

Review

Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review

Shaohua Qiu , Xiaopeng Cui, Zuowei Ping *, Nanliang Shan, Zhong Li, Xianqiang Bao and Xinghua Xu

National Key Laboratory of Science and Technology on Vessel Integrated Power System,
Naval University of Engineering, Wuhan 430033, China

* Correspondence: pingzuowei@hust.edu.cn

Abstract: Fault diagnosis and prognosis (FDP) tries to recognize and locate the faults from the captured sensory data, and also predict their failures in advance, which can greatly help to take appropriate actions for maintenance and avoid serious consequences in industrial systems. In recent years, deep learning methods are being widely introduced into FDP due to the powerful feature representation ability, and its rapid development is bringing new opportunities to the promotion of FDP. In order to facilitate the related research, we give a summary of recent advances in deep learning techniques for industrial FDP in this paper. Related concepts and formulations of FDP are firstly given. Seven commonly used deep learning architectures, especially the emerging generative adversarial network, transformer, and graph neural network, are reviewed. Finally, we give insights into the challenges in current applications of deep learning-based methods from four different aspects of imbalanced data, compound fault types, multimodal data fusion, and edge device implementation, and provide possible solutions, respectively. This paper tries to give a comprehensive guideline for further research into the problem of intelligent industrial FDP for the community.

Keywords: fault diagnosis; fault prognosis; machine learning; deep learning; industrial systems



Citation: Qiu, S.; Cui, X.; Ping, Z.; Shan, N.; Li, Z.; Bao, X.; Xu, X. Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review. *Sensors* **2023**, *23*, 1305. <https://doi.org/10.3390/s23031305>

Academic Editor: Jongmyon Kim

Received: 26 November 2022

Revised: 23 December 2022

Accepted: 18 January 2023

Published: 23 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

Industrial systems are typical complex systems with various subsystems and device types of mechanical system, power system, information system, electronic system, or their combinations. They are playing an increasingly important role in the economy, such as manufacturing industry, energy industry and chemical industry, which are now developed with more functions, more sophisticated structures, and larger scales [1]. Reliability issues have gradually become the key of whether many modern industrial systems can be truly practical. Once a failure occurs, it may affect the safe and stable operation of the entire system, i.e., reducing the efficiency of the system, and causing system breakdown or damage in severe cases [2]. It may also endanger personnel safety, and cause other catastrophic consequences. Therefore, the early identification of faults in advance can greatly help to take appropriate actions of maintenance to avoid the undesired consequences.

Driven by demand, prognostics and health management (PHM) [3] technology, firstly originated from engine health monitoring systems [4], has gained increasingly more attention. PHM is an expansion of the traditional reliability or predictive maintenance concept oriented for complex industrial systems. It realizes the development from the initial condition monitoring and fault diagnosis that aims to estimate health status, to health management that aims at formulating the countermeasures based on the results of monitoring, diagnosis, and prognosis.

In practical scenes, it is often difficult or even impossible to establish mathematical models of complex components or systems [5], in order to trace and analyze faults. Therefore, a large amount of historical data that were collected in the process of system operation

and maintenance have become the major method by which to evaluate the system's health status. As the core part of PHM technology, the fault diagnosis and prognosis (FDP) technique based on data-driven machine learning (ML) methods recognizes or learns the health features of the system from historical data, and tries to discover and mine the information hidden in the data, so that it can accurately analyze and predict future system behavior without precisely knowing the forward physical model. ML methods generally have a more powerful capacity for FDP without the assumption of data distribution, smoother and more intelligent FDP processes with fewer processing stages and less human intervention, and, moreover, less prior-knowledge requirements for more complex components or systems to be modeled [6].

Consequently, data-driven ML methods have long been applied in various industrial FDP applications. A typical ML pipeline generally consists of three steps [7], i.e., data preprocessing, feature extraction and classification or regression. The performance of ML heavily depends on the manually predefined feature extraction rules. In the past decade, with the great development of mega-scale open datasets [8], evolutionary computing capacity of new GPU architectures [9] and innovative neural network training methods [10], deep learning [11] can hierarchically extract highly-abstract features in an end-to-end way from the labeled training dataset. Due to its superior performance over ML methods, deep learning (DL) has gained remarkable success in the tasks of computer vision, natural-language processing, etc. In the community of industrial FDP, researchers have also made great efforts to introduce DL techniques into different and unique industrial FDP scenarios, and tremendous progress has been witnessed.

At present in the era of Industry 4.0 [12], the emerging of Big Data [1,13], Internet of Things (IoT) [14,15], and artificial intelligence (AI) technology [16,17] are now promoting the transformation of PHM (specifically FDP in this paper) from traditional single-sensor-oriented diagnosis to system-wise intelligent diagnosis and prognosis. When the traditional physical model-based PHM technology is progressing slowly in the face of unprecedented complex systems, the scientific "The Fourth Paradigm" [18] based on Big Data collected from IoT and supported by modern AI technology is also making industrial systems truly intelligent.

1.2. A Survey of Relevant Reviews

To summarize the current research of intelligent FDP, there are a number of outstanding surveys on the topic of intelligent FDP [1,7,19–28]. They conduct extensive review on existing literature quantitatively and qualitatively from their unique viewpoints, and identify the trends and ideas of FDP methods for different scenarios.

Xu et al. [1] analyzed existing issues and challenges in the Big Data era from different driving factors, such as data quality and cost balance, method selection, application problems, and deep utilization. Li et al. [19] summarized the common fault types of sensors in monitoring and control systems and presented the latest fault diagnosis methods that combined different advanced technologies. Furthermore, Tang et al. [27] reviewed the DL applications toward fault diagnosis methods for rotating machinery according to its major components, including bearing, gear, and pumps. A comprehensive review of Big Data-driven intelligent FDP for mechanical systems was given by Lei et al. [28], wherein the latest cutting-edge research results are focused, e.g., deep transfer learning-based FD, Big Data-driven RUL prediction, data-model fusion prognosis, etc. In addition, Fernandes et al. [20] provided a systematic literature review of ML methods for mechanical FDP in manufacturing. They examined and characterized the research in more details based on five basic research questions.

1.3. Motivation

The aforementioned review work provides a very good foundation for the work in this paper. Some surveys concentrate on FDP for specific type of device, e.g., machinery [20–24,27,28], wind power converter [25], lithium-ion battery system [26], while some focus on specific FDP

method, e.g., deep domain adaptation [21], attention mechanism [22], recurrent neural network (RNN) [23], etc. Most of these reviews cover the data-driven ML techniques, but few of them give a comprehensive overview of the generic DL techniques used for industrial FDP. Moreover, due to the rapid development and iteration of DL techniques in recent years, a large number of excellent DL architectures and algorithms have emerged, bringing new opportunities to the promotion of FDP. The most up-to-date trends of recent a couple of years in industrial FDP, especially about emerging DL architectures, as well as the future trends in the next few years, are rarely covered by relevant reviews. To the best of our knowledge, there is currently no review paper of the Transformer technique's application in intelligent FDP.

Therefore, a review to comprehensively cover the latest development of DL techniques for intelligent industrial FDP is still left blank but desired. In order to track the latest achievement of DL techniques for intelligent industrial FDP, we conduct a comprehensive survey on relevant literature of the past 5 years in this paper. The main contributions of this paper are as follows:

1. From a different viewpoint of data analysis, we provide a generalized definition and mathematical formulations for FDP problems compared to previous work.
2. We collect and summarize recent advances of recent 5 years for intelligent industrial FDP, review and analyze them from the perspective of DL techniques.
3. The emerging DL architectures, including generative adversarial network, and transformer and graph neural network, are investigated in the survey to provide an up-to-date view of the latest research trends of intelligent FDP.
4. Challenges encountered in current research are discussed from the aspects of data imbalance, compound faults, multimodal fusion and edge implementation, which are seldom analyzed by other literature. Possible solutions are also provided.

The rest of this paper is organized as follows. Section 2 gives the problem formulations. In Section 3, we elaborate the FDP methods of emerging DL techniques. Its detailed analyses are given in the followed Section 4 and Section 5. In Section 6, the major problems encountered in the current research are summarized and the trend is prospected. The conclusions are finally drawn in Section 7.

2. Problem Formulation

Different from previous work that deals with specific industrial faults and analyzes them from the aspect of physical model or fault mechanism, we analyze the problem of FDP from a novel viewpoint of data analysis. In this section, we give the generalized definitions of faults and the mathematical formulations of FDP problems.

2.1. Definitions of Faults

In general, the condition monitoring results of certain object in industrial systems experiences changes all the time, and not all changes in sensory data are failures or faults. Here are some common senses:

- Changes caused by random noise are not necessarily faults, but when the variance of the noise changes, it is generally considered to be a fault.
- Fluctuation within a stable range in a certain operation condition is not a malfunction. In different operating conditions, this fluctuation may be different.
- A change that breaks the current pattern is a fault.

Figure 1 gives a comparison of the normal three-phase current waveform and the current waveform of interturn short-circuit fault under the same working condition. At no point does the current amplitude exceeds the working condition mode range, but the (blue) curve pattern of $t > 125$ ms changes and it is a fault. Therefore, we consider that the core part of FDP is to discriminate the faulty patterns from normal working patterns which are represented in sensory data, and to build a health index that indicates the changing trend in working patterns.

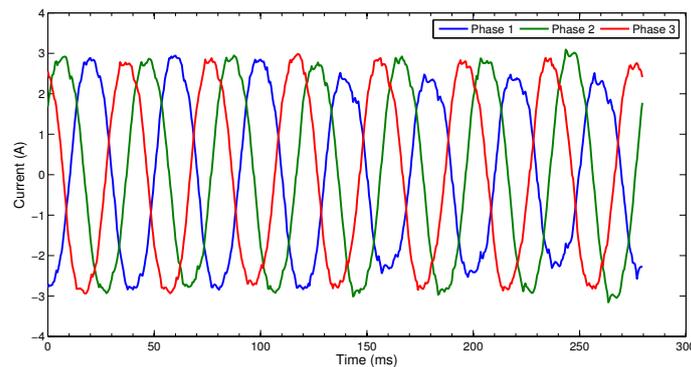


Figure 1. An example of three-phase current waveform.

2.2. Mathematical Formulations of Fault Diagnosis

Given N physical variables (such as pressure, current, temperature) within a specific time range $T = [t_1, t_2]$ measured by a number of sensors (such as strain gauges, Hall sensors, temperature sensors, etc.) at a specific position of a specific device, we set $\mathbf{M}(T) = \{m_i(T) | i = 1, 2, \dots, N\}$. When the current operating condition is p , the fault indicator function $f_\theta(\mathbf{M}(T), p)$ is to judge whether the current state s as in Equation (1) is normal or not, its value range of f_θ is $\{0, 1\}$, and θ is the parameter of f .

$$s = f_\theta(\mathbf{M}(T), p) \quad (1)$$

when the monitoring variable $\mathbf{M}(T)$ and the working mode p are known, the corresponding fault state is also determined theoretically, i.e., for a certain type of device, its fault indicator function f is determined.

In this way, the problem of fault diagnosis becomes the process of solving the parameter θ of the fault indicator function f . The determination of function parameters θ can be explicitly solved by forward modeling of physical models, but it is often too complicated or even unsolvable. The data-driven fault diagnosis methods make use of the existing data, and tries to mine the parameter θ of f backward from the data [7]. It then becomes the following problem as in Equation (2), that is, searching for a certain point θ' in the parameter space Θ , so that its output pattern on a large number of data samples is the least different from the real situation, thereby turning it into an optimization problem:

$$\arg \min_{\theta' \in \Theta} \|s' - f_{\theta'}(\mathbf{M}'(T), p')\|. \quad (2)$$

Among them, s' and $(\mathbf{M}'(T), p')$ are the labels and data vectors in the known sample set.

If the current device status is judged as fault, the fault can then be classified. The current pattern is compared with the fault patterns in the fault database, the smallest deviation degree between the current fault and each fault pattern can be searched. It is worth noting that since the original data $\mathbf{M}(T)$ used for diagnosis is usually high dimensional and redundant in feature spaces, it is usually necessary to perform feature selection, feature extraction or feature fusion on the original data to reduce the data dimension.

2.3. Mathematical Formulations of Fault Prognosis

One major challenging problem in fault prognosis is the remaining useful life (RUL) estimation of the device whose specific meaning is shown in Figure 2. It is necessary to select an appropriate health indicator for RUL estimation, which can well reflect the change in the degradation degree of device health, and there is a corresponding threshold to indicate when will the device reach a functional failure.

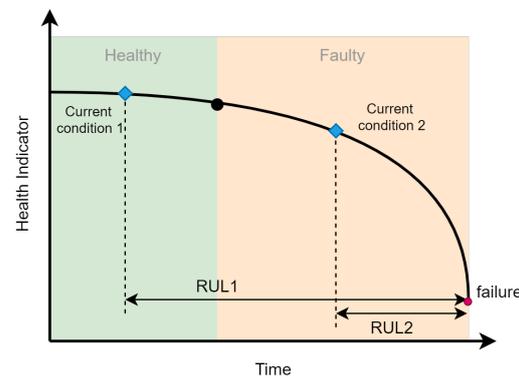


Figure 2. Schematic diagram of life cycle.

Given k known historical data and their corresponding health feature sequence $\{f_i(n)|i = 1, 2, \dots, k, n = 1, 2, \dots, N\}$, where N is the length of the known health feature sequence, the dataset $\{T_i(l), f_i(l)\}$ can be formed according to all the historical data and the corresponding sequence of health indices. According to the determined device-life degradation model g , we can perform fitting via regression on $\{T_i(l), f_i(l)\}$ to determine the model parameters of the degradation model g . Given the current observation data health indicator sequence $f(n)$, the degradation model g is used to extrapolation predict and estimate the evolution trend \hat{f} of the predicted features. The estimated evolution curve \hat{f} obtained is then compared with the failure threshold. When \hat{f} exceeds the failure threshold for the first time at time T_f , the device fails. Assuming that T_N is the time length of known observation data, RUL of the device is

$$RUL = T_f - T_N. \quad (3)$$

The key point of fault prognosis is the choice of degradation model. The factors considered include the global degradation mode, short-term degradation characteristics, the amount of data available for modeling and the data noise level, etc.

3. Modern Deep Learning Techniques for Intelligent Industrial FDP

3.1. Modern Deep Learning Techniques

As a young and developing field of AI, ML techniques try to discover knowledge from a large amount of historical data for prediction or classification on new data. More specifically, it is designed to find a projection to fit the input data for desired results, which is often too complex to be explicitly formulated. In terms of application purposes, supervised machine learning is mainly divided into two categories [29]: classification and regression. The former learns the boundaries between categories to achieve classification of new data [30]. The latter fits regularities to the data to predict the properties of new data points. Correspondingly, fault diagnosis is actually a classification problem, and fault prognosis is a regression problem.

As a subset of ML, the emerging DL is currently the hottest topic in AI. It is originated from the paper [10] published in 2006 by Hinton et al. This paper reveals two characteristics of deep learning. The first is that the neural network with multiple hidden layers has excellent potential for learning more representative features from raw data which are generally designed manually in traditional ML methods. The second is that the difficulty of training deep neural networks can be overcome by layer-by-layer pre-training using the method of unsupervised learning in the Restricted Boltzmann Machine (RBM).

The concept “deep” in deep learning is compared to traditional machine learning algorithms, such as SVM, ANN, and other shallow learning methods, in which there are more layers of non-linear functions in deep learning methods. In traditional shallow neural learning methods, data sample features need to be manually extracted. Conversely, DL automatically learns to obtain feature representations by performing layer-by-layer

feature transformation on original data via back-propagation, and these hierarchical feature representations are highly abstract and task-oriented. One of its major merits is that it can complete the learning in an end-to-end way directly from raw data to results of classification and regression tasks.

Typical DL architectures include deep belief network (DBN) [31], autoencoder (AE) [32], convolutional neural network (CNN) [33], and RNN [34]. With the rapid development of DL techniques in these years, many new architectures have been proposed and introduced into the tasks of intelligent industrial FDP. Examples are generative adversarial network (GAN) [35], transformer [36], and graph neural network (GNN) [37]. Similarly, CNN is prospering again, due to the progress made in the fields of computer vision in recent years.

3.2. Categorization and Literature Trends of DL Techniques for Industrial FDP

Figure 3 shows the categorization of major DL-based approaches used in intelligent FDP. According to the supervision type, they can be divided into unsupervised methods and supervised methods. The former tries to find the inherent common pattern within data which are unlabeled, while the latter refers to methods that learn highly non-linear relationship between the input data and its paired labeled output. More specifically, the supervised methods can be further divided into processing of specific data types or extraction of distinctive features, depending on their objectives. Their detailed introductions will be expanded in the following sections.

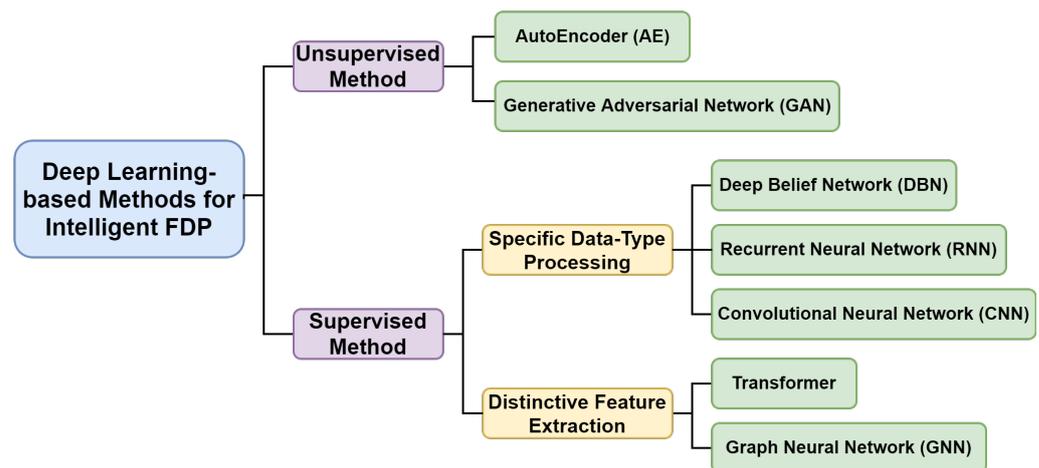


Figure 3. The categorization of deep learning techniques in intelligent FDP.

Figure 4 illustrates the number of journal publications of deep learning methods in intelligent FDP from January 2013 to September 2022 on Web of Knowledge. As can be seen, the number of papers published is increasing year by year, and CNN-based FDP methods account for the majority of all methods. The publication number of typical DL architectures, such as DBN and AE, are stable or growing with relatively slower speed. Note that emerging network architectures are also gradually attracting the attention of researchers.

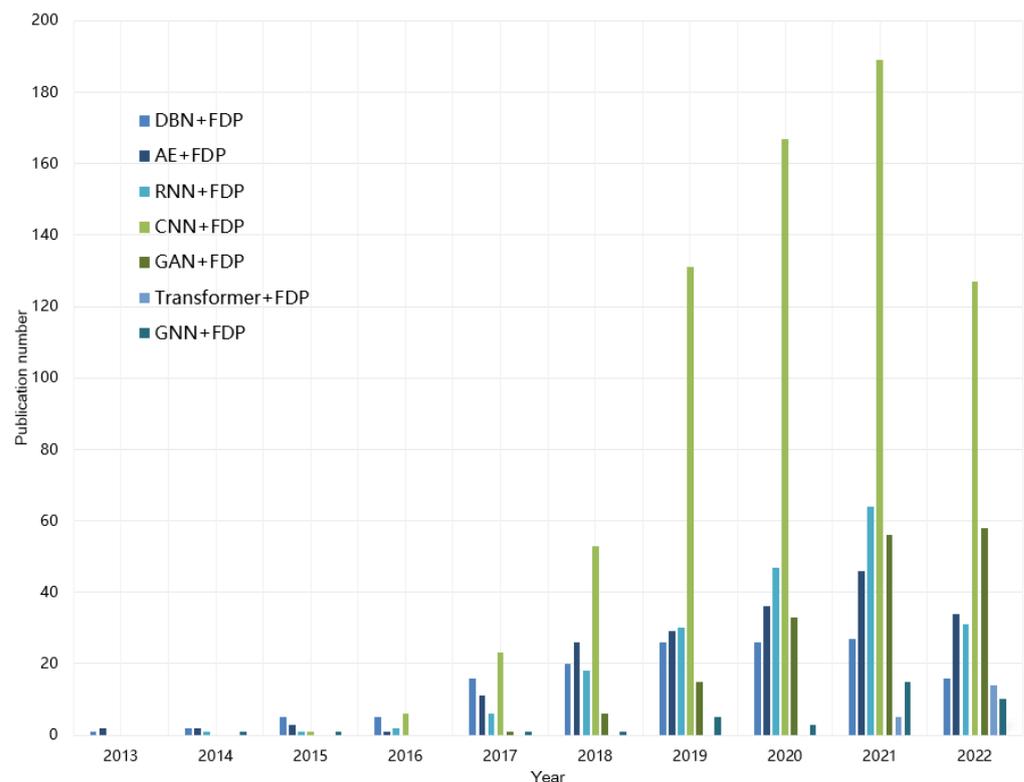


Figure 4. Publication trends of deep learning methods in intelligent industrial FDP.

4. Part I: Unsupervised DL Methods for Intelligent Industrial FDP

Unsupervised DL methods are not fed with labeled information, so it is necessary for them to mine the inherent structure and pattern within data. Unsupervised DL methods generally does not solve the tasks of FDP in a direct way, but also serve for peripheral tasks that are also crucial, such as feature reduction and data generation.

4.1. Autoencoder (AE) for High-Dimensional Feature Reduction

Autoencoder (AE) is an unsupervised architecture which assumes that the output being encoded and decoded is the same with the input. In this sense, the encoder part can be used for feature reduction where high-dimensional input data can be converted into low-dimensional encoded vectors. The idea of an encoder–decoder is also widely adopted by other DL architectures such as CNNs. A simple architecture of AE is illustrated in Figure 5. AE can also be divided into standard AE [38–40], denoising AE [41], sparse AE [42], variational AE [43] and contractive AE [44], etc.

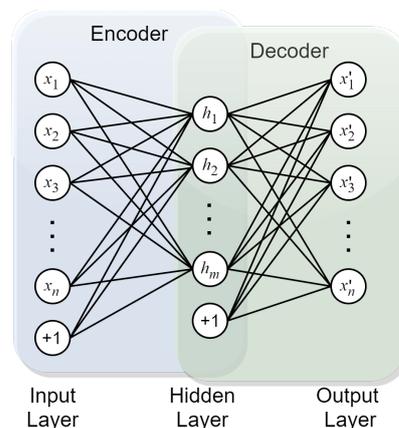


Figure 5. Basic structure of AE.

AEs have been widely used for feature extraction and fault classification, and have demonstrated powerful feature extraction and non-linear dimensionality-reduction capabilities and robustness in practical FDP applications. In [45], a sparse AE is designed to automatically extract degradation indicators for followed fault detection in multi-component system. Ref. [46] use multi-layer sparse AE as a multi-sensor feature fusion and extraction method combined with DBN for bearing fault diagnosis. A list of recent publications of AE-based intelligent FDP are given in Table 1. As seen, in order to obtain better performance, stacked AEs are preferred to be used in different scenarios, while the borders between different types of AEs are breaking down and leading to fused architectures, e.g., sparse denoising AE. Despite the above advantages, it still suffer from the situation that meaningful features sometimes cannot be easily extracted due to the inherent properties of AEs. Moreover, its capability is generally highly correlated to its training samples.

Table 1. Recent publications of intelligent FDP methods based on AEs.

Type	Reference	Year	Method	Object
Standard AE	[39]	2019	A stacked AE for compressing the feature depth	high-voltage circuit breakers
	[47]	2020	1-D residual convolutional AE for learning features from vibration signals directly in an unsupervised-learning way	machinery
	[40]	2022	AE with adaptive Morlet wavelet to establish accurate mapping hidden in the fused health index	aeroengine
	[38]	2022	Stacked AE to establish an accurate non-linear mapping between the raw data and different fault states	rotating machinery
Denoising AE	[41]	2018	Stacked denoising AE to extract useful feature and reduce the dimension of vibration signal to 2 or 3 dimensions	bearing
Sparse AE	[42]	2022	Sparse representation convolutional AE to extract impulsive components of vibration signals	rotating machinery
Sparse denoising AE	[48]	2019	A sparse stacked denoising AE is proposed for feature extraction	bearing
Variational AE	[43]	2022	A convolutional variational AE with attention mechanism providing better spatial distributions of features	aeroengine
Contractive AE	[44]	2018	Stacked contractive AE for automatic robust features extraction	rotating machinery

4.2. Generative Adversarial Network (GAN) for Data Generation

An important requisition for supervised deep learning methods is the massive amount of training samples. However, in many practical scenarios, training data collected at hand are scarce and imbalanced, which is reflected on the ratio of numbers of positive and negative samples, as well as the known fault patterns. It is a well-known problem of small sample or small data. Traditional over-sampling techniques can hardly capture the data distribution and will easily lead to over-fitting [49]. Firstly, succeed in computer vision from 2014 by Goodfellow, generative adversarial network (GAN) [35] is an unsupervised method that is able to generate realistic samples via a minimax game between two networks. It consists of a generator network to generate samples and a discriminator network to judge the likeness of the generated samples. The generated realistic fake data fit within the distribution of the training data, which outperforms the traditional over-sampling methods, such as synthetic minority oversampling technique (SMOTE) [50], by a large margin. As a result, GAN has shown outstanding performance in many areas beyond computer vision. In the field of FDP, GANs have gradually been adopted, and it has show promising results compared with other architectures. The basic idea of data augmentation using GAN is illustrated in Figure 6.

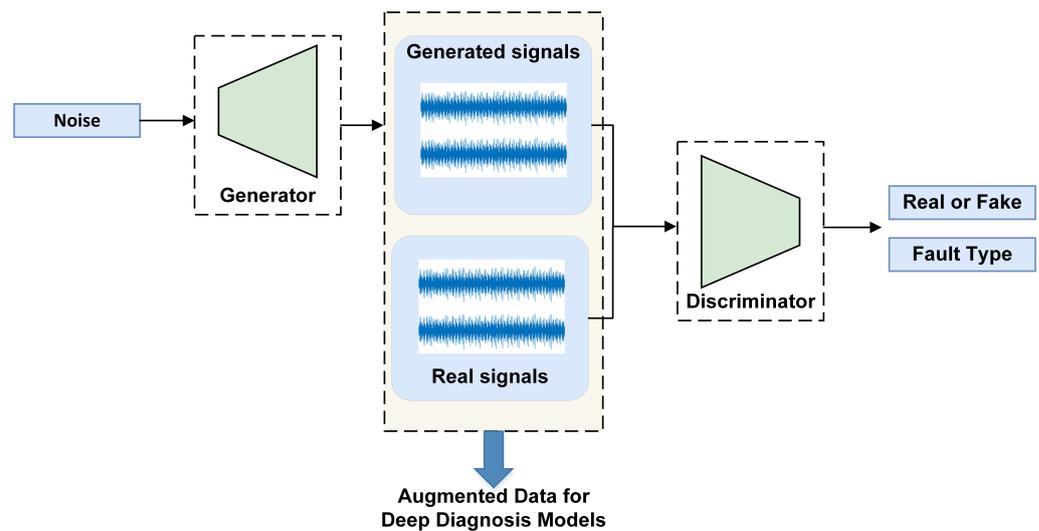


Figure 6. A basic example of data augmentation using GAN.

Initially, GAN is mainly adopted for normal or faulty sample generation, either for images or for signals. Figure 6 is an example GAN for data augmentation for the training of deep fault diagnosis models. Usually, the capacity of modeling data distribution in GAN can be further extended for fault diagnosis. For example, the trained generator can be used to fix a faulty sample, and the fault can then be located by sample comparison [51]. Moreover, this adversarial learning strategy of GAN has also been widely implemented to tackle the problem of domain shift of data distribution for fault diagnosis under different working conditions or environments, i.e., the distribution of available training data in the source domain is different from that of data to be tested in the target domain, making the trained model hard to be generalized [52]. It is a very challenging issue usually faced by industrial applications.

Due to its special and excellent property, GAN has, consequentially, received significant attention when dealing with intelligent FDP of real industrial systems. A list of recent methods based on GANs are given in Table 2 for more comprehensive and detailed information. The current work mainly focuses on the gaming strategy of GAN to achieve the goal of more realistic sample generation and cross domain adaption for intelligent FDP. Ref. [49] set up an infoGAN-based failure-prediction algorithm, and it uses an auxiliary GAN to enforce consistency of the generated samples and their corresponding labels. Ref. [53] propose to use deep feature enhanced GAN to ensure the accuracy and diversity of synthesized samples, thereby improving the performance of rolling bearing imbalanced fault diagnosis. Aiming at the problem that in real industries only data in machine healthy condition can be collected in advance, literature [54] propose a multilabel 1-D GAN to generate damage data of industry equipment, and the fault diagnosis accuracy was improved with these generated data. Ref. [55] jointly use labeled samples in auxiliary domain and unlabeled samples in target domain via domain-adversarial training in order to enhance the adaptability of samples in auxiliary domain to target domain and improve the transfer performance.

Despite the fact that GANs can generate samples with the same distribution, it is still difficult to judge or evaluate the quality of generated 1-D signals, as opposed to the image generation. Moreover, how to ensure that the adversarial training process converges to the desired destination is also a challenge. Lastly, as faulty sample generation is always on the top of the objective list, the way of combining prior knowledge from experts to improve the generation is also an important issue to be explored for real industrial applications.

Table 2. Some of recent intelligent FDP methods based on GANs.

Type	Reference	Year	Method	Object
Data generation	[56]	2019	GAN is used to refine the rough fault data more similar with real data.	wind turbine
	[57]	2019	An auxiliary classifier GAN-based framework to learn from mechanical sensor signals and generate realistic one-dimensional raw data.	induction motor
	[58]	2020	GAN to generate new samples similar to the simulation and measurement fault samples in order to enlarge datasets.	bearing
	[59]	2021	GANs is used to acquire abundant synthetic samples generated from the simulation and measurement samples, which aims to expand fault samples.	rotor-bearing systems
	[60]	2021	DCGAN is employed to produce new face-portraits of the nominal and failure behaviors.	ball-bearing joints
	[53]	2022	GAN to enhance the deep features of real signals.	rolling bearings
	[61]	2022	GAN uses available time series degradation data to generate synthetic degradation data.	bearing
	[62]	2022	A Wasserstein conditional GAN constrain the data generation characteristics to improve the validity of data.	rolling bearings
Local domain FD	[63]	2020	A semi-supervised multi-scale convolutional GAN to learn discriminativity from unlabeled data.	rolling bearings
	[64]	2022	Stepwise GAN trains multistage with unlabeled normal data and fuses multi-source information at feature level and aggregating neighboring information at decision level	liquid rocket engine
Cross domain FD	[58]	2020	Domain adversarial transfer network exploits task-specific feature learning networks and domain adversarial training techniques for handling large distribution discrepancy across domains.	rotating machinery
	[55]	2021	A deep transfer learning model based on an adversarial learning strategy to effectively separate multiple unlabeled new fault types.	mechanical equipment
	[65]	2022	A one-class GAN based on semi-supervised learning to learn one-class latent knowledge for dealing with multiple semi-supervised fault diagnosis tasks.	industrial robot

5. Part II: Supervised DL Methods for Intelligent Industrial FDP

Different from the unsupervised learning way that does not utilize labeled input data, supervised learning methods use a training set with inputs and correct outputs to teach models to yield the desired output. For intelligent FDP, supervised learning methods can be used to extract distinctive features for the specific task from specific types of sensory data.

5.1. Deep Belief Network (DBN) for Fault Features Mining

The traditional neural network is more computationally efficient when it has only few hidden layers, so it is mostly used to solve some relatively simple mapping modeling problems. DBN is a network constructed by stacking RBM which is a special type of generative stochastic neural network, including visible units and hidden units, and a basic example of DBN with two hidden layers is shown in Figure 7. It can be trained through pre-training the stacked RBMs. Based on DBN with multiple hidden layers, it can remove the dependence on prior-knowledge and adaptively extract fault features for diagnosis. It is also able to process non-linear high-dimensional data, thereby effectively avoiding problems, such as dimensional disaster. Therefore, DBNs are well suited for dealing with fault diagnosis of industrial Big Data.

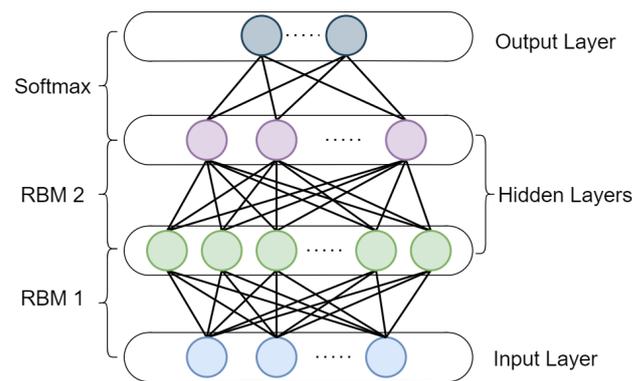


Figure 7. Basic structure of DBN.

Until now, plenty DBN-based researches have been carried out in this area, and widely used in fault diagnosis of aircraft engines [66], reciprocating compressors [67,68], gearboxes [69–72], rolling bearings [73–76], power transformers [77,78], etc. Current studies generally either use DBN as a classifier in a supervised way, or replace traditional signal processing methods to mine fault features in an unsupervised way. A compilation of recent work on DBNs for intelligent FDP are given in Table 3 from the classification of five aspects, along with their objects.

As a very classical technique in DL, DBN maintains a great deal of parameters to be set, and once inappropriately handled, it will affect its generalization and limit the accuracy, especially compared with other modern DL techniques. As a result, DBN is now being widely combined with other architectures, e.g., CNN, to achieve better performance, which can also be observed in Table 3.

Table 3. A compilation of recent intelligent FDP methods based on DBNs.

Purpose	Reference	Year	Method	Object
Classification	[74]	2019	Convolutional DBN based on Fisher parameter optimization	rolling bearings
	[79]	2020	DBN optimized by quantum-inspired differential evolution	rolling bearings
	[80]	2022	DBN classifies features from wavelet energy entropy	robot joint bearing
	[81]	2022	Gaussian convolutional DBN for classification	rotor bearing system
Feature Extraction	[82]	2020	Multi-scale cascading DBN for feature extraction	rotating machinery
	[68]	2020	Convolutional DBN for feature extraction	reciprocating compressors
	[83]	2022	Dilated convolution DBN to extract transferable characteristics	roller bearing
Feature Fusion	[71]	2019	DBN for feature fusion and classification	wind turbine gearbox
	[84]	2022	DBN fuses multivariables for parameter estimation	deep-sea human occupied vehicle
Index Regression	[66]	2019	DBN to construct health indicator for RUL prediction	aircraft engine
	[85]	2020	Median filtering DBN to extract health indicator	bearings
	[86]	2021	DBN to eliminate health indicator curve oscillation	bearings
Pretraining	[72]	2015	DBN to pretrain multilayer neural network	gearbox

5.2. Recurrent Neural Network (RNN) for Time-Series Data Processing

Compared with other architectures, recurrent neural network (RNN) [34] assumes that the input and output are not independent of each other, i.e., it tries to learn long-term dependencies from sequential or time-series input data. RNN contains non-linear recurrent units with directed cycles, combined with unit hidden states, so that time-series information can be preserved. Due to this structure, the state of the hidden layer is not only affected by the input data, but also by the previous calculation results, showing better dynamic characteristics. RNN is theoretically an ideal non-linear time-series forecasting tool and a universal approximator for dynamic systems. Common RNNs include gated recurrent unit (GRU) [87,88] and long short-term memory networks (LSTM) [89–91], which have become one of the most effective FDP methods for time-series data at present. Their basic unit comparison of them are given in Figure 8.

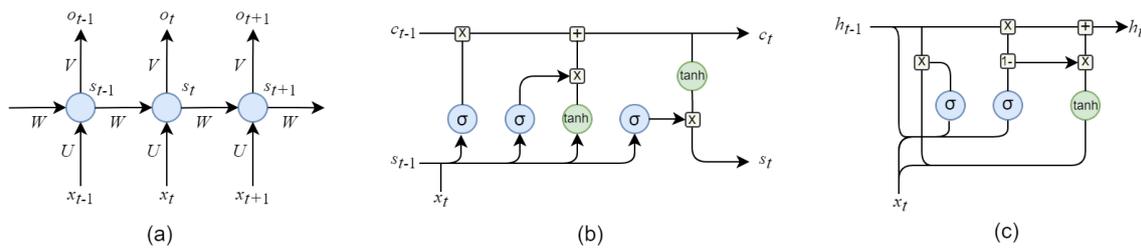


Figure 8. Unit comparison of (a) basic RNN, (b) LSTM, and (c) GRU.

Since long-term condition monitoring data are collected, RNN-based methods are in great demand in intelligent FDP. Ref. [91] proposes a convolutional LSTM that simultaneously extracts time-frequency domain features and models their long-term dependencies of vibration signals from bearing. The work in [92] utilized LSTM for fault diagnosis and RUL estimation on time-series aeroengine data. Ref. [93] use a RNN to implement early warning in the fault creep period for nuclear power machinery, together with principal component analysis, wavelet analysis, and Bayesian inference model. Ref. [34] design a fault prognosis approach with the degradation sequence of equipment based on LSTM, which uses the concatenated feature and operation state indicator for RUL estimation. Some of recent methods based on RNNs are listed in Table 4 according to their RNN types and purposes, e.g., fault diagnosis and RUL estimation.

Table 4. Recent publications of intelligent FDP methods based on RNNs.

Type	Reference	Year	Purpose	Method	Object
basic RNN	[93]	2020	Fault prediction	A fully connected RNN to predict faults from signal data dimensionally reduced.	nuclear power machinery
	[94]	2022	Fault diagnosis	RNN to identify different relevant types of faults, based on the past 24h of satellite measurements without on-site sensors.	photovoltaic systems
GRU	[95]	2020	RUL estimation	GRU to construct health indicator from sensitive fetures.	rolling element bearings
	[96]	2021	Fault diagnosis	GRU to exploit temporal information of time-series data and learn representative features from constructed signal images.	rotating machinery
	[87]	2021	Fault diagnosis	RNN with GRU and LSTM to capture the hidden patterns of vibration time series.	power transformer
	[88]	2022	Fault diagnosis	GRUs to understand whether data in a time series is crucial enough to preserve or forget.	bearings of wind turbines
LSTM	[97]	2019	Fault diagnosis	LSTM to capture long-term dependencies through recurrent behaviour.	wind turbines
	[98]	2020	RUL estimation	A LSTM model fuses multi-sensor monitoring signals to discover the hidden long-term dependencies among sensor time series signals.	turbofan engine
	[34]	2020	Fault diagnosis	LSTM learns long-term dependencies from the concatenated feature and operation state indicator of the equipment.	aircraft turbofan engines
	[91]	2021	RUL estimation	Convolution-based LSTM to capture long-term dependencies and extract features from the time-frequency domain at the same time.	rotating machinery
	[90]	2021	RUL estimation	Dual LSTM to characterize both long and short-term dependencies from historical information.	turbofan engine
	[99]	2022	Fault diagnosis	CNN to determine spatial correlations between two measurements within one time step, and LSTM to identify temporal dependencies between two adjacent time steps.	planetary gearbox

On one hand, the special structure of recurrent units with directed cycles enable RNN to better modeling time-series information and on the other hand, it makes that the training of RNN is generally much slower than that of other architectures such as CNNs, which poses a great computational requirement for industrial computing centers. Meanwhile, similar to CNN, RNN is also sensitive to training data, and when the fault feature is weak or distorted by noise, it is also hard to maintain good performance.

5.3. Convolutional Neural Network (CNN) for Image Fault Diagnosis

The convolutional neural network (CNN) is inspired by biological visual perception mechanism. It has unique structural characteristics, such as local connection, weight sharing, and pooling, which enables CNN with strong feature learning and representation ability. At present, CNN are mainly used in fault diagnosis, and it can hardly realize the status trends analysis of equipment or fault prognosis. In the field of intelligent FDP, there are generally three situations. A list of recent publications on intelligent FDP based on CNN architectures are given in Table 5. Details are described in the following subsections.

Table 5. A list of recent intelligent FDP methods based on CNNs.

Type	Reference	Year	Method	Object
Camera sensors	[100]	2019	CNN for feature extraction and classification	cooling radiator
	[101]	2020	CNN extracts fault features from infrared thermal images	rotating machinery
	[102]	2021	Mask rcnn for detection	power transformers
Signals to images	[103]	2019	Wavelet transform is adopted to extract 2-D time-frequency features from raw 1-D vibration signals	gearboxes
	[104]	2020	Continuous wavelet transform (CWT) converts signals into images	aeroengine control system
	[105]	2020	Sensor signals are converted to time-frequency distribution by wavelet transform	induction motor
	[106]	2021	1-D vibration signals are converted to 2-D grayscale vibration images	rolling element bearing
	[107]	2021	Vibration signals are first transformed into angular domain and then converted to corresponding envelope and squared envelope spectrum features, which are fused into RGB color image form	mechanical rotating components
	[108]	2022	CWT converts the vibratory time-series signals to the scalogram feature images	induction motors
	[109]	2022	A conversion method based on principal component analysis is applied to fuse multisignal data into three-channel RGB images	mechanical manufacturing systems
1-D CNN	[110]	2018	1-D CNN learns features adaptively from raw mechanical data without prior knowledge	motor bearing
	[111]	2019	Adaptive 1-D CNN for real-time and highly accurate circuit monitoring system	modular multilevel converter
	[112]	2020	Multi-attention 1-D CNN to diagnose faults	rolling bearing
	[113]	2021	1-D CNN to learn feature from the high-frequency components	high-speed train bogie
	[114]	2022	1-D CNN to establish model for fault diagnosis	UAV rotor
	[115]	2022	Multi-level features fusion 1-D CNN for good performance of feature extraction on vibration signals	bearing

5.3.1. The Monitoring Sensors Are Cameras

When the device fault can be captured by camera, i.e., there are evidences reflected at pixel level, the CNN-based methods can obtain better diagnosis results, such as in the fields of machinery and circuits. Ref. [101] proposes a fault diagnosis strategy for rotating machinery based on CNN using infrared thermal images. Ref. [116] integrates an attention mechanism into CNN to efficiently extract the fault features of analog circuit. Similarly, Ref. [117] use an encoder–decoder-like CNN to find cracks on device surface in complex background. The diagnosis of such image data generally can hardly achieve precise quantitative description of the faults, it can usually only obtain the qualitative trend of the device faults.

5.3.2. Conversion from Other Sensory Data into Images

Usually the monitoring variable observed by the sensor is a one-dimensional signal, which is different from a two-dimensional image. In order to leverage the powerful feature learning ability of CNN, many researchers consider converting one-dimensional signals into two-dimensional images, and then input them into CNN for classification or recognition. For example, Ref. [104] propose an intelligent fault diagnosis method for aeroengine sensors combining a CNN with time-frequency analysis wherein the signal recognition problem is transformed into an image-recognition problem. An example pipeline is illustrated in Figure 9. Many of these work puts their main focus on how to convert to two-dimensional

images. Common methods include wavelet transform [102,104,108], S-transform [118], phase space reconstruction [119], etc. These two-dimensional time-frequency distribution images generated by transformation often have simpler backgrounds than natural images. The quality of these transformation methods directly affects the performance of CNN. If there is little difference between the two-dimensional images of fault and non-fault signals, the accuracy of CNN classification will also be unsatisfactory.

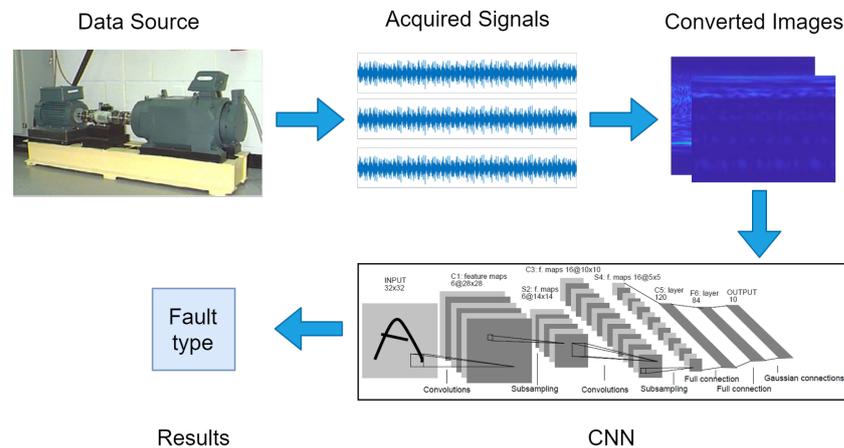


Figure 9. A typical fault diagnosis pipeline based on signal-to-image conversion and CNN.

5.3.3. 1-D CNN for Signal Processing

Actually, two-dimensional convolution operations can also be decomposed into two one-dimensional convolutions vertically and horizontally. Therefore, another attempt is that tries to fit two-dimensional CNN to one-dimensional data, i.e., 1-D CNN [54,113], which is specialized for temporal signals [120]. This operation is inherently suitable for sensory data, and has been widely used for intelligent FDP in recent years. For example, Ref. [121] presents a 1-D CNN-based approach to automatically learn features for rub-impact fault diagnosis from the raw vibration signals of a rotor system, and [114] establish a fault identification model based on the powerful feature extraction and complex data analysis abilities of 1D-CNN. Due to its inherent properties, many modern techniques for 2-D CNN can be imported into 1-D CNN for better signal feature extraction, such as attention [112], lightweight design [122], and dilated convolution [123].

Although CNN has provided an alternative way to process different types of condition monitoring data, there are still limitations. Firstly, the conversion from signal data to image is equivalent to the quantization process of imaging, which means that important details of signal intensity can be naturally omitted when projecting to pixel bins. In this way, subtle abnormality in the early stages can easily be ignored by convolution and pooling operations. Lastly, the methods for conversion should also be carefully designed to prevent overfitting. Furthermore, it is also a challenge for CNN-based FDP methods to achieve real-time diagnosis since they are with relatively high computational overheads for image data.

5.4. Transformer for Self-Attention Feature Extraction

Initially designed in natural-language processing, attention mechanism is a technique that can model sequence dependencies, which allow a model to focus only on a set of elements and to decompose a problem into a sequence of attention-based reasoning tasks [124,125]. The attention mechanism now has been adopted in various deep learning architectures, such as CNNs and RNNs. Transformer architecture [126] abandons all the recurrent and convolutional structures, and only contains multi-head self-attention (MSA), multi-layer perceptron (MLP), and a basic fully connected layer [127] to capture the long-term dependencies between elements in a sequence without considering their distance, which can consider the global information comprehensively.

In Figure 10, we illustrate an example of fault diagnosis pipeline using transformer. The captured signals are firstly cropped into signal subsequences according to their original positions, which is then mapped into a high-dimensional vector through linear embedding and followed by trainable position encoding to retain the position information of the signal. Vectors are then fed into multiple stacked transformer blocks for long-distance modeling through layer-normalized MSAs and MLPs. Finally, the extracted features are input into the MLP head, i.e., fully-connected layer, for the classification results. Common loss functions for other classification tasks are also used.

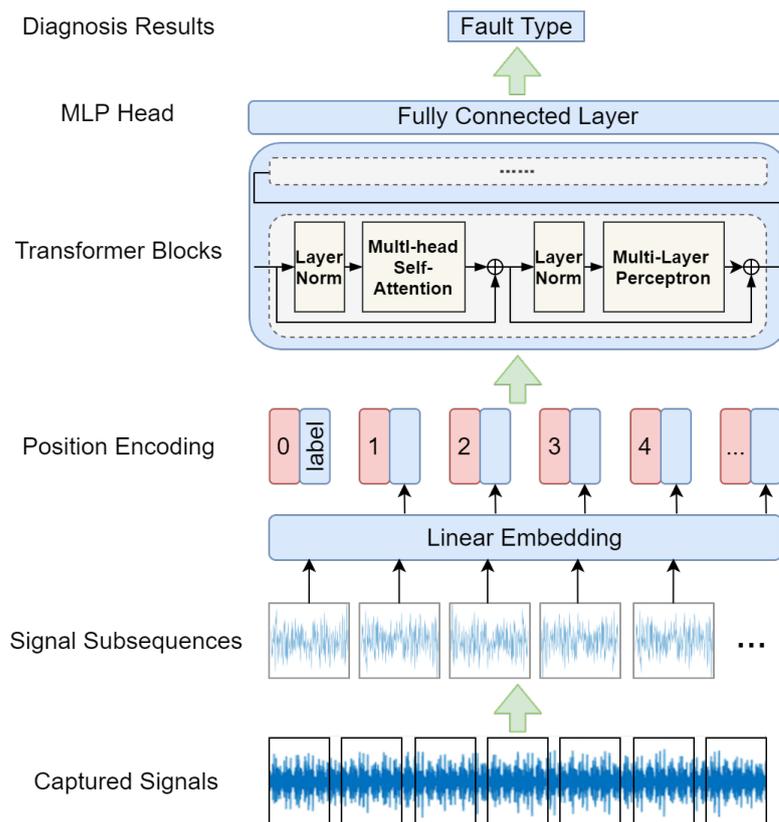


Figure 10. An example pipeline of transformer-based fault diagnosis.

Due to the outstanding global information modeling ability, transformer has outperformed other architectures in feature extraction for many tasks, and is a hot research topic of FDP in these two years. Ref. [127] proposes a time-series transformer which utilizes raw vibration signals for the rotating machinery fault diagnosis, and it tries to capture translation invariance and long-term dependencies with a new time-series tokenizer. Different from [127], Ref. [128] designs a time-frequency transformer with a fresh tokenizer and encoder module to extract effective abstractions from the time–frequency representation of vibration signals. Ref. [36] use an integrated vision transformer (ViT) based on the soft voting fusion method to diagnose the bearing fault with high accuracy and generalization. For RUL prediction, Ref. [129] propose a transformer-based encoder–decoder structure with a dual-aspect encoders design to extract features from the sensor and time step simultaneously, while adaptively learning to focus on more important part of input and processing long data sequences.

Some recent work of these two years for intelligent FDP based on a transformer are given in Table 6. As can be observed, transformer-based FDP methods are gradually being used as excellent feature extractors and for time-series data processing, due to their outstanding performance in modeling long-distance information in input data, compared with CNNs and RNNs.

Table 6. Some of recent intelligent FDP methods based on a transformer.

Type	Reference	Year	Method	Object
Fault diagnosis	[130]	2021	Linear embedding sequence of signal patches is used as an input to a Transformer encoder, CNN is used as decoder and classifier.	bearing and gearbox datasets
	[128]	2022	A time-frequency Transformer model with a new tokenizer and encoder module to extract effective abstractions from the time-frequency representation of vibration signals.	bearing
	[131]	2022	The weight parameters of self-extracted features of SPBO-SDAE network are optimized through the self-attention mechanism of transformer to retain the target features and filter the redundant features.	rotating machinery
	[132]	2022	A lightweight transformer based on convolutional embedding and linear self-attention to deal with the challenges of limited samples, noise interference, and lightweight.	rotating machinery
	[133]	2022	Convformer-NSE to extract robust features that integrate both global and local information under heavy noise.	gearbox systems
	[127]	2022	Time series transformer with a tokens sequences generation method handling data in 1D format.	rotating machinery
	[134]	2022	Transformer is built to extract temporal features.	electromagnetic systems
Fault prediction	[135]	2022	Transformer architecture is employed to diagnose the simultaneous faults with time-series data.	on-site air handling unit
	[136]	2021	As a variant of transformer, Informer is used for Long sequence time-series prediction.	nuclear power valves
	[137]	2022	Informer is introduced to solve the problem of error accumulation caused by the conventional methods of time series forecasting of motor bearing vibration.	bearing
RUL prediction	[138]	2022	A self-attention module is designed by adopting the attention mechanism into ConvLSTM cell to focus on the degraded data that is beneficial to the prediction result, and suppressing less useful ones.	bearing
	[139]	2022	Convolutional transformer combines the global context capturing of attention mechanism with the local dependencies modeling of convolutional operation	bearing

Owing to the ability of long-range modeling of data, its side-effect is that its local information modeling ability is relatively lower than CNNs and RNNs, and there are also attempts to make up the shortcoming through combining transformer with CNN or RNN. The second limitation is its computational efficiency because of its special structure, and it is undoubtedly the current hot spot for DL community. However, then again, there is still much to be further explored on this topic.

5.5. Graph Neural Network (GNN) for Relationship Modeling

Although the above deep learning techniques can effectively capture the hidden features or model the inherent knowledge from input data in an end-to-end way, most of them ignore the inter-dependencies between data or various physical measurements of multiple sensors [140]. Since [141] first applied neural networks to directed acyclic graphs, graph neural networks (GNN) have successfully handled data characterized by complex spatiotemporal relationships [142]. Although deep learning effectively captures the hidden patterns in Euclidean domains, more data are generated from non-Euclidean domains and represented as graphs with complex spatiotemporal relationships among objects. GNN tries to model the relationships with graph representations, i.e., feature node and adjacency edge, and concentrate on the tasks of node classification (node level), edge classification and link prediction (edge level), and graph classification (graph level) [140,143]. GNN can be integrated with other architectures and extended to graph convolutional networks (GCNs) [144], graph attention networks (GATs) [145], graph autoencoders (GAEs) [146], etc.

A graph structure in GNN can be generally represented by a node feature matrix, an adjacency matrix and a set of weighted edges. It can propagate the node information through the edges of a graph via graph operations, such as graph convolutions, and learn a promising node or graph representations. The most commonly used GNN is GCN, and many operations in GCN can find their similar counterparts in CNN, such as convolutions

on nodes to aggregate the information of connected neighbor nodes along the weighted edges, Relu function for non-linear activation and pooling layer to reduce dimensions, though there are very small differences in operations in practice.

Owing to the capability to model relationships in data, GNN has been receiving attentions from researchers in the FDP community recently, and the challenges faced in FDP are the appropriate way of constructing and realizing the graph [142]. Figure 11 gives an example diagnosis pipeline based on GCN. Similarly, [144] present a GCN-based fault diagnosis method that uses a association graph constructed from prediagnostic results and adjust the graph via using a hybrid of measurements and prior knowledge, which obtained good diagnosis results. When dealing with time-series data, the work in [140] constructs three kinds of graphs for fault diagnosis and prognosis according to the time-series subsample types as univariate and multivariate data, respectively. Ref. [147] proposes an interaction-aware GNN for fault diagnosis of complex industrial process, which transforms sensor signals into a heterogeneous graph with multiple edge types and employ a GNN to extract fault feature of one edge type, so it can learn implicit interactions between sensor signals.

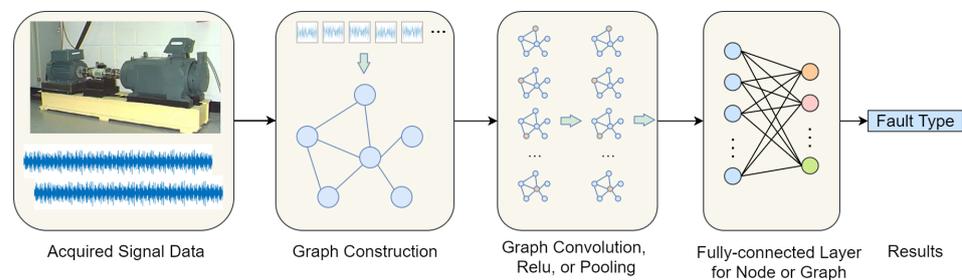


Figure 11. An example pipeline of fault diagnosis using GCN.

In Table 7, more recent GNN-based intelligent FDP methods are listed for the references of readers. It can be observed that GNNs has a high popularity in the last two years. On the basis of knowledge graph, GNN is recognized to reason or infer knowledge, which realizes the promotion from perception to cognition of AI. As a result, at current stage of research, the explicit incorporation of (prior) knowledge for constructing graphs in GNN instead of currently using a large amount of training data, and more generalized knowledge inference are desired and beneficial for FDP. GNN is expected to show greater potential in subsequent studies for intelligent industrial FDP.

Table 7. Some of recent intelligent FDP methods based on GNNs.

Type	Reference	Year	Method	Object
GCN	[148]	2020	A deep GCN based on graph theory transforms data into graphs of geometric structures with weights representing the similarity between connected vertices.	roller bearings
	[149]	2021	Semi-supervised GCN constructs all samples into an undirected and weighted k-nearest neighbor graph, which is trained using both labeled and unlabeled samples.	rotating machinery
	[150]	2021	GCN incorporates the weighted horizontal visibility graph to transform time series to graph data, and uses graph isomorphism network to learn the graph representation and perform fault classification.	bearing
	[151]	2021	GCN decomposes signals to present frequency feature as graph and extract the features of points with a large span of the defined graph samples.	wind turbine
	[144]	2021	A structure analysis-based GCN integrates the measurement and the prior knowledge of the system of interest and introduces a weight coefficient to adjust their influence.	rectifier
	[152]	2022	Multi-scale cluster-GCN is proposed to learn the representation feature extracted by AE layer.	gearbox and bearing
	[153]	2022	Edge connections of the input static graph are updated according to the relationship among high-level features extracted by GCN.	rotating machinery

Table 7. Cont.

Type	Reference	Year	Method	Object
GAT	[145]	2021	A semi-supervised conditional random field-based GAT learns the effective node representations and models the label dependency through assigning adaptive weights to different neighbors.	motor
	[154]	2022	A triplet metric driven multi-head GNN combines deep metric learning and improves triplet loss to convert signals into graph structure, and introduces multi-head attention to reduce interference of heterogeneous vertices.	rolling bearing
GAE	[146]	2022	Graph dynamic AE uses graph convolution to avoid the dimensionality increase problem of classic dynamic methods, and a weighted adjacency matrix to adaptively assign weights to the temporal samples.	Tennessee Eastman process
	[155]	2022	Sparse AE and GNN are combined to effectively capture inter-dependencies in high-dimensional sensor data with few anomalies.	cyber-physical systems

6. Challenges and Possible Solutions

This paper has provided a systematic literature review of deep learning based intelligent industrial FDP. It can be concluded that there are a lot of interest in using CNN, DBN, or RNN for fault diagnosis purposes, but when architectures develop, more complicated but powerful methods have been introduced into FDP. GNN, Transformer, and GAN are gradually receiving attention and their performance has also begun to surpass traditional methods. Although the deep learning methods have been applied in the intelligent FDP of industrial systems, there are still several challenges that need to be explored and solved. In this section, we analyze the open challenges from the four aspects of data imbalance, compound fault type, multimodal data fusion, and edge device implementation, and provide possible solutions.

6.1. Imbalance Problem in Industrial Applications

In practical industrial applications, the acquisition of typical data (including historical health data, fault data, and simulation data) of some devices is usually expensive, labor-intensive, and sometimes impossible [156]. Even if the state data of the system can be acquired, it often has strong uncertainty and incompleteness, these problems increase the difficulty of FDP. At present, the total amount of existing data can only support the implementation of traditional methods or shallow machine learning methods. It is still a challenge to train robust intelligent FDP models with limited data and works well under complex working conditions. The second problem [157] is the imbalance data that (1) there are too few fault samples and too much duplicated normal data samples; and (2) there is an open set of fault modes that many of the modes may not be encountered in operation.

One possible way is to run long-term laboratory tests or simulation for every single device and the whole system, in order to simulate various working conditions in the laboratory, and find all possible fault modes of devices and the system. However, obtaining complete fault data of the entire system sometimes is expensive and infeasible [156]. In terms of intelligent FDP techniques, it could be solved from the following aspects.

6.1.1. Task-Level Transfer Learning

Despite the imbalance in local systems, there are a large number of similar devices or subsystems in other industrial, mechanical, power grid systems, etc. These devices and subsystems share the similar architecture or composition, and they have accumulated a certain amount of historical health data. The utilization of these large amounts of useful data or knowledge from other systems for the FDP of local industrial system, i.e., task-level transfer learning, becomes an efficient and promising approach. It emphasizes the transformation data, feature, knowledge or model to different fields. At present, transfer learning-based methods have been implemented in other fields such as image recognition, and several pioneering work has been completed for intelligent FDP. Ref. [158] adopt the

knowledge transfer scheme and use a multi-input multi-output convolutional network to extract domain-invariant feature representations and classifiers from the labeled dataset from scientific test rigs and the unlabeled dataset from industrial application to be tested.

6.1.2. Data-Level Augmentation

One direct way is to generate more balanced/diverse data to enhance the training sets for FDP models. Traditional data augmentation through transformations, such as translation, deformation, and scaling, has low computational cost and is easy to implement, which is a simple and efficient way to generate a large amount of labeled samples to improve FDP performance with limited data. However, the generated samples can be considered as local distortions of existing labeled sample points in high-dimensional space, i.e., they are still with limited diversity. GANs offer a good option to generate more realistic or vivid data samples with the same original data distribution of minor fault patterns for both 2-D image data and 1-D timer-series signal data, as we have analyzed in Section 4.2.

6.1.3. Model-Level Meta-Learning

Meta-learning is a flexible framework which can learn to obtain the ability of extracting meta-knowledge from multiple relevant tasks to gain generalization on various tasks, in order to guide the learning and improve its performance on target tasks without training from scratch [159,160]. Currently, the studies of model-level meta-learning for intelligent FDP with imbalanced data are still in their earlier stages. Some work [159], mostly based on metric-based meta-learning, has explored its implementation in industrial FDP, and shown excellent accuracy and robustness on public datasets. However, it needs further development and verification in operational industrial systems.

6.2. Lifting Diagnosis from Single Faults to Compound Faults

Most of the modern deep learning-based intelligent FDP methods are only applied in the single-fault diagnosis. However, in actual complex industrial systems, several kinds of single faults may exist simultaneously, which means several components or devices may break down together, resulting in compound-fault modes [103]. Usually, these faults are related to each other and affect each other at the same time. The signals captured by sensors may be coupled with multiple fault signals, and the generic FDP methods that work for one single fault will inevitably fail in compound-fault modes. In addition, the compound-fault samples are also difficult to collect and label, which further limits the application of the existing deep learning-based methods [161]. In operational complex industrial systems, compound faults are generally more dangerous and harmful than a single fault [162]. It has, therefore, become a key issue to be solved for complex industrial systems.

Traditional compound-fault-diagnosis methods rely heavily on either prior knowledge inference or signal analysis [161], which is difficult to be applied in operational industrial systems. Identifying and decoupling the compound fault are still a great challenge for intelligent FDP. The effective separation of fault characteristic components is the core of compound-fault diagnosis [163]. Ref. [103] uses a multi-label CNN to achieve compound fault diagnosis based on the 2-D time-frequency features in an end-to-end way. Ref. [164] propose a deep ensemble capsule network that combines multiple decoupling capsule network individually trained on one sensory data in a way of ensemble learning to effectively decouple the compound fault into individual faults. In [162], a decoupling classifier is designed to decouple the compound fault into single faults by outputting multiple labels for samples.

Considering that the compound-fault-sample data are always scarce, it is also important to use the single fault data to train the compound-fault decoupling model with the help of the knowledge learned from the single fault mode data. The decoupling classifier in [162] is trained on a dataset only containing normal and single fault samples. To address the problem of identifying unknown compound faults, Ref. [161] present a zero-shot learning model which classifies the compound faults according to the similarity measure between the signal features

and the semantic features of the compound faults to identify the categories of unknown compound faults. Actually, the scarce of compound fault samples is a key issue to improve the practicability of the intelligent compound-fault-diagnosis methods.

6.3. Boosting Intelligent FDP with Multimodal Fusion

On one hand, an individual sensor can hardly provide the complementary and thorough information of complex industrial devices, and various signal transfer paths from the fault point to the location of sensor, so it is necessary to place several sensors at different places to capture more comprehensive and accurate information for the faults [164]. Therefore, in industrial systems, there are always multisensory data used for intelligent FDP. In recent years, intelligent FDP based on the fusion of multi-source homogeneous information has been thoroughly explored and discussed. On the other hand, a fault can be reflected in several relevant sensors with heterogeneous platforms simultaneously, such as current, voltage, temperature, etc. The fusion of sensory data from heterogeneous platforms, i.e., multimodal fusion, is for the purpose that complementary information could be extracted from each modality, thus yielding a richer representation that could be used to achieve higher-quality intelligent FDP, compared to using only a single modality [165]. The efficient fusion of multimodal sensory data remains challenging for the community.

Early stage of multimodal fusion mainly are at data-level, i.e., representing the fused data in a lower-dimensional subspace, in which principal component analysis is commonly used. It is then extended to feature-based fusion that features extracted from each model for each modality is fused, and decision-based fusion which makes a weighted fusion decision for the outputs of those models [166]. For example, [167] use a coupling AE to find a joint feature between vibration and acoustic signals for health-state classification, and [168] propose to extract the multiscale features of vibration and torque signals through a three-stage feature fusion method for the fault diagnosis of bearings. In [169], a multimodal decision-fusion model is built to achieve comprehensive fault diagnosis for rotor-bearing systems.

As can be observed in the related literatures of multimodal fusion for intelligent FDP, current modalities used mostly are derived from similar mechanisms, such as acceleration signals and acoustic signals formed by vibration, and voltage and current signals formed by electronics. They are generally with the same data representation and can easily be fused through data transformations. The modalities derived from different mechanisms are merely used, for example the fusion of vibration signals and 2-D images, temperature signals and current signals, or even text descriptions and images. Therefore, there is still room for the fusion of these modalities to boost the performance and applicability of intelligent FDP in complex industrial systems.

6.4. Intelligent FDP Acceleration for Edge Implementation

Industrial IoT and AI have been playing highly significant roles in modern industrial systems, more and more sensors are installed, generating massive amounts of sensory data. With the increase in data scale, the response delay of data transmission and calculation cannot be guaranteed, which brings great challenges to the computing center-based industrial systems. Moreover, modern, intelligent FDP algorithms based on deep learning are generally computationally intensive, i.e., with huge parameters and deep architectures.

To tackle this problem, an emerging computing paradigm, edge computing, has been widely recognized as a promising solution [170]. In the edge computing paradigm, model training is performed by the center, and models are deployed and runs on the edge nodes, such as gateway, smart devices, and the way of bringing data and computation closer to where data are produced can help to save the response time and bandwidth, as well as energy consumption [171].

However, edge ends are always constrained by resources, which means their power supply and computing capability are limited and heavy deep learning models can hardly adapt to these platforms. Therefore, it brings great challenges to the intelligent FDP algorithms in turn. Models that are computationally lightweight and of high accuracy are

preferred for the edge implementation [172]. In the field of computer vision, the lightweight design of deep learning models has been a hot research spot for edge implementation, and typical methods are network pruning [173] and knowledge distillation [174]. Currently, some pioneer work [175,176] has been conducted and shown promising results for intelligent FDP on edge ends.

7. Conclusions

The diagnosis and prognosis of faults are important for the operation of industrial systems. This paper mainly reviews the development of deep learning techniques in intelligent FDP for industrial systems. The tasks of fault diagnosis and fault prognosis are firstly defined mathematically. An overview of deep learning architectures that are commonly used for intelligent FDP are then summarized. To be specific, the architectures of DBN, CNN, AE, RNN, GAN, Transformer, and GNN are introduced, along with their applications. Finally, we prospect four future directions from the aspects of data imbalance, compound fault type, multimodal data fusion, and edge implementation, and possible solutions are also provided. This survey is expected to comprehensively present the development of deep learning techniques used in intelligent FDP for industrial systems and provide possible guidelines for the research in the community.

Early detection, isolation, and identification of different faults enabled with DL techniques will help to greatly improve the efficiency, reliability, and repeatability of industrial systems. With the fast development and evolution of DL and related techniques, in near future many fundamental problems, such as the mentioned open challenges, are very likely to be addressed. As for the research trends, the borders between different DL architectures are being broken down and a hybrid architecture that takes both advantages is expected to produce better flexibility and performance. In addition, physics-informed DL techniques based on the physical characteristics and related physical models of the industrial system will be an important future direction.

Author Contributions: Conceptualization, S.Q., X.C., Z.P., and X.B.; investigation and analysis, S.Q., N.S., Z.L. and Z.P.; writing—original draft preparation, S.Q., X.C., N.S., Z.L. and X.B.; visualization, S.Q., X.B. and Z.P.; supervision, S.Q. and X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 41901376, Hubei Provincial Natural Science Foundation of China under Grant 2022CFB989, the Foundation for the National Key Laboratory of Science and Technology under Grant 6142217210503 and 614221720190507, and the Project Foundation of University (NUE) under Grant 202250E050.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xu, Y.; Sun, Y.; Wan, J.; Liu, X.; Song, Z. Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications. *IEEE Access* **2017**, *5*, 17368–17380. [\[CrossRef\]](#)
2. Dash, S.; Venkatasubramanian, V. Challenges in the industrial applications of fault diagnostic systems. *Comput. Chem. Eng.* **2000**, *24*, 785–791. [\[CrossRef\]](#)
3. Zio, E. Prognostics and health management of industrial equipment. In *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques*; IGI-Global: Hershey, PA, USA, 2013; pp. 333–356.
4. Tumer, I.; Bajwa, A. A survey of aircraft engine health monitoring systems. In Proceedings of the 35th Joint Propulsion Conference and Exhibit, Los Angeles, CA, USA, 20–24 June 1999; p. 2528.
5. Zhong, K.; Han, M.; Han, B. Data-driven based fault prognosis for industrial systems: A concise overview. *IEEE/CAA J. Autom. Sin.* **2019**, *7*, 330–345. [\[CrossRef\]](#)
6. Tsui, K.L.; Chen, N.; Zhou, Q.; Hai, Y.; Wang, W. Prognostics and health management: A review on data driven approaches. *Math. Probl. Eng.* **2015**, *2015*, 793161. [\[CrossRef\]](#)

7. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
8. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
9. NVIDIA. NVIDIA Technologies and GPU Architectures. Available online: <https://www.nvidia.com/en-us/technologies/> (accessed on 4 September 2022).
10. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef]
11. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
12. Lasi, H.; Fettke, P.; Kemper, H.G.; Feld, T.; Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **2014**, *6*, 239–242. [CrossRef]
13. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3137–3147. [CrossRef]
14. Boyes, H.; Hallaq, B.; Cunningham, J.; Watson, T. The industrial internet of things (IIoT): An analysis framework. *Comput. Ind.* **2018**, *101*, 1–12. [CrossRef]
15. Marino, R.; Wisultschew, C.; Otero, A.; Lanza-Gutierrez, J.M.; Torre, E. A Machine-Learning-Based Distributed System for Fault Diagnosis With Scalable Detection Quality in Industrial IoT. *IEEE Internet Things J.* **2021**, *8*, 4339–4352. [CrossRef]
16. Shen, S.; Lu, H.; Sadoughi, M.; Hu, C.; Kenny, S. A physics-informed deep learning approach for bearing fault detection. *Eng. Appl. Artif. Intell.* **2021**, *103*, 104295. [CrossRef]
17. Miao, H.; He, D. Deep Learning Based Approach for Bearing Fault Diagnosis. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3057–3065.
18. Hey, A.J.; Tansley, S.; Tolle, K.M. *The Fourth Paradigm: Data-Intensive Scientific Discovery*; Microsoft Research: Redmond, WA, USA, 2009; Volume 1.
19. Li, D.; Wang, Y.; Wang, J.; Wang, C.; Duan, Y. Recent advances in sensor fault diagnosis: A review. *Sensors Actuators A Phys.* **2020**, *309*, 111990. [CrossRef]
20. Fernandes, M.; Corchado, J.M.; Marreiros, G. Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: A systematic literature review. *Appl. Intell.* **2022**, *52*, 14246–14280. [CrossRef]
21. Zhang, S.; Su, L.; Gu, J.; Li, K.; Zhou, L.; Pecht, M. Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey. *Chin. J. Aeronaut.* **2021**. [CrossRef]
22. Lv, H.; Chen, J.; Pan, T.; Zhang, T.; Feng, Y.; Liu, S. Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application. *Measurement* **2022**, *199*, 111594. [CrossRef]
23. Zhu, J.; Jiang, Q.; Shen, Y.; Qian, C.; Xu, F.; Zhu, Q. Application of recurrent neural network to mechanical fault diagnosis: A review. *J. Mech. Sci. Technol.* **2022**, *36*, 527–542. [CrossRef]
24. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
25. Liang, J.; Zhang, K.; Al-Durra, A.; Muyeen, S.M.; Zhou, D. A state-of-the-art review on wind power converter fault diagnosis. *Energy Rep.* **2022**, *8*, 5341–5369. [CrossRef]
26. Hu, X.; Zhang, K.; Liu, K.; Lin, X.; Dey, S.; Onori, S. Advanced Fault Diagnosis for Lithium-Ion Battery Systems: A Review of Fault Mechanisms, Fault Features, and Diagnosis Procedures. *IEEE Ind. Electron. Mag.* **2020**, *14*, 65–91. [CrossRef]
27. Tang, S.; Yuan, S.; Zhu, Y. Deep Learning-Based Intelligent Fault Diagnosis Methods Toward Rotating Machinery. *IEEE Access* **2020**, *8*, 9335–9346. [CrossRef]
28. Lei, Y.; Li, N.; Li, X. *Big Data-Driven Intelligent Fault Diagnosis and Prognosis for Mechanical Systems*; Springer: Singapore, 2023; p. 281.
29. Jain, A.K.; Duin, R.P.W.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37. [CrossRef]
30. Zhao, B.; Zhang, X.; Zhan, Z.; Wu, Q. Deep multi-scale adversarial network with attention: A novel domain adaptation method for intelligent fault diagnosis. *J. Manuf. Syst.* **2021**, *59*, 565–576. [CrossRef]
31. Xie, J.; Du, G.; Shen, C.; Chen, N.; Chen, L.; Zhu, Z. An end-to-end model based on improved adaptive deep belief network and its application to bearing fault diagnosis. *IEEE Access* **2018**, *6*, 63584–63596. [CrossRef]
32. Mao, W.; Feng, W.; Liu, Y.; Zhang, D.; Liang, X. A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mech. Syst. Signal Process.* **2021**, *150*, 107233. [CrossRef]
33. Huang, Y.C.; Wang, P.J. Infrared Air Turbine Dental Handpiece Rotor Fault Diagnosis with Convolutional Neural Network. *Sens. Mater.* **2020**, *32*, 3545–3558. [CrossRef]
34. Wu, Q.; Ding, K.; Huang, B. Approach for fault prognosis using recurrent neural network. *J. Intell. Manuf.* **2020**, *31*, 1621–1633. [CrossRef]
35. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
36. Tang, X.; Xu, Z.; Wang, Z. A Novel Fault Diagnosis Method of Rolling Bearing Based on Integrated Vision Transformer Model. *Sensors* **2022**, *22*, 3878. [CrossRef]

37. Li, T.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Multireceptive field graph convolutional networks for machine fault diagnosis. *IEEE Trans. Ind. Electron.* **2020**, *68*, 12739–12749. [\[CrossRef\]](#)
38. Shao, H.; Xia, M.; Wan, J.; de Silva, C.W. Modified Stacked Autoencoder Using Adaptive Morlet Wavelet for Intelligent Fault Diagnosis of Rotating Machinery. *IEEE-ASME Trans. Mechatronics* **2022**, *27*, 24–33. [\[CrossRef\]](#)
39. Ma, S.; Chen, M.; Wu, J.; Wang, Y.; Jia, B.; Jiang, Y. High-Voltage Circuit Breaker Fault Diagnosis Using a Hybrid Feature Transformation Approach Based on Random Forest and Stacked Autoencoder. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9777–9788. [\[CrossRef\]](#)
40. He, Z.; Shao, H.; Ding, Z.; Jiang, H.; Cheng, J. Modified Deep Autoencoder Driven by Multisource Parameters for Fault Transfer Prognosis of Aeroengine. *IEEE Trans. Ind. Electron.* **2022**, *69*, 845–855. [\[CrossRef\]](#)
41. Xu, F.; Tse, W.T.P.; Tse, Y.L. Roller bearing fault diagnosis using stacked denoising autoencoder in deep learning and Gath-Geva clustering algorithm without principal component analysis and data label. *Appl. Soft Comput.* **2018**, *73*, 898–913. [\[CrossRef\]](#)
42. Miao, M.; Sun, Y.; Yu, J. Sparse Representation Convolutional Autoencoder for Feature Learning of Vibration Signals and its Applications in Machinery Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2022**, *69*, 13565–13575. [\[CrossRef\]](#)
43. Remadna, I.; Terrissa, L.S.; Al Masry, Z.; Zerhouni, N. RUL Prediction Using a Fusion of Attention-Based Convolutional Variational AutoEncoder and Ensemble Learning Classifier. *IEEE Trans. Reliab.* **2022**. [\[CrossRef\]](#)
44. Shen, C.; Qi, Y.; Wang, J.; Cai, G.; Zhu, Z. An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive. *Eng. Appl. Artif. Intell.* **2018**, *76*, 170–184. [\[CrossRef\]](#)
45. Yang, Z.; Baraldi, P.; Zio, E. A method for fault detection in multi-component systems based on sparse autoencoder-based deep neural networks. *Reliab. Eng. Syst. Saf.* **2022**, *220*, 108278. [\[CrossRef\]](#)
46. Chen, Z.; Li, W. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1693–1702. [\[CrossRef\]](#)
47. Yu, J.; Zhou, X. One-Dimensional Residual Convolutional Autoencoder Based Feature Learning for Gearbox Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6347–6358. [\[CrossRef\]](#)
48. Sun, M.; Wang, H.; Liu, P.; Huang, S.; Fan, P. A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings. *Measurement* **2019**, *146*, 305–314. [\[CrossRef\]](#)
49. Zheng, S.; Farahat, A.; Gupta, C. Generative adversarial networks for failure prediction. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 621–637.
50. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [\[CrossRef\]](#)
51. Zhao, Z.; Li, B.; Dong, R.; Zhao, P. A surface defect detection method based on positive samples. In *Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, Nanjing, China, 28–31 August 2018; Volume 11013, pp. 473–481. [\[CrossRef\]](#)
52. Pan, T.; Chen, J.; Zhang, T.; Liu, S.; He, S.; Lv, H. Generative adversarial network in mechanical fault diagnosis under small sample: A systematic review on applications and future perspectives. *ISA Trans.* **2022**, *128*, 1–10. [\[CrossRef\]](#)
53. Liu, S.; Jiang, H.; Wu, Z.; Li, X. Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. *Mech. Syst. Signal Process.* **2022**, *163*, 108139. [\[CrossRef\]](#)
54. Guo, Q.; Li, Y.; Song, Y.; Wang, D.; Chen, W. Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network. *IEEE Trans. Ind. Informatics* **2019**, *16*, 2044–2053. [\[CrossRef\]](#)
55. Li, F.; Tang, T.; Tang, B.; He, Q. Deep convolution domain-adversarial transfer learning for fault diagnosis of rolling bearings. *Measurement* **2021**, *169*, 108339. [\[CrossRef\]](#)
56. Liu, J.; Qu, F.; Hong, X.; Zhang, H. A Small-Sample Wind Turbine Fault Detection Method With Synthetic Fault Data Using Generative Adversarial Nets. *IEEE Trans. Ind. Inform.* **2019**, *15*, 3877–3888. [\[CrossRef\]](#)
57. Shao, S.; Wang, P.; Yan, R. Generative adversarial networks for data augmentation in machine fault diagnosis. *Comput. Ind.* **2019**, *106*, 85–93. [\[CrossRef\]](#)
58. Chen, Z.; He, G.; Li, J.; Liao, Y.; Gryllias, K.; Li, W. Domain Adversarial Transfer Network for Cross-Domain Fault Diagnosis of Rotary Machinery. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8702–8712. [\[CrossRef\]](#)
59. Gao, Y.; Liu, X.; Huang, H.; Xiang, J. A hybrid of FEM simulations and generative adversarial networks to classify faults in rotor-bearing systems. *ISA Trans.* **2021**, *108*, 356–366. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Viola, J.; Chen, Y.; Wang, J. FaultFace: Deep Convolutional Generative Adversarial Network (DCGAN) based Ball-Bearing failure detection method. *Inf. Sci.* **2021**, *542*, 195–211. [\[CrossRef\]](#)
61. Lu, H.; Barzegar, V.; Nemani, V.P.; Hu, C.; Laflamme, S.; Zimmerman, A.T. Joint training of a predictor network and a generative adversarial network for time series forecasting: A case study of bearing prognostics. *Expert Syst. Appl.* **2022**, *203*, 117415. [\[CrossRef\]](#)
62. Peng, Y.; Wang, Y.; Shao, Y. A novel bearing imbalance Fault-diagnosis method based on a Wasserstein conditional generative adversarial network. *Measurement* **2022**, *192*, 110924. [\[CrossRef\]](#)
63. Pan, T.; Chen, J.; Xie, J.; Chang, Y.; Zhou, Z. Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples. *ISA Trans.* **2020**, *101*, 379–389. [\[CrossRef\]](#)
64. Feng, Y.; Liu, Z.; Chen, J.; Lv, H.; Wang, J.; Yuan, J. Make the Rocket Intelligent at IoT Edge: Stepwise GAN for Anomaly Detection of LRE With Multisource Fusion. *IEEE Internet Things J.* **2022**, *9*, 3135–3149. [\[CrossRef\]](#)

65. Pu, Z.; Cabrera, D.; Bai, Y.; Li, C. A One-Class Generative Adversarial Detection Framework for Multifunctional Fault Diagnoses. *IEEE Trans. Ind. Electron.* **2022**, *69*, 8411–8419. [[CrossRef](#)]
66. Peng, K.; Jiao, R.; Dong, J.; Pi, Y. A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter. *Neurocomputing* **2019**, *361*, 19–28. [[CrossRef](#)]
67. Zhang, Y.; Ji, J.; Ma, B. Fault diagnosis of reciprocating compressor using a novel ensemble empirical mode decomposition-convolutional deep belief network. *Measurement* **2020**, *156*, 107619. [[CrossRef](#)]
68. Zhang, Y.; Ji, J.; Ma, B. Reciprocating compressor fault diagnosis using an optimized convolutional deep belief network. *J. Vib. Control* **2020**, *26*, 1538–1548. [[CrossRef](#)]
69. Yu, J.; Liu, G. Knowledge extraction and insertion to deep belief network for gearbox fault diagnosis. *Knowl.-Based Syst.* **2020**, *197*, 105883. [[CrossRef](#)]
70. Chen, Z.; Chen, X.; Li, C.; Sanchez, R.V.; Qin, H. Vibration-based gearbox fault diagnosis using deep neural networks. *J. Vibroengineering* **2017**, *19*, 2475–2496. [[CrossRef](#)]
71. Jiang, G.; Zhao, J.; Jia, C.; He, Q.; Xie, P.; Meng, Z. Intelligent Fault Diagnosis of Gearbox Based on Vibration and Current Signals: A Multimodal Deep Learning Approach. In Proceedings of the 10th IEEE Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China, 25–29 October 2019; pp. 1–6.
72. Chen, Z.; Li, C.; Sanchez, R.V. Multi-layer neural network with deep belief network for gearbox fault diagnosis. *J. Vibroeng.* **2015**, *17*, 2379–2392.
73. He, X.; Ma, J. Weak fault diagnosis of rolling bearing based on FRFT and DBN. *Syst. Sci. Control. Eng.* **2020**, *8*, 57–66. [[CrossRef](#)]
74. Zhao, X.; Jia, M. A new Local-Global Deep Neural Network and its application in rotating machinery fault diagnosis. *Neurocomputing* **2019**, *366*, 215–233. [[CrossRef](#)]
75. Gao, S.; Xu, L.; Zhang, Y.; Pei, Z. Rolling bearing fault diagnosis based on intelligent optimized self-adaptive deep belief network. *Meas. Sci. Technol.* **2020**, *31*, 055009. [[CrossRef](#)]
76. Gao, S.; Xu, L.; Zhang, Y.; Pei, Z. Rolling bearing fault diagnosis based on SSA optimized self-adaptive DBN. *ISA Trans.* **2021**, *128*, 485–502. [[CrossRef](#)]
77. Zhang, C.; He, Y.; Jiang, S.; Wang, T.; Yuan, L.; Li, B. Transformer Fault Diagnosis Method Based on Self-Powered RFID Sensor Tag, DBN, and MKSVM. *IEEE Sens. J.* **2019**, *19*, 8202–8214. [[CrossRef](#)]
78. Lin, J.; Su, L.; Yan, Y.; Sheng, G.; Xie, D.; Jiang, X. Prediction Method for Power Transformer Running State Based on LSTM_DBN Network. *Energies* **2018**, *11*, 1880. [[CrossRef](#)]
79. Deng, W.; Liu, H.; Xu, J.; Zhao, H.; Song, Y. An Improved Quantum-Inspired Differential Evolution Algorithm for Deep Belief Network. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 7319–7327. [[CrossRef](#)]
80. Jiao, J.; Zheng, X.J. Fault Diagnosis Method for Industrial Robots Based on DBN Joint Information Fusion Technology. *Comput. Intell. Neurosci.* **2022**, *2022*, 4340817. [[CrossRef](#)]
81. Xin, L.; Haidong, S.; Hongkai, J.; Jiawei, X. Modified Gaussian convolutional deep belief network and infrared thermal imaging for intelligent fault diagnosis of rotor-bearing system under time-varying speeds. *Struct. Health-Monit.-Int. J.* **2022**, *21*, 339–353. [[CrossRef](#)]
82. Yan, X.; Liu, Y.; Jia, M. Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions. *Knowl.-Based Syst.* **2020**, *193*, 105484. [[CrossRef](#)]
83. Qin, B.; Luo, Q.; Li, Z.; Zhang, C.; Wang, H.; Liu, W. Data Screening Based on Correlation Energy Fluctuation Coefficient and Deep Learning for Fault Diagnosis of Rolling Bearings. *Energies* **2022**, *15*, 2707. [[CrossRef](#)]
84. Zhu, D.; Cheng, X.; Yang, L.; Chen, Y.; Yang, S.X. Information Fusion Fault Diagnosis Method for Deep-Sea Human Occupied Vehicle Thruster Based on Deep Belief Network. *IEEE Trans. Cybern.* **2022**, *52*, 9414–9427. [[CrossRef](#)]
85. Xu, F.; Fang, Z.; Tang, R.; Li, X.; Tsui, K.L. An unsupervised and enhanced deep belief network for bearing performance degradation assessment. *Measurement* **2020**, *162*, 107902. [[CrossRef](#)]
86. Xu, F.; Shu, X.; Li, X.; Tang, R. Health indicator construction for roller bearing based on an unsupervised deep belief network with a novel sigmoid zero local minimum point model. *Struct. Health-Monit.-Int. J.* **2021**, *20*, 2110–2123. [[CrossRef](#)]
87. Zollanvari, A.; Kunanbayev, K.; Bitaghsir, S.A.; Bagheri, M. Transformer Fault Prognosis Using Deep Recurrent Neural Network Over Vibration Signals. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 2502011. [[CrossRef](#)]
88. Encalada-Davila, A.; Moyon, L.; Tutiven, C.; Puruncajas, B.; Vidal, Y. Early Fault Detection in the Main Bearing of Wind Turbines Based on Gated Recurrent Unit (GRU) Neural Networks and SCADA Data. *IEEE-ASME Trans. Mechatronics* **2022**, *27*, 5583–5593. [[CrossRef](#)]
89. Hao, S.; Ge, F.X.; Li, Y.; Jiang, J. Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks. *Measurement* **2020**, *159*, 107802. [[CrossRef](#)]
90. Shi, Z.; Chehade, A. A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107257. [[CrossRef](#)]
91. Ma, M.; Mao, Z. Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction. *IEEE Trans. Ind. Inform.* **2021**, *17*, 1658–1667. [[CrossRef](#)]
92. Yuan, M.; Wu, Y.; Lin, L. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. In Proceedings of the 2016 IEEE International Conference on Aircraft Utility Systems (AUS), Beijing, China, 10–12 October 2016; pp. 135–140. [[CrossRef](#)]

93. Ling, J.; Liu, G.J.; Li, J.L.; Shen, X.C.; You, D.D. Fault prediction method for nuclear power machinery based on Bayesian PPCA recurrent neural network model. *Nucl. Sci. Tech.* **2020**, *31*, 75. [[CrossRef](#)]
94. Van Gompel, J.; Spina, D.; Develder, C. Satellite based fault diagnosis of photovoltaic systems using recurrent neural networks. *Appl. Energy* **2022**, *305*, 117874. [[CrossRef](#)]
95. Xiao, L.; Liu, Z.; Zhang, Y.; Zheng, Y.; Cheng, C. Degradation assessment of bearings with trend-reconstruct-based features selection and gated recurrent unit network. *Measurement* **2020**, *165*, 108064. [[CrossRef](#)]
96. Zhang, Y.; Zhou, T.; Huang, X.; Cao, L.; Zhou, Q. Fault diagnosis of rotating machinery based on recurrent neural networks. *Measurement* **2021**, *171*, 108774. [[CrossRef](#)]
97. Lei, J.; Liu, C.; Jiang, D. Fault diagnosis of wind turbine based on Long Short-term memory networks. *Renew. Energy* **2019**, *133*, 422–432. [[CrossRef](#)]
98. Wu, J.; Hu, K.; Cheng, Y.; Zhu, H.; Shao, X.; Wang, Y. Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network. *ISA Trans.* **2020**, *97*, 241–250. [[CrossRef](#)]
99. Shi, J.; Peng, D.; Peng, Z.; Zhang, Z.; Goebel, K.; Wu, D. Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks. *Mech. Syst. Signal Process.* **2022**, *162*, 107996. [[CrossRef](#)]
100. Nasiri, A.; Taheri-Garavand, A.; Omid, M.; Carlomagno, G.M. Intelligent fault diagnosis of cooling radiator based on deep learning analysis of infrared thermal images. *Appl. Therm. Eng.* **2019**, *163*, 114410. [[CrossRef](#)]
101. Yongbo, L.; Xiaoqiang, D.; Fangyi, W.; Xianzhi, W.; Huangchao, Y. Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging. *Chin. J. Aeronaut.* **2020**, *33*, 427–438.
102. Jiang, J.; Bie, Y.; Li, J.; Yang, X.; Ma, G.; Lu, Y.; Zhang, C. Fault diagnosis of the bushing infrared images based on mask R-CNN and improved PCNN joint algorithm. *High Volt.* **2021**, *6*, 116–124. [[CrossRef](#)]
103. Liang, P.; Deng, C.; Wu, J.; Yang, Z.; Zhu, J.; Zhang, Z. Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform. *Comput. Ind.* **2019**, *113*, 103132. [[CrossRef](#)]
104. Gou, L.; Li, H.; Zheng, H.; Li, H.; Pei, X. Aeroengine control system sensor fault diagnosis based on CWT and CNN. *Math. Probl. Eng.* **2020**, *2020*, 5357146. [[CrossRef](#)]
105. Shao, S.; Yan, R.; Lu, Y.; Wang, P.; Gao, R.X. DCNN-Based Multi-Signal Induction Motor Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 2658–2669. [[CrossRef](#)]
106. Ahmed, H.O.A.; Nandi, A.K. Connected Components-based Colour Image Representations of Vibrations for a Two-stage Fault Diagnosis of Roller Bearings Using Convolutional Neural Networks. *Chin. J. Mech. Eng.* **2021**, *34*, 37. [[CrossRef](#)]
107. Miao, J.; Wang, J.; Miao, Q. An Enhanced Multifeature Fusion Method for Rotating Component Fault Diagnosis in Different Working Conditions. *IEEE Trans. Reliab.* **2021**, *70*, 1611–1620. [[CrossRef](#)]
108. Minh-Quang, T.; Liu, M.K.; Quoc-Viet, T.; Toan-Khoa, N. Effective Fault Diagnosis Based on Wavelet and Convolutional Attention Neural Network for Induction Motors. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3501613. [[CrossRef](#)]
109. Xie, T.; Huang, X.; Choi, S.K. Intelligent Mechanical Fault Diagnosis Using Multisensor Fusion and Convolution Neural Network. *IEEE Trans. Ind. Inform.* **2022**, *18*, 3213–3223. [[CrossRef](#)]
110. Pan, J.; Zi, Y.; Chen, J.; Zhou, Z.; Wang, B. LiftingNet: A Novel Deep Learning Network With Layerwise Feature Learning From Noisy Mechanical Data for Fault Classification. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4973–4982. [[CrossRef](#)]
111. Kiranyaz, S.; Gastli, A.; Ben-Brahim, L.; Al-Emadi, N.; Gabbouj, M. Real-Time Fault Detection and Identification for MMC Using 1-D Convolutional Neural Networks. *IEEE Trans. Ind. Electron.* **2019**, *66*, 8760–8771. [[CrossRef](#)]
112. Wang, H.; Liu, Z.; Peng, D.; Qin, Y. Understanding and Learning Discriminant Features based on Multiattention 1DCNN for Wheelset Bearing Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5735–5745. [[CrossRef](#)]
113. Huang, D.; Li, S.; Qin, N.; Zhang, Y. Fault diagnosis of high-speed train bogie based on the improved-CEEMDAN and 1-D CNN algorithms. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3508811. [[CrossRef](#)]
114. Du, C.; Zhang, X.; Zhong, R.; Li, F.; Yu, F.; Rong, Y.; Gong, Y. Unmanned aerial vehicle rotor fault diagnosis based on interval sampling reconstruction of vibration signals and a one-dimensional convolutional neural network deep learning method. *Meas. Sci. Technol.* **2022**, *33*, 065003. [[CrossRef](#)]
115. Ye, Z.; Yu, J. Multi-level features fusion network-based feature learning for machinery fault diagnosis. *Appl. Soft Comput.* **2022**, *122*, 108900. [[CrossRef](#)]
116. Gong, B.; Du, X. Research on analog circuit fault diagnosis based on CBAM-CNN. In Proceedings of the 2021 IEEE International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 27–29 August 2021; pp. 258–261.
117. Ran, R.; Xu, X.; Qiu, S.; Cui, X.; Wu, F. Crack-SegNet: Surface Crack Detection in Complex Background Using Encoder-Decoder Architecture. In Proceedings of the 2021 4th International Conference on Sensors, Signal and Image Processing, Nanjing China, 15–17 October 2021; pp. 15–22.
118. Meng, S.; Kang, J.; Chi, K.; Die, X. Intelligent Fault Diagnosis of Gearbox based on Multiple Synchrosqueezing S-Transform and Convolutional Neural Networks. *Int. J. Perform. Eng.* **2020**, *16*, 528–536. [[CrossRef](#)]
119. Chen, Y.L.; Chiang, Y.; Chiu, P.H.; Huang, I.; Xiao, Y.B.; Chang, S.W.; Huang, C.W.; et al. High-Dimensional Phase Space Reconstruction with a Convolutional Neural Network for Structural Health Monitoring. *Sensors* **2021**, *21*, 3514. [[CrossRef](#)]
120. Ince, T.; Kiranyaz, S.; Eren, L.; Askar, M.; Gabbouj, M. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7067–7075. [[CrossRef](#)]

121. Wu, X.; Peng, Z.; Ren, J.; Cheng, C.; Zhang, W.; Wang, D. Rub-Impact Fault Diagnosis of Rotating Machinery Based on 1-D Convolutional Neural Networks. *IEEE Sens. J.* **2020**, *20*, 8349–8363. [[CrossRef](#)]
122. Jimenez-Guarneros, M.; Morales-Perez, C.; de Jesus Rangel-Magdaleno, J. Diagnostic of Combined Mechanical and Electrical Faults in ASD-Powered Induction Motor Using MODWT and a Lightweight 1-D CNN. *IEEE Trans. Ind. Inform.* **2022**, *18*, 4688–4697. [[CrossRef](#)]
123. Khan, M.A.; Kim, Y.H.; Choo, J. Intelligent fault detection using raw vibration signals via dilated convolutional neural networks. *J. Supercomput.* **2020**, *76*, 8086–8100. [[CrossRef](#)]
124. Hudson, D.A.; Manning, C.D. Compositional attention networks for machine reasoning. *arXiv Prepr.* **2018**, arXiv:1803.03067.
125. Hernández, A.; Amigó, J.M. Attention mechanisms and their applications to complex systems. *Entropy* **2021**, *23*, 283. [[CrossRef](#)]
126. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 2017 Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
127. Jin, Y.; Hou, L.; Chen, Y. A Time Series Transformer based method for the rotating machinery fault diagnosis. *Neurocomputing* **2022**, *494*, 379–395. [[CrossRef](#)]
128. Ding, Y.; Jia, M.; Miao, Q.; Cao, Y. A novel time–frequency Transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings. *Mech. Syst. Signal Process.* **2022**, *168*, 108616. [[CrossRef](#)]
129. Zhang, Z.; Song, W.; Li, Q. Dual-Aspect Self-Attention Based on Transformer for Remaining Useful Life Prediction. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2505711. [[CrossRef](#)]
130. Pei, X.; Zheng, X.; Wu, J. Rotating Machinery Fault Diagnosis Through a Transformer Convolution Network Subjected to Transfer Learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 2515611. [[CrossRef](#)]
131. Du, X.; Jia, L.; Ul Haq, I. Fault diagnosis based on SPBO-SDAE and transformer neural network for rotating machinery. *Measurement* **2022**, *188*, 110545. [[CrossRef](#)]
132. Fang, H.; Deng, J.; Bai, Y.; Feng, B.; Li, S.; Shao, S.; Chen, D. CLFormer: A Lightweight Transformer Based on Convolutional Embedding and Linear Self-Attention With Strong Robustness for Bearing Fault Diagnosis Under Limited Sample Conditions. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3504608. [[CrossRef](#)]
133. Han, S.; Shao, H.; Cheng, J.; Yang, X.; Cai, B. Convformer-NSE: A Novel End-to-End Gearbox Fault Diagnosis Framework under Heavy Noise Using Joint Global and Local Information. *IEEE-ASME Trans. Mechatronics* **2022**. [[CrossRef](#)]
134. Li, Z.; Ouyang, B.; Cui, X.; Xu, X.; Qiu, S. Fault Diagnosis Method of Electromagnetic Launch and Recovery Systems Based on Large-Scale Time Series Similarity Search. *IEEE Trans. Plasma Sci.* **2022**, *50*, 2293–2304. [[CrossRef](#)]
135. Wu, B.; Cai, W.; Cheng, F.; Chen, H. Simultaneous-fault diagnosis considering time series with a deep learning transformer architecture for air handling units. *Energy Build.* **2022**, *257*, 111608. [[CrossRef](#)]
136. Li, B.; Tang, B.; Deng, L.; Zhao, M. Self-Attention ConvLSTM and Its Application in RUL Prediction of Rolling Bearings. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3518811. [[CrossRef](#)]
137. Ding, Y.; Jia, M. Convolutional Transformer: An Enhanced Attention Mechanism Architecture for Remaining Useful Life Estimation of Bearings. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3515010. [[CrossRef](#)]
138. An, Z.; Cheng, L.; Guo, Y.; Ren, M.; Feng, W.; Sun, B.; Ling, J.; Chen, H.; Chen, W.; Luo, Y.; et al. A Novel Principal Component Analysis-Informer Model for Fault Prediction of Nuclear Valves. *Machines* **2022**, *10*, 240. [[CrossRef](#)]
139. Yang, Z.; Liu, L.; Li, N.; Tian, J. Time Series Forecasting of Motor Bearing Vibration Based on Informer. *Sensors* **2022**, *22*, 5858. [[CrossRef](#)]
140. Li, T.; Zhou, Z.; Li, S.; Sun, C.; Yan, R.; Chen, X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mech. Syst. Signal Process.* **2022**, *168*, 108653. [[CrossRef](#)]
141. Sperduti, A.; Starita, A. Supervised neural networks for the classification of structures. *IEEE Trans. Neural Netw.* **1997**, *8*, 714–735. [[CrossRef](#)]
142. Chen, Z.; Xu, J.; Alippi, C.; Ding, S.X.; Shardt, Y.; Peng, T.; Yang, C. Graph neural network-based fault diagnosis: A review. *arXiv Prepr.* **2021**, arXiv:2111.08185.
143. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Philip, S.Y. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [[CrossRef](#)]
144. Chen, Z.; Xu, J.; Peng, T.; Yang, C. Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge. *IEEE Trans. Cybern.* **2021**, *52*, 9157–9169. [[CrossRef](#)]
145. Tang, Y.; Zhang, X.; Zhai, Y.; Qin, G.; Song, D.; Huang, S.; Long, Z. Rotating machine systems fault diagnosis using semisupervised conditional random field-based graph attention network. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [[CrossRef](#)]
146. Liu, L.; Zhao, H.; Hu, Z. Graph dynamic autoencoder for fault detection. *Chem. Eng. Sci.* **2022**, *254*, 117637. [[CrossRef](#)]
147. Chen, D.; Liu, R.; Hu, Q.; Ding, S.X. Interaction-Aware Graph Neural Networks for Fault Diagnosis of Complex Industrial Processes. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–14. [[CrossRef](#)]
148. Zhang, D.; Stewart, E.; Entezami, M.; Roberts, C.; Yu, D. Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network. *Measurement* **2020**, *156*, 107585. [[CrossRef](#)]
149. Gao, Y.; Chen, M.; Yu, D. Semi-supervised graph convolutional network and its application in intelligent fault diagnosis of rotating machinery. *Measurement* **2021**, *186*, 110084. [[CrossRef](#)]
150. Li, C.; Mo, L.; Yan, R. Fault Diagnosis of Rolling Bearing Based on WHVG and GCN. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [[CrossRef](#)]

151. Yu, X.; Tang, B.; Zhang, K. Fault Diagnosis of Wind Turbine Gearbox Using a Novel Method of Fast Deep Graph Convolutional Networks. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 6502714. [[CrossRef](#)]
152. Sun, K.; Huang, Z.; Mao, H.; Qin, A.; Li, X.; Tang, W.; Xiong, J. Multi-Scale Cluster-Graph Convolution Network With Multi-Channel Residual Network for Intelligent Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2502612. [[CrossRef](#)]
153. Zhou, K.; Yang, C.; Liu, J.; Xu, Q. Dynamic Graph-Based Feature Learning With Few Edges Considering Noisy Samples for Rotating Machinery Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2022**, *69*, 10595–10604. [[CrossRef](#)]
154. Zhang, K.; Chen, J.; He, S.; Li, F.; Feng, Y.; Zhou, Z. Triplet metric driven multi-head GNN augmented with decoupling adversarial learning for intelligent fault diagnosis of machines under varying working condition. *J. Manuf. Syst.* **2022**, *62*, 1–16. [[CrossRef](#)]
155. Han, S.; Woo, S.S. Learning Sparse Latent Graph Representations for Anomaly Detection in Multivariate Time Series. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 2977–2986. [[CrossRef](#)]
156. Zhang, T.; Chen, J.; Li, F.; Zhang, K.; Lv, H.; He, S.; Xu, E. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Trans.* **2022**, *119*, 152–171. [[PubMed](#)]
157. Wu, Z.; Guo, Y.; Lin, W.; Yu, S.; Ji, Y. A Weighted Deep Representation Learning Model for Imbalanced Fault Diagnosis in Cyber-Physical Systems. *Sensors* **2018**, *18*, 1096. [[CrossRef](#)] [[PubMed](#)]
158. Cao, X.; Wang, Y.; Chen, B.; Zeng, N. Domain-adaptive intelligence for fault diagnosis based on deep transfer learning from scientific test rigs to industrial applications. *Neural Comput. Appl.* **2021**, *33*, 4483–4499. [[CrossRef](#)]
159. Li, C.; Li, S.; Zhang, A.; He, Q.; Liao, Z.; Hu, J. Meta-learning for few-shot bearing fault diagnosis under complex working conditions. *Neurocomputing* **2021**, *439*, 197–211. [[CrossRef](#)]
160. Vilalta, R.; Drissi, Y. A perspective view and survey of meta-learning. *Artif. Intell. Rev.* **2002**, *18*, 77–95. [[CrossRef](#)]
161. Xu, J.; Zhou, L.; Zhao, W.; Fan, Y.; Ding, X.; Yuan, X. Zero-shot learning for compound fault diagnosis of bearings. *Expert Syst. Appl.* **2022**, *190*, 116197. [[CrossRef](#)]
162. Huang, R.; Liao, Y.; Zhang, S.; Li, W. Deep decoupling convolutional neural network for intelligent compound fault diagnosis. *IEEE Access* **2018**, *7*, 1848–1858. [[CrossRef](#)]
163. Deng, W.; Li, Z.; Li, X.; Chen, H.; Zhao, H. Compound fault diagnosis using optimized MCKD and sparse representation for rolling bearings. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3508509. [[CrossRef](#)]
164. Huang, R.; Li, J.; Li, W.; Cui, L. Deep ensemble capsule network for intelligent compound fault diagnosis using multisensory data. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 2304–2314. [[CrossRef](#)]
165. Ramachandram, D.; Taylor, G.W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **2017**, *34*, 96–108. [[CrossRef](#)]
166. Che, C.; Wang, H.; Ni, X.; Lin, R. Hybrid multimodal fusion with deep learning for rolling bearing fault diagnosis. *Measurement* **2021**, *173*, 108655. [[CrossRef](#)]
167. Ma, M.; Sun, C.; Chen, X. Deep Coupling Autoencoder for Fault Diagnosis With Multimodal Sensory Data. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1137–1145. [[CrossRef](#)]
168. Wang, D.; Li, Y.; Jia, L.; Song, Y.; Liu, Y. Novel Three-Stage Feature Fusion Method of Multimodal Data for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [[CrossRef](#)]
169. Ma, S.; Chu, F. Ensemble deep learning-based fault diagnosis of rotor bearing systems. *Comput. Ind.* **2019**, *105*, 143–152. [[CrossRef](#)]
170. Wang, X.; Yang, B.; Wang, Z.; Liu, Q.; Chen, C.; Guan, X. A compressed sensing and CNN-based method for fault diagnosis of photovoltaic inverters in edge computing scenarios. *IET Renew. Power Gener.* **2022**, *16*, 1434–1444. [[CrossRef](#)]
171. Li, H.; Hu, G.; Li, J.; Zhou, M. Intelligent Fault Diagnosis for Large-Scale Rotating Machines Using Binarized Deep Neural Networks and Random Forests. *IEEE Trans. Autom. Sci. Eng.* **2022**, *19*, 1109–1119. [[CrossRef](#)]
172. Imamura, L.Y.; Avila, S.L.; Pacheco, F.S.; Salles, M.B.C.; Jablon, L.S. Diagnosis of Unbalance in Lightweight Rotating Machines Using a Recurrent Neural Network Suitable for an Edge-Computing Framework. *J. Control Autom. Electr. Syst.* **2022**, *33*, 1272–1285. [[CrossRef](#)]
173. Liu, Z.; Sun, M.; Zhou, T.; Huang, G.; Darrell, T. Rethinking the value of network pruning. *arXiv Prepr.* **2018**, arXiv:1810.05270.
174. Gou, J.; Yu, B.; Maybank, S.J.; Tao, D. Knowledge distillation: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 1789–1819. [[CrossRef](#)]
175. Shan, N.; Xu, X.; Bao, X.; Qiu, S. Fast Fault Diagnosis in Industrial Embedded Systems Based on Compressed Sensing and Deep Kernel Extreme Learning Machines. *Sensors* **2022**, *22*, 3997. [[CrossRef](#)]
176. Wu, Y.; Tang, B.; Deng, L.; Li, Q. Distillation-enhanced fast neural architecture search method for edge-side fault diagnosis of wind turbine gearboxes. *Expert Syst. Appl.* **2022**, *208*, 118049. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.