial intelligence for advanced non-conventional machining processes. *Procedia Manuf* 41:453–459. <https://doi.org/10.1016/j.promfg.2019.09.032>. (8th Manufacturing Engineering Society International Conference, MESIC 2019, 19–21 June 2019, Madrid, Spain)
- Wang X, Shen CQ, Xia M, Wang D, Zhu J, Zhu ZK (2020) Multi-scale deep intra-class transfer learning for bearing fault diagnosis. *Reliab Eng Syst Saf* 202:107050
- Wen L, Li X, Gao L, Zhang Y (2018) A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans Ind Electron* 65(7):5990–5998
- Wu JY, Zhao ZB, Sun C, Yan RQ, Chen XF (2020a) Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection. *IEEE Trans Ind Inform* 16(12):7479–7488
- Wu JY, Zhao ZB, Sun C, Yan RQ, Chen XF (2020b) Few-shot transfer learning for intelligent fault diagnosis of machine. *Measurement* 166:108202
- Wu YC, Zhao RZ, Jin WY, He TJ, Ma SC, Shi MK (2021) Intelligent fault diagnosis of rolling bearings using a semi-supervised convolutional neural network. *Appl Intell* 51:2144–2160
- Xu X, Cao D, Zhou Y, Gao J (2020a) Application of neural network algorithm in fault diagnosis of mechanical intelligence. *Mech Syst Signal Process* 141:106625
- Xu XW, Tao ZR, Ming WW, An QL, Chen M (2020b) Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion. *Measurement* 165:108086
- Yang B, Lei YG, Jia F, Li NP, Du ZJ (2020) A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines. *IEEE Trans Ind Electron* 67(11):9747–9757
- Yang C, Zhou K, Liu J (2022) SuperGraph: spatial-temporal graph-based feature extraction for rotating machinery diagnosis. *IEEE Trans Ind Electron* 69(4):4167–4176
- You W, Shen C, Guo X, Jiang X, Shi J, Zhu Z (2017) A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery. *Adv Mech Eng* 9(6):1687814017704146
- Yu K, Han HZ, Fu Q, Ma H, Zeng J (2020) Symmetric co-training based unsupervised domain adaptation approach for intelligent fault diagnosis of rolling bearing. *Meas Sci Technol* 31(11):115008
- Yu X, Tang B, Zhang K (2021) Fault diagnosis of wind turbine gearbox using a novel method of fast deep graph convolutional networks. *IEEE Trans Instrum Meas* 70:1–14
- Zhang B, Zhang S, Li W (2019) Bearing performance degradation assessment using long short-term memory recurrent network. *Comput Ind* 106:14–29. <https://doi.org/10.1016/j.compind.2018.12.016>
- Zhang D, Stewart E, Entezami M, Roberts C, Yu D (2020) Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network. *Measurement* 156:107585

- Zhang D, Chen Y, Guo F, Karimi HR, Dong H, Xuan Q (2021a) A new interpretable learning method for fault diagnosis of rolling bearings. *IEEE Trans Instrum Meas* 70:1–10. <https://doi.org/10.1109/TIM.2020.3043873>
- Zhang Y, Zhou T, Huang X, Cao L, Zhou Q (2021b) Fault diagnosis of rotating machinery based on recurrent neural networks. *Measurement* 171:108774
- Zhao X, Jia M, Bin J, Wang T, Liu Z (2021) Multiple-order graphical deep extreme learning machine for unsupervised fault diagnosis of rolling bearing. *IEEE Trans Instrum Meas* 70:1–12
- Zhao X, Jia M, Liu Z (2021) Semisupervised graph convolution deep belief network for fault diagnosis of electromechanical system with limited labeled data. *IEEE Trans Ind Inform* 17(8):5450–5460
- Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2921–2929
- Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M (2020) Graph neural networks: a review of methods and applications. *AI Open* 1:57–81
- Zhou K, Yang C, Liu J, Xu Q (2022) Deep graph feature learning-based diagnosis approach for rotating machinery using multi-sensor data. *J Intell Manuf*. <https://doi.org/10.1007/s10845-021-01884-y>
- Zhu XX, Luo XZ, Zhao JH, Hou DN, Han ZH, Wang Y (2020) Research on deep feature learning and condition recognition method for bearing vibration. *Appl Acoust* 168:107435
- Zhu Z, Wang L, Peng G, Li S (2021) Wda: an improved wasserstein distance-based transfer learning fault diagnosis method. *Sensors*. <https://doi.org/10.3390/s21134394>

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