



A systematic review of data-driven approaches to fault diagnosis and early warning

Peng Jieyang^{1,2} · Andreas Kimmig² · Wang Dongkun³ · Zhibin Niu⁵ · Fan Zhi⁶ · Wang Jiahai¹ · Xiufeng Liu⁴ · Jivka Ovtcharova²

Received: 31 January 2022 / Accepted: 1 September 2022 / Published online: 15 September 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

As an important stage of life cycle management, machinery PHM (prognostics and health management), an emerging subject in mechanical engineering, has seen a huge amount of research. Here the authors present a comprehensive overview that details previous and current efforts in PHM from an industrial big data perspective. The authors first analyze the historical development of industrial big data and its distinction from big data of other domains and summarize the sources, types, and processing modes of industrial big data. Then, the authors provide an overview of common representation and fusion (data pre-processing) methods of industrial big data. Next, the authors comprehensively review common PHM methods in the data-driven context, focusing on the application of deep learning. Finally, two industrial cases from our previous studies are included in this paper to demonstrate how the PHM technique may facilitate the manufacturing industry. Furthermore, a visual bibliography is developed for displaying current results of PHM in an appropriate theme. The bibliography is open source at “<https://mango-hund.github.io/>”. The authors believe that future research endeavors will require an understanding of this previous work, and our efforts in this paper will make it possible to customize and integrate PHM systems quickly for a variety of applications.

Keywords Industrial big data · Prognostics and health management · Deep learning · Industry 4.0 · Data visualization · Industrial Internet of Things

Introduction

Since the mid-20th century, with the emergence of a new generation of information technology, the manufacturing

industry has gradually entered the era of cyber physics: at the physical level, the large-scale application of sensors and SCADA systems enables production data to be monitored and stored in real time. At the information level, data mining methods make workshop data association analysis possible. Wireless communication technology enables distributed manufacturing resources to cooperate with each other. With the further development and integration of the physical layer and the information layer, the concept of “Industrial Internet of Things (IIoT)” came into being (Negri et al., 2017). An Industrial Internet of Things system is an interconnected system between the physical environment and computing capabilities of management equipment. Through the integration of computing capabilities and physical environment, IIoT intelligently connects independent devices to achieve self-adaptive interaction, so that the sub-systems can be aware of each other, achieving effective coordination, thus automatically adjusting and configuring the operational logic according to the tasks requirements.

✉ Zhibin Niu
zniu@tju.edu.cn

- ¹ Advanced Manufacturing Technology Center, Tongji University, Shanghai 200092, People’s Republic of China
- ² Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
- ³ State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau 999078, SAR, China
- ⁴ Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
- ⁵ College of Intelligence and Computing, Tianjin University, 300354 Tianjin, China
- ⁶ Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Fraunhofer, 70569 Stuttgart, Germany

Generally speaking, the two main functions of IIoT are:

- Efficient information connection at the physical layer, that is, to ensure real-time data collection in the physical environment and information feedback in cyberspace;
- Intelligent data management, analysis and calculation capabilities in cyberspace.

Intelligent analytics and intelligent equipments are cooperating to realize a new approach of industrial transformation using the above-mentioned IIoT technologies. Various signals, such as vibration, pressure, acoustic emission, etc, can be retrieved by proper sensor installations, which sparked the demand for big data analytics in the industrial environment. Large volume, multi-variety, high velocity and veracity constitute the 4V characteristics of big data. In the industrial field, the quality of data is also a priority to be considered. The industrial data in the field of fault diagnosis is usually more structured, with stronger correlation and sequentiality. This is due to the fact that industrial data is usually generated by automated equipment in a pre-designed production process. Often there is a logical association between equipment failure phenomena and causes, which is directly manifested in the data. In addition to the 4V characteristics of big data, there are specific characteristics of data in the field of fault diagnosis, specifically:

- Context of the data: Industrial Big Data is particularly concerned with the meaning of the data and its association with failures and mechanisms. This means that the analysis of industrial big data requires more domain knowledge (e.g., failure mechanisms, process knowledge, etc.).
- Data integrity: Compared with other fields, the demand for data integrity in industrial big data exceeds the demand for data volume. In fault-oriented data analysis, it requires the data sample to be as complete as possible, i.e., the data should include all possible fault types and ensure the comprehensiveness of the information extracted from the data reflecting the true status of the fault.
- Temporal validity of data: In industrial big data, the validity of data is reflected not only in the completeness of data but also in the instantaneity of data, which is caused by the distinctive spatio-temporal dynamic characteristics of data in the field of fault diagnosis. Analysis of low-quality or erroneous data cannot reflect the correct failure mechanism behind the variables and has a disastrous impact on prediction accuracy.
- Label of data: To provide usable insights, fault diagnosis and health assessment models usually rely on accurate, clean, and frequently adequately labelled (in supervised techniques) training data, making data quality a vital aspect for the industrial success of these solutions (Peres et al., 2020).

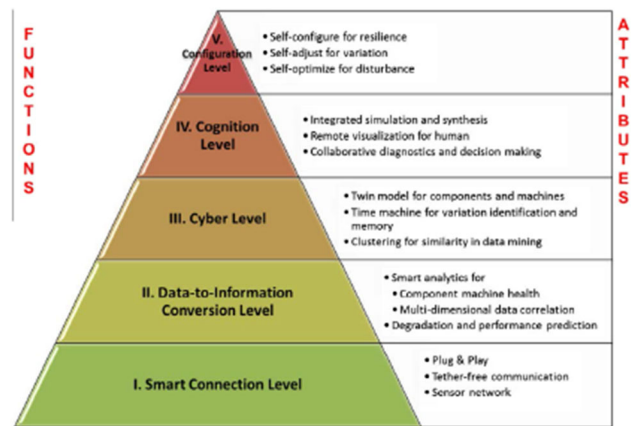


Fig. 1 5C architecture for implementation of Industrial Internet of Things (Lee et al., 2015)

Based on the above characteristics, Prof. Jay Lee suggests that the quality of industrial big data should also be evaluated (Lee et al., 2002). According to Prof. Jay Lee, data quality could be divided into four categories: (1) Intrinsic, which refers to characteristics native to the data itself, such as timeliness, completeness, accuracy, and consistency; (2) Contextual, which refers to attributes that are dependent on the context of the task at hand, such as relevancy and quantity; (3) Representational, which emphasizes that systems should present data in a way that is understandable and represented concisely and consistently; (4) Accessibility, which emphasizes that systems should be accessible to all users (Lee, 2020). In general, industrial big data is far less tolerant of error in prediction and analysis results than the internet and finance fields. Data in the internet and finance domains are predicted and analyzed by considering only whether the variables are statistically significantly correlated with each other, while noise and individual sample differences can be ignored when the sample size is large enough. However, simply applying the processing methods of big data to the industrial field cannot meet the analysis needs of industrial big data. Industrial Big Data requires specialized domain knowledge, a clear understanding of the mechanistic meaning reflected in the data, and the visualization and presentation of the results to make rational decisions (Song et al., 2021).

Predictive technologies can further entangle intelligent algorithms with electronics and wireless connected intelligence as more IIoT software and industrial big data-based intelligence methods are integrated into industrial products and systems. On this basis, some scholars have extended the traditional 3C (Computer, Communication, Control) concept to a 5C architecture (Lee et al., 2015), as shown in Fig. 1.

In the context of the 5C architecture, data-driven PHM technologies have emerged (Xu & Saleh, 2021). The authors summarize the data-driven PHM process in Fig. 2. In Fig. 2, the key components of PHM include Health Assessment,

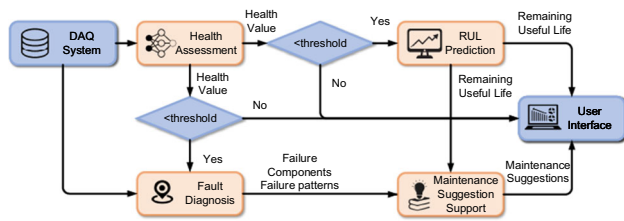


Fig. 2 Four key components in the PHM domain

RUL Prediction, Fault Diagnosis, and Maintenance Suggestion Support. With the operation of the equipment, the PHM system will initially assess the health condition of the equipment based on the data provided by the Industrial Data Acquisition System. The health value is then compared with pre-defined health thresholds (usually determined by an experienced maintenance engineer). If the health value is dangerously low, a prediction of the remaining useful life of the key components of the machine (usually moving parts such as spindles, servo motors, gearboxes, etc.) is carried out, as well as fault localization and fault diagnosis. The results of the RUL prediction and fault diagnosis are delivered to the Maintenance Suggestion Support module, which provides appropriate maintenance recommendations based on the current state of the equipment. Current applications of data-driven PHM technologies include embedded maintenance and cloud maintenance. At present, embedded maintenance systems have been widely used in the automotive electronics industry, such as the application of On-Board Diagnostics (ODB), which has greatly simplified the process of automotive fault diagnosis (Merkisz et al., 2001; Klinedinst & King, 2016); a successful case of cloud maintenance is Rolls-Royce in the UK. As the largest aero-engine company in Europe, each engine produced by Rolls-Royce continuously sends back status parameter information to the data center during service, thus realizing online engine status monitoring and preventive maintenance.

In addition, data-driven PHM technologies can provide considerable benefits to the manufacturing industry. According to a 2014 study by Accenture, with the further application of preventive maintenance technologies in manufacturing, 12% of unexpected failures in the U.S. manufacturing industry could be avoided, shop floor maintenance costs could be reduced by 30%, and equipment downtime losses would be reduced by 70% (Daugherty et al., 2015). The essence of the data-driven method is to classify a large amount of production data and use machine learning algorithms to mine rules to identify and or predict the status of equipment. Its intelligence is reflected in the ability of a learning algorithm to adaptively construct a suitable model for the equipment, and the model can be continuously adjusted and revised along with the production process. For data-driven fault diagnosis, the emergence of industrial big data provides a large number

of data sources for the technology on the one hand. On the other hand, the heterogeneous data types and huge amount of data also pose new challenges to data-driven fault diagnosis methods.

With these concerns in mind, the purpose of this study is to review data-driven predictive maintenance and fault diagnosis methods which are evolving in preparation for the forthcoming industrial big data environment, as well as to provide a long-term perspective on future developments in the industry. The remainder of this paper is organized as follows: Chapter 2 provides an overview of the development and sources of industrial big data. Chapter 3 briefly describes the classification of fault diagnosis methods and visualises the relationship between the current research object, data pre-processing and algorithmic models. Then, the authors select the three most widely applied algorithms from the last three years and present their highlights in Chapter 4. In Chapter 5 the authors present a visual bibliography of Prognostics and Health Management, and in chapter 6 the authors summarise the current situation and provide an outlook on future developments.

Big data in industry

In this chapter, the authors first introduce the sources of industrial big data, and comprehensively review the current prevailing industrial data processing models and existing industrial big data platforms, then elaborate the relationship between industrial big data and fault diagnosis, and introduce the topic of this paper - data-driven fault diagnosis in the context of Industry 4.0.

The development of industrial big data

Industrial big data refers to the general term of all kinds of data generated in the entire product life cycle from order, research and development(R&D), planning, technology, manufacturing, and sales in the industrial field. Compared with big data in the internet field, industrial big data has stronger structure, relevance and timing (Lee et al., 2015), and is more suitable for large-scale storage.

At the beginning of this century, with the maturity of information technology such as data mining and databases, data processing methods were initially applied to the fields of industrial software and industrial knowledge management. At this time, industrial big data is characterized by a relatively small amount of data and a relatively high value density in the data. In 2012, General Electric put forward the concept of “Industrial Big Data” for the first time, and pointed out that the current manufacturing industry should focus on the massive data generated during the production process. With the application of the Internet of

Things and the development of data mining algorithms, manufacturing giants have stepped into the practical stage of industrial big data. In August 2015, General Electric launched the first cloud platform Predix (General Electric Company, 2015) for industrial big data analysis, whose architecture is shown in Fig. 3. As a pioneer of industrial big data platform, Predix's functions include: industrial data management, industrial data analysis, cloud computing, etc. By connecting industrial assets and equipment to the cloud, the Predix platform provides asset performance management (APM) and operation optimization services based on big data analysis. In April 2016, Siemens Germany released the industrial cloud platform MindSphere (Germany, 2016). As an application of cloud computing technology in the industrial field, its main functions include: managing and optimizing plant assets, a plug-and-play data gateway (MindConnect), and the integration of data sources in different systems. In China, the China Electronics Standardization Institute and the National Information Technology Standardization Technical Committee jointly released the "Industrial Big Data White Paper (Electronics Standardization Institute China, 2017)" in February 2017. This document sorts out the definition and characteristics of industrial big data, and focuses on the relationship between industrial big data and smart manufacturing, industrial Internet, and smart factory. The major events of industrial big data development are summarized in Fig. 4 in this paper.

Sources of industrial big data

Industrial big data usually comes from various units in the entire manufacturing field, such as sensors, smart devices, logistics, electrical, and workshop staff. Generally speaking, these distributed data can be classified into several sources:

- Device/sensor data: process data, operating data, environmental data, etc. of equipment in the context of IoT (e.g. sensors, actuators, RFID devices and SCADA data, etc.)
- Enterprise operation data: mainly includes business management, marketing, procurement, inventory, business plan, etc. Analysis of procurement, storage, sales and other data help to reduce inventory and achieve lean production.
- Manufacturing chain data: including customer relationship management (CRM), supply chain management (SCM), environmental management system (EMS), and product market and macroeconomic data. Analysis of Manufacturing Chain Data helps enterprises to improve strategic planning, providing enterprise managers with a new perspective of corporate strategy development.

The complexity of acquiring data from electromechanical systems in an industrial context rises in proportional proportion to the size and accessibility of the automated sector, resulting in a PHM system that is less interpolated (Rohan, 2022). Therefore, developing a precise PHM system for each industrial system needs the acquisition of a unique dataset under specific conditions. Obtaining this one-of-a-kind dataset is challenging in most instances, and the resulting dataset has a major imbalance, a lack of some valuable information, and multi-domain knowledge. Although the above data source can cover some of the failure cases, researchers have to deal with a lack of availability of actual fault data, that can be used to train fault diagnostic models. Strategies to compensate for this include: Generative Adversarial Networks (GANs) (Luo et al., 2021), Transfer Learning (Li et al., 2022), Failure Modes and Effects Analysis Knowledge Graphs (FMEA KGs) (Duan et al., 2021), Digital Twin Fault Provocation (Martínez López, 2021). In Chapter 4, we present the application of transfer learning in fault diagnosis.

Collecting data from naturally deteriorated components, however, is a time-consuming procedure. Some institutions and research institutes have managed to develop datasets and provide open access to scholars all around the world, allowing them to test and deploy diagnosis algorithms. NASA Database (NASA USA, 2021) is one of the most famous databases in the industry. The various data sets provided by NASA can be used to perform pattern identification, health management, remaining useful life forecasting, etc. The Milling Data set provided by the University of Berkeley California contains wear data of milling machine tools under different speeds, feed rates, cutting depths (University of Berkeley California, 2021). Center for Intelligent Maintenance Systems (IMS) of Cincinnati University provides a data set of bearing degradation (Cincinnati University, 2021a). The University of Besanson designed experiments to study the relationship between the bearing acceleration and the remaining service life (France University of Besanson, 2021). Some of the published industrial data sets are shown in Table 1.

The relationship between industrial big data and PHM

Nowadays, an increasing study is being directed toward prognostics and health management, which focuses on detecting incipient failure, assessing present health, and predicting remaining service life. However, maintenance procedures should be adjusted in diverse maintenance scenarios with varying levels of data complexity and unpredictability. The maintenance transformation map (Fig. 5) depicts several maintenance procedures concerning the data's complexity and dynamicity.

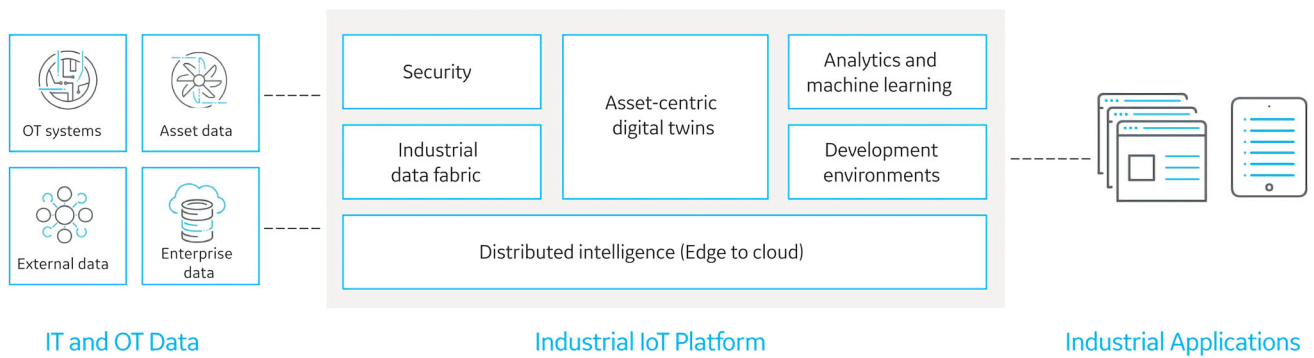


Fig. 3 Industrial Big Data Integration Platform Predix

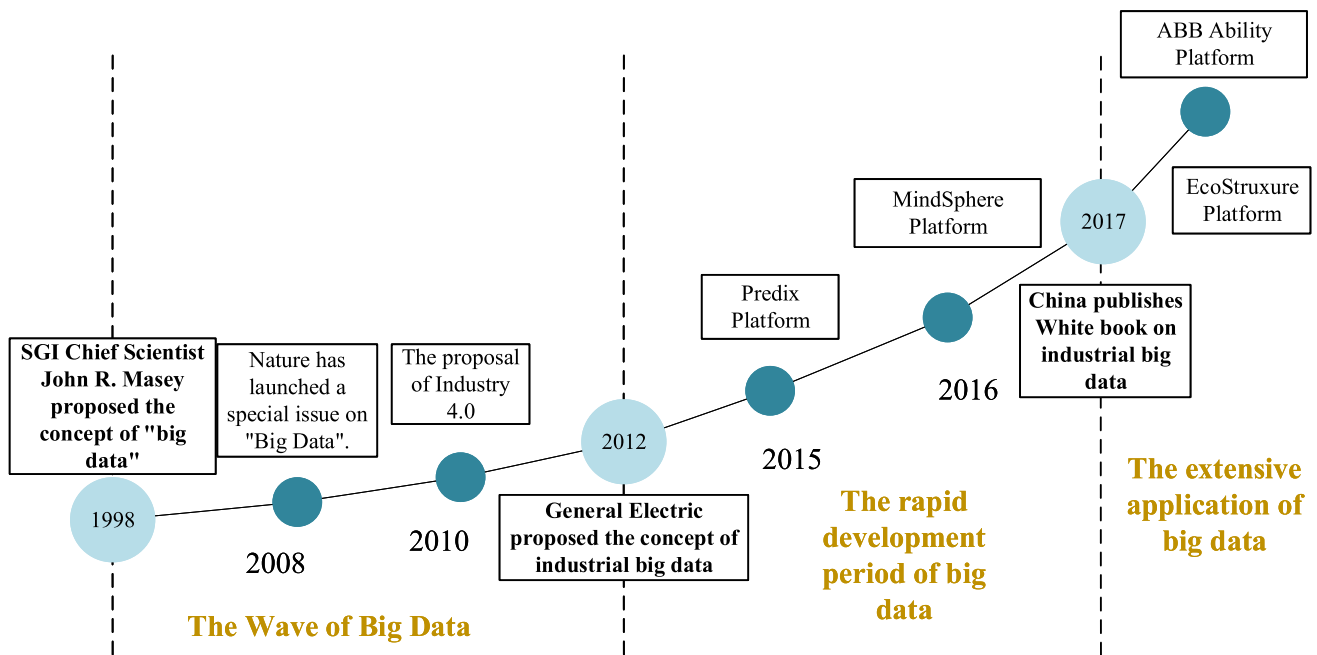


Fig. 4 Development and major events of industrial big data

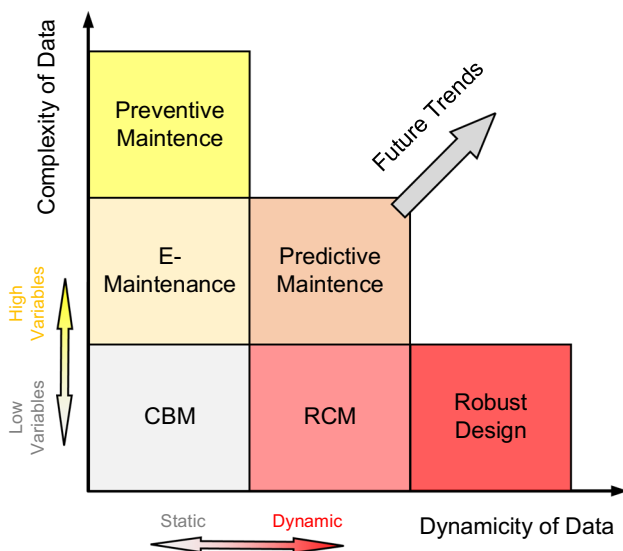
Figure 5 illustrates the variants of the PHM strategy in the industrial Big Data environment with different characteristics. Among them, CBM (condition based maintenance) can be used in systems that are deterministic to some extent, and for which raw signal variables can be extracted and regard as health indicators despite low data dimensionality. If the data is uncertain, the output cannot be predicted by using a predefined relationship model. As a result, future behavior cannot be accurately projected on current data. For example, the reliability of the cutting tool is determined by a variety of factors such as cutting material, cutting speed, feed speed and depth of cut, all of which are highly dynamic and therefore difficult to construct a perfect RUL model. Reliability-centered maintenance is more appropriate in such an unpredictable system. RCM focuses on a capacity to maintain predicted reliability over a certain duration of time, and it uses statistical methods such as failure modes and effects criticality analysis

(FMECA) to retrieve the information of possible durations before each of the failure patterns can happen. Robust design should be considered if the data uncertainty is more intricate, such as airspace engine with highly dynamic characteristics. In this circumstance, resources should be allocated to the robust design processes, rather than relying on inspection to ensure quality.

Along the other axis in Fig. 5, if a system has a high level of complexity and a large number of sensor data that reflect different characteristics of the system, E-Maintenance is recommended because CBM procedures typically deal with low dimension data. A typical high complexity system is the wind turbine. A wind turbine is usually fitted with more than 100 sensors such as generator speed, three-phase grid voltage, wind speed, grid frequency, nacelle position, pitch angle, reactive power, cable twist angle, etc (Peng et al., 2022). E-Maintenance may also align the maintenance pro-

Table 1 Comparison of prognostics datasets

Dataset	Sensors type	Research object
Case Western Reserve University (CaseWestern Reserve University, 2021b)	Accelerometer	Bearing
UC Berkeley (Agogino & Goebel, 2007)	Acoustic emission, Vibration	Milling
NASA (Saxena & Goebel, 2008)	Temperature, Pressure, Fan speed	Aircraft engine
Paderborn University (Lessmeier, 2021)	Accelerometer, Current, Thermocouple	Bearing
Intelligent Maintenance Systems (Cincinnati University, 2021a)	Accelerometer	Bearing
Paderbc (Nectoux et al., 2012)	Accelerometer, thermocouple	Bearing

**Fig. 5** The relationship between industrial big data and PHM (Lee et al., 2014)

cess with the business and operational processes to achieve optimal integrated production and asset management by connecting with industrial business systems. When a system contains a large amount of heterogeneous data from multiple sources, the complexity of the data reflects not only in quantity but also in quality. Preventive maintenance may be chosen rather than CBM or E-Maintenance which require systems to be heavily instrumented. Preventive maintenance is a time-driven maintenance strategy that schedules maintenance for a machine or component based on the experience of the mean time between failures (MTBF). Preventive maintenance follows strong assumptions that the machine is working under deterministic and static conditions, and therefore cannot be applied to system that operates in dynamic working regimes (Lee et al., 2014).

The fast growth of Information and Communication Technologies (ICT) has made wireless communication devices, and remote computing solutions available in recent years. The emergence of these technologies has enabled the expansion of industrial data in both complexity and dynamism dimensions. Therefore, for devices integrating the new gen-

eration of information technology systems, PHM strategies need to have the ability to handle data with both complex and dynamic characteristics. Consequently, the system must be self-predictive to some extent to reduce/avoid possible failures or even achieve zero downtime production.

Process models for industrial big data

Industrial big data analysis refers to the process of excavating potential knowledge in industrial data using machine learning method. According to the characteristics of industry data, there are three types of processing models of industrial data:

- **Batch processing:** In the batch processing, the data is stored based on the time window, and then the data within each time window is analyzed. The batch model is suitable for mining the time series knowledge in industrial data to generate an accurate predictive model. Currently used batch architecture includes MapReduce models (Dean & Ghemawat, 2008) and Spark models.
- **Stream processing:** Stream processing is a novel data processing paradigm that refers to the processing of data in motion, or in other words, computing on data directly as it is produced or received. Stream processing usually applies to device health assessments. Representative open source models of stream processing architecture include SparkStreaming, Strom and Kafka (Stonebraker et al., 2005; Hiranman et al., 2018; Lopez et al., 2016).

According to its characteristics, industrial big data can adopt different processing architectures as shown in Table 2 (O'Donovan et al., 2015).

Overview of preventive maintenance techniques

This section provides a summary of our literature study on the frequency with which data-driven technologies are used in PHM (Farsi & Zio, 2019; Guibing et al., 2021). The key application areas of data-driven approaches are identified,

Table 2 Common types, interfaces and processing methods of workshop data (O'Donovan et al., 2015)

Data source	Data type	Data interface	Processing model
High frequency sensor	Vibration, rotation speed, temperature, etc	ODBC, OLEDB, JDBC, etc	Stream processing
Manufacturing execution system(MES)	Equipment Information, Advanced Planning & Scheduling (APS), etc	ODBC, OLEDB, JDBC, etc	Batch processing
Programmable logic controller (PLC)	Control signal	OPCUA, Mt Connect, Bacnet, Modbus, etc	Stream processing
Gateway	Multiple types of data	HTTP, OPCUA, Modbus, I/O, etc	Batch processing/stream processing
Intelligent equipment and Internet of Things	Multiple types of data	MQTT, COAP, HTTP	Stream processing/stream processing

as well as how they are connected and what the current hot themes are.

Three models of preventive maintenance techniques

Generally speaking, preventive maintenance methods can be divided into three categories, namely, analytical model-based methods, knowledge-based methods and data-driven methods.

The analytical model-based approach constructs an analytical model from the failure mechanism of the equipment. However, this method has certain limitations in engineering applications because it needs to consider the relationship between stress, fatigue, material, structure, and service life when the model is built, while the actual equipment has considerable complexity. At present, the prediction methods based on analytical models are mainly focused on chips, integrated circuits, and aerospace components (Vachtsevanos & Vachtsevanos, 2006; Ye et al., 2015; Shlyannikov et al., 2021; Xudong et al., 2015).

The knowledge-based approach (Yun et al., 2019; Su et al., 2020; Lv et al., 2020; Yang et al., 2020) is a typical interactive query method, whose specific applications include knowledge graphs, expert systems, etc. The key of a knowledge-based approach lies in the representation and storage of knowledge. The knowledge-based approach has the advantages of transparency (the reasoning process is clear and explanatory) and flexibility (the knowledge in the knowledge base should be easy to modify and supplement). A deficiency of the knowledge-based approach is its high requirement for completeness of the knowledge base. Furthermore, the existing knowledge association is simple and the knowledge retrieval strategies strategy is simple (mainly by depth-first- and breadth-first-search).

In the data-driven approach, experts use the IoT of the plant floor to collect data, followed by data fusion and feature extraction. Finally, a decision model is trained and fitted. When new data is input, the system can use the existing

model to make decisions directly (Chen et al., 2020; Zhong-Xu et al., 2019; Khorasgani et al., 2018). The data-driven approach is capable of adaptively extracting decision rules from data with no reliance of precise analytical models and domain knowledge, and has become a popular application and research in the current industrial big data context. In this paper, the advantages, disadvantages and common scenarios of the above methods are summarized in Table 3. And the types, fault characteristics of mechanical faults of common components are summarized as shown in Table 4.

Statistical distribution of current researches

Figure 6 shows the statistical distribution of data-driven methodologies in various industrial research objects. A histogram of categorization of the used methods may be seen at the top of the graphic. Another histogram follows the study object and particular model from (“Data-driven approaches of fault diagnosis and health management” section) on the right side of the picture. A 2D histogram (heat map) in the figure’s centre section correlates these two frequencies in a single view. Each cell is colored to correlate to the number of data-driven methods it contains (darker cells indicate higher quantities). In this statistical survey, due to the popularity of deep learning in recent years, the authors divided traditional machine learning into two categories, “deep learning” and “shallow learning”, and separated them into two statistical categories. Deep Learning’s main distinguishing feature from Shallow Learning is that Deep Learning methods derive their own features directly from data (feature learning), while Shallow Learning relies on handcrafted features based upon heuristics of the target problem. In addition, although “ensemble learning” and “transfer learning”, as two training strategies, are not specific algorithm models, the authors statistically separate them from deep learning due to their promising applications in the PHM field. Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of

Table 3 Advantages, disadvantages and applicable scenarios of common methods in the field of fault prediction

	Advantage	Disadvantage	Scenario
Analytical model-based approach (Wang et al., 2021)	Model can accurately reflect the operation process and mechanism	Engineers need to understand the fault mechanism, and a large quantitative calculation resource is required	Chip, integrated circuit, aerospace components
Knowledge-based approach (Zhang et al., 2021a; Zhou et al., 2021)	Complex fault mechanism is not required, the fault diagnosis process is done by the inference engine	High requirement for completeness of knowledge base; current knowledge association method is simple and query strategy is inflexible	CNC system, electrical system
Data-driven approach (Yang et al., 2021; Xiaolei et al., 2021)	The model has self-learning capability and is capable of mining the data for potential fault knowledge	Models are often specific to a certain object and need to be retrained for different diagnostic objects; Models cannot clearly interpret the fault mechanism	Spindle, motor, bearing, gear, pump (refer to Table 4)

Table 4 Summarization of industrial data of critical components

Component	Common failure	Common measures	Common features
Bearing	Outer-race, inner-race, roller and cage failures	Vibration, Oil debris, Acoustic emission	Vibration characteristic frequency, Time domain statistical characteristics, Metallic debris shape, Sharp pulses and rate of development of stress-waves propagation
Gear	Tooth missing, Tooth pitting, Gear crack, Gear fatigue	Vibration, Oil debris, Acoustic emission	Time domain statistical features, Vibration signature frequencies, Oil debris quantity and chemical analysis
Shaft	Unbalance, BMisalignment, Rub	Vibration	Vibration characteristic frequency, Time domain statistical characteristics, System modal characteristics
Pump	Valve impact, Fracture, Piston slap, Defective bearing and revolving crank, Hydraulic problem	Vibration, Pressure, Acoustic, Emission	Vibration characteristic frequency, Pressure time domain statistical characteristics, Sharp pulses and rate of development of stress-waves propagation
Alternator	Stator faults, Rotor electrical faults, Rotor mechanical faults	Stator currents and voltages, magnetic fields and frame vibrations	Specific harmonic components, Sideband components

the constituent learning algorithms alone. Transfer learning focuses on storing knowledge gained while solving one problem and applying it to a different but related problem, which has the potential to significantly improve the sample efficiency of a reinforcement learning agent. That is, the “deep learning” category contains only those studies that directly apply deep learning algorithms, while “ensemble learning” and “transfer learning” refer to those deep learning studies that apply the above two strategies.

The histogram on the top of Fig. 6 has maximum values at “Deep learning”. This sub-phase is responsible for 51% of all data-driven applications. The second most common sub-phase is referred to as “Shallow learning”. When ensemble learning and transfer learning, which are based on deep learning algorithms, are taken into account, the proportion of machine learning cases in PHM research reaches 86%. This means that traditional signal decom-

position and model-based approaches are gradually being superseded by machine learning methods. Behind the prosperity of machine learning approaches, it is also important to be cautionary of the large number of similar studies in data-driven approaches. For example, those cases on bearings, as the focus of PHM technology, accounts for 41.8% of all researches. And nearly half of the research data on bearings is predicated on a dataset from the Case Western Reserve University (Case Western Reserve University, 2021b) with a huge number of cases from different research institutions applying similar deep learning frameworks to achieve very high classification accuracy. This is due to the fact that the data set is collected from the test bench with human-made defective bearings and is consequently inherently quite highly classifiable, which requires only a brief data pre-processing and can be processed by almost any classifier directly. Therefore, further data-driven based researches should focus on the

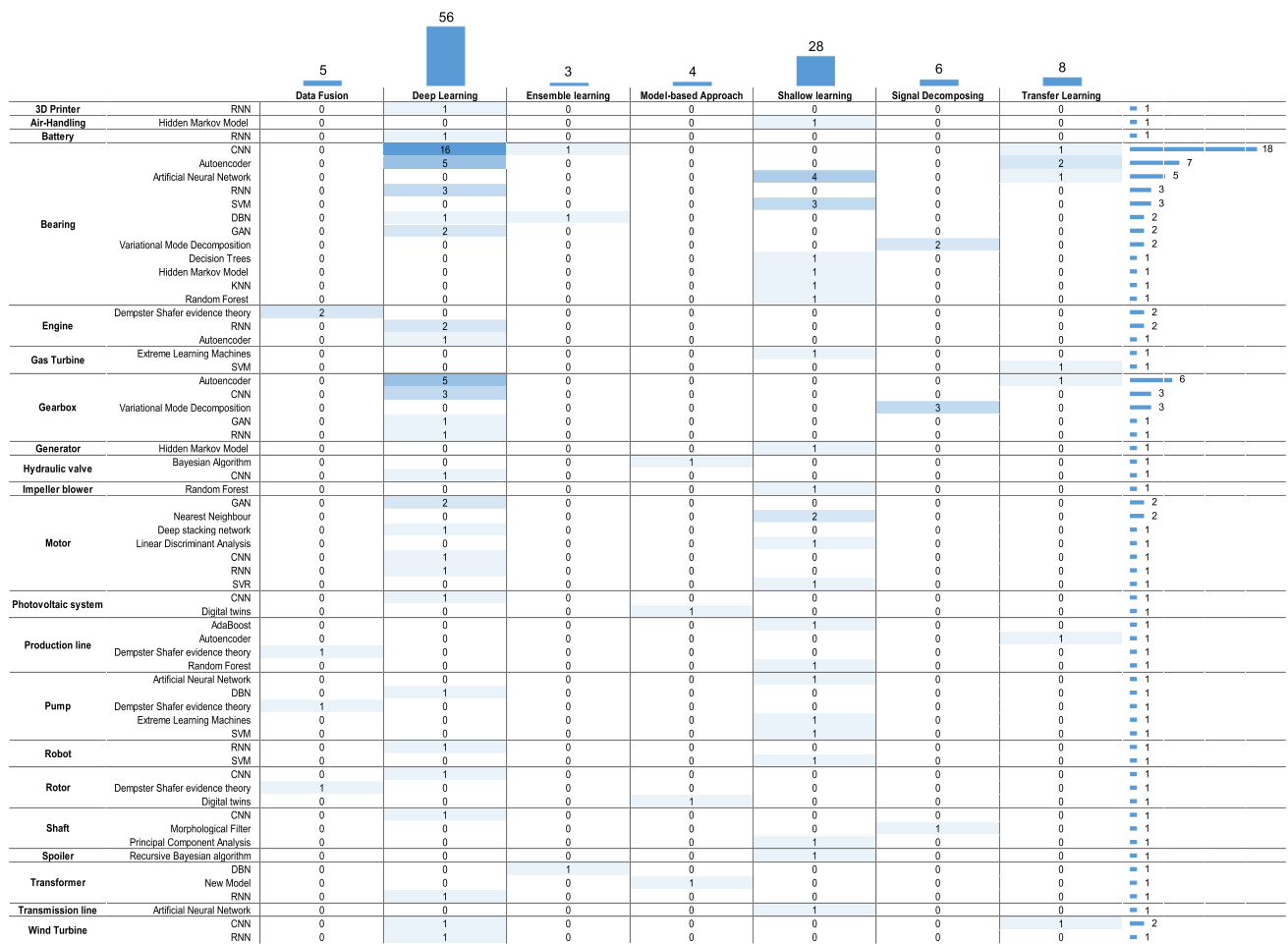


Fig. 6 Heat map and histograms showing the statistical distribution of the surveyed applications of data-driven preventive maintenance technology. The horizontal axis contains classification of the applied algorithms and the vertical axis follows the research object and spe-

cific model from (“Data-driven approaches of fault diagnosis and health management” section). Each heat map cell indicates its corresponding number of applied applications. Darker cells indicate higher quantities

interpretability of existing machine learning models rather than on developing more classification models.

Figure 7 describes correlations among research objects, sensors, pre-processing methods, and methodology in preventive maintenance. Flow widths indicate frequency of use and colors distinguish different techniques. As can be seen from Fig. 7, rotating machinery such as bearings, gearboxes and motors account for more than 80% of the total research, and bearings in particular account for around 50% of the research on rotating equipment. As mentioned above, this is partially due to the importance of rotating machinery, but also due to the high classifiability.

From a data-driven perspective, the high degree of classifiability can be explained by the vibration signals. Studies have shown that more than 80% of failures in rotating machinery are caused by three failures: rotor imbalance, axis misalignment and bearing instability (NASA, 1994), all of which are directly related to vibration signals. Consequently, the vast

majority of rotating machines in Fig. 7 are studied in terms of vibration signals, and vibration sensors account for 67% of the total study.

In Fig. 7 the authors can also notice that deep learning methods are less dependent on pre-processing of data than shallow learning, which is due to the fact that deep learning models are capable of adaptively extracting abstract features by themselves while in traditional machine learning approach, features need to be artificially identified. Those data pre-processing methods for deep learning studies are mainly aimed at converting the data into a format that conforms to the requirements of the deep learning model. For example, standard CNNs cannot handle 1-dimensional data. As a result, some studies have exploited signal processing to transform the original signal into a 2-dimensional matrix, which can then be identified by CNNs. Whereas most of the data pre-processing methods for shallow learning are derived

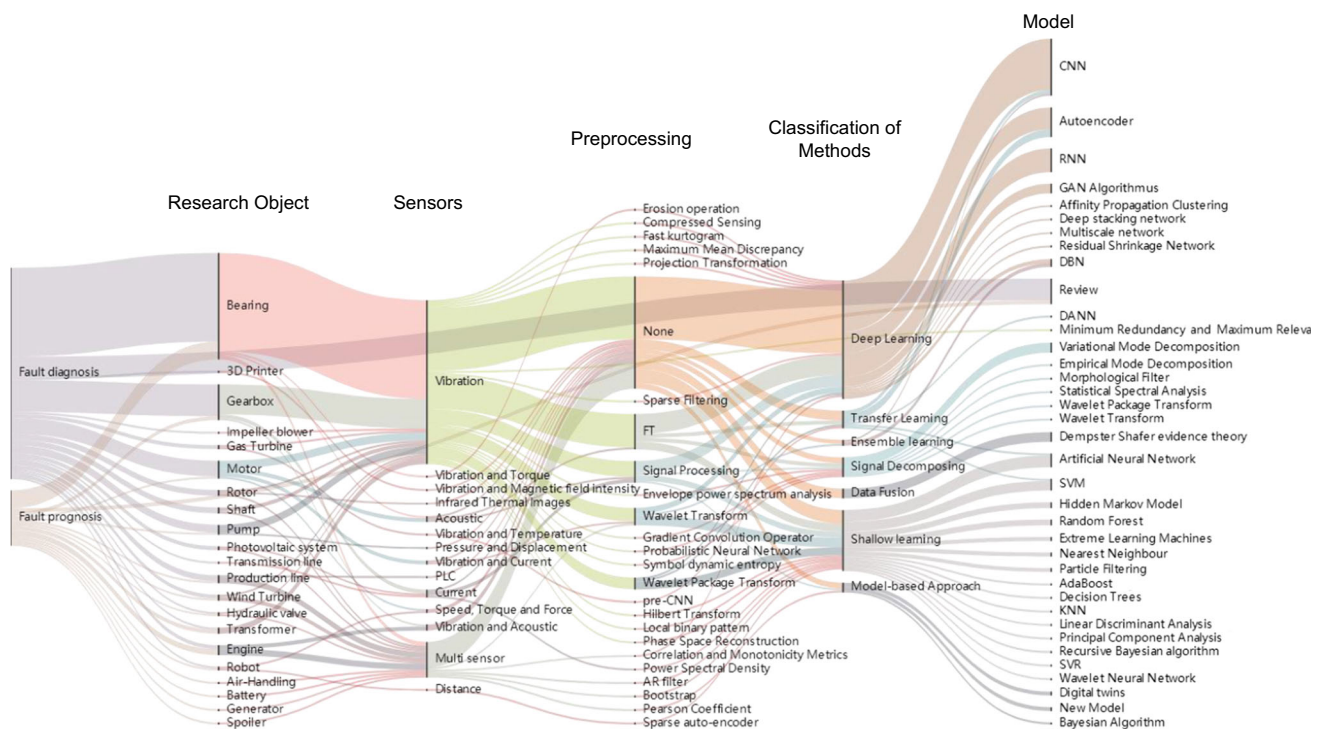


Fig. 7 Correlations among research objects, sensors, preprocessing methods, and methodology in fault diagnosis. Flow widths indicate frequency of use and colors distinguish different techniques

from wave packet transform, Fourier transform and other signal processing methods.

Benefiting from extensive studies in deep learning, convolutional neural networks, recurrent neural networks and auto-encoders are the most popular methods for fault diagnosis and prediction. Therefore, in Chapter 4, the authors will concentrate on an review of these three deep learning methods. According to the current research trend, the following conclusions can be drawn:

- Deep learning models are derived from image recognition and natural language processing. Although deep learning can adaptively mine potential features, the mined features are not necessarily the valid “fault features” for fault diagnosis (as deep learning models are designed to mine image/semantic features). Therefore, when deep learning is applied to fault diagnosis, its generalization ability needs to be verified.
- Signal processing methods, such as the Fourier transform and wave packet decomposition, are the most common pre-processing tools in fault diagnosis. This is due to the ability of signal processing to identify fault features in the frequency domain through time-frequency transformations.
- Future research should focus on fault mechanism analysis, i.e. what manifestations of specific faults are available and which sensors can be used for fault warn-

ing/diagnosis, especially for non-rotating machinery, while for rotating machines, the requirement for algorithmic interpretability is much more urgent than developing new models.

Data-driven approaches of fault diagnosis and health management

Equipment fault diagnosis and health management is generally achieved by analysing the time-series data in process and using machine learning algorithms to implement time-series prediction or pattern recognition of the collected data. The common algorithms currently applied for industrial data are shown in Table 5.

The development of Internet technology has enabled the collection and transmission of massive amounts of industrial data, which has raised the demand for understanding the correlation between data and the meaning of the data itself. For industrial big data, traditional shallow learning methods are less resource intensive, faster and relatively more interpretable. Shallow learning is autonomous but highly susceptible to errors. Deep learning is capable of adaptively extracting features from data, with the drawback of having more parameters and requiring high computational costs. Therefore, traditional machine learning methods (Zhou et al., 2020) are suitable for situations with low data volumes

Table 5 Common preventive maintenance algorithm

	Machine learning algorithm	Advantage	Disadvantage
Shallow learning	Logistic regression	<ol style="list-style-type: none"> 1. Easier to implement, interpret 2. Very efficient to train 	<ol style="list-style-type: none"> 1. The assumption of linearity between the dependent variable and the independent variables 2. Requires average or no multicollinearity between independent variables
	Artificial Neural Network	<ol style="list-style-type: none"> 1. Storing information on the entire network 2. Robust to noise in the training data 	<ol style="list-style-type: none"> 1. Unexplained behavior of the network 2. Assurance of proper network structure
	Support vector machine	<ol style="list-style-type: none"> 1. SVM more effective in high dimensional spaces 2. Effective in cases where the number of dimensions is greater than the number of samples 	<ol style="list-style-type: none"> 1. An Optimal kernel needs to be selected 2. Not suitable for large data sets 3. Does not perform very well when the data set has more noise i.e. target classes are overlapping
	Deep Belief Networks	<ol style="list-style-type: none"> 1. Only needs a small labeled dataset 2. Training time is fairly short on GPU powered machines 3. Very accurate compared to a shallow net 4. Is a solution to the vanishing gradient problem 	<ol style="list-style-type: none"> 1. Unexplained behavior of the network 2. Determination of proper network structure
	Convolutional Neural Network	<ol style="list-style-type: none"> 1. Very High accuracy in image recognition problems 2. Automatically detects the important features without any human supervision 3. Weight sharing 	<ol style="list-style-type: none"> 1. Data requirements leading to overfitting & underfitting 2. Parameter-to-memory requirements 3. Computationally expensive
Deep learning	Recurrent neural network	<ol style="list-style-type: none"> 1. An RNN remembers each and every information through time 2. Suitable for analyzing industrial time series data 	<ol style="list-style-type: none"> 1. Gradient vanishing and exploding problems 2. It cannot process very long sequences if using tanh or relu as an activation function
Other models	Hidden Markov Model	<ol style="list-style-type: none"> 1. HMMs are able to represent the variance of appliances' power demands through probability distributions 2. HMMs capture the dependencies between consecutive measurements, as defined by Hart as the switch continuity principle 	<ol style="list-style-type: none"> 1. The HMM needs to be trained on a set of seed sequences and generally requires a larger seed than the simple Markov models 2. The Viterbi algorithm is expensive, both in terms of memory and compute time
Visual analysis method	Cluster visualization	<ol style="list-style-type: none"> 1. Data labels are not required and the results are intuitive 	<ol style="list-style-type: none"> 1. Domain knowledge is required to further analyze the results 2. There is no standard method to determine model parameters

and obvious fault characteristics, while deep learning methods are suitable for situations with large data volumes and complex fault types. In addition, clustering algorithms such as self-organising mapping neural networks also play an important role in the field of fault diagnosis as a means of visual analysis of equipment deterioration trends. This chapter focuses on the applicable scenarios, characteristics and drawbacks of different prediction methods.

Shallow learning method of fault diagnosis

Regression Analysis is a method that studies the relationship between the independent and dependent variables and predicts the trend of the dependent variable by revealing the relationship between the independent and dependent variables. Pandya et al. (2014). extracted energy and Kurtosis features from ball bearing vibration signals and used multinomial logistic regression to identify local defects at the micron level. Zhang et al. (2018) proposed a fault diagnosis method based on multi-class Logistic regression for various types of bearing faults, and an example is given to verify that the proposed method can effectively classify the normal working state and fault types. Jiang and Yin (2017) investigates the practical difficulties of the vehicular CPSs online implementation, and based on that proposes a new recursive total principle component regression based design and implementation approach for efficient data-driven fault detection.

Support Vector Machines are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces, which plays an important role in the analysis of high frequency signals such as vibrations and energy.

SVM parameter selection is key for accurate model classification. Wumaier Tuerxun et al. (2021) uses the sparrow search algorithm(SSA) to optimize the penalty factor and kernel function parameter of SVM and construct the SSA-SVM model for wind turbine fault diagnosis. Since the fault diagnosis of wind turbines is a nonlinear problem, nonlinear mapping ϕ is used to map the data samples onto high-dimensional space for classification. The kernel function $K(x_i, y_i)$ is used to solve the problem of high-dimensional space. Meanwhile, a relaxation variable ξ_i ($\xi_i > 0$) is introduced to weight the classification plane, and a penalty factor C is introduced to weight the penalty degree of the relaxation variable ξ_i . Then, the optimization problem can be expressed as:

$$\min \phi(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (1)$$

The Lagrange multiplier α is introduced to transform the optimal hyperplane problem constructed by the constraints in Eq. (1) into a dual quadratic programming problem:

$$\min L(\alpha) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^n \alpha_i \quad (2)$$

By solving Eq. (2), the optimal classification decision function is finally obtained as:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n \alpha_i y_i K(x_i \cdot x_j) + b \right] \quad (3)$$

To obtain the appropriate kernel functions, these kernel functions are tested with relevant data. The results show that the classification achieves the best effect when the radial basis kernel function is used. $K(x_i \cdot x_j)$ can be expressed as:

$$K(x_i \cdot x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2g^2} \right) \quad (4)$$

The penalty factor C and kernel function parameter g must be given in the SVM model because the classification performance of the SVM is closely related to both C and g . Tuerxun proposes a parameter optimization method based on the sparrow search algorithm to select the appropriate penalty factor and kernel function parameter. The steps to optimize the relevant parameters are as follows: First, pre-process the original SCADA data, select the characters and normalize them, and import the data into the algorithm. Then, initialize SVM parameters C , g and the parameters of the SSA, including sparrow flock size N , the maximum number of iterations M , etc. Next, calculate the sparrow's position to obtain the current new position, and update the position if the new position is better. The classification accuracy of the SVM fault model is the current fitness value of each sparrow, and the maximum fitness value is updated in real-time. If the current fitness value of a sparrow is greater than the saved fitness value, the original fitness value is updated; otherwise, the original fitness value remains unchanged. Finally, save the corresponding SVM parameter combination (C , g) according to the optimal fitness value, and output the optimal parameter combination (C , g) and its corresponding classification results. The operational flow of the SSA-SVM model is shown in Fig. 8. Experiments show that the SSA-SVM diagnostic model effectively improves the accuracy of wind turbine fault diagnosis compared with the traditional SVM models and has fast convergence speed and strong optimization ability.

In addition, Han et al. (2019) proposes a least squares support vector machine model optimized by cross validation to implement fault detection on a 90-ton centrifugal chiller.

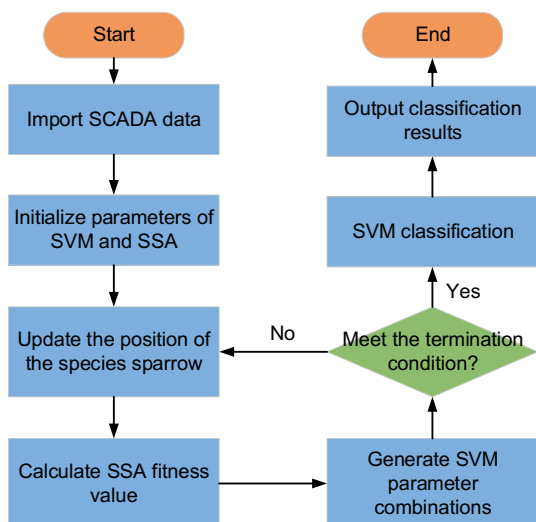


Fig. 8 Operational flow of the SSA-SVM model (Wumaier Tuerxun et al., 2021)

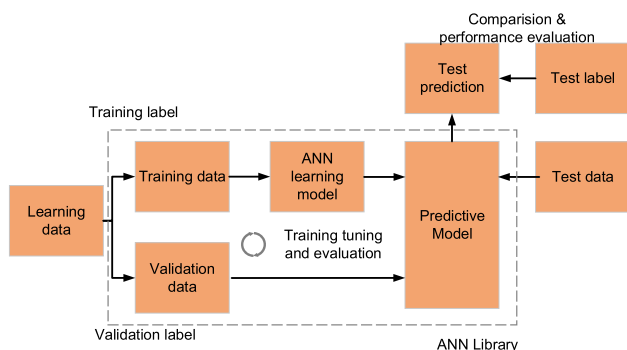


Fig. 9 Flowchart of the general training, validation, and test procedure of the proposed NN model (Wumaier Tuerxun et al., 2021)

Wang et al. (2020) proposed an intelligent fault diagnosis method is proposed for wind turbine rolling bearings based on Beetle Antennae Search based Support Vector Machine. The performance of the proposed fault diagnosis method was confirmed by conducting a fault diagnosis experiment of wind turbine rolling bearings.

Shallow Neural Networks is a term used to describe neural networks that usually have only few hidden layer as opposed to deep neural network which have several hidden layers, often of various types. Cho et al. (2021) developed a hybrid fault diagnosis algorithm based on a Kalman filter and artificial neural network to conduct effective fault diagnosis for the blade pitch system of wind turbines. One of the main parts of the proposed algorithm is estimation of the parameters of blade pitch systems, and these parameters indicate a healthy or faulty system using the Kalman filter with system input and output variables.

Figure 9 shows the flowchart of the training, validation, and test procedures. After the learning process including

training and validation procedures, the ANN library builds a predictive model that learns certain properties from a training dataset to make those predictions. During the training procedure at each hidden layer, the rectified linear unit (ReLU) function is used as the activation function. The activation function with ReLU prevents overfitting and the gradient vanishing problem. Equation (5) defines the function, where h_i is the i – th hidden layer value.

$$Relu(h_i) = \max(0, h_i) \tag{5}$$

Equation (6) shows the Softmax function $S(y_i)$, which is used for normalizing the values between 0 and 1, where the sum of the values is 1.

$$S(y_{NN,i}) = \frac{\exp(y_{NN,i})}{\sum_{j=1}^n y_{NN,j}}, i = 1, 2, \dots, n \tag{6}$$

The training uses the cross-entropy loss function $L_{CE}(S, y_d)$, where n is the number of fault types, S_i is the i – th value in the output of the Softmax function, and $y_{d,i}$ is the i – th desired value in the labeled data.

$$L_{CE}(S, y_d) = -\frac{1}{n} \sum_{i=1}^n (y_{d,i} \log S_i + (1 - \log(1 - S_i))) \tag{7}$$

The hybrid approach has many advantages because the Kalman filter involves less computational costs in the detection and the artificial neural network has high accuracy in the diagnosis procedure.

However, shallow networks are very good at memorization but not so good at generalization (Liu et al., 2021a). Iannace et al. (2019) built a model based on artificial neural network algorithms to detect unbalanced blades in propeller of unmanned aerial vehicles. Malik and Mishra (2017) presented an artificial neural network based condition monitoring approach of a wind turbine under six distinct conditions, i.e. aerodynamic asymmetry, rotor-furl imbalance, tail-furl imbalance, blade imbalance, nacelle-yaw imbalance and normal operating scenarios. Xianzhen et al. (2020) briefly gives the superiority of fuzzy neural network technology application in equipment fault diagnosis, and expounds the combinations of fuzzy theory and neural network technology.

Although shallow learning methods such as artificial neural networks are capable of handling fault identification and pattern classification tasks, in the face of increasingly complex equipment structures and expanding data volumes, shallow learning algorithms can no longer meet the requirements of intelligent fault diagnosis, and fault diagnosis research is moving towards deep neural networks with more hidden layers.

Fault diagnosis based on deep learning

The advantage of deep learning with more multiple layers is that they can learn features at various levels of abstraction, while Shallow Learning relies on handcrafted features based upon heuristics of the target problem. So they are much better at generalizing because they learn all the intermediate features between the raw data and the high-level classification (Ellefsen et al., 2019). Deep learning models that have been applied to the field of fault diagnosis include deep belief networks (DBN), convolutional neural networks (CNN), and recurrent neural nets (RNN).

Deep belief network A deep belief network is a generative graphical model, that is constructed of several layers of latent variables with connections between them but not between the units within each layer. DBNs are composed of basic, unsupervised networks such restricted Boltzmann machines (RBM) or autoencoders, with the hidden layer of each sub-network serving as the visible layer for the next. DBN can be trained layer by layer by greedy learning algorithms to obtain high level feature representations directly from the original signal, without the extraction of features manually when processing industrial signals, thus eliminating the loss of information caused by traditional manual feature selection and enhancing the self-adaptability of the recognition process (Miao et al., 2021).

DBNs can be used for both unsupervised and supervised learning. When used as an unsupervised learning algorithm, DBNs are similar to a self-coding machine that aims to reduce feature dimensionality while retaining as much of the original data as possible; from a supervised learning perspective, DBNs are often used as classifiers. Whether it is supervised or unsupervised learning, DBN is essentially a process of feature learning, i.e. how to get a better representation of the raw data.

For its benefits such as rapid inference and the capacity to represent deeper and higher order network structures, DBN has lately become a prominent technique in fault diagnosis. A typical example of a DBN as a classifier is that Tamilselvan and Wang (2013) proposed a multi-sensor health diagnosis methodology using DBN based state classification. The iterative learning process for one RBM unit, is depicted in Fig. 10. The synaptic weights and biases of all neurons in each RBM layer are initialized at the start of the training phase. The RBM unit will be trained periodically with the input training data after initialization. Transforming the states of neurons in the visible layer with matching synaptic weights and the biases of hidden layer neurons using a transformation function produces the states of neurons in the RBM hidden layer. There are two stages to each training epoch: a positive phase and a negative phase. The positive phase converts input data from the visible layer to the hidden layer, whereas the negative phase reconstructs the neurons from the previous visible

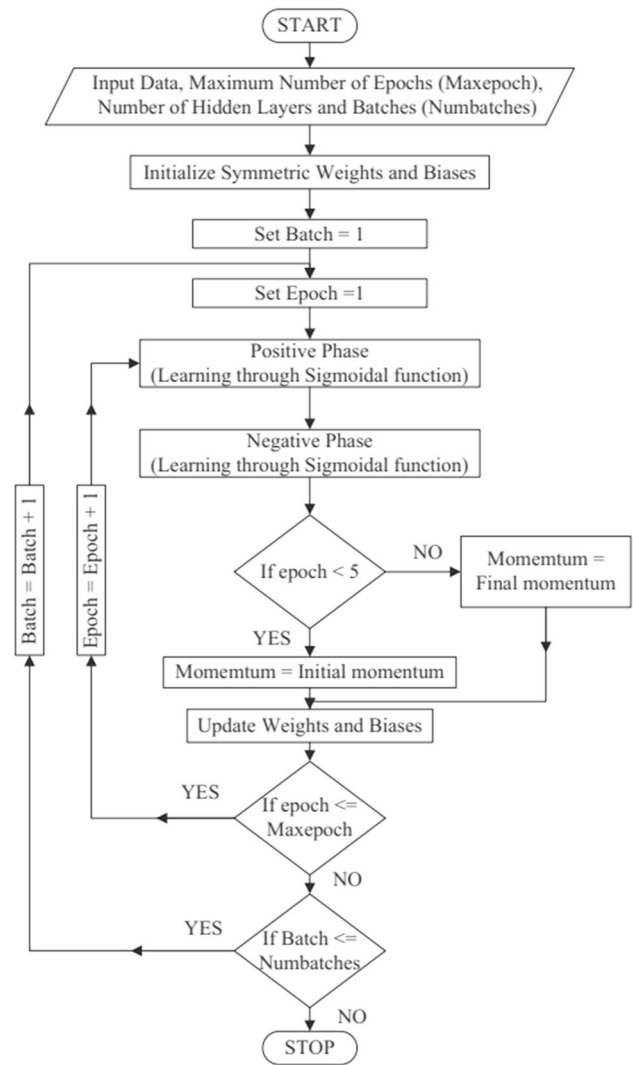


Fig. 10 Optimized RBM iterative learning process (Tamilselvan & Wang, 2013)

layer. Equations (8) and (9) represent the positive and negative phases of RBM learning analytically.

$$P(h_j = 1 | v) = \text{sigm} \left(-b_j - \sum_k v_k w_{jk} \right) C \quad (8)$$

$$P(v_k = 1 | h) = \text{sigm} \left(-b_k - \sum_j h_j w_{jk} \right) \quad (9)$$

where h_j and v_k are the states of the j th hidden layer neuron and the k th visible layer neuron, respectively. The positive phase learning process is represented by Eq. (8), in which the states of neurons in the hidden layer are determined by sigmoid transformation, whereas the negative phase learning process is represented by Eq. (9), in which the states of neurons in the visible layer are reconfigured. Synaptic weights

and biases may be adjusted based on state vectors of neurons in both hidden and visible layers after the learning process for both positive and negative phases (Hinton, 2012). The synaptic weight update, w_{jk} , may be represented as

$$\Delta w_{jk} = \delta (\langle v_k h_j \rangle_{\text{data}} - \langle v_k h_j \rangle_{\text{recon}}) \quad (10)$$

where δ denotes the learning rate; $\langle v_k h_j \rangle_{\text{data}}$ denotes the pairwise product of the state vectors for the j th hidden layer neuron and the k th visible layer neuron after the positive phase learning process, whereas $\langle v_k h_j \rangle_{\text{recon}}$ denotes the pairwise product after the negative phase learning process for visible layer reconstruction. The results of the aero engine research demonstrate that the suggested proposed DBN-based diagnosis technique provides better accurate classification rates in general. This is mostly due to its capacity to learn high-complexity non-linear connections between input sensory signals and various health states by embedding richer and higher-order network structures in the deep learning process, which may be performed both supervised and unsupervised.

Furthermore, Ma et al. (2017) proposed a novel method based on discriminative deep belief networks (DDBN) and ant colony optimization (ACO) to predict health status of bearing. Chen et al. (2018) applied a deep belief network with various parameters, such as cutting force, vibration, and acoustic emission (AE) to predict the flank wear of a cutting tool. Qin et al. (2018) applied an intelligent and integrated approach based on deep belief networks (DBNs), improved logistic Sigmoid (Isigmoid) units, and impulsive features to solve the vanishing gradient problem, which is an inherent drawback of conventional Sigmoid units, and usually occurs in the backpropagation process of DBNs, resulting in that the training is considerably slowed down and the classification rate is reduced.

When DBN is used as an unsupervised learning algorithm, Dong et al. (2017) proposed a deep hybrid model named Stochastic Convolutional and Deep Belief Network (SCDBN), which assembles unsupervised CNN with DBN, to realize the goal of bearing fault diagnosis. An improved unsupervised deep belief network (DBN), namely median filtering deep belief network (MFDBN) model is proposed by Xu et al. (2020) through median filtering (MF) for bearing performance degradation. For the unsupervised fault diagnosis of a gear transmission chain, He et al. (2017) developed a new deep belief network. In comparison to traditional methods, the suggested method uses unsupervised feature learning to adaptively exploit robust characteristics associated to defects, requiring less previous knowledge of signal processing techniques and diagnostic expertise.

Compared with traditional fault diagnosis methods, a large amount of fault diagnosis experience is not required in DBN, and it is independent of the subjectivity of human extraction

of fault features and can achieve an adaptive extraction of fault features. In addition, the method does not require a strict periodicity of the signal when processing timing signals, and has a strong generalization capability.

Convolutional Neural Network CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons are typically completely connected networks, meaning that each neuron in one layer is linked to all neurons in the following layer. In comparison to other image classification algorithms, CNNs need much less pre-processing for handling industrial large data. This means that the network learns to optimize the filters (or kernels) by automatic learning, as opposed to hand-engineered filters in traditional methods. On the other hand, because of their “full connectivity” these networks are prone to data overfitting.

When CNNs are applied to fault diagnosis, a high learning rate could increase loss error and could also lead to excessive weight fluctuations. Guo et al. (2016) proposed a novel hierarchical learning rate adaptive deep convolution neural network based on an improved algorithm, and investigated its use to diagnose bearing faults and determine their severity.

To extract the most typical features and obtain the best recognition results, Guo introduced an adaptive learning rate. For each w_{ij}^l (the weight of the kernel connecting the i th feature map of the $l-1$ th layer with the j th feature map of the l th layer) update, the parameter α or β is used to update the learning rate, respectively, as follows:

$$\lambda = \begin{cases} 1 & \text{if } \text{sgn} \left(\text{grad}_{ij}^k \text{grad}_{ij}^{k-1} \right) = 1 \\ -1 & \text{if } \text{sgn} \left(\text{grad}_{ij}^k \text{grad}_{ij}^{k-1} \right) = 0 \end{cases} \quad (11)$$

$$r(k) = \alpha^\lambda r(k-1) \quad (12)$$

where grad_{ij}^k denotes the weight gradient after the k th training. Considering λ is either -1 or 1 , the value of α should be as $1 < \alpha < 2$ to ensure that α^λ is within the range of $(0-1)$ or $(1-2)$. The function $\text{sgn}(o)$ is used to determine the orientation of two adjacent gradients. As a result, when the loss function drops over time, the learning rate rises to speed up convergence; as the loss function rises too quickly, the learning rate decreases. The vanishing (or exploding) gradient problem (Bengio, 2009) is avoided with this technique, which explains why many deep-learning models fail when trained using conventional back-propagation networks.

Without considering experience, i.e. the gradient orientation of the preceding moment, the gradient-descent algorithm always updates weight, w_{ij}^k , toward a negative gradient direction. During training, this frequently results in oscillation and delayed convergence. As a result, a momentum item has been introduced. As a result, weight in terms of loss-function gra-

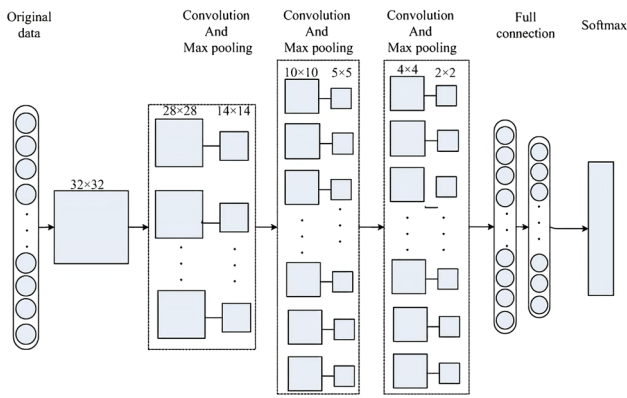


Fig. 11 Architectural hierarchy of the proposed Adaptive CNNs (Guo et al., 2016)

dent is calculated as follows:

$$W_{ij}^k = W_{ij}^{k-1} - r(k) \left[(1 - \beta) \frac{\partial E}{\partial W_{ij}^k} + \beta \frac{\partial E}{\partial W_{ij}^{k-1}} \right] \quad (13)$$

According to the weight distribution of the last two training effects, β should be in the range 0–1. Figure 11 illustrates the proposed Adaptive CNN’s architecture. Each plane is a feature map with a set of units that must have their weights determined. There are seven primary layers in the proposed structure. The signal-vector input is converted into a matrix in the first layer, which is the standard CNN input format. CNN layers are the following three layers, each with a convolution layer and a max-pooling layer. In the first, second, and third CNN layers, there are 5, 10, and 10 feature filters, respectively. Full connection layers are the following two layers, and they prepare characteristics for categorization. The final layer is a logistic-regression layer that employs the softmax classification algorithm. Each layer’s weight is initialized at random and then trained for optimization. The test samples are sent into the logistic-regression layer after training, and the output is a set of class labels corresponding to the samples.

In other studies, Chen et al. (2019) propose an effective and reliable method based on CNNs and discrete wavelet transformation (DWT) to identify the fault conditions of planetary gearboxes, considering that planetary gearbox vibration signals are highly non-stationary and non-linear due to wind turbines (WTs) often working under time-varying running conditions. Wang et al. (2016) presents an intelligent fault diagnosis method based on convolution neural network for rotor system. Using envelope spectrums and a convolutional neural network, Appana et al. (2018) presents an acoustic emission analysis-based bearing defect detection that is invariant to rotational speed variations using CNNs. For fault diagnosis, Wen et al. (2017) suggested a novel CNN based on LeNet-5, which extracts the character-

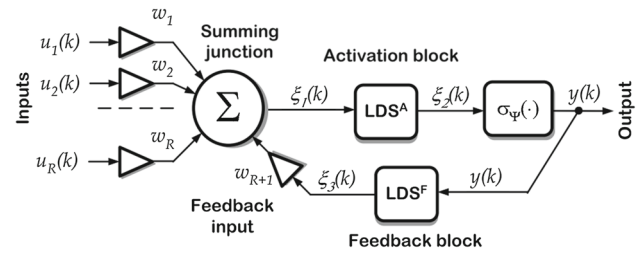


Fig. 12 Dynamic neural unit with linear dynamic systems (LDS) embedded in the activation and feedback blocks (Przystała & Moczulski, 2015)

istics of converted 2-D pictures and eliminates the influence of handmade features using a conversion method that converts signals into two-dimensional (2-D) images. Islam and Kim (2019) presents an adaptive deep convolutional neural network that uses a two-dimensional visualization of the raw acoustic emission data to give bearing health condition information, allowing the feature extraction and classification procedure to be automated and generalized.

In general, CNN models generally perform superior compared to the manual extraction process. However, it takes a very long time to train a convolutional neural network, especially with large industrial datasets. Besides, while CNNs are translation-invariant, they are generally bad at handling rotation and scale-invariance without explicit data augmentation.

Recurrent neural network The connections between nodes in a recurrent neural network (RNN) create a directed graph along a temporal sequence. This enables it to behave in a temporally dynamic manner. RNNs, which are derived from feedforward neural networks, can handle variable length sequences of inputs by using their internal state (memory). This makes RNNs particularly suitable for industrial time-series data (Abiodun et al., 2018; Chen et al., 2019).

In the industrial field, Przystała and Moczulski (2015) utilize recurrent neural networks and chaotic engineering to solve the challenge of resilient defect detection. The highlights of the method are a modified neural processor that can be investigated as an extension of two well-known fault diagnosis units, namely a dynamic neural unit with an infinite impulse response filter in the activation block and a dynamic neural unit with an infinite impulse response filter in the detection block (Patan, 2008) and Frasconi Gori Sodas neuron (Frasconi et al., 1992). The unit’s scheme is seen in Fig. 12.

The properties of this mechanism can be described by the following equation:

$$\xi_1(k) = \sum_{i=1}^{r_n} w_i u_i(k) + w_{(r_n+1)} \xi_3(k) \quad (14)$$

where $w_1, w_2, \dots, w_{r_n+1}$ are the external and feedback input weights, $u_i(k)$ is the unit's i -th external input, and $\xi_3(k)$ is the neuron's feedback state. The activation state of the unit at discrete time k is defined as Eq. (15) using the backward shift operator q^{-1} .

$$A(q^{-1})\xi_2(k) = B(q^{-1})\xi_1(k) + C(q^{-1})\phi_A(k) \tag{15}$$

In the same approach, the state in the feedback block can be written as:

$$D(q^{-1})\xi_3(k) = E(q^{-1})y(k) + G(q^{-1})\phi_F(k) \tag{16}$$

Finally, at discrete time k , the output of the unit is derived from:

$$y(k) = f(\xi_2(k) + b) \tag{17}$$

where $A(q^{-1}), B(q^{-1}), \dots, G(q^{-1})$ denotes polynomials of the delay operator q^{-1} , terms ϕ_A, ϕ_F represents random processes with mean values μ_a, μ_a , and the neuron's output function is δ_A^2, δ_F^2 , and the bias is b . When compared to previous artificial neurons in the literature, this simple processor has two novel properties. For one instance, the neuron's output may be stochastic; for another, the neuron could be defined exclusively in deterministic terms, but its behavior would be impossible to anticipate. For industrial data, a preliminary verification of the established approach in modeling tasks was carried out. The obtained findings support the efficiency of the recommended strategy.

In other researches, Zhang et al. (2021) presented a new approach for identifying defect types in rotating equipment based on recurrent neural networks. The Gated Recurrent Unit (GRU) is developed to accommodate use of temporal information and learn representative characteristics from the raw data. Shahnazari (2020) provides a fault detection and isolation (FDI) technique that allows for the diagnosis of single, multiple, and concurrent actuator and sensor failures. Qiao et al. (2020) introduced a dual-input model based on long-short term memory (LSTM) neural network to investigate the difficulties of rolling bearing defect detection under various sounds and loads. Jung (2020) established an automated residual design for grey-box recurrent neural networks based on a bipartite graph representation of the system model, which he tested using a real-world case study. The potentials of integrating machine learning with model-based fault diagnostic approaches are demonstrated using data from an internal combustion engine test bench. In the form of an autoencoder, Liu et al. (2018) developed a unique approach for bearing defect detection using RNN.

Deep transfer learning Deep Transfer Learning (DTL) is a new machine learning paradigm that combines the advantages of Transfer Learning in knowledge transfer with the

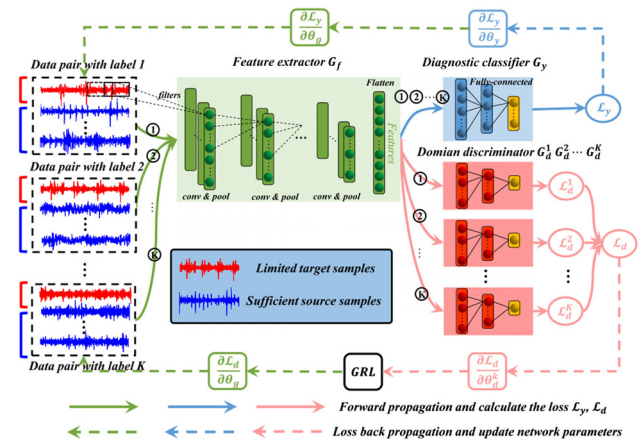


Fig. 13 Deep transfer learning structure for machinery fault diagnosis (Przystała & Moczulski, 2015)

advantages of Deep Learning in feature representation. As a result, DTL techniques have been widely developed and researched in the field of intelligent fault detection, and they may improve the reliability, robustness, and adaptability of DL-based fault diagnostic methods. Deep transfer learning research for mechanical fault diagnosis can overcome the limitations of large amounts of training data and accelerates the implementation of diagnostic algorithms (Li et al., 2022).

Han et al. (2021) presents a novel paradigm for resolving the problem of sparse target data transfer diagnostics. The fundamental concept is to match the source and target data with the same machine condition and perform individual domain adaptation to address the shortage of target data, reduce distribution discrepancy, and avoid negative transfer. The detailed architecture of the model is shown in the Fig. 13.

First of all, Both the massive source samples $\{(x_i^S, y_i^S)\}_{i=1}^{N_S}$ and sparse target samples $\{(x_i^T, y_i^T)\}_{i=1}^{N_T}$ are fed into the feature extractor G_f for achieving abstract feature representation $f = G_f(x)$. Then, the extracted features are sent to the classifier G_y and multiple domain discriminators $G_d^k, k = 1, 2, \dots, K$, respectively. Once the network parameters are initialized, the source samples along with the sparse target samples are used to train the classifier G_y . The classification loss of the G_y is denoted as:

$$L_y = \frac{1}{N_s} \sum_{x_i \in D_s} L_y(G_y(G_f(x_i)), y_i) + \frac{1}{N_t} \sum_{x_i \in D_t} L_y(G_y(G_f(x_j)), y_j) \tag{18}$$

In the process of distribution alignment, pairing data from separate domains but with the same machine condition effectively compensates for the unclear feature distribution provided by the sparse target data. As a result, with target data, a trained classifier with both source and target supervi-

sion has sufficient generalization potential. The domain loss of all K domain discriminators is defined as:

$$L_d = \frac{1}{N_s + N_t} \sum_{k=1}^K \sum_{x_i \in D_s \cup D_t} L_d^k(G_d^k(G_f(x_i)), d_i) \quad (19)$$

where G_d^k is the k th domain discriminator associated with label k , L_d^k denotes the binary cross entropy loss (BCE loss) for G_d^k . d_i is the domain label. The BCE loss L_d^k for G_d^k is formulated as:

$$\begin{aligned} \mathcal{L}_d^k(G_d^k(G_f(x_i)), d_i) &= -d_i \log(G_d^k(G_f(x_i))) \\ &\quad - (1 - d_i) \log(1 - G_d^k(G_f(x_i))) \end{aligned} \quad (20)$$

Instead of overfitting, adversarial training inside a data pair of the same machine condition encourages the use of transferable features from relevant source data and the learning of an adapted feature representation for target tasks. More importantly, this method considerably reduces the poor adaptation induced by major distribution shifts and label space mismatching.

Furthermore, for fault diagnosis in building chillers, Liu et al. (2021) suggests a transfer-learning-based technique with consideration of different transfer learning tasks, training instances, and learning situations. Li et al. (2021) suggested a fault diagnostic approach for wind turbines with small-scale data based on parameter-based transfer learning and convolutional autoencoder. Knowledge from comparable wind turbines can be transferred to the target wind turbine using the suggested technique. Qian et al. (2021) proposes a novel deep transfer learning network based on a convolutional auto-encoder to execute mechanical fault diagnosis in the target domain without labeled data, considering the influence of noise on transfer fault diagnosis. The suggested transfer learning model outperforms the competition in terms of anti-noise performance.

Fault diagnosis based on data fusion

Several approaches for data fusion are presented in the literature (Wang et al., 2019). Data fusion methods can be classified into three categories: data layer fusion, feature layer fusion, and decision layer fusion. Data layer fusion is a method that directly integrates raw data such as sensors and PLC, which aims at performing data fusion before industrial data is analyzed. Kalman filtering is a common data layer fusion method. Kalman filtering (KF) is an algorithm that uses a series of measurements observed over time, including statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate

than those based on a single measurement alone, by estimating a joint probability distribution over the variables for each timeframe. Willner et al. (1976) examine several Kalman filter algorithms that can be used for state estimation with a multiple sensor system. El Madbouly et al. (2009) propose a new adaptive Kalman filter-based multisensor fusion to satisfy the real time performance requirements.

The feature layer data fusion method first performs feature extraction on the original data, and then performs fusion on the extracted features. The advantage of this approach is information compression. When combined with edge computing technology, data cleaning and feature extraction are usually performed at the “edge” side (i.e., the device side), which can effectively save data storage and transmission costs; the deficiency is, that the information compression process can lead to information loss. Feature layer fusion techniques are mainly used in neural networks, feature clustering and other pattern recognition tools. Literature (Shao et al., 2017) proposed an enhancement deep feature fusion method for rotating machinery fault diagnosis. In the experiment, locality preserving projection (LPP) is adopted to fuse the deep features to further improve the quality of the learned features, and the developed method is applied to the fault diagnosis of rotor and bearing.

Decision-level data fusion is a distributed data fusion approach. Data cleaning, feature extraction and pattern recognition are first performed separately for each data source. Then the initial conclusions related to the faults from each data source are given independently. Then the final determination is made at the decision level by information fusion.

- Bayes’ theorem (Dimaio et al., 2021; Ji et al., 2022; Kim et al., 2021), and fuzzy set theory are all traditional approaches to decision-level data fusion. Bayes’ theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. With Bayesian probability interpretation (Amin et al., 2018), the theorem expresses how a degree of belief, expressed as a probability, should rationally change to account for the availability of related evidence. Literature (Zhang & Ji, 2006) propose an information fusion framework based on dynamic Bayesian networks to provide dynamic information fusion in order to arrive at a reliable conclusion with reasonable time and limited resources.
- Dempster–Shafer theory (D–S theory) is a general framework for reasoning with uncertainty, with understood connections to other frameworks such as probability, possibility and imprecise probability theories. Ji et al. (2020) developed an intelligent fault diagnosis approach based on Dempster–Shafer (DS) theory specifically for detecting several faults occurred in hydraulic valves. Gong et

al. (2018) presented a new fault diagnosis method for the main coolant system of nuclear power plant based on D–S evidence theory. However, the computational complexity of D–S theory grows exponentially with the number of data sources (Chen & Que, 2004).

- Fuzzy logic is a form of many-valued logic in which the truth value of variables may be any real number between 0 and 1. It is employed to handle the concept of partial truth, where the truth value may range between completely true and completely false (Su & Bougiouklis, 2007). The theory allows one to combine evidence from different sources and arrive at a degree of belief (represented by a mathematical object called belief function) that takes into account all the available evidence. However, how to elicit fuzzy data, and how to validate the accuracy of the data is still an ongoing effort strongly related to the application of fuzzy logic. Demidova et al. (2021) suggests that new defect detection algorithms based on fuzzy logic approaches should be developed to assist designers of energy conversion systems to establish reliability factors for apparatus such as electrical machines and power electronics subsystems. In addition, Mangai et al. (2010) conducted a Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification.

Fault diagnosis based on visual analysis of data

Usually, the process of data analysis is often inseparable from the mutual collaboration and complementary strengths of machines and humans. From this point of view, theoretical research on big data analysis can be developed from two perspectives: firstly, from a computer-centric perspective, where the efficient computing power of computers is emphasised and various advanced data mining methods are developed and applied. For example, big data processing methods based on Predix and MindSphere Industrial cloud platform (General Electric Company, 2015; Germany, 2016; Belhadi et al., 2019) and various data analysis methods for industrial big data, which are currently the mainstream methods for industrial big data analysis; the other perspective considers the human as the subject and emphasises analysis methods that conform to human cognitive laws, with the intention of incorporating cognitive abilities possessed by humans, which machines are not specialized in, into the analysis process. This branch of research is represented by the visual analytics of big data (Keim et al., 2013; Li & Zhou, 2017; Andrienko et al., 2020).

Self-organizing mapping neural network Self-organizing map (SOM) is an unsupervised machine learning technique used to produce a low-dimensional representation of a higher dimensional data set while preserving the topological structure of the data. An SOM is a type of artificial neural network

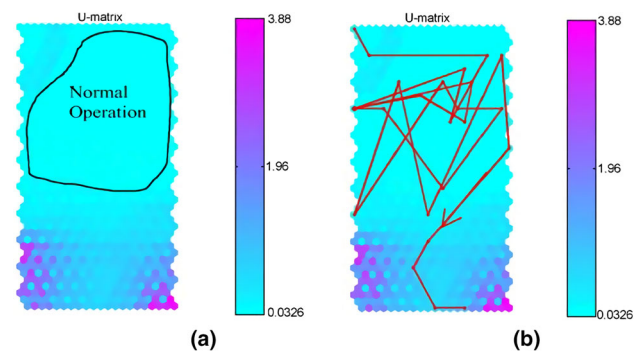


Fig. 14 The U-matrix of SOM and trajectory of degradation process of bearing

but is trained using competitive learning rather than the error-correction learning.

Zhong et al. (2005) focuses on the application of self-organizing maps in motor bearing fault diagnosis and presents an approach for motor rolling bearing fault diagnosis using SOM neural networks and time/frequency-domain bearing analysis. Peng et al. (2007) constructed a self-organizing neural network model to achieve automatization of fault diagnosis by automatically clustering dynamometer cards and solved dynamometer cards auto recognition problem in suck rod pumping system. Bossio et al. (2017) presented a SOM scheme for the identification and quantification of faults in induction motors. From the clusters generated by the neural network, the quantization error of each cluster is used to determine the fault magnitude. Wei-Peng and Yan (2019) proposed a method based on kernel Fisher vector and self-organizing map networks (KFV-SOM) to improve the visualization of process monitoring. Result shows that KFV-SOM can effectively visualize monitoring, and it is capable of showing the operation states of normal state and different kinds of faults on the output map of the SOM network.

An example of bearing fault diagnosis results based on SOM is shown in Fig. 14. The trained SOM's U-matrix map is shown in Fig. 14. Three clusters can be detected by observing at the gray color difference on the U-matrix map, as shown in Fig. 14. Those clusters found in the U-matrix can be linked to three stages in the run-to-failure test: normal, deterioration, and failure. The usual working state might be cluster one, the largest cluster in the upper left region.

Radviz Visualization RadViz is a multivariate data visualization algorithm that plots each feature dimension uniformly around the circumference of a circle then plots points on the interior of the circle such that the point normalizes its values on the axes from the center to each arc. This mechanism allows as many dimensions as will easily fit on a circle, greatly expanding the dimensionality of the visualization.

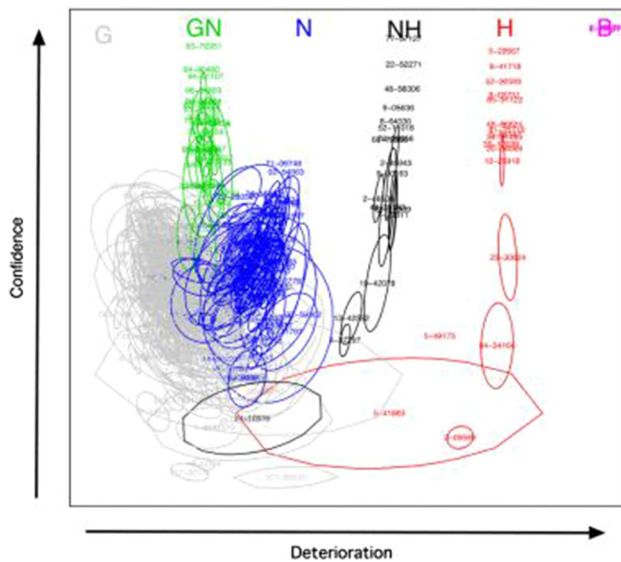


Fig. 15 Compared health of compressors in a fleet. *G* good, *GN* good to normal, *N* normal, *NH* normal to high deterioration, *HD* high deterioration and *B* bad (Martínez et al., 2013)

Novikova et al. (2019) suggests a visualization-driven approach to the analysis of the data from heating, ventilating and conditioning system. XU et al. (2009) investigated radial visualization(Radviz) technique, and This study demonstrates how basic two-dimensional graphs like the Radviz graph may successfully distinguish diagnostic groups in the Tennessee Eastman process. Martínez et al. (2013) proposed a Radviz-based algorithm for assessment of Engine Health Monitoring data in aircraft. By extending the Radviz visualization technique to fuzzy data, the activation of the rules is represented in a 2D map, as shown in Fig. 15.

Visual analysis of workshop data An example of visual analysis based on shop floor data is the introduction of the Marey graph by Panpan et al. (2016) to visualise shop floor data to reflect information on faulty workstations. The Marey graph is a classic representation of bus or railway timetables. A parallel arrangement of time axes is used. A rail or bus stop is represented by each time axis. According to the timetables, polylines linking the time points on the axes show when the buses or trains are anticipated to arrive at a stop. If the authors view each work station on the assembly line as a bus or train stop, and the time when the components are transferred onto each work station as the time in bus or train timetables, this visual encoding may be easily applied to manufacturing processes data, as shown in Fig. 16. Out-of-order processes are visually represented by line segments crossing each other between the time axes, and abnormal delays are visually indicated by line segments stretching considerably longer than the others between two time axes in Marey's graph.

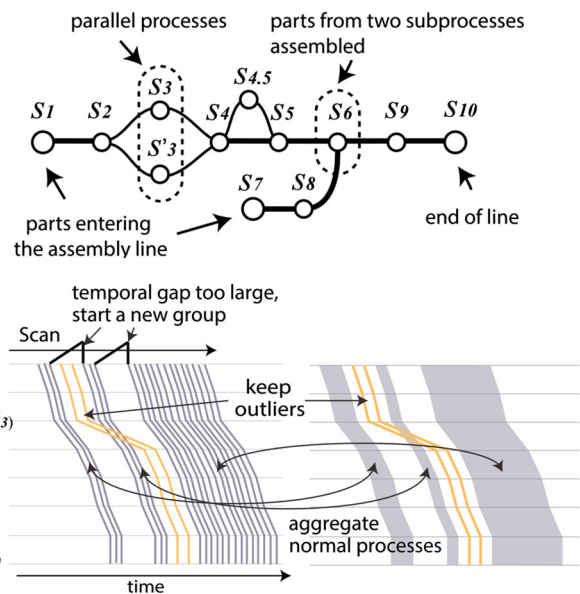


Fig. 16 The schematic view of an assembly line as a directed acyclic graph (DAG). The parts are moved among the stations following predefined paths(upper). Marey's graph and the relevant visual patterns when applied to the manufacturing process data(lower) (Panpan et al., 2016)

Case study

In this section, the authors demonstrate two representative cases from our previous scientific research as references for the applications of the above theories.

Support vector machine based fault prognosis of shredder

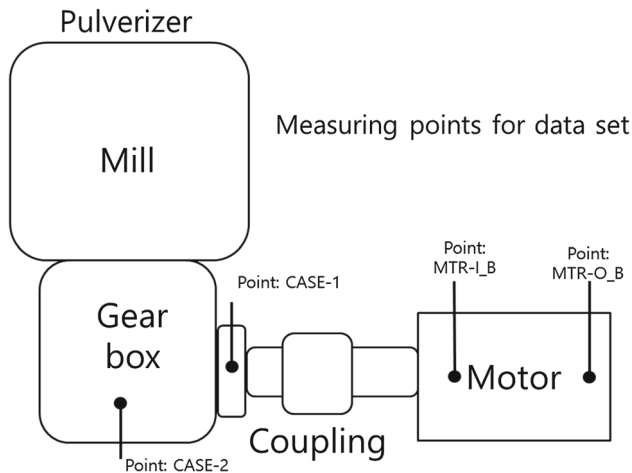
Steam power generation requires the burning of coal as an energy source. In order to make full use of coal, it is necessary to crush it.

The research object of this case is a shredder used to grind coal. The power source of the shredder comes from an electric gearbox. As the transmission mechanism of the shredder, the motor and gearbox of the shredder are prone to failure. In this case, three types of data of the gearbox of the shredder are collected: vibration acceleration data, displacement data and velocity data.

The types and locations of the sensors are illustrated in Fig. 17: Body-CASE-1 sensor is placed on the coupling, Body-Case-2 sensor corresponds to the gearbox. $MTR - I_B$ and $MTR - O_B$ sensors collect data from different locations of the motor respectively. Body-Case sensors can collect acceleration and velocity data, $MTR - I_B$ sensors collect acceleration and displacement data, and $MTR - O_B$ data can only collect displacement data. There are three types of acquisition frequencies: 7680, 2049, and 5128, they correspond to acquisition durations of 0.533 s, 2 s, and 0.88 s,

Table 6 Data set of pulverizer A

Sensor type	Sensor location	Number of data sets	Duration of collection for each data set
Acceleration	Body-CASE-1	28	0.533/2
	Body-CASE-2	27	0.533/2
Displacement	MTR-I_B	14	0.533
	MTR-I_B	14	0.8
	MTR-O_B	14	0.8
Velocity	Body-CASE-1	15	0.533
	Body-CASE-2	15	0.533
	MTR-I_B	14	0.533

**Fig. 17** Sensors of the shredder

respectively. The detailed information of the data set is shown in Table 6.

The dataset is provided for a period of two years. In terms of data type, the dataset belongs to high-frequency sensor data. Therefore, based on the correspondence between data types and processing modes in Table 2, a stream processing model was developed to identify the operational state of the pulverizer in real-time. There are two major challenges when dealing with this dataset: fusion of data from multiple sources and imbalance of positive and negative samples. In terms of data fusion, the data is derived from four different sensors and contains three data types (acceleration/velocity/displacement). These sensors are located on three different components (gearbox/coupling/motor). In terms of data types, the number of normal operation days in the training data set is 100, but the fault data are only from the 81st day before the fault until the last day before the fault for fault warning, the normal and fault samples ratio is almost 50:1. In this case, the dimensionality of the data is relatively high and the stream processing requires high efficiency in classifying the data. Considering that SVM works relatively well when there is a clear margin of separation between classes and is more effective in high dimensional

spaces, the authors employ SVM as a classifier that classifies each group of sensor data separately and develops data fusion rules for fault diagnosis.

Solution and Processing flow

- **Data pre-processing:** According to the time and space span, data type and collection frequency of the data set, data cleaning is first required. The main purpose of data cleaning is to divide the data set according to the type of the data and sensors, collection frequency. For example, Body-CASE-1 sensor collects acceleration data with the collection frequency of 0.533/2s; the velocity data with the collection time of 0.533s, in this way, it can be divided into three classes of data: Body-CASE-1-acceleration-0.533s, Body-CASE-acceleration-2s and Body-CASE-velocity-0.533s. Similarly, the entire data set can be divided into 10 subsets.
- **Feature extraction:** The data set included the time series data of acceleration, velocity and displacement. The maximum value, mean value, peak value, absolute mean value, root mean square value, variance, crest factor, kurtosis factor, kurtosis index, form factor, impulse index, margin coefficient, skewness, wavelet energy (wavelet basis function 'db3', three-layer decomposition) are chosen as the features. Then combined with the sliding time window method to convert the time domain signal into a 19-dimensional feature vector as the feature vector of the original signal.
- **Classifier training:** Support vector machine (SVM) is chosen as the base learner in the integrated classifier, and trains respectively 10 weak classifiers for 10 subsets. For the problem of unbalanced data sets, "smote oversampling" is used to expand the set of counterexamples and adjust the penalty coefficients of positive and negative examples of the SVM classifier. SVM uses a linear kernel, because the Gaussian kernel has strong fitting ability, is prone to overfitting, and fails to identify the failure date for some test sets. At the same time, the posterior probability output by the SVM classifier is used as a proven basis for fault diagnosis.

Table 7 Failure prediction results of pulverizer A

Pulverizer A	Body-case-1 acceleration/2s	Body-case-1 acceleration/0.533s	Body-case-2 acceleration/0.533s	Body-case-2 acceleration/2s	MTR-I_B displacement	MTR-I_B displacement	MTR-O_B displacement	Body-case-1 velocity/2s	Body-case-2 velocity/0.533s	MTR-I_B velocity
Number of eigenvectors	29	7	7	29	7	11	11	7	7	7
Positive example	0	7	7	11	7	11	5	0	7	7
Negative example	29	0	0	18	0	0	6	7	0	0
Posterior mean of positive example	1.00E-07	0.018809673	0.994385122	0.369382652	0.999970592	0.943131466	0.5346611	0.01692006	0.952166752	0.999999325
Posterior mean of negative example	0.9999999	0.981190327	0.005614878	0.630617348	2.94E-05	0.056868534	0.4653389	0.98307994	0.047833248	6.75E-07
Classification result	Fault	Fault	Normal operation	Fault	Normal operation	Normal operation	Normal operation	Fault	Normal operation	Normal operation

- Comprehensive fault diagnosis: After the model training, the real-time data is as input into the combined classifier. In each weak classifier, a relative majority voting method is adopted, and the recognition result of the majority in the feature vector is used as the recognition result of the classifier on that day. The judgment principles are as follows: 1. If 4 or more of the 10 classifier results are counterexamples, it is considered that the equipment is about to fail on that day; 2. If 3 or more of the 10 classification results are counterexamples, at the same time at least two classifiers of the counterexamples are from the same sensor, it can be determined that the device is about to fail, especially the sensor corresponding component is about to fail. Table 7 shows the classification results of shredder A on May 18, 2012. Among the 10 classifications, 4 of them are counterexamples. According to principle 1, it can be preliminarily judged that the equipment is about to fail, and 3 of them are in the position of Body-Case-1, which means that the coupling is about to fail.

Deep learning network based expert system for fault prognosis

The research object of the case is the Siemens SINUMERIK808D CNC system. Through knowledge representation of the related cases and one intelligent fault reasoning algorithm is developed. The objective of this case is to quickly match similar cases based on the input of machine failure phenomenon. On the basis of more than 3000 failure cases provided by Siemens, the author divided the failure cases into 7 types, which are NC system failure, input and output module failure, servo system failure, PLC failure, cable failure, tool changer failure, CNC machine tool failure, among them part of cases of NC system failure are shown in Table 8.

The dataset comes mainly from the manufacturing fault execution system (MES). Based on Table 2, the authors developed an expert system based on a batch process for fault prognosis with a back-propagation neural network as the interpreter algorithm and a cyclic neural network as the inference engine algorithm. The inference engine evaluates the current state of the knowledge base, applies relevant rules, and then performs fault inference. The explanation facility can explain to a user the chain of reasoning used to arrive at a particular conclusion by tracing back over the firing of rules that resulted in the assertion.

The core matching mechanism of the expert system is composed of an interpreter and an inference engine. The interpreter is composed of a four-layer neural network. After the input layer, the information passes through two hidden layer with ReLU as excitation function, and finally is sent to the output layer with Softmax as excitation function. The neu-

Table 8 Failure records of machine tool

Fault phenomenon	Fault position	Fault cause	Solution
NC system cannot be activated	Software	loss of NC data	Upgrade the system and restore the new data
Abnormal pulse signal voltage at the X53 terminal of the PPU	PPU	Fault in X53 terminal	Replace PPU
MCP's multiplier switch shows 0%	MCP	The MCP's multiplier switch has failed	Replace MCP
The x-axis drive displays an A05 alarm and the x-axis cannot be moved	Cable	Short circuit in the cable of the motor powering the X-axis	Replacing power cables
Drive failure, power cable and motor cannot be separated	Motor	Water in the motor, oxidation of the joints	Motor replacement
There is no voltage at terminal 65 of the drive	circuit board	False connection of the circuit board	Fastening of the circuit board
Collision of machine tools	Z-axis	The Z-axis zero return block is loose, resulting in a deviation of the machine coordinates	Fastening block

ral network uses the back propagation algorithm to update the parameter matrix, and the number of hidden layer nodes is 64. After the interpreter receives the data of failure, it will assist the inference engine, which means help the user understand the fault diagnosis solutions given by the inference engine.

The inference engine uses a three-layer neural network structure. After the input layer, the information passes through the hidden layer of the recurrent neural network with 256 nodes, and then sends it to the output layer with Softmax as excitation function. Although from the view of structure, the neural network of the inference engine has only three layers, after unfolding the calculation process of the recurrent neural network, each individual input feature can be regarded as a separate layer of neural network, so the number of neural network structure will change, when the input data changes.

The output of the inference engine is a solution with a higher probability after calculation. When an unexpected fault occurs, it is impossible to directly solve the fault with only one solution. Therefore, a most likely solution and several alternatives are provided to increase the troubleshooting rate.

Figure 18 shows the user interface of the expert system for fault prognosis. The user describes the failure phenomena of the equipment to match cases, as shown in Fig. 18a. Figure 18b shows matching results of the related failures.

A visual bibliography of prognostics and health management

Based on the above work, the authors have developed a web-based paper retrieval tool to realize a swift search for achievements in the field of PHM. The retrieval tool is open source at "<https://mango-hund.github.io/>". The main page of the website is shown in Fig. 19.

In Fig. 19, the PHM-related scientific achievements collected in this Bibliography are presented in the form of headings and photo abstracts in a columnar layout. Users can choose the achievements of interest to access their details.

Users can overview important information about the paper, such as authors, abstract, keywords, etc. The source of the paper is presented as a link at the bottom of the page. In addition, the authors labeled each paper with several tags in order to categorize and filter the results. These labels are usually related to the paper's methodology, the subject of the research and the journal. The label for each paper is displayed above the paper title. All tags and the papers associated with each tag are displayed in the "Keywords" interface, allowing users to search by specific research subjects or algorithms.

In addition to retrieval by tags, users can also search for papers according to the time of publication and the type of paper, including "Review", "Remaining Useful Life prediction", "Fault Diagnosis", "Industrial Data Cleaning", "Health Assessment", etc.

Current trends and future work

This chapter summarizes the current progress and difficulties in the field of fault diagnosis, and provides an overview of the future evolution.

Progress and difficulties

The current application of PHM technology is gradually expanding from traditional mechanical components to large-scale and complex equipment such as electrical equipment, aerospace equipment and wind power generators, and fault diagnosis is constantly evolving from a traditional mechanical specialty to an integrated discipline with multiple spe-

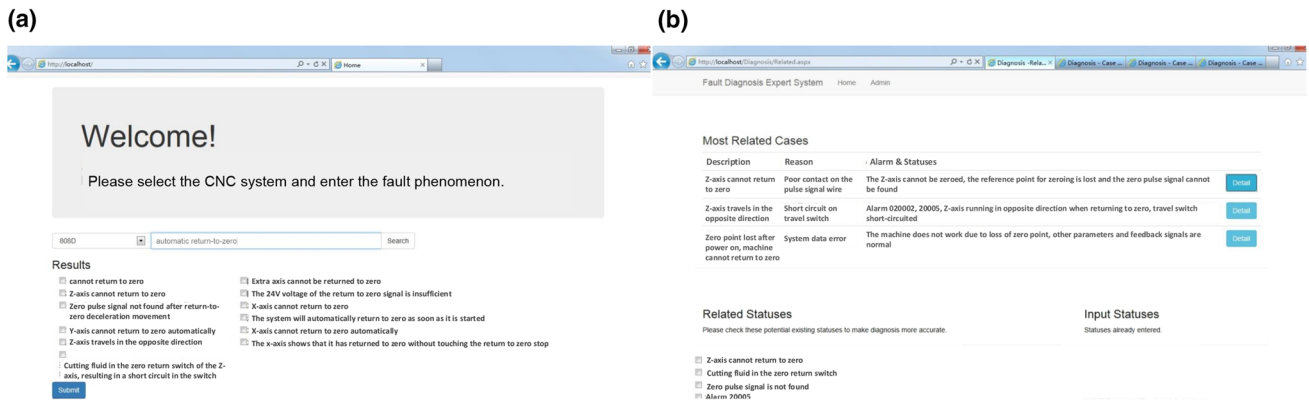


Fig. 18 Expert system search interface (a) and diagnostic results interface (b)

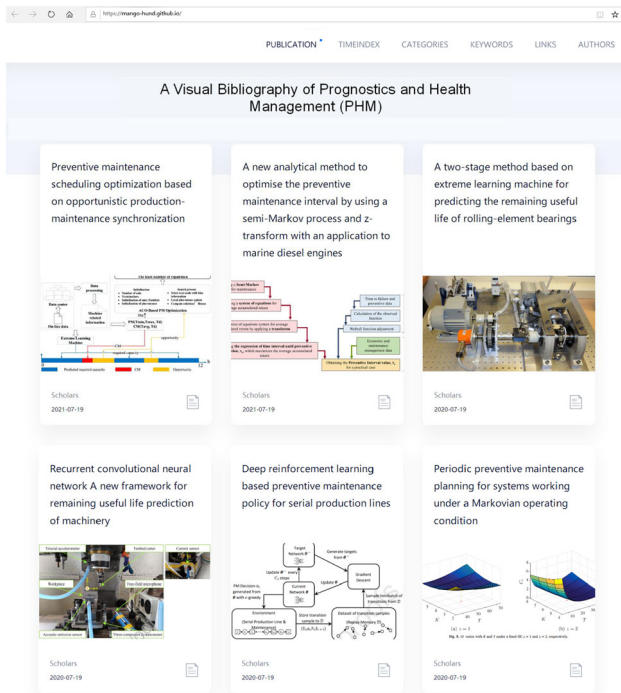


Fig. 19 A visual bibliography of prognostics and health management—publication

cialties such as mechanics, communication and computers. In addition, deep learning algorithms in the field of fault diagnosis rely on a large amount of standard training data with labels to train models. However, in practical applications, the data may not have a uniform format, and its structure and its intrinsic association may be more complex, so unsupervised learning is another important research direction for PHM technologies.

Limited by the computer, hardware and other factors, the current data-driven fault diagnosis methods still face some difficulties to be solved, such as:

- Reduce the time cost of a data-driven approach. As the volume of data increases, the contradiction between the

time sensitivity of big data and the time-consuming data-driven methods becomes increasingly prominent. This is constrained by the development of computer hardware, thus needing to make improvements to the computational model of big data analysis algorithms. Distributed computing, a computational model that emerged at the beginning of this century, can reduce the time cost of computing to a certain extent through the distributed allocation of resources. In the field of deep learning, Dean et al. (2012), the architect of Google MapReduce, proposes a distributed training strategy for deep network inference that employs large-scale clusters of machines. The proposed framework allows for model parallelism both within a computer (by multithreading) and across machines, with the framework handling the intricacies of parallelism, synchronization, and communication.

- Reliance of data-driven methods on high-quality data. In industrial applications, heterogeneous shop floor data from multiple sources often suffers from data fragmentation and sample imbalance. At the hardware level, this is contingent on the development of standardized sensing systems and the further popularization of shop floor IoT platforms; at the algorithm level, the problem can also be solved to some extent by improving the deep learning models themselves. For example, Yang et al. (2017) proposes a comprehensive model reuse approach that takes advantage of deep models' learning power to implicitly extract useful discriminative information from fixed models that have been widely employed in general tasks. Because of the reuse of fixed model in the suggested technique, the amount of training data required can be drastically decreased.
- Fleet-based fault diagnosis: Mass production is usually accomplished by several production lines, where a large number of machines with the same or similar functions are produced simultaneously. Currently, traditional fault diagnosis methods are usually studied based on a specific machine or solving a specific problem, without studying

machines as a fleet and collecting valuable knowledge from different instances for collaborative diagnosis and synthesis reference (Lee et al., 2015). How to utilize historical failure cases/failure cases of other devices to achieve autonomous fault diagnosis and even solve unprecedented failures is another challenge in the future.

Future trends

As the process of intellectualization and cyberization accelerates, technologies for data-driven fault diagnosis will be facilitated by developments in the following areas:

- Standardization of transmission protocols of industrial data: Due to the complexity of the equipment, different types of sensors are sometimes applied, which subsequently brings about different data acquisition and transmission protocols. Therefore, unified standards for different transmission protocols should be developed. There are several current industrial standards available, such as OPC, OPC-UA, MT-Connect, etc.
- Data compression and feature extraction: The industrial data will be transmitted to the central server through the IOT platform. If the raw data is transmitted directly, the huge amount of data and the unreasonable allocation of network resources will lead to transmission delays. Therefore, data compression and feature extraction are needed to pre-process the data. Unlike data compression in the computer field, the emphasis here is on data compression that combines engineering knowledge and diagnostic requirements, i.e. “engineering feature-based data compression”. For example:
 - (a) Using the data to construct device deterioration models and transfer only the key coefficients of the models to the server.
 - (b) Extracting statistical information from the raw sensing data.
 - (c) Developing dynamic thresholds that can be adjusted in real time according to operating conditions and using them to filter raw data.

Conclusion

The authors review the common algorithms for fault diagnosis by focusing on industrial big data, and highlight the application of deep learning. In addition, the authors developed a visual bibliography and analysed two fault diagnosis cases in the field of PHM, and concluded with a synopsis and outlook of the key technologies.

In summary, the big data era has seen the emergence of a variety of data-driven fault diagnostic methods. However, the

main bottleneck in this area is the lack of availability of actual fault data which can be used to teach fault diagnostic systems and the interpretability of data-driven methods. Therefore, future research should evolve from the development of new models to the practical oriented analysis of the correlation between models and failure principles.

Acknowledgements The research is partially supported by National Key R&D Program of China (No. 2017YFE0101400), National Natural Science Foundation of China (No. 61802278), China Scholarship Council, and EU H2020 Research and Innovation Program under the Marie Skłodowska-Curie Grant Agreement (754462). The research is also supported by the German KIT- internal research project Wertstromkinematik (Value Stream Kinematics).

References

- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), e00938.
- Agogino, A., & Goebel, K. (2007). *Ucberkeley*. Retrieved from <https://ti.arc.nasa.gov/m/project/prognostic-repository/mill.zip>.
- Andrienko, G., Andrienko, N., Drucker, S., Fekete, J.-D., Fisher, D., Idreos, S., Kraska, T., Li, G., Ma, K.-L., Mackinlay, J. et al. (2020). Big data visualization and analytics: Future research challenges and emerging applications. In *BigVis 2020-3rd International Workshop on Big Data Visual Exploration and Analytics*.
- Appana, D. K., Prosvirin, A., & Kim, J.-M. (2018). Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks. *Soft Computing*, 22(20), 6719–6729.
- Belhadi, A., Zkik, K., Cherrafi, A., Sha’ri, M. Y., et al. (2019). Understanding big data analytics for manufacturing processes: insights from literature review and multiple case studies. *Computers & Industrial Engineering*, 137, 106099.
- Bengio, Y. (2009). *Learning deep architectures for AI*. Now Publishers Inc.
- Bossio, J. M., Bossio, G. R., & De Angelo, C. H. (2017). Fault diagnosis in induction motors using self-organizing neural networks and quantization error. In *2017 XVII Workshop on Information Processing and Control (RPIC)* (pp. 1–6). IEEE.
- CaseWestern Reserve University. (2021b). *Casewestern*. Retrieved from <https://csegroups.case.edu/bearingdatacenter/>.
- Chen, H., Jiang, B., Ding, S. X., & Huang, B. (2020). Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Transactions on Intelligent Transportation Systems*.
- Chen, R., Huang, X., Yang, L., Xiangyang, X., Zhang, X., & Zhang, Y. (2019a). Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Computers in Industry*, 106, 48–59.
- Chen, J., Jing, H., Chang, Y., & Liu, Q. (2019b). Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. *Reliability Engineering & System Safety*, 185, 372–382.
- Chen, Y., Jin, Y., & Jiri, G. (2018). Predicting tool wear with multi-sensor data using deep belief networks. *The International Journal of Advanced Manufacturing Technology*, 99(5), 1917–1926.
- Chen, T. L., & Que, P. W. (2004). Research on invalidation of ds evidential theory in data fusion. *Journal of Transducer Technology*, 23, 25–27.
- Cho, S., Choi, M., Gao, Z., & Moan, T. (2021). Fault detection and diagnosis of a blade pitch system in a floating wind turbine based

- on kalman filters and artificial neural networks. *Renewable Energy*, 169, 1–13.
- Cincinnati University. (2021a). *Ims*. Retrieved from <https://best.berkeley.edu/>.
- Daugherty, P., Banerjee, P., Negm, W., & Alter, A. E. (2015). Driving unconventional growth through the industrial internet of things. *Accenture Technology*.
- Dean, J., & Ghemawat, S. (2008). Mapreduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113.
- Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., et al. (2012). Large scale distributed deep networks. *Advances in Neural Information Processing Systems*, 25, 1223–1231.
- Demidova, G., Rassölkin, A., Vaimann, T., Kallaste, A., Zakis, J., & Suzdalenko, A. (2021). An overview of fuzzy logic approaches for fault diagnosis in energy conversion devices. In *2021 28th International Workshop on Electric Drives: Improving Reliability of Electric Drives (IWED)* (pp. 1–7). IEEE.
- Di Liu, Wang, S., & Cui, X. (2021a). An artificial neural network supported wiener process based reliability estimation method considering individual difference and measurement error. *Reliability Engineering & System Safety*, 108162.
- Dimairo, F., Scapinello, O., Zio, E., Ciarapica, C., Cincotta, S., Crivellari, A., et al. (2021). Accounting for safety barriers degradation in the risk assessment of oil and gas systems by multistate bayesian networks. *Reliability Engineering & System Safety*, 216, 107943.
- Dong, S., Zhang, Z., Wen, G., & Wen, G. (2017). Design and application of unsupervised convolutional neural networks integrated with deep belief networks for mechanical fault diagnosis. In *2017 Prognostics and system health management conference (PHM-Harbin)* (pp. 1–7). IEEE.
- Duan, C., Zhu, M., Wang, K., & Zhou, W. (2021). Reliability analysis of intelligent manufacturing systems based on improved fmea combined with machine learning. *Researchsquare*.
- El Madbouly, E. E., Abdalla, A. E., & El Banby, G. M. (2009). Fuzzy adaptive kalman filter for multi-sensor system. In *2009 International Conference on Networking and Media Convergence* (pp. 141–145). IEEE.
- Electronics Standardization Institute China. (2017). *White book on industrial big data*. Retrieved from <http://www.cesi.cn/201703/2250.html>.
- Ellefsen, A. L., Bjørlykhaug, E., Æsøy, V., Ushakov, S., & Zhang, H. (2019). Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliability Engineering & System Safety*, 183, 240–251.
- Farsi, M. A., & Zio, E. (2019). Industry 4.0: Some challenges and opportunities for reliability engineering. *International Journal of Reliability, Risk and Safety: Theory and Application*, 2(1), 23–34.
- France University of Besanson. (2021). *Besanson*. Retrieved from <https://www.femto-st.fr/en/>.
- Frasconi, P., Gori, M., & Soda, G. (1992). Local feedback multilayered networks. *Neural Computation*, 4(1), 120–130.
- General Electric Company. (2015). *Predix: The industrial iot application platform*. Retrieved from https://www.ge.com/digital/sites/default/files/download_assets/Predix-The-Industrial-Internet-Platform-Brief.pdf.
- Germany, S. (2016). *Mindsphere: An industrial iot as a service solution to build on*. Retrieved from <https://siemens.mindsphere.io/en/about/for-users>.
- Gong, Y., Xiaoyan, S., Qian, H., & Yang, N. (2018). Research on fault diagnosis methods for the reactor coolant system of nuclear power plant based on ds evidence theory. *Annals of Nuclear Energy*, 112, 395–399.
- Guibing, G. A. O., Dengming, Z. H. O. U., Hao, T. A. N. G., & Xin, H. U. (2021). An intelligent health diagnosis and maintenance decision-making approach in smart manufacturing. *Reliability Engineering & System Safety*, 216, 107965.
- Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement*, 93, 490–502.
- Han, H., Cui, X., Fan, Y., & Qing, H. (2019). Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features. *Applied Thermal Engineering*, 154, 540–547.
- Han, T., Liu, C., Rui, W., & Jiang, D. (2021). Deep transfer learning with limited data for machinery fault diagnosis. *Applied Soft Computing*, 103, 107150.
- He, J., Yang, S., & Gan, C. (2017). Unsupervised fault diagnosis of a gear transmission chain using a deep belief network. *Sensors*, 17(7), 1564.
- Hinton, G.E. (2012). A practical guide to training restricted boltzmann machines. In *Neural networks: Tricks of the trade* (pp. 599–619). Springer.
- Hiraman, B. R. et al. (2018). A study of apache kafka in big data stream processing. In *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)*, (pp. 1–3). IEEE.
- Iannace, G., Ciaburro, G., & Trematerra, A. (2019). Fault diagnosis for UAV blades using artificial neural network. *Robotics*, 8(3), 59.
- Islam, M. M. M., & Kim, J.-M. (2019). Automated bearing fault diagnosis scheme using 2d representation of wavelet packet transform and deep convolutional neural network. *Computers in Industry*, 106, 142–153.
- Jiang, Y., & Yin, S. (2017). Recursive total principle component regression based fault detection and its application to vehicular cyber-physical systems. *IEEE Transactions on Industrial Informatics*, 14(4), 1415–1423.
- Ji, C., Xing, S., Qin, Z., & Nawaz, A. (2022). Probability analysis of construction risk based on noisy-or gate bayesian networks. *Reliability Engineering & System Safety*, 217, 107974.
- Ji, X., Ren, Y., Tang, H., Shi, C., & Xiang, J. (2020). An intelligent fault diagnosis approach based on Dempster-Shafer theory for hydraulic valves. *Measurement*, 165, 108129.
- Jung, D. (2020). Residual generation using physically-based grey-box recurrent neural networks for engine fault diagnosis. arXiv preprint [arXiv:2008.04644](https://arxiv.org/abs/2008.04644).
- Keim, D., Huamin, Q., & Ma, K.-L. (2013). Big-data visualization. *IEEE Computer Graphics and Applications*, 33(4), 20–21.
- Khorasgani, H., Farahat, A., Ristovski, K., Gupta, C., & Biswas, G. (2018). A framework for unifying model-based and data-driven fault diagnosis. In *PHM society conference* (Vol. 10).
- Kim, J., Zhao, X., Shah, A. A., & Kang, H. G. (2021). System risk quantification and decision making support using functional modeling and dynamic bayesian network. *Reliability Engineering and System Safety*, 215, 107880.
- Klinedinst, D., & King, C. (2016). *On board diagnostics: Risks and vulnerabilities of the connected vehicle*. CERT Coordination Center, Tech: Rep.
- Lee, J. (2020). *Industrial AI: Applications with sustainable performance*. Springer.
- Lee, J., Ardakani, H. D., Yang, S., & Bagheri, B. (2015). Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. *Procedia Cirp*, 38, 3–7.
- Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3, 18–23.
- Lee, J., Fangji, W., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 42(1–2), 314–334.
- Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). Aimq: A methodology for information quality assessment. *Information & Management*, 40(2), 133–146.

- Lessmeier, C. K. (2021). *Kat-datacenter*. Retrieved from <https://mb.uni-paderborn.de/kat/>.
- Li, M., & Zhou, Q. (2017). Industrial big data visualization: A case study using flight data recordings to discover the factors affecting the airplane fuel efficiency. In *2017 IEEE Trustcom/BigDataSE/ICSS* (pp. 853–858). IEEE.
- Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., et al. (2022). A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mechanical Systems and Signal Processing*, *167*, 108487.
- Li, Y., Jiang, W., Zhang, G., & Shu, L. (2021). Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data. *Renewable Energy*, *171*, 103–115.
- Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions*, *77*, 167–178.
- Liu, J., Zhang, Q., Li, X., Li, G., Liu, Z., Xie, Y., et al. (2021). Transfer learning-based strategies for fault diagnosis in building energy systems. *Energy and Buildings*, *250*, 111256.
- Lopez, M. A., Lobato, A. G. P., & Duarte, O. C. M. B. (2016). A performance comparison of open-source stream processing platforms. In *2016 IEEE Global Communications Conference (GLOBECOM)* (pp. 1–6). IEEE.
- Luo, J., Huang, J., & Li, H. (2021). A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis. *Journal of Intelligent Manufacturing*, *32*(2), 407–425.
- Lv, K., Gao, C., Si, J., Feng, H., & Cao, W. (2020). Fault coil location of inter-turn short-circuit for direct-drive permanent magnet synchronous motor using knowledge graph. *IET Electric Power Applications*, *14*(9), 1712–1721.
- Malik, H., & Mishra, S. (2017). Artificial neural network and empirical mode decomposition based imbalance fault diagnosis of wind turbine using turbsim, fast and simulink. *IET Renewable Power Generation*, *11*(6), 889–902.
- Mangai, U. G., Samanta, S., Das, S., & Chowdhury, P. R. (2010). A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Technical Review*, *27*(4), 293–307.
- Martínez, A., Sánchez, L., & Couso, I. (2013). Engine health monitoring for engine fleets using fuzzy radviz. In *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1–8). IEEE.
- Martínez López, K. (2021). *Development of machine learning strategies for fault diagnosis in virtual plants (digital twin)*. Master's thesis, Universitat Politècnica de Catalunya.
- Ma, M., Sun, C., & Chen, X. (2017). Discriminative deep belief networks with ant colony optimization for health status assessment of machine. *IEEE Transactions on Instrumentation and Measurement*, *66*(12), 3115–3125.
- Md Amin, T., Khan, F., & Imtiaz, S. (2018). Dynamic availability assessment of safety critical systems using a dynamic bayesian network. *Reliability Engineering & System Safety*, *178*, 108–117.
- Merkisz, J., Bogus, P., & Grzeszczyk, R. (2001). Overview of engine misfire detection methods used in on board diagnostics. *Journal of Kones, Combustion Engines*, *8*(1–2), 326–341.
- Miao, M., Yu, J., & Zhao, Z. (2021). A sparse domain adaption network for remaining useful life prediction of rolling bearings under different working conditions. *Reliability Engineering & System Safety*, 108259.
- NASA. (1994). *Nasa public lessons*. Retrieved from <https://llis.nasa.gov/lesson/845>.
- NASA USA. (2021). *Pcoe datasets*. Retrieved from <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>.
- Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). Pronostia: An experimental platform for bearings accelerated degradation tests. In *IEEE International Conference on Prognostics and Health Management, PHM'12*. (pp. 1–8). IEEE Catalog Number: CPF12PHM-CDR.
- Negri, E., Fumagalli, L., & Macchi, M. (2017). A review of the roles of digital twin in cps-based production systems. *Procedia Manufacturing*, *11*, 939–948.
- Novikova, E., Bestuzhev, M., & Kotenko, I. (2019). Anomaly detection in the hvac system operation by a radviz based visualization-driven approach. In *Computer Security* (pp. 402–418). Springer.
- O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. (2015). An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *Journal of Big Data*, *2*(1), 1–26.
- Pandya, D. H., Upadhyay, S. H., & Harsha, S. P. (2014). Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform. *Soft Computing*, *18*(2), 255–266.
- Panpan, X., Mei, H., Ren, L., & Chen, W. (2016). Vidix: Visual diagnostics of assembly line performance in smart factories. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 291–300.
- Patan, K. (2008). *Artificial neural networks for the modelling and fault diagnosis of technical processes*. Springer.
- Peng, J., Kimmig, A., Niu, Z., Wang, J., Liu, X., Wang, D., & Ovtcharova, J. (2022). Wind turbine failure prediction and health assessment based on adaptive maximum mean discrepancy. *International Journal of Electrical Power & Energy Systems*, *134*, 107391.
- Peng, X., Shijin, X., & Yin, H. (2007). Application of self-organizing competitive neural network in fault diagnosis of suck rod pumping system. *Journal of Petroleum Science and Engineering*, *58*(1–2), 43–48.
- Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial artificial intelligence in industry 4.0-systematic review, challenges and outlook. *IEEE Access*, *8*, 220121–220139.
- Przystałka, P., & Moczulski, W. (2015). Methodology of neural modelling in fault detection with the use of chaos engineering. *Engineering Applications of Artificial Intelligence*, *41*, 25–40.
- Qian, Q., Qin, Y., Wang, Y., & Liu, F. (2021). A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis. *Measurement*, *178*, 109352.
- Qiao, M., Yan, S., Tang, X., & Chengkuan, X. (2020). Deep convolutional and LSTM recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads. *IEEE Access*, *8*, 66257–66269.
- Qin, Y., Wang, X., & Zou, J. (2018). The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines. *IEEE Transactions on Industrial Electronics*, *66*(5), 3814–3824.
- Rohan, A. (2022). Holistic fault detection and diagnosis system in imbalanced, scarce, multi-domain (ISMD) data setting for component-level prognostics and health management (phm). arXiv preprint [arXiv:2204.02969](https://arxiv.org/abs/2204.02969).
- Saxena, A., & Goebel, K. (2008). *Nasa2*. Retrieved from <https://ti.arc.nasa.gov/m/project/prognostic-repository/CMAPSSData.zip>.
- Shahnazari, H. (2020). Fault diagnosis of nonlinear systems using recurrent neural networks. *Chemical Engineering Research and Design*, *153*, 233–245.
- Shao, H., Jiang, H., Wang, F., & Zhao, H. (2017). An enhancement deep feature fusion method for rotating machinery fault diagnosis. *Knowledge-Based Systems*, *119*, 200–220.
- Shlyannikov, V., Yarullin, R., Yakovlev, M., Giannella, V., & Citarella, R. (2021). Mixed-mode crack growth simulation in aviation engine compressor disk. *Engineering Fracture Mechanics*, *246*, 107617.
- Song, F., Zhang, Y., Lin, L., Zhao, M., & Zhong, S. (2021). Deep residual lstm with domain-invariance for remaining useful life prediction across domains. *Reliability Engineering & System Safety*, *216*, 108012.

- Stonebraker, M., Çetintemel, U., & Zdonik, S. (2005). The 8 requirements of real-time stream processing. *ACM Sigmod Record*, 34(4), 42–47.
- Su, L., Wang, Z., Ji, Y., & Guo, X. (2020). A survey based on knowledge graph in fault diagnosis, analysis and prediction: key technologies and challenges. In *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)* (pp. 458–462). IEEE.
- Su, W., & Bougiouklis, T. C. (2007). Data fusion algorithms in cluster-based wireless sensor networks using fuzzy logic theory. In *Proceedings of the 11th WSEAS international conference on communications*.
- Tamilselvan, P., & Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety*, 115, 124–135.
- University of Berkeley California. (2021). berkeley. Retrieved from <https://best.berkeley.edu/>.
- Vachtsevanos, G. J., & Vachtsevanos, G. J. (2006). *Intelligent fault diagnosis and prognosis for engineering systems* (Vol. 456). Wiley.
- Wang, J., Zhuang, J., Duan, L., & Cheng, W. (2016). A multi-scale convolution neural network for featureless fault diagnosis. In *2016 International Symposium on Flexible Automation (ISFA)* (pp. 65–70). IEEE.
- Wang, L., Hodges, J., Yu, D., & Fearing, R. S. (2021). Automatic modeling and fault diagnosis of car production lines based on first-principle qualitative mechanics and semantic web technology. *Advanced Engineering Informatics*, 49, 101248.
- Wang, P., Yang, L. T., Li, J., Chen, J., & Hu, S. (2019). Data fusion in cyber-physical-social systems: State-of-the-art and perspectives. *Information Fusion*, 51, 42–57.
- Wang, Z., Yao, L., Cai, Y., & Zhang, J. (2020). Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis. *Renewable Energy*, 155, 1312–1327.
- Wei-Peng, L., & Yan, X.-F. (2019). Visual monitoring of industrial operation states based on kernel fisher vector and self-organizing map networks. *International Journal of Control, Automation and Systems*, 17(6), 1535–1546.
- Wen, L., Li, X., Gao, L., & Zhang, Y. (2017). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, 65(7), 5990–5998.
- Willner, D., Chang, C. B., & Dunn, K. P. (1976). Kalman filter algorithms for a multi-sensor system. In *1976 IEEE conference on decision and control including the 15th symposium on adaptive processes* (pp. 570–574). IEEE.
- Wumaier Tuerxun, X., Chang, G. H., Zhijie, J., & Huajian, Z. (2021). Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm. *Ieee Access*, 9, 69307–69315.
- Xiaolei, Y., Zhao, Z., Zhang, X., Zhang, Q., Liu, Y., Sun, C., & Chen, X. (2021). Deep-learning-based open set fault diagnosis by extreme value theory. *IEEE Transactions on Industrial Informatics*, 18(1), 185–196.
- Xu, F., Fang, Z., Tang, R., Li, X., & Tsui, K. L. (2020). An unsupervised and enhanced deep belief network for bearing performance degradation assessment. *Measurement*, 162, 107902.
- Xu, X., Cao, D., Zhou, Y., & Gao, J. (2020). Application of neural network algorithm in fault diagnosis of mechanical intelligence. *Mechanical Systems and Signal Processing*, 141, 106625.
- Xu, Z., & Saleh, J. H. (2021). Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. *Reliability Engineering & System Safety*, 107530.
- Xu, Y.-H., Hong, W.-X., & Chen, M.-M. (2009). Visualized fault diagnosis method based on radviz and its optimization. *Application Research of Computers*, 3.
- Xudong, S., Hongguang, L., Ruowen, W., & Meng, X. (2015). Modeling and simulation of aviation engine ignition spark frequency disorder. *The Open Electrical & Electronic Engineering Journal*, 9(1).
- Yang, N., Zhang, G., & Wang, J. (2020). Research on knowledge graph and bayesian network in fault diagnosis of steam turbine. In *2020 Global Reliability and Prognostics and Health Management (PHM-Shanghai)* (pp. 1–6). IEEE.
- Yang, X., Li, Z., Wang, S., Li, W., Sarkodie-Gyan, T., & Feng, S. (2021). A hybrid deep-learning model for fault diagnosis of rolling bearings. *Measurement*, 169, 108502.
- Yang, Y., Zhan, D.-C., Fan, Y., Jiang, Y., & Zhou, Z.-H. (2017). Deep learning for fixed model reuse. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- Ye, S., Wang, J., Zhang, G., Su, H., Yao, Q., & Huang, J. (2015). Simulation and fault diagnosis for aviation engine starting system based on simulink. In *2015 Chinese Automation Congress (CAC)* (pp. 1834–1839). IEEE.
- Yun, F., Feng, Z., Baofeng, L., & Yongfeng, C. (2019). Research on intelligent fault diagnosis of power acquisition based on knowledge graph. In *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)* (pp. 1737–1740). IEEE.
- Zhang, H., Chen, P., & Wang, Q. (2018). Fault diagnosis method based on EEMD and multi-class logistic regression. In *2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE)* (pp. 859–863). IEEE.
- Zhang, S., Zhang, Y., Yang, Y., Cheng, W., Zhao, H., & Li, Y. (2021a). Knowledge graph construction for fault diagnosis of aircraft environmental control system. In *2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing)* (pp. 1–5). IEEE.
- Zhang, Y., & Ji, Q. (2006). Active and dynamic information fusion for multisensor systems with dynamic bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics Part B (Cybernetics)*, 36(2), 467–472.
- Zhang, Y., Zhou, T., Huang, X., Cao, L., & Zhou, Q. (2021). Fault diagnosis of rotating machinery based on recurrent neural networks. *Measurement*, 171, 108774.
- Zhong, F., Shi, T., & He, T. (2005). Fault diagnosis of motor bearing using self-organizing maps. In *2005 International Conference on Electrical Machines and Systems* (Vol. 3, pp. 2411–2414). IEEE.
- Zhong-Xu, H., Wang, Y., Ge, M.-F., & Liu, J. (2019). Data-driven fault diagnosis method based on compressed sensing and improved multiscale network. *IEEE Transactions on Industrial Electronics*, 67(4), 3216–3225.
- Zhou, J., Wang, T., & Deng, J. (2021). Corpus construction and entity recognition for the field of industrial robot fault diagnosis. In *2021 13th International Conference on Machine Learning and Computing* (pp. 410–416).
- Zhou, X., Pan, L., Zheng, Z., Tolliver, D., & Keramati, A. (2020). Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety*, 200, 106931.

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.