This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Deep Learning for Time-Series Prediction in IIoT: Progress, Challenges, and Prospects

Lei Ren, *Member, IEEE*, Zidi Jia, *Graduate Student Member, IEEE*, Yuanjun Laili, *Member, IEEE*, and Di Huang, *Senior Member, IEEE*

***Abstract*— Time-series prediction plays a crucial role in the Industrial Internet of Things (IIoT) to enable intelligent process control, analysis, and management, such as complex equipment maintenance, product quality management, and dynamic process monitoring. Traditional methods face challenges in obtaining latent insights due to the growing complexity of IIoT. Recently, the latest development of deep learning provides innovative solutions for IIoT time-series prediction. In this survey, we analyze the existing deep learning-based time-series prediction methods and present the main challenges of time-series prediction in IIoT. Furthermore, we propose a framework of state-of-the-art solutions to overcome the challenges of time-series prediction in IIoT and summarize its application in practical scenarios, such as predictive maintenance, product quality prediction, and supply chain management. Finally, we conclude with comments on possible future directions for the development of time-series prediction to enable extensible knowledge mining for complex tasks in IIoT.**

***Index Terms*— Deep learning, industrial intelligence, Industrial Internet of Things (IIoT), neural network, time-series prediction.**

I. Introduction

INDUSTRIAL Internet of Things (IIoT) [1], [2] has emerged as a powerful tool for aggregating data from various production devices and smart terminals, with the continued development of cutting-edge information and communication technologies, including cloud computing [3], big data [4], artificial intelligence (AI) [5], and 5G [6]. By connecting cloud datacenter, edge computing technologies, and industrial control networks, IIoT is able to provide sufficient computing capability and in-time manufacturing services, and further enable more efficient and flexible production.

Since time series is the most collected data from IIoT, such as sensor signals and smart terminal monitoring signals, time-series prediction is emerging as a vital component in IIoT. The key tasks that require time-series prediction include effective process and condition monitoring, such as prognostic and health management (PHM) [7], fault diagnosis (FD) [8], product quality prediction [9], and product lifecycle management (PLM) [10]. By modeling, analyzing, and predicting IIoT time-series, the insights of complex industrial issues can be captured. Therefore, there is an increasing need to explore the patterns and laws underlying industrial time series to provide intelligent insights and solutions.

In recent years, deep learning-based time-series prediction methods have emerged as a promising alternative to conventional data analysis techniques, because of several key technological breakthroughs. These methods offer several advantages to IIoT, including end-to-end processing that extracts implicit features, eliminating the need for mechanism modeling [11], [12]. These methods do not rely on expert experience and only require sufficient training data to extract deep, abstract information from the data and generate accurate representations. The model's representation capability can be further enhanced by increasing the width and depth of the neural network (although this necessitates additional data and computational resources). As a result, deep learning has garnered significant interest among researchers in the field. Deep learning models enable the extraction and analysis of interfeature spatiotemporal correlations from IIoT temporal data. These models capture a range of temporal correlations, including periodic temporal correlations, temporal degradations, intervariable correlations, and so on. Through their ability to identify and analyze these correlations, deep learning models provide a powerful means of understanding the complex relationships between different features of IIoT temporal data.

Although deep learning methods have demonstrated significant potential in IIoT time-series prediction, the increasingly complex nature of modern manufacturing tasks, equipment, and processes has posed significant challenges. As the industry demands greater flexibility in manufacturing, tasks are subject to constant change. Multivariety, small-lot mixed-line manufacturing has become more common, involving multiple, ever-changing processes [13]. Consequently, monitoring data have become increasingly heterogeneous, nonlinear, high noise, and non-Gaussian, thereby making associated monitoring, detection, and scheduling tasks more challenging. Moreover, in contrast to IoT, IIoT temporal data are broadly sourced, high-throughput, low-value intensive, and highly

Manuscript received 23 December 2022; revised 27 May 2023; accepted 20 June 2023. This work was supported by the National Science Foundation of China under Project 62225302, Project 92167108, and Project 62173023. *(Corresponding authors: Lei Ren; Zidi Jia.)*

Lei Ren and Yuanjun Laili are with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China, also with the Zhongguancun Laboratory, Beijing 100094, China, and also with the State Key Laboratory of Intelligent Manufacturing System Technology, Beijing 100854, China (e-mail: renlei@buaa.edu.cn; lailiyuanjun@buaa.edu.cn).

Zidi Jia is with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: jiazidi@buaa.edu.cn).

Di Huang is with the School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: dhuang@buaa.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3291371.

Digital Object Identifier 10.1109/TNNLS.2023.3291371

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
COMPARISON WITH THE EXISTING RELATED SURVEYS

Refs	Time-Series	Methods	IIoT	Applicaitons
[14]	✓	Deep Learning		
[15]	✓	Deep Learning		
[16]		Deep Learning	✓	Predictive Maintenance
[17]		Data-driven methods	✓	Fault Diagnosis
[18]		Generative Learning	✓	Overall
[19]		Transfer Learning	✓	Overall
Ours	✓	Deep Learning	✓	Overall

dynamic. These data often exhibit unsatisfactory quality, characterized by imbalances, missing, unlabeled, unseen, and heterogeneous information. For instance, data from complex industrial processes may be heterogeneous and time-varying, while data from smart grids often suffer from widespread missing data. Similarly, operation and maintenance data from mechanical equipment may exhibit imbalances and lack labels.

Conventional methods struggle to effectively handle the complexity, multimodality, and dynamic nature of IIoT temporal data. However, as IIoT-enabled industrial applications demand greater accuracy, security, and low latency, intelligent instruments are required to process data accurately and efficiently. Specifically, there is a need for models that are highly generalizable, robust, and evolvable. Deep learning methods possess these characteristics and offer a promising solution to these challenges. Deep learning methods are particularly well-suited to address the unique demands of IIoT temporal data, thanks to their ability to extract and analyze complex features and correlations from this type of data. These characteristics make deep learning models highly effective in handling the complexity and dynamism of IIoT temporal data. In Section II, we will introduce commonly adopted models and paradigms for deep learning-based solutions to these problems, followed by a specific description of their applications in industrial contexts.

After reviewing the existing literature on deep learning-based IIoT applications and time-series prediction [14], [15], [16], [17], [18], [19], we have identified a gap in the literature regarding the application of deep learning-based time-series prediction in IIoT, particularly in its specific challenges. While some studies have focused on time-series prediction methods, some have focused on a particular method applied in industry, and others have focused on the application of deep learning in specific industrial settings, none have specifically examined the challenges and prospects of using deep learning for time-series prediction in IIoT. Therefore, this survey aims to address this gap by exploring the potential of deep learning methods in this area, as well as the challenges associated with low-quality data. We provide a comparison between this survey and existing surveys in Table I.

The subsequent sections are organized as follows. In Section II, the commonly applied deep learning methods and modeling paradigms in general time-series prediction tasks are presented. In Section III, the main challenges of low-quality IIoT temporal data are presented. In Section IV, various challenges in IIoT time-series prediction and their solutions are comprehensively reviewed. In Section V, the typically applications are presented. The discussion and prospects are given in Section VI. Section VII is conclusion.

II. General Deep Learning Models and Paradigms for Time-Series Prediction

A. General Deep Learning-Based Models

Deep learning has gained widespread attention in time-series prediction applications, with typical representative methods, including convolutional neural networks (CNNs) [20], [21], recurrent neural networks (RNNs) [22], autoencoders (AEs) [8], [23], restricted Boltzmann machines (RBMs) [24], [25], attention-based neural networks [26], [27], and graph neural networks (GNNs) [28], [29], among others. Compared with traditional methods, deep learning-based approaches have demonstrated superior performance in terms of capturing tendencies, singularities, and scale similarities [30].

1) CNN-Based Models: CNN has been extensively researched and widely employed in various deep-learning methods, making it a prominent technique of time-series prediction. Its ability to capture local data characteristics and exhibit strong generalization capabilities has contributed to its popularity. Moreover, CNN is particularly effective when dealing with process data structured in an array-like topology. By integrating convolutional computations of multiple scales and channel numbers in a sequential and parallel manner, it is possible to construct a neural network capable of capturing periodic and trending temporal features across different scales. This enables efficient processing of temporal data. In recent years, temporal convolutional networks (TCNs) have emerged as a specialized approach for handling temporal data [31]. The utilization of CNN-based methods for temporal data prediction has become a significant research focus within the field.

2) RNN-Based Models: RNNs can effectively model sequential data by employing directed graphs [32], generating outputs for each input time step. RNNs are particularly adept at capturing intrinsic information from time series. To overcome the challenge of handling long-term dependencies, several variants of RNNs have been introduced, including long short-term memory (LSTM) and gated recurrent unit (GRU). These variants incorporate gating mechanisms to selectively retain or discard long-term memory information, enabling the maintenance of dependencies overextended time periods within the hidden states. Currently, there remains a fervent interest in academic research focused on the application of RNN-like neural networks [33], [34].

3) AE-Based Models: AE is an unsupervised neural network model capable of extracting implicit features from data without relying on labeled information [35]. Comprising an encoder and a decoder, AE extracts significant features from the data while reconstructing the original input data. The reconstructive nature of the decoder aids the encoder in capturing more accurate representations of the original data. AE's modular structure enhances interpretability compared with other models. Furthermore, stacked AEs (SAEs) [23], [36] can be created by layering multiple AEs, enabling the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

extraction of deeper levels of hidden information. Various AE variants, such as sparse AEs [37], [38], VAE [39], [40], and denoising AEs (DAE) [41], [42], are employed in time-series prediction. AE has gained popularity as a common model for time-series prediction [43]. It facilitates the exploration of implicit representations associated with temporal aspects.

4) RBM-Based Models: The restricted Boltzmann machine (RBM) [44] is another deep learning method widely used. Its objective is to reduce the output of the visible layer as much as possible to the original input, making the hidden layer an alternative representation of the visible layer. This allows the RBM to extract essential features from the training dataset and avoid the problem of local minima. Extensions of the generic RBM include deep belief networks (DBNs) and deep Boltzmann machines (DBMs). RBM technology is invaluable for time-series prediction, with its efficiency and accuracy greatly improving the latter in terms of production, operational safety, and overall system reliability [45], [46], [47].

5) Attention-Based Models: The attention mechanism enables models to allocate computational resources to more crucial information. Recently, the attention mechanism has attracted considerable interest from scholars in the field of deep learning due to the proposal of transformer [48], a neural network that utilizes this mechanism and has demonstrated superior performance. As a result, the attention mechanism has become a highly popular topic in deep learning research [49], [50]. Apparently, attention-based neural network mechanisms have gradually become the hottest research in academia for time-series prediction issues in recent years.

6) GNN-Based Models: Big data are usually represented in the form of graphs. Due to the complexity of graph data, it is difficult for most existing deep learning methods to process it effectively. GNNs can obtain information about the associative features of data from non-Euclidean spaces and be used to capture knowledge about the correlation between data features [51], thus gaining the interests of scholars in academia. Actually, by capturing correlations among features and constructing association graphs, GNN performs well in processing sequence data [29], [52].

B. Modeling Paradigms

Regular time-series prediction issues can be effectively addressed using discriminative models, which involve directly mapping the prediction target to the data features. Discriminative models are widely utilized and encompass a broad range of prediction and classification tasks. Thus, this survey focuses on other modeling approaches and does not delve into discriminative models separately. With the development of advanced AI technologies, a series of novel neural network modeling paradigms have been developed, such as transfer learning [53], [54], [55], generative learning [56], contrastive learning [57], [58], and adversarial learning [59], [60], which provide solutions for the increasingly complex applications. The architectures and objectives of these paradigms are shown in Fig. 1.

1) Transfer Learning: Actually, transfer learning has become a popular and promising area in time-series prediction due to its wide application prospects [61], [62].

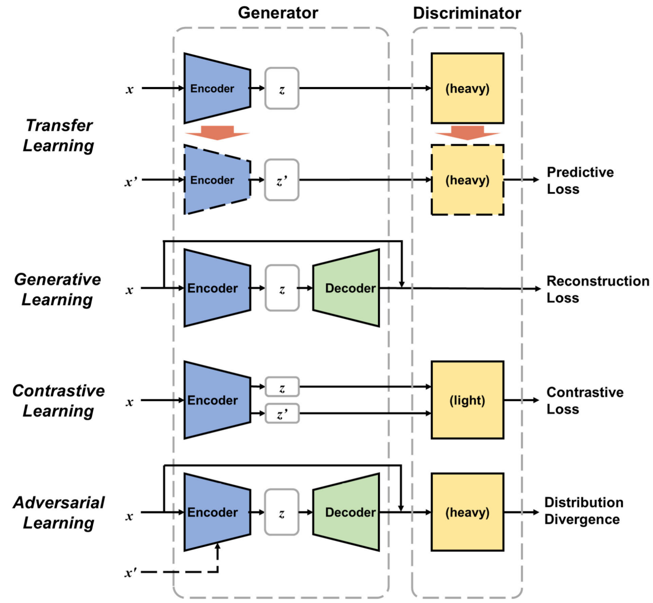

Fig. 1. Architectures and objectives of the modeling paradigms. Among them, transfer learning enables knowledge transfer by migrating modules or parameters of the model; generative learning encodes the input into an explicit vector and reconstructs the input with it; contrastive learning captures insights by contrasting the similarity of the latent vectors; and adversarial learning improves model performance by adversarial gaming among modules or optimizing targets.

Transfer learning aims to improve the performance of the learner in the target domain by leveraging knowledge from different but correlated source domains. As a result, transfer learning has become a popular modeling paradigm. Transfer learning can be classified into two types: data-based and model-based [63].

Model-based transfer learning, on the other hand, focuses on knowledge transferring through the adaptation or replacement of neural network modules. The submodules, such as classifiers, extractors, or encoders, may be changed by model-based transfer learning.

Data-based transfer learning aims to transfer knowledge by adapting and transforming data. This type of transfer learning is primarily used to transfer knowledge between application scenarios. Homogeneous knowledge migration is a common approach that focuses on reducing distribution discrepancies between the instances of the source and target domains. There are typically two strategies to achieve this goal: instance weighting and feature transformation. A representative weighting strategy can be expressed as follows [64]:

$$\begin{aligned}\mathbb{E}_{(x,y)\sim P^T}[\mathcal{L}(x,y;f)] &= \mathbb{E}_{(x,y)\sim P^S}\left[\frac{P^T(x,y)}{P^S(x,y)}\mathcal{L}(x,y;f)\right] \\ &= \mathbb{E}_{(x,y)\sim P^S}\left[\frac{P^T(x)}{P^S(x)}\mathcal{L}(x,y;f)\right] \quad (1)\end{aligned}$$

where P^S and P^T are the distribution probabilities of the source and target domain data in the feature space, respectively. The upper and lower parts of the equation represent supervised and unsupervised policies, respectively.

Feature transformation approaches intend to find shared latent features and use them as a medium for transferring

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

knowledge. Such approaches transform each original feature into a new feature representation for knowledge transfer. The objectives of constructing representations include minimizing marginal and conditional distribution differences, preserving the attributes or underlying structure of the data, and finding correspondences among features. One of the main objectives is to reduce the discrepancies in the distribution of source and target domain instances. Maximum mean discrepancy (MMD) [65] is an extensively applied measure.

2) Generative Learning: Generative learning methods, which can approximate and generate a joint distribution of target and training data, generate samples similar to real data. It has, therefore, become an important branch in time-series prediction [7], [66]. Generative learning allows modeling the substrate distribution of real data and generating data or features in an unsupervised manner for augmentation of real samples [18]. Generative models estimate the approximate distribution of the data by conditional density. The probabilistic generated model parameters allow the uncertainty of the data to be captured. By combining the probabilistic distribution of the data with the generative process, generative models can provide data support for inference, prediction, and decision making. Such modeling paradigms are also robust to uncertainty and the time-varying nature of data distribution.

Generative learning captures the uncertainty of the data by estimating the approximate distribution of the data to model $p(x|y)$ and $p(y)$. Commonly adopted generative learning models are variational AEs (VAE) and generative adversarial networks (GANs) [67].

VAE assumes that both the prior $p(z)$ and the approximate posterior $q(z|x)$ follow Gaussian distributions. VAE is an extension of AE that models the output of the encoder as the mean and variance of the target distribution. The VAE comprises two encoders, for computing the mean and the variance, respectively. The output of the decoder is expected to be variance-free, and co-evolution is achieved through an implicit confrontation between the decoder and encoder 2. In variational inference, the evidence lower bound on the log-likelihood of the data is maximized

$$\log p(x) \geq \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)\|p(z)) \tag{2}$$

where $p(x)$, $p(z)$, and $p(x|z)$ are evidence probability, prior, and likelihood probability, respectively.

GAN comprises the generator and discriminator modules. The discriminator aids the generator in obtaining realistic data generation by playing against it. The generator captures the data distribution and aims to maximize the discriminator's error probability. Through iterative adversarial training, both modules, especially the generator, improve their performance. The discriminator attempts to minimize its loss $D(x)$, while the generator attempts to maximize the loss of the discriminator $1-G(D(z))$. It can be described by the following equation:

$$\max_G \min_D \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - G(D(z)))]. \tag{3}$$

3) Contrastive Learning: Currently, methods based on contrastive learning are extensively utilized for time-series prediction [68], [69], [70]. Contrastive learning is a typical discriminant self-supervised learning, thereby mining properties of the dataset itself to help the model learn without labels [71], [72]. It aims to learn similarities and differences among samples through representation learning [73], [74], [75]. Contrastive learning obtains mappings of certain pseudolabels by comparing input samples and their reconstructions. On the other hand, contrastive learning can learn by comparing multiple samples of the same signal [67].

Contrastive learning objectives can be divided into two types, i.e., context-instance contrast and instance–instance contrast. Context-instance contrast, also known as global–local contrast, aims to model the attribution relationship between the local and global representations of the sample. A representative optimization objective for context-instance contrast is mutual information (MI). The objective is to capture the direct attribution of local features to global semantics. MI is obtained by maximizing the association between two relevant variables. Noise contrastive estimation (NCE) [76] is one of the typical loss functions. NCE and its variant InfoNCE [71] can be utilized to learn MI

$$\mathcal{L}_{\text{NCE}} = -\log \frac{e^{\text{sim}(z_i,z_j)/\tau}}{\sum_{k=1}^{2N} \mathbb{I}_{k\neq i} e^{\text{sim}(z_i,z_k)/\tau}}. \tag{4}$$

Each sample x is mapped into two augmented views (i.e., x^+ and x^-); thus, there are $2N$ augmented pairs, in which there are $2N-1$ negative instances correspondingly. In this way, the latent representations of similar samples can be kept close to each other in the feature space, and the latent representations of dissimilar samples can be made away from each other. The mathematical representation of InfoNCE loss is as follows, which distinguishes z_i from its k negative pairs:

$$\mathcal{L}_{\text{InfoNCE}} = -\log \frac{e^{\text{sim}(z_i,z^+)/\tau}}{e^{\text{sim}(z_i,z^+)/\tau} + \sum_{j=1}^{k} e^{\text{sim}(z_i,z_j^-)/\tau}}. \tag{5}$$

Instance–instance contrastive learning directly studies the relationship between instance-level local representations of various samples, i.e., metric learning. For example, the multivariate monitoring signals associated with predictive maintenance of equipment may contain information about the remaining useful life (RUL) of bearings and information about possible faults. For the RUL prediction task, however, the fault type information is relatively unimportant. The model should focus on important sensing signals or sequence segments. One representative approach is MoCo [74], which distinguishes instances by momentum contrast. For a sample x, an instinctive representation $q = f_q(x)$ is learned by a query encoder $f_q(\cdot)$ that distinguishes x from other instances. For the other samples, an asynchronously updated key encoder $fk(\cdot)$ is used to generate $k^+ = f_k(x)$ and $k_i = f_k(x_i)$ with the following optimization objectives:

$$\mathcal{L}_{\text{MoCo}} = -\log \frac{e^{q\cdot k^+/\tau}}{\sum_{i=1}^{K} e^{q\cdot k_i/\tau}}. \tag{6}$$

4) Adversarial Learning: Due to the demand for time-series prediction, adversarial learning has gradually gained the spotlight of scholars in related fields in recent years.

TABLE II
ADVANTAGES AND DISADVANTAGES OF GENERAL DEEP LEARNING MODELS AND PARADIGMS FOR TIME-SERIES PREDICTION

	Characteristics	Advantages	Disadvantages
CNN	Supervised	Process array-like data; present local correlation; robust; sparse interactions; parameters sharing; equivariant representation	Complex computations; slow parameter updates for deep networks
RNN	Supervised	Process sequence data	Non-parallel computing
AE	Unsupervised	Low complexity; denoising and generation; modular structural	Outputs may lack practical utility
RBM	Unsupervised	Robust to ambiguous data; label not required	High computation cost
Attention	Unsupervised	Present correlation among arbitrary locations; parallelization; creditable	Slow convergence; non-computationally universal
GNN	Supervised	Process graph data; present inter-correlation of variables	Difficult to model
Transfer	-	Transfer knowledge among domains	-
Generative	Self-supervised	Generate data in expected distribution	May exhibits limited relevance to the main task
Contrastive	Self-supervised	Present mutual information	Need large batch size; overlook details
Adversarial	-	Auxiliary to other paradigms	Rarely applied alone

Adversarial learning is a widespread research strategy for neural network modeling in deep learning, which applies the ideas of game theory in order to make full utilization of data resources [77] and to improve model robustness and completeness. Adversarial training is generally implemented in two ways. One is neural networks represented by GAN [78], domain adversarial neural networks (DANNs) [79], and so on, which are implemented by adversarial gaming between various modules or optimization objectives. It is noted that there is no difference between the GAN here and the previous GAN in the generative learning part, where there are both generative and adversarial learning paradigms in GAN. Such adversarial learning generally acts as an auxiliary to other tasks, such as transfer learning or generative learning, improving the performance of the model for the main task. The other is represented by strategies, such as projected gradient descent (PGD) [80] and fast gradient method (FGM) [81], which are realized with adversarial samples [82]. Such adversarial tasks aim at improving the model's robustness with minor perturbations.

C. Summary

In general, most of the time-series prediction methods are constructed with these extensively adopted models and modeling paradigms. Deep learning models for time-series analytics applications have unique strengths and constraints, as shown in Table II, that can make it challenging to handle complex time-series prediction tasks. CNNs capture local feature correlations well and are robust, but deeper CNNs can have slow convergence and high computational complexity. RNNs excel in processing sequence data but have difficulty with long-term dependencies and are computationally inefficient for parallel computing. AEs have low complexity and are robust to noise and missingness, but may not be useful for problem-oriented applications. RBMs are unsupervised and have robustness to ambiguous data, but have excessive computational loss. Attention-based neural networks have strong representation abilities, capture correlations between features at arbitrary locations, and have some interpretability, but have slow convergence and are not universal computation. In some scenarios, conventional deep learning methods may be invalid, and advanced modeling paradigms need to be employed. Specifically, transfer learning is the most prevalent paradigm for distribution shifting. Domain knowledge from different data domains can be fused with each other to transfer the complete solution to the scenario with low-quality data. However, it requires high-quality source data related to the target domain data. Generative learning and contrastive learning, two paradigms of self-supervised learning, can be applied to obtain insights by learning the features of the data itself. However, it is difficult to obtain favorable results when the self-supervised task is not well defined or has little suitability to the downstream task. Adversarial learning is generally used as an auxiliary to other modeling paradigms and is rarely used alone. It is worth noting that these paradigms are not mutually exclusive and more powerful models can be constructed through their collaboration. Although each model offers distinct advantages, they may face challenges when addressing complex time-series prediction tasks.

III. MAIN CHALLENGES OF TIME-SERIES PREDICTION IN IIoT

IIoT time-series refers to the collection of historical data over a period of time in an industrial setting, such as a manufacturing plant, power station, or chemical plant. These data are often gathered by sensors and other measurement devices that monitor various aspects of the industrial processes, such as temperature, pressure, flow rate, and vibration.

One of the primary challenges in effectively representing IIoT time series lies in its complex characteristics, which are generally nonlinear and non-Gaussian. Besides, errors and malfunctions in sensor operations, data collection, transmission, storage, and environmental factors may contribute to low-quality industrial data that is highly noisy and incomplete. Another significant challenge in time-series prediction in IIoT applications is the high variability and unpredictability of industrial processes. These factors can lead to imbalanced, unlabeled, and unseen industrial temporal data, further complicating the prediction process. An additional challenge lies in addressing the heterogeneity and multisource of the data. Due to the involvement of multiple sensor types and devices that generate large volumes of temporal data in various formats and from diverse sources, fusing time-series data in IIoT can be challenging.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

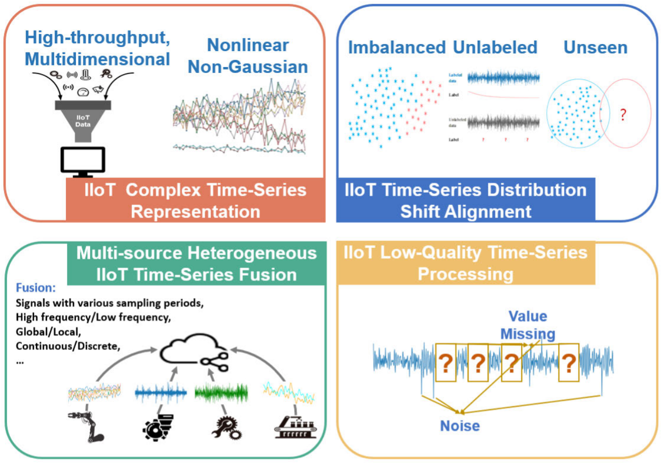

Fig. 2. Main challenges of time-series prediction in IIoT.

These issues of low-quality data-related challenges of time-series prediction in IIoT can be summarized as follows and shown in Fig. 2.

A. Challenges in IIoT Multisource Heterogeneous Time-Series Fusion

The data involved in industrial intelligence applications are generally characterized by multisources, heterogeneity, and multimodal, due to the gradual development of intelligent manufacturing toward complexity, distribution, and dynamism. Data from various sources with various structures are to be processed jointly. For example, the monitoring signals in the manufacturing workshop come from various sensors, such as sound, light, electricity, acceleration, and pressure. The sensors are placed at various locations, and the signals may be heterogeneous, for instance, signals may be with different sampling periods, with large frequency differences, carry different global or local information, and so on. If these signals are processed separately, the coupling and correlation between them will be ignored. Intelligent methods for efficient fusion of industrial multisource heterogeneous data are urgently needed [83], [84]. Industrial data fusion models need to be transitioned to hybrid, dynamic and distributed [85], and the efficient fusion and accurate characterization of heterogeneous data in industry is an essential task.

B. Challenges in IIoT Low-Quality Time-Series Processing

High-quality data is not always available in IIoT due to various factors, such as environmental conditions, equipment failure, sensor malfunctions, data transmission errors, incompatible data formats, and maintenance and downtime. For instance, in power production and transmission processes, monitoring data may have missing elements due to various factors, such as communication interruptions, sensor failures, power outages, or equipment maintenance. And the high-intensity power fluctuations bring about unfilterable noise. The noise and missing values can significantly degrade the efficiency and effectiveness of subsequent data prediction and analysis. Besides, inaccurate or incomplete data not only increases the risk of errors or malfunctions in industrial systems but also reduces their overall reliability. Therefore, productive processing of noisy and missing data has become hit subject in IIoT temporal data processing [86]. To provide high-quality data input for subsequent intelligent industrial time-series prediction models, efficient processing of industrial data is necessary after data acquisition to improve and enhance data quality.

C. Challenges in IIoT Complex Time-Series Representation

The complexity of industrial scenarios, equipment, and processes is on the rise, leading to increasingly complex sensing signals collected in IIoT. These data are typically high-throughput, multidimensional, nonlinear, and non-Gaussian, making it challenging to extract the insights hidden in industrial data. For instance, in the iron and steel smelting process, which involves high temperatures, high pressures, and complex chemical reactions, monitoring data encompasses various dimensions such as temperature, pressure, flow, and chemical composition. These data often exhibit nonlinear and non-Gaussian distributions, exemplified by variations in oxygen content and furnace temperature during converter steelmaking. Traditional statistical and signal processing methods may analyze a specific statistical indicator or signal trend. These methods may struggle to describe the temporal and spatial correlations present in raw sensor data. As a result, there is a growing need to develop high-performance representation learning methods for time-series in IIoT.

D. Challenges in IIoT Time-Series Distribution Shift Alignment

The multivariate, small-batch customization characteristic of modern discrete manufacturing makes the operating conditions of industrial processes highly variable, the independent, and identically distributed assumption does not always hold in for the monitoring data. Besides, as Industrial Internet-enabled manufacturing paradigms, such as cloud manufacturing, are often loosely federated, the data may not be publicly available [87]. For example, the lifecycle of complex equipment is usually in normal operation; thus, fault data is scarce for FD tasks. Besides, the security of production data is also a limit on modeling [87]. Thus, if industrial AI is deployed in a new scenario, it may not maintain satisfied effectiveness. Moreover, the different operating environments and equipment conditions often result in expensive and inaccessible data, which may also be unlabeled [45] and imbalanced [88]. These challenges can be attributed to the distribution shifts of industrial data, making it difficult for models to accurately capture the discrepancies in data domains and leading to poor robustness of IIoT time-series prediction models [89]. Conventional methods heavily rely on large amounts of high-quality data and may not be applicable in such scenarios.

IV. Current Solutions to the Challenges of Time-Series Prediction in IIoT

Sufficient high-quality industrial data are not always available. For instance, in time-varying industrial systems, due to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

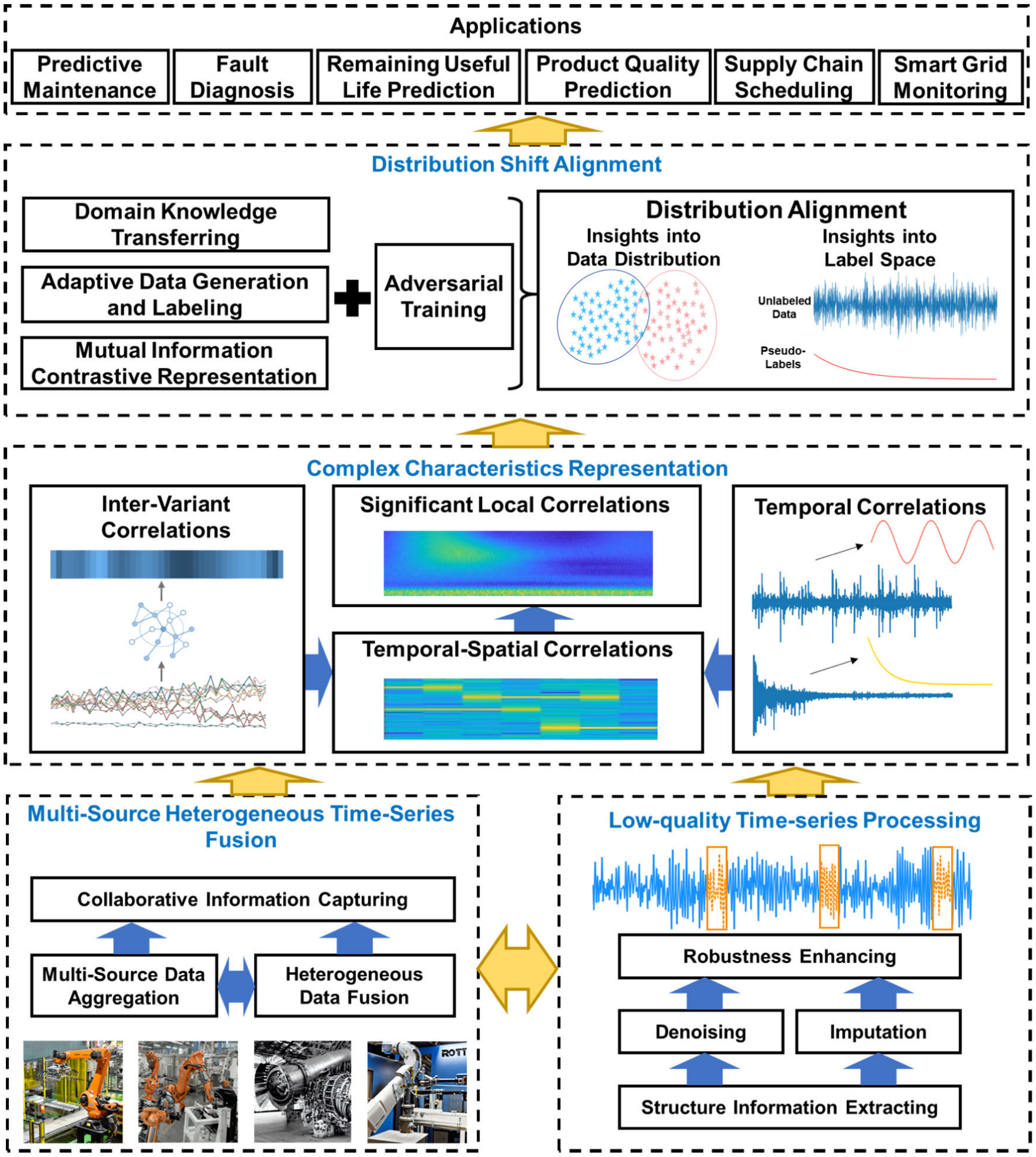

Fig. 3. Framework of state-of-the-art solutions to the challenges of time-series prediction in IIoT.

variations in equipment, operating, environments, and production tasks, monitoring data varies accordingly, making it difficult to collect sufficient and well-labeled historical data, resulting in unlabeled and imbalanced data. In power systems, unexpected data transmission interruptions, sensor failures, network delays, and extreme environmental factors may result in high-noise and widespread data missing. In a loose industrial alliance like cloud manufacturing, production-related industrial AI needs to be mutually supported and maintained by individual users, yet their respective data is unseen to each other. Supply chain data can come from different sources such as suppliers, logistics providers, and shipping companies, which may have various structures and formats. These problems make it difficult for conventional deep learning methods to meet the needs of current temporal data prediction issues in IIoT.

Nevertheless, with the develop of neural networks and advanced paradigms, these issues can be effectively managed. This section analyses the corresponding deep-learning solutions for these issues. The main challenges and their solutions are shown in Fig. 3 and Table III.

A. Methods for IIoT Multisource Heterogeneous Time-Series Fusion

Since the data collected by IIoT are usually multisource and heterogeneous, it is difficult to model accurately by simple models. It is clearly unreasonable to treat data of different modalities or structures separately, because highly correlated data between different modalities may reserve valuable insights. Thus, multisource heterogeneous time-series need to be effectively aggregated and fused.

1) Aggregation of Multisource Time-Series: Monitoring data may be collected from various sources in IIoT. And constructing and training industrial data analysis models with data from a single source may prove challenging. Therefore, integrating data from multiple sources is necessary to build more comprehensive models that can generalize well.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
TYPICAL STATE-OF-THE-ART SOLUTIONS TO THE CHALLENGES OF TIME-SERIES PREDICTION IN IIoT

Challenges	Tasks	Typical state-of-the-art methods
Multi-source heterogeneous time-series fusion	Aggregation of multi-source time-series	MSAN [90], LM-CNN [91], MLMF-CR [92], MSUDA [93], FTL-CDP [94], etc.
	Fusion of heterogeneous time-series	WDS [9], STS-D network [95], ADL-FDI4 [96], DCNN [97], MTS-DCGAN [66], COCOA [98], etc.
Low-quality time-series processing	Processing of time-series with high-noise	RDL [99], DGRU [100], MF-DRCN [101], SSDAE [37], CC-Net [102], etc.
	Processing of time-series with missing value	FIGAN [103], AM-DAE [104], SVAE + WGAN [105], GNN [106], CARMA-DACA [107], etc.
Complex time-series representation	Inter-variants correlations representation	SQAE [23], GNBN [108], SSDFE [109], GCN-SA [29], etc.
	Temporal correlations representation	LSTM + GPR [110], MDGRU [111], DeepHealth [112], etc.
	Temporal-spatial correlations representation	SuperGraph [113], MCTAN [114], STA-LSTM [22], etc.
	Significant local correlations representation	PGA-Net [26], AAE [115], GCT [116], etc.
Time-series distribution shift alignment	Processing of imbalanced data	HGTL [117], PCMMD+MIMOCN [118], DB-CGAN [119], GAN-DSM [120], VLSTM [33], CR-GAN [7], CIATL [121], CDG [70], etc.
	Processing of unlabeled data	CDAE [122], DATN [123], SSPCL [68], CADA [57], SSCDGM [56], LSTM-AE + OCSVM [34], AMBi-GAN [124], Fed-SAR [125], etc.
	Processing of unseen data	DIDBN [25], BCNN [21], AMINet [126], SBO-SCL-CL [127], Meta-GENE [128], EACRL [129], ADGN [130], etc.

Guo et al. [90] proposed a method for extracting features from multisource data and predicting the RUL of cutting tools in real time, in which the data are fed into a multiscale convolutional attention network (MSAN) to learn the features and fuse the multisource data. Ren et al. [91] proposed an FD method LM-CNN, for multisource data fusion. A multiinput multioutput strategy is introduced for implicit data augmentation, and a label-spit method is proposed for label sampling space enlarging to fuse the latent insights of the multisource data.

The effective resolution of conflicting information from multiple sources, while leveraging the complementary nature of heterogeneous information sources, is anticipated. An important aspect is to analyze the relative relationship of data from each source in the feature space. For instance, Wang et al. [92] proposed a bearing FD method based on multilocal model decision conflict resolution (MLMF-CR). The high conflict situation that may occur in the decision fusion process is considered, and the trust degree distribution is introduced to reduce the information conflict.

To address the challenges of complex operating conditions in variable sources, transfer learning can be adopted in knowledge transferring among the sources. For instance, Zhou et al. [93] proposed a dynamic transfer learning approach for industrial process prediction with limited data. The historical data from similar equipment or conditions are utilized as an auxiliary in target models' training. A multisource transfer learning framework with dynamic maximum mean difference loss is built based on the distribution distance among each historic data. Yang et al. [131] propose a transfer learning-enabled edge CNN framework for 5G industrial edge networks with privacy-preserving characteristics. The proposed architecture allows for fine-tuning of the edge CNN model without training it from scratch. Wang et al. [94] proposed a federated transfer learning framework with a central server and smart devices to address data scarcity and privacy challenges in cross-domain predictions in smart manufacturing. The central server shares knowledge and edge devices transform the base model into target domain models with task-specific data.

2) Fusion of Heterogeneous Time-Series: Advanced solutions to such problems are generally based on data fusion and multimodule collaboration [84]. Heterogeneous temporal data are extensively fused by hybrid models, in order to create a more complete and accurate view of a process or system. For instance, Ren et al. [9] propose a wide-deep-sequence (WDS) model-based data-driven prediction approach. This approach allows the fusion representation of heterogeneous features, such as multiprocess and multidimensional data of production processes, and multicategory data originating from the supply chain. It provides reliable quality predictions for industrial processes with different types of industrial data. Zhang et al. [95] proposed a Siamese time-series and difference network (STS-D network), which includes the Siamese time-series and various modules, for froth flotation process performance monitoring. The former module is adopted to extract valid and uniform representations for the input time series at the current and previous moments; afterward, the later combines the representations of the two inputs with the tagged performance at the previous moment to predict the performance at the current moment in an incremental manner. Mezair et al. [96] propose a deep learning framework for processing heterogeneous data for the task of fault detection of multisensor signals, combining LSTM, CNN, and graph convolution, and using a branch-and-bound procedure to guide the inference process. These collaborative approaches can process data in multiple formats on a single device. Shao et al. [97] proposed a multisignal FD method based on wavelet transform and CNN, capable of simultaneously taking sensor signals from multiple modalities. The acquired sensor signals are converted into time-frequency signals by wavelet transform as input to an FD model, after which a CNN is used to obtain representations of the faults from the time-frequency maps to predict the condition of the motor.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Generative learning and contrastive learning can be adopted as an auxiliary for knowledge fusion in heterogeneous data. For example, Liang et al. [66] proposed a multitimescale deep convolutional GAN to handle anomaly detection of industrial time series. The multivariate time series is first transformed into multichannel signature matrices by mutual correlation computation, and then the hidden features of the multichannel signature matrices are captured by unsupervised adversarial training. Deldari et al. [98] propose a cross modality sensor data contrastive learning method, with a novel objective function that computes intercorrelations between multisensor data to learn qualitative representations of different data patterns and minimizes similarities between uncorrelated instances.

B. Methods for IIoT Low-Quality Time-Series Processing

1) Processing of High Noise in IIoT Time-Series: The IIoT time-series data often contains a significant amount of noise, which can obscure useful information and degrade the performance of data analysis models. There are typically two approaches to addressing this problem: one is to filter out noise through data preprocessing. Yao et al. [99] developed a mechanical fault detection method that employs a recursive learning strategy to effectively mitigate noise. Their approach introduces a novel multilevel attention mechanism, which recursively tracks the noise components and gradually denoises them in a coarse-to-fine manner, leading to satisfied noise suppression performance. Yan et al. [100] introduced a denoising GRU (DGRU) [100] to mitigate the effect of noise by adding a denoising gate to the GRU and applied it for sinter ore burn-through point prediction.

The other is to reduce the impact of noise by enhancing the robustness of the model to noise. For instance, Xu et al. [101] proposed a multireceptive field denoising (MFD) residual convolutional network (MF-DRCN) for FD in strong noise conditions, in which an MFD block is designed to enhance the deep features extracted by the CNN model and filter out the interference feature information. Liu et al. [37] propose a complex industrial process fault identification scheme, stacked sparse denoised AE (SSDAE)-Softmax, which can automatically and adaptively learn the potential intrinsic features of fault data with noise contamination. Chang et al. [102] proposed a CC-Net-based soft sensor method for denitrification and desulfurization processes. Contrastive learning is adopted for separate normal samples and anomalies to obtain robustness to the anomaly.

2) Processing of Missing Value in IIoT Time Series: The most intuitive solution for missing values in data is imputation, i.e., artificially generated data to fill in the missing positions. Traditional data imputation is generally achieved by statistical methods. However, these methods usually require a defined optimization objective. If the optimization objective is deviated, then such imputation methods will be ineffective. Besides, such imputation methods can also reduce the robustness of the model. Thus, more "intelligent" methods are needed.

The paradigm of generative learning becomes a viable solution. It is also the most extensively adopted method. Yao and Zhao [103] propose a downstream task-oriented customized soft sensor data imputation method, fine-tuned imputation GAN. This method customizes data imputation by discriminating key variables of quality relevance and interpolating them precisely. Pseudolabeling is applied to overcome the problems of interactive optimization of data imputation and label prediction. Gao et al. [105] propose a soft sensor data augmentation method combining stacked variational AEs (SVAEs) and Wasserstein GANs (WGANs) for the process industry. Depth features are extracted by stacking of SVAE and generative models are constructed by combining SVAE and WGAN for generating missing data and prediction. Pan et al. [104] proposed an adaptive-learned median-filled deep AE for missing values imputation, which continuously replaces the missing values by the median of the input data and its reconstruction. And it pays more attention to the reconstruction learning of nonmissing values or missing values in different iteration periods.

Graph-based techniques can enhance the robustness of models to missing data and mitigate their impact on downstream tasks by extracting structural information. For instance, Kang [106] proposed a graph-based method that employs GNNs for data-driven product fault detection with missing values. The graph represents the variables and their pairwise relationships to improve the robustness of the prediction model. Kavianpour et al. [107] proposed a semisupervised approach based on ARMA graph convolution, adversarial adaptive, and multilayer multikernel local MMD. These techniques extract structural information from data, align classes, reduce differences in the structural distribution in domains, and address the problem of missing data.

C. Methods for IIoT Complex Time-Series Representation

1) Representation of Intervariants Correlations: In manufacturing systems, IoT aggregates a wide range of dissimilar data. However, the excessive size of the data can lead to a waste of computational and storage resources, while redundant information can deteriorate the performance of data analysis models [132]. For instance, Yuan et al. [23] proposed a product quality prediction method adopting a stacked quality-driven self-encoder (SQAE) to generate quality-dependent feature representations. With a quality-driven AE, irrelevant information from raw input data is reduced through reconstruction using input and quality data. This approach enables a deep SQAE network to learn hierarchical quality-related features by progressively reducing irrelevant features layer by layer. Wang et al. [108] proposed an optimized condition prediction method based on a Gaussian–Bernoulli DBN (GDBM) model. The model introduces extreme value perturbations and simple particle swarm optimization methods to optimize the model hyperparameters. This method can balance efficiency and accuracy to some extent.

Moreover, the various signals may be highly coupled and correlated. While the information may be redundant, it can provide important industrial insights that are essential for industrial AI. Therefore, it is necessary to perform feature extraction and representational modeling of high-dimensional industrial data to compress the volume of data while capturing

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

as many industrial insights as possible. For instance, Yao and Ge [109] proposed an industrial quality predictor based on cooperative depth dynamic feature extraction and variable time delay estimation with a semisupervised dynamic feature extraction (SSDFE) network to extract nonlinear dynamic features to build a regression model for quality prediction. Chen et al. proposed a fault prediagnosis framework to obtain fault association graphs and constructed FD models using graph convolutional networks with the weight coefficients [29].

2) Representation of Temporal Correlations: The periodic changes and degradation of time-series signals collected in IIoT can be applied to interpret the state change trends of equipment. These time-series are crucial for improving the accuracy of analytical models, making it extremely important to capture temporal features in IIoT time-series. The temporal correlations can be captured by several neural network modules, such as CNN, LSTM, and attention mechanism. Singh Chadha et al. [133] proposed an FD method based on CNN, exploiting the temporal dependencies in process data. Liu et al. [110] propose a future capacity method for lithium-ion batteries RUL prediction by fusing LSTM and Gaussian process regression (LSTM + GPR). The method decomposes raw battery capacity data using empirical mode decomposition (EMD) into intrinsic mode functions (IMFs) and a residual. An LSTM submodel estimates the residual, while a GPR submodel fits the IMFs with uncertainty estimation. Ren et al. [111] proposed a multiscale dense GRU (MDGRU) for the RUL prediction of rotating bearings, which captures serial attributes and integrates information from different time scales. Song et al. [186] introduced an RUL prediction method based on attention mechanisms, which captures long-term dependencies within sequence data by modeling the relationships between different time series. Zhang et al. [112] introduced DeepHealth, a self-attention-based framework for predictive operation and maintenance. DeepHealth comprises two submodels for health perception and sequence prediction. To capture temporal correlations, an enhanced attention mechanism captures global dependencies in vibration signals, facilitating both long-term and short-term sequence predictions for timely maintenance decisions.

3) Representation of Temporal–Spatial Correlations: The simultaneous acquisition of intervariants and sequence correlation in industrial multivariate time-series is also a feasible and effective method. Yang et al. [113] proposed a spatiotemporal graph-based feature extraction method called SuperGraph for rotating machinery FD. The spatiotemporal graph is constructed with spectral analysis and feature vectors based on Laplace matrices are extracted from it. Each node of the graph represents a spatiotemporal graph, which is called SuperGraph. Classifying the nodes in SuperGraph enables graph classification. Ren et al. [114] proposed a multichannel temporal attention-based network (MCTAN) for health indicator prediction of aircraft engines. The contribution of data collected by various sensors is measured by channel attention and the potential long-range temporal relationships are effectively extracted by a multiheaded local attention mechanism. Yuan et al. [22] proposed a spatiotemporal attention-based LSTM for quality prediction. To select variables associated with quality prediction at each time step and identify hidden states for quality prediction at various time steps, the spatial and temporal variable attention are introduced.

4) Representation of Significant Local Correlations: The information in IIoT data is often sparse, and for a given industrial issue, critical insights may exist mainly in data related to a specific time and space. Therefore, it is expected to focus more on obtaining the crucial feature significantly relevant to the current task, which is extremely important for both the accuracy and speed of industrial prediction models. For instance, Dong et al. [26] proposed PGA-Net, a pyramid feature fusion and global contextual attention network. Multiscale features are extracted and then fused into five resolutions with dense skip connections in a pyramid feature fusion module. The global contextual attention module facilitates effective information propagation from low-resolution to high-resolution fused feature maps. Jang et al. [115] proposed an adversarial AE (AAE)-based process monitoring system capable of generating features that adhere to a specified prior distribution, capturing high-dimensional data information manifold. Geng et al. [116] introduced a gated convolutional neural network-based transformer (GCT), a dynamic soft sensor, considering the nonlinearity, dynamics, and noise present in industrial time-series data. GCT utilizes a modified gated convolutional neural network to adaptively filter important features and employs a multihead attention mechanism to model correlations between different time moments.

D. Methods for IIoT Time-Series Distribution Shift Alignment

Gaining insights into data distribution and labeling space is crucial for effectively aligning distribution shifts in IIoT time series. Thus, addressing the challenges of data imbalance, unlabeled, and unseen data requires considering both aspects.

1) Processing of Imbalanced Data: Data imbalanced is one of the most general issues in industrial data analysis, especially in IIoT time-series analysis. Traditional methods are usually realized by simple downsampling and oversampling. Obviously, these methods may lead to bias in resampling samples. There are two broad advanced solutions to this problem, one is to improve the robustness and generalization of the model by learning domain invariance or domain generalization, and the other is to change the distribution of data by data augmentation with "pseudodata." The former is generally realized with transfer learning and the latter with generative learning.

Transfer learning is one of the favored approaches to data imbalance, by decreasing the dependence of the constructed deep learning models on the huge amount of target domain data. For instance, Chen et al. [117] proposed a hierarchy-guided transfer learning (HGTL) framework for small sample fault identification. It constructs category affinities using domain knowledge, label semantics, and interclass distances, forming a hierarchical cluster structure. The hierarchical feature learning network is pretrained on source domain majority class samples with high similarity to extract transferable fault information. By leveraging information about similar faults, this framework enables feature extraction from minority sample classes. Cao et al. [118] proposed a transfer learning model combining pseudocategorized maximum

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

mean difference (PCMMD) and multiinput multioutput convolutional networks (MIMOCNs). A domain-shared encoder and classifier are trained with labeled data from the source domain. Then, the MIMOCN is trained with both the source data and target data, incorporating pseudolabeling to assess cross-domain distribution differences within classes.

Generative learning-based approaches for data generation and data augmentation have been seen as a viable way to address data imbalanced. For instance, Zhou et al. [119] proposed a distribution bias-aware collaborative GAN, to mitigate distribution bias in imbalanced IIoT data. It incorporates a supplementary classifier for data augmentation and a conditional generator with random labels to augment samples in the minor classes. Generators, discriminators, and classifiers are collaborative adversarial trained using a weight-sharing scheme. Yang et al. proposed a GAN-based data generation method to address the problem of data imbalance in FD [120]. Various GANs were used to generate various fault spectrum data, and a data selection module (DSM) was used to filter and purify the data. The filtered generated data is combined with real data to form a balanced dataset for FD. To address the inconsistency between dimensionality reduction and feature retention in imbalanced IIoT time-series, Zhou et al. [33] propose a variational LSTM (VLSTM) model for intelligent anomaly detection using reconstructed feature representation. Zhang et al. [7] propose an RUL prediction framework with data self-generation for acyclic and cyclic degenerate models. By employing a convolutional recursive GAN, photorealistic time-series data are generated. Furthermore, a hierarchical framework is introduced to integrate the generated data into the existing RUL estimation method.

Contrastive learning also can be adopted in concert with transfer learning to extract similarity features cross domain to deal with imbalanced data. For instance, Kuang et al. [121] proposed a class imbalance adversarial transfer learning (CIATL) network to learn domain invariants and knowledge for FD tasks with imbalanced data. Zhang et al. [70] propose CDG, a contrastive decoder generator for Few-shot quality prediction in multistage manufacturing processes. CDG addresses the lack of annotation and handles previously unseen tasks without additional training. It comprises a machine feature encoder for encoding machine feature dependencies, a contrastive stage and task feature generator for self-supervised generation of task and stage vectors using contrastive learning, and an instance-specific decoder generator for generating weight and deviation parameters based on query-support vector correlation, facilitating quality prediction.

2) Processing of Unlabeled Data: Industrial data unlabeled is a typical task for transfer learning. In traditional approaches, unlabeled data are generally processed by clustering methods. It is inefficient and the inconsistency of clustering goals and downstream tasks may lead to unsatisfied performance. Thus, some semisupervised and self-supervised methods have become of interest to researchers.

By applying semisupervised or self-supervised methods, domain knowledge in the source domain can be transferred to the target domain. Thus, the most extensively applied method for data unlabeled is transfer learning. For instance, Mao et al. [122] propose a transfer learning method for RUL prediction of rolling bearings across operating conditions. After the prediction model is trained in the original domain, transfer component analysis is applied to calibrate the original model to a model applicable to the target domain. Chen et al. [123] proposed DATN, a domain adversarial transfer network, to transfer fault-related knowledge across large domain shifts. It employs two asymmetric encoder complexes with deep CNN to learn hierarchical representations from the source and target domains. The network trained on the source task is transferred to enhance training on the target task. To minimize interdomain differences, domain adversarial training with inverted label loss is incorporated. However, though transfer learning addresses the transfer of knowledge among different scenarios or tasks, the diversity of devices and sensors in IIoT causes the types of data and labels in the origin and target domains to be potentially different [134].

Contrastive learning is another viable solution [135]. It can learn usable features from unlabeled data on its own. Ding et al. [68] propose a bearing fault prognosis method with self-supervised pretraining using contrastive learning. Their method utilizes SSPCL representation learning to enhance generalization and obtain discriminative depth features from unlabeled bearing datasets. Based on this, a semisupervised intelligent early fault detection method is proposed to further enhance generalization by leveraging a large amount of unlabeled data. Ragab et al. [57] propose a cross-domain mechanical RUL prediction method based on contrast adversarial domain adaptive (CADA). InfoNCE [72] is used as one of the indicators of domain alignment to maximize the representation of the MI among domains.

Generative learning-based methods can also be an effective approach to such issues. For instance, Ko and Kim [56] proposed a semisupervised deep generative model for fault classification with unlabeled data. To decrease the reliance on labeled anomaly data and leverage the abundance of normal data in normal working conditions, Zuo et al. [34] propose a semisupervised multivariate time-series prediction method combining LSTM-AE and OCSVM. Kong et al. [124] proposed an AMBi-GAN-based deep generative model for industrial multidimensional time-series anomaly with unlabeled data. Abdel-Basset et al. [125] proposed a privacy-preserving federated semisupervised class-rebalanced framework for anomaly detection, in which a semisupervised generative network is introduced to enhance the generated samples and model the relationships between labeled and unlabeled data.

3) Processing of Unseen Data: In general, unseen industrial data is addressed through domain randomization or domain generalization. The former focuses on data by eliminating domain specificity through augmentation, while the later emphasizes model robustness across multiple domains via specific construction ideas and optimization objectives like domain alignment, semantic matching, and regularization.

Domain randomization can be realized by generative learning and contrastive learning, by generating synthetic data and capturing the MI of the tasks and the generated data to prevent unknown domain shifts. For instance, Xing et al. [25]

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

proposed distribution invariant DBN (DIDBN) for machinery FD, which can directly extract distribution-invariant features from raw machinery vibration data. Yu and Zhao [21] proposed a broad CNN-based incremental FD model to represent the fault process and update the diagnosis model to include emerging anomalous samples and fault classes. Zhao and Shen [126] proposed an adversarial MI-guided approach for FD in unknown conditions. A domain generation module is utilized to generate synthetic target data, enhancing generalization capability. And MI between the domain generation module and task diagnosis module is captured to mitigate unknown domain shifts. Peng et al. [127] tackle the challenge of classifying unknown faults as normal conditions by employing supervised contrastive learning. This approach enables the learning of discriminative and compact embeddings for known normal and fault conditions. The method involves comparing given samples as well as comparing normal samples with their own negative augmentations generated through soft Brownian offset sampling.

The domain generalization-based methods can be induced into unsupervised transfer learning without target data. In these methods, contrastive learning can also be adopted to capture MI capturing and domain invariants. For instance, Ren et al. [128] proposed a domain generalization framework for FD in an unseen condition called Meta-GENE. A gradient aligning algorithm is introduced to learn domain-invariant strategy for robust prediction and a semantic matching technique is proposed to alleviate low-resource problems. Sun et al. [129] proposed an environmentally adaptive and contrastive representation learning method for condition monitoring in variational environments and volatile operations. Li et al. [130] proposed an adversarial domain generalization network to diagnose faults in unknown operating environments. The classification boundary is detected by maximizing the classifier differences, and better feature mapping functions and domain-invariant features are obtained by adversarial training.

V. Typical Applications of Time-Series Prediction in IIoT

A. Prognostics and Health Management of Complex Equipment

By analyzing process monitoring data from the production and manufacturing line, as well as real-time equipment status data, PHM aims to uncover system failure characteristics, proactively detect potential anomalies during system operation, diagnose the root cause of anomalies, and predict the RUL of the equipment [136], [137]. It is crucial for the safe and reliable operation of the equipment Fig. 4 shows a schematic architecture of complex equipment PHM system [138]. Ren et al. [138] proposed an IIoT complex equipment RUL prediction framework. The edge plane collects the monitoring data and preprocesses the data for denoise and imputation. Real-time prediction results are obtained on the edge plane, higher accuracy prediction results are obtained by historical information on the cloud plane, and parameters are updated through continuous learning.

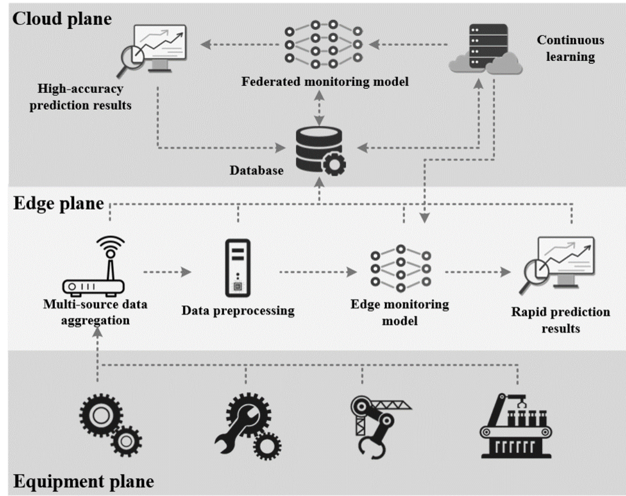

Fig. 4. Schematic architecture of IIoT complex equipment PHM system [138].

Deep learning has eliminated the need for separate data feature extraction and subsequent analysis, which were previously required for PHM tasks. This has allowed for the direct use of collected data to predict health indicators. Besides, the complexity of industrial mechanisms presents challenges in discovering and accurately modeling them. As a result, traditional methods often struggle to gain insights into industrial processes and make precise analyses and predictions of the health indicators. The industrial data analyzed in PHM tasks include vibration signals, acceleration signals, temperature data, optical signals, acoustic signals, electrical properties of equipment, and so on. These signals are generally collected and processed as time-series data, making PHM one of the main application scenarios for time-series prediction in IIoT [139], [140]. Zhu et al. [141] proposed an RUL estimation method with time-frequency representation and multiscale CNN. Wen et al. [142] developed a reinforcement learning-based learning rate scheduler for efficiently and automatically scheduling the learning rate of fault classification models, which can adaptively implement the training of PHM models. Liu et al. [143] developed a fault diagnosis framework based on a dislocated time-series convolutional neural network (DTS-CNN). Dislocation layers were added to the neural network to extract relationships between signals in different intervals of periodic mechanical signals, allowing multiscale periodic features in signals with nonstationary conditions to be captured.

B. Product Quality Management

In PLM, it is of significance to utilize product manufacturing process data and quality inspection data to manage product quality, achieve product production traceability, and optimize processes [140], [144]. Fig. 5 shows a typical product quality prediction framework. By analyzing the correlation of product quality with processing temporal signals, such as raw material quality, equipment status parameters, process flow, and workshop environment, the main factors affecting product quality are identified. Furthermore, by constructing linear and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

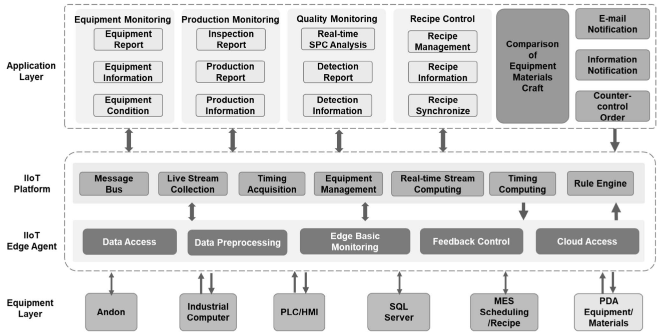

Fig. 5. Typical product quality prediction framework in IIoT.

nonlinear mapping relationships between product quality and these factors, accurate predictions of product quality can be made. In addition, intelligent optimization algorithms can be used to optimize the production process and achieve adaptive product quality monitoring and optimization control.

In order to predict product quality, two main tasks are typically undertaken. First, correlation analysis is conducted between influencing factors and product quality indicators to select features for big data analysis, resulting in a low-cost and efficient prediction. Second, a data-driven model is constructed for the final quality indicator prediction. Traditional data processing methods are used to analyze the correlations between various factors and quality indicators to reveal the causes of product quality problems. However, these methods use artificially defined correlation metrics that may not provide a comprehensive and in-depth interpretation of the relationship between factors and quality metrics. As a result, the correlations analyzed may be biased, leading to inaccurate identification of key factors. Deep learning techniques can adaptively extract correlations between influencing factors and product quality metrics, and are, therefore, of wide interest in this area. Liu et al. [145] developed a domain-adaptive extreme learning machine as a basis for soft sensor modeling for product prediction for multilevel processes with limited labeling data. Zhang et al. [146] proposed a material removal rate prediction method for the chemical mechanical polishing process with the use of residual CNN.

C. Supply Chain Optimization and Production Scheduling

The supply chain of a product involves the whole process from production to sales and service, and the optimization of the supply chain is of great value to reduce costs and improve efficiency and is one of the dominant application scenarios of IIoT time-series prediction [147]. Supply chain optimization and prediction, as a system-level complex issue, contains subproblems, such as demand prediction [148], risk prediction [149], production resource scheduling [150], and production planning scheduling [151]. Fig. 6 shows a typical framework of a supply chain system introduced by Ivanov et al. [152] data from each node of the supply chain is collected and fused for analysis, optimization, simulation, and control of each link in the supply chain system, thereby supporting the supply, manufacturing, and sales services of the supply chain system.

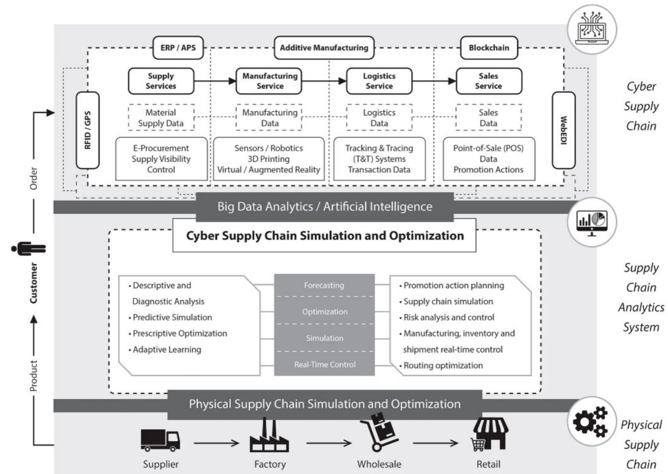

Fig. 6. Typical framework of supply chain system in IIoT [152].

Supply chain optimization and prediction is one of the core problems of enterprise operation and management. A high-quality supply chain management system can allocate production resources effectively and can demonstrate high robustness to various unexpected situations. However, traditional prediction and scheduling methods focus on the optimal utilization of performance and resources in static environments, which may lead to the deterioration or failure of static progress.

Deep learning-based approaches [153] provide a series of effective methods and tools for the optimization of production planning and scheduling in dynamic environments. Deep learning techniques can provide comprehensive information support for supply chain prediction and scheduling decisions. For example, Lima-Junior and Carpinetti [154] proposed an artificial neural network-based prediction system to make prospective diagnoses of supply chain performance and facilitate rational decision making. Shi et al. [155] proposed a deep reinforcement learning method to schedule automated production lines to avoid manual feature extraction and overcome the lack of structured datasets, which improves the adaptability and flexibility of automated production lines. Tong et al. [156] proposed an effective task scheduling method DDQN-TS to achieve high-quality service with limited resources, which exploits the adaptive learning capability of dual DQN (DDQN) to explore the optimal task scheduling policy.

D. Smart Grid

The application of time-series prediction for smart grids has a wide range of benefits, including fault and anomaly induced outage detection, load and generation prediction, load management with demand response, and asset management [157]. These capabilities have the potential to optimize energy delivery by reducing costs and improving the overall quality of energy services [158]. Fig. 7 shows a typical model of a smart grid analytic system [159]. In the system proposed in [159], IIoT time series are collected and analyzed. The analysis results can be used for advanced maintenance, such

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

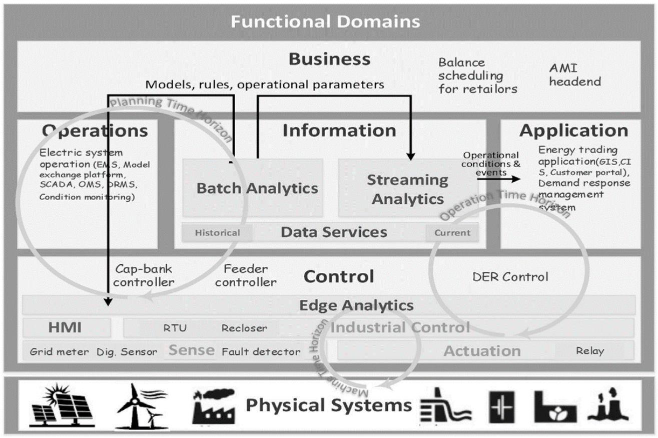

Fig. 7. Typical model of smart grid analytic system in IIoT [159].

as automatic fault identification and preventive maintenance, or to optimize the performance of assets and machines.

A smart grid is crucial for business and government decision-makers. Smart grid management presents several challenges. In addition to the high energy demands, speed requirements, and the need for high-capacity storage, the data resources in the smart grid are not centrally distributed and the collected data is often heterogeneous. Traditional methods can hardly address these issues. Many studies on deep learning-based solutions have been conducted in this area. For instance, a novel GRU model coupling two new mechanisms of selective state updating and adaptive mixed gradient optimization for accurate power time-series prediction was proposed by Zheng and Chen [160]. Jia et al. [161] proposed a retroactive scheduling regime in handling heterogeneous schedulable sources in small-scale microgrids.

VI. Discussion and Prospects

Although deep learning has a wide range of applications for time-series prediction in IIoT, current research still has much room for improvement. Existing deep learning methods, which still fail to meet the needs of industrial applications, face many challenges and require urgent attention from researchers.

A. Lightweight Model for Industrial Edge Intelligence

Numerous time-series prediction models for IIoT applications have been proposed, primarily focusing on accuracy in specific industrial applications. However, these models often overlook the importance of efficiency and computational resource consumption [162]. In industrial settings, AI applications often require low latency. For example, a high-temperature furnace refinery monitoring system needs to make predictions within tens of milliseconds. Additionally, computing devices in industrial sites typically have limited storage and computational capacity, often in the range of megabytes, which makes it challenging to accommodate large-scale industrial AI models that require extensive computational and storage resources. Therefore, there has been a significant research and development trend in the field of industrial intelligence toward lightweight and efficient deep learning models. These models aim to strike a balance between accuracy and resource efficiency, enabling effective deployment in industrial environments. Currently, several compression methods for large neural networks are being explored, including parameter compression [163], pruning [164], and distillation [165]. These techniques aim to reduce the model size and computational requirements without significantly sacrificing predictive performance.

Knowledge distillation can be seen as one of the mainstream methods of neural network "slimming" in IIoT. For instance, as industrial edge devices are limited in terms of computational and storage capacity, Xu et al. [165] proposed a novel knowledge distillation framework called KDnet-RUL for compressing complex LSTM models into structurally simple CNN models without loss of accuracy. Fang et al. [166] proposed an on-demand DNN model inference system for industrial edge devices, called knowledge distillation and early exit on edge (EdgeKE), in which large complex models are distilled and supervised into compact edge models, and an early exit strategy is utilized to provide flexibility to meet different latency or accuracy requirements of edge applications.

The dynamic neural network presents a novel solution to the issue. It possesses the ability to adaptively adjust its own depth and width based on the difficulty of feature extraction for each sample. This dynamic adjustment enables the simplification of the inference process while maintaining satisfactory characterization capability. Meanwhile, this dynamic selection mechanism of network modules indirectly contributes to the interpretability of the model. Consequently, this has made it attractive to researchers. Although the application of this method in IIoT data processing is not yet widespread, the authors believe it holds significant value and represents an interesting area of research with broad potential.

Applying a fusion of these methods is the key to addressing the lightweighting of industrial intelligence models. Lightweight construction and real-time computation algorithms for industrial intelligence models will be essential research areas in the future.

B. Cloud-Edge Distributed Industrial Intelligence

With the emergence of new manufacturing modalities and paradigms, such as cloud manufacturing [167], industrial applications have evolved toward a more distributed and collaborative approach. Consequently, the edge layer is shouldering an increasing number of functional and performance tasks [168]. To address this challenge, the collaboration between industrial cloud and edge computing has become a critical technology and essential support in the current Industrial Internet [169]. In response, academia has proposed various cloud, edge, and fog computing-based industrial neural networks, which have garnered significant interest from the industry [138], [170], [171], [172].

However, despite the growing interest in distributed industrial intelligence models, current research is still in its early stages. Designing such models that cater to the unique characteristics of industrial applications and the cloud-edge distributed model of Industrial Internet data remains a challenge. On the one hand, in the industrial area, cloud-edge

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

collaboration is still at the stage of conceptual framework and lacks theoretical architecture for in-depth integration of industrial AI. On the other hand, current distributed models take less account of the specificity requirements of multiple edges in the Industrial Internet and the distributed requirements of the cloud and the corresponding multiple edges for data analysis models. Therefore, the convergence of intelligence in the cloud edge of the Industrial Internet is also an area of significant research in the future.

C. Adaptive Learning for Varying Scenarios in Industry

Application scenarios in industrial manufacturing are highly diverse and prone to various uncertainties. For example, similar manufacturing tasks can yield varying outcomes under various working conditions. Previously validated predictive models for intelligent analysis of industrial data may not be applicable to new working conditions when they are altered. While training proprietary industrial deep learning models with data specific to each scenario is a viable solution, it can be resource-intensive in terms of manpower, funds, and time. Moreover, it can lead to a degradation in the ability to analyze previous problems encountered in different application scenarios while deploying new industrial data intelligence models, which is not desirable for users.

It requires deep learning models to be adaptive to the attributes and regularities of new application scenarios while maintaining the adaptive capabilities of previously existing application scenarios [173], [174], [175]. Therefore, the industrial models with adaptability and generalization capabilities would greatly facilitate the application of IIoT time-series prediction in the ever-changing industrial application scenarios. One notable approach that has been extensively studied in the field of deep learning is "learning without forgetting" [174], [176]. The aforementioned industrial intelligence model, which exhibits strong generalization and self-adaptive capabilities, can be seen as an integration and innovation of transfer learning, multitask learning, and continual learning. It represents a crucial research direction for the future.

D. Interpretable/Reliable/Credible Industrial Model

Many advanced deep learning techniques widely adopted in industries today are considered "black-box" models. These models offer superior performance in terms of accuracy and deployment cost; however, they lack interpretability when it comes to understanding their internal operations. With "black-box" models, we can observe the input and output data, but we cannot gain detailed insights into the workings of the deep neural networks themselves. The complexity and opacity of these industrial intelligence models make it challenging to assess their reliability and trustworthiness [177]. In sectors, such as aviation and nuclear industries, where reliability and trustworthiness are paramount, even minor errors can lead to severe consequences. Consequently, the widespread adoption of highly complex and uncertain neural network models becomes a challenge. Establishing trustworthiness in AI models is a consensus in various fields, not just limited to industrial manufacturing [178], [179].

As a result, there is a growing demand for industrial AI models to be highly interpretable and credible. In recent years, the physics-informed neural network (PINN) approach has gained significant attention among researchers [180]. PINN combines physical mechanics principles with neural networks, offering a promising avenue to explore the credibility and interpretability of industrial intelligence models.

E. Industrial Foundation Model

The various industrial intelligence technologies that have emerged in recent years are mostly "mini-model" with a strong correlation to domain-specific issues that originate from the application needs of specific industrial fields. With the integration of advanced information technology and industrial manufacturing, the formation of increasingly complex industrial cyber–physical systems (CPSs), and AI in industrial manufacturing will face increasingly complex problems at the system level. For instance, in the Industrial Internet system, the analysis, prediction, and optimization issues of the supply chain system of the large-scale intelligent manufacturing industry chain, involve cross-industry, cross-enterprise, and cross-process multilayer industrial manufacturing systems with complex correlation relationships. The acquisition of the intrinsic operating mechanisms and pattern laws of such complex systems will confront the problem of large-scale integration of deep learning of large-scale subproblems driven by super large-scale multisource heterogeneous industrial data. The existing industrial intelligence "mini model" techniques for single-subproblems in specific fields are not able to cope with this, and there is an urgent need to explore new industrial intelligence techniques for such large and complex industrial system problems.

The foundation model technique provides a potential solution to this problem [181], [182]. It should be noted that the philosophy of the foundation model here is different from ChatGPT, where the foundation model in the industry focuses on common knowledge in cross-domain tasks. As the latest research hotspot in the field of deep learning, the foundation model provides a fundamental and homogeneous "cornerstone" modeling technique that can support model pretraining and fast adapting to downstream tasks and can break through the accuracy limitations of current small models at the structural level. The foundation models have continued to break through thousands, millions, and even trillions of model parameter scales in recent years and have been applied with initial success in areas, such as natural language processing and biocomputing. Examples of proposed models, such as GPT-3 [183], switch transformer [184], and Alphafold [185], have shown phenomenal astonishing capabilities. Currently, there is still a gap in the exploration of foundation models in industrial manufacturing, and foundation models for industrial intelligence are expected to be one of the disruptive technologies to solve various highly complex problems in the process of industrial manufacturing intelligence in the future.

VII. CONCLUSION

Deep learning has made remarkable achievements in the field of industrial intelligence, especially in IIoT time-series

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

prediction. In this survey, deep learning-based approaches to the IIoT time-series prediction issues are discussed in the following aspects.

1) Existing deep learning-based time-series prediction methods are illustrated and summarized.
2) The practical challenges of time-series prediction in IIoT are presented and analyzed.
3) A framework of state-of-the-art solutions for time-series prediction in IIoT is proposed to address the challenges and typical related works are categorized and reviewed.
4) Typical industrial application scenarios of time-series prediction in IIoT are summarized and discussed.
5) Possible future directions and prospects of this promising area of industrial intelligence are put forward and discussed.

It is anticipated that this survey will serve as a catalyst for IIoT and deep learning researchers to delve deeper into this fascinating research field and innovate more advanced and sophisticated deep learning models tailored to IIoT applications.

References

[1] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial Internet of Things (IIoT): An analysis framework," *Comput. Ind.*, vol. 101, pp. 1–12, Oct. 2018.

[2] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, opportunities, and directions," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4724–4734, Nov. 2018.

[3] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," *Inf. Sci.*, vol. 305, pp. 357–383, Jun. 2015.

[4] K. Gu, D. Tao, J. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.

[5] B. Cai et al., "Artificial intelligence enhanced reliability assessment methodology with small samples," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Nov. 25, 2021, doi: 10.1109/TNNLS.2021.3128514.

[6] L. Chettri and R. Bera, "A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 16–32, Jan. 2020.

[7] X. Zhang, Y. Qin, C. Yuen, L. Jayasinghe, and X. Liu, "Time-series regeneration with convolutional recurrent generative adversarial network for remaining useful life estimation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6820–6831, Oct. 2021.

[8] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Gearbox fault diagnosis using a deep learning model with limited data sample," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6263–6271, Oct. 2020.

[9] L. Ren, Z. Meng, X. Wang, R. Lu, and L. T. Yang, "A wide-deep-sequence model-based quality prediction method in industrial process analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3721–3731, Sep. 2020.

[10] S. Singh, S. C. Misra, and S. Kumar, "Identification and ranking of the risk factors involved in PLM implementation," *Int. J. Prod. Econ.*, vol. 222, Apr. 2020, Art. no. 107496.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] L. Zhou, L. Zhang, and N. Konz, "Computer vision techniques in manufacturing," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 1, pp. 105–117, Jan. 2023, doi: 10.1109/TSMC.2022.3166397.

[13] I. Makhdoom, M. Abolhasan, J. Lipman, R. P. Liu, and W. Ni, "Anatomy of threats to the Internet of Things," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1636–1675, 2nd Quart., 2019.

[14] J. C. B. Gamboa, "Deep learning for time-series analysis," 2017, *arXiv:1701.01887*.

[15] Q. Wen et al., "Time series data augmentation for deep learning: A survey," 2020, *arXiv:2002.12478*.

[16] W. Zhang, D. Yang, and H. Wang, "Data-driven methods for predictive maintenance of industrial equipment: A survey," *IEEE Syst. J.*, vol. 13, no. 3, pp. 2213–2227, Sep. 2019.

[17] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[18] S. De, M. Bermudez-Edo, H. Xu, and Z. Cai, "Deep generative models in the industrial Internet of Things: A survey," *IEEE Trans. Ind. Informat.*, vol. 18, no. 9, pp. 5728–5737, Sep. 2022.

[19] B. Maschler and M. Weyrich, "Deep transfer learning for industrial automation: A review and discussion of new techniques for data-driven machine learning," *IEEE Ind. Electron. Mag.*, vol. 15, no. 2, pp. 65–75, Jun. 2021.

[20] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9521–9530, Dec. 2019.

[21] W. Yu and C. Zhao, "Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability," *IEEE Trans. Ind. Electron.*, vol. 67, no. 6, pp. 5081–5091, Jun. 2020.

[22] X. Yuan, L. Li, Y. A. W. Shardt, Y. Wang, and C. Yang, "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4404–4414, May 2021.

[23] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang, and W. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3721–3730, Jun. 2020.

[24] G. Wang, J. Qiao, J. Bi, Q. Jia, and M. Zhou, "An adaptive deep belief network with sparse restricted Boltzmann machines," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4217–4228, Oct. 2020.

[25] S. Xing, Y. Lei, S. Wang, and F. Jia, "Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2617–2625, Mar. 2021.

[26] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, "PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7448–7458, Dec. 2020.

[27] Y. Qin, S. Xiang, Y. Chai, and H. Chen, "Macroscopic-microscopic attention in LSTM networks based on fusion features for gear remaining life prediction," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10865–10875, Dec. 2020.

[28] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Multireceptive field graph convolutional networks for machine fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12739–12749, Dec. 2021.

[29] Z. Chen, J. Xu, T. Peng, and C. Yang, "Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9157–9169, Sep. 2022.

[30] L. Guo, Y. Lei, N. Li, T. Yan, and N. Li, "Machinery health indicator construction based on convolutional neural networks considering trend burr," *Neurocomputing*, vol. 292, pp. 142–150, May 2018.

[31] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[32] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.

[33] X. Zhou, Y. Hu, W. Liang, J. Ma, and Q. Jin, "Variational LSTM enhanced anomaly detection for industrial big data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3469–3477, May 2021.

[34] Z. Zuo, L. Ma, S. Liang, J. Liang, H. Zhang, and T. Liu, "A semi-supervised leakage detection method driven by multivariate time series for natural gas gathering pipeline," *Process Saf. Environ. Protection*, vol. 164, pp. 468–478, Aug. 2022.

[35] Y. Zheng, S. Chen, Y. Xue, and J. Xue, "A pythagorean-type fuzzy deep denoising autoencoder for industrial accident early warning," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1561–1575, Dec. 2017.

[36] Y. Wang, C. Liu, and X. Yuan, "Stacked locality preserving autoencoder for feature extraction and its application for industrial process data modeling," *Chemometric Intell. Lab. Syst.*, vol. 203, Aug. 2020, Art. no. 104086.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[37] J. Liu et al., "Toward robust fault identification of complex industrial processes using stacked sparse-denoising autoencoder with softmax classifier," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 428–442, Jan. 2023.

[38] J. Yu, X. Liu, and L. Ye, "Convolutional long short-term memory autoencoder-based feature learning for fault detection in industrial processes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021.

[39] R. Xie, N. M. Jan, K. Hao, L. Chen, and B. Huang, "Supervised variational autoencoders for soft sensor modeling with missing data," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2820–2828, Apr. 2020.

[40] M. Keshk, B. Turnbull, N. Moustafa, D. Vatsalan, and K. R. Choo, "A privacy-preserving-framework-based blockchain and deep learning for protecting smart power networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5110–5118, Aug. 2020.

[41] W. Yu and C. Zhao, "Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 3, pp. 1083–1091, May 2020.

[42] J. Yu, X. Zheng, and J. Liu, "Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map," *Comput. Ind.*, vol. 109, pp. 121–133, Aug. 2019.

[43] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2022.

[44] N. Zhang, S. Ding, J. Zhang, and Y. Xue, "An overview on restricted Boltzmann machines," *Neurocomputing*, vol. 275, pp. 1186–1199, Jan. 2018.

[45] R. A. Khalil, N. Saeed, M. Masood, Y. M. Fard, M. Alouini, and T. Y. Al-Naffouri, "Deep learning in the industrial Internet of Things: Potentials, challenges, and emerging applications," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11016–11040, Jul. 2021.

[46] X. Yuan, Y. Gu, and Y. Wang, "Supervised deep belief network for quality prediction in industrial processes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[47] V. Pareek, S. Chaudhury, and S. Singh, "Hybrid 3DCNN-RBM network for gas mixture concentration estimation with sensor array," *IEEE Sensors J.*, vol. 21, no. 21, pp. 24263–24273, Nov. 2021.

[48] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[49] H. Zhang et al., "ResNeSt: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 2736–2746.

[50] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[51] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[52] F. Zhou, Q. Yang, T. Zhong, D. Chen, and N. Zhang, "Variational graph neural networks for road traffic prediction in intelligent transportation systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2802–2812, Apr. 2021.

[53] X. Li and W. Zhang, "Deep learning-based partial domain adaptation method on intelligent machinery fault diagnostics," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4351–4361, May 2021.

[54] Q. Wang, G. Michau, and O. Fink, "Missing-class-robust domain adaptation by unilateral alignment," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 663–671, Jan. 2021.

[55] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5525–5534, Jul. 2019.

[56] T. Ko and H. Kim, "Fault classification in high-dimensional complex processes using semi-supervised deep convolutional generative models," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2868–2877, Apr. 2020.

[57] M. Ragab et al., "Contrastive adversarial domain adaptation for machine remaining useful life prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5239–5249, Aug. 2021.

[58] C. Huang, Z. Wu, J. Wen, Y. Xu, Q. Jiang, and Y. Wang, "Abnormal event detection using deep contrastive learning for intelligent video surveillance system," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5171–5179, Aug. 2022.

[59] Z. Chai and C. Zhao, "A fine-grained adversarial network method for cross-domain industrial fault diagnosis," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1432–1442, Jul. 2020.

[60] T. Pan, J. Chen, J. Xie, Y. Chang, and Z. Zhou, "Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples," *ISA Trans.*, vol. 101, pp. 379–389, Jun. 2020.

[61] G. Xu, M. Liu, Z. Jiang, W. Shen, and C. Huang, "Online fault diagnosis method based on transfer convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 509–520, Feb. 2020.

[62] Z. Chen, K. Gryllias, and W. Li, "Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 339–349, Jan. 2020.

[63] F. Zhuang et al., "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[64] J. Huang, A. Gretton, K. Borgwardt, B. Scholkopf, and A. Smola, "Correcting sample selection bias by unlabeled data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2006, pp. 1–8.

[65] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. 49–57, Jul. 2006.

[66] H. Liang, L. Song, J. Wang, L. Guo, X. Li, and J. Liang, "Robust unsupervised anomaly detection via multi-time scale DCGANs with forgetting mechanism for industrial multivariate time series," *Neurocomputing*, vol. 423, pp. 444–462, Jan. 2021.

[67] X. Liu et al., "Self-supervised learning: Generative or contrastive," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 857–876, Jan. 2023.

[68] Y. Ding, J. Zhuang, P. Ding, and M. Jia, "Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings," *Rel. Eng. Syst. Saf.*, vol. 218, Feb. 2022, Art. no. 108126.

[69] A. L. Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov, and H. Zhang, "Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture," *Rel. Eng. Syst. Saf.*, vol. 183, pp. 240–251, Mar. 2019.

[70] D. Zhang, Z. Liu, W. Jia, H. Liu, and J. Tan, "Contrastive decoder generator for few-shot learning in product quality prediction," *IEEE Trans. Ind. Informat.*, early access, Jul. 13, 2022, doi: 10.1109/TII.2022.3190554.

[71] A. Van Den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[72] O. Henaff, "Data-efficient image recognition with contrastive predictive coding," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4182–4192.

[73] P. Bachman, R. D. Hjelm, and W. Buchwalter, "Learning representations by maximizing mutual information across views," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[74] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[75] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[76] M. Gutmann and A. Hyvarinen, "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 297–304.

[77] N. Ye and Z. Zhu, "Bayesian adversarial learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–10.

[78] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[79] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, Apr. 2016.

[80] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," 2017, *arXiv:1706.06083*.

[81] O. Bastani, Y. Ioannou, L. Lampropoulos, D. Vytiniotis, A. Nori, and A. Criminisi, "Measuring neural net robustness with constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[82] M. Khoda, T. Imam, J. Kamruzzaman, I. Gondal, and A. Rahman, "Robust malware defense in industrial IoT applications using machine learning with selective adversarial samples," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 4415–4424, Jul. 2020.

[83] Z. Yang, P. Baraldi, and E. Zio, "A multi-branch deep neural network model for failure prognostics based on multimodal data," *J. Manuf. Syst.*, vol. 59, pp. 42–50, Apr. 2021.

[84] A. Diez-Olivan, J. Del Ser, D. Galar, and B. Sierra, "Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0," *Inf. Fusion*, vol. 50, pp. 92–111, Oct. 2019.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[85] D. Mourtzis, E. Vlachou, and N. Milas, "Industrial big data as a result of IoT adoption in manufacturing," *Proc. CIRP*, vol. 55, pp. 290–295, Jan. 2016.

[86] J. Du, M. Hu, and W. Zhang, "Missing data problem in the monitoring system: A review," *IEEE Sensors J.*, vol. 20, no. 23, pp. 13984–13998, Dec. 2020.

[87] M. Serror, S. Hack, M. Henze, M. Schuba, and K. Wehrle, "Challenges and opportunities in securing the industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 2985–2996, May 2021.

[88] A. Duan, L. Guo, H. Gao, X. Wu, and X. Dong, "Deep focus parallel convolutional neural network for imbalanced classification of machinery fault diagnostics," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8680–8689, Nov. 2020.

[89] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4396–4415, Apr. 2023.

[90] L. Guo, Y. Yu, H. Gao, T. Feng, and Y. Liu, "Online remaining useful life prediction of milling cutters based on multisource data and feature learning," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5199–5208, Aug. 2022.

[91] L. Ren, Z. Jia, T. Wang, Y. Ma, and L. Wang, "LM-CNN: A cloud-edge collaborative method for adaptive fault diagnosis with label sampling space enlarging," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9057–9067, Dec. 2022.

[92] X. Wang, A. Li, and G. Han, "A deep-learning-based fault diagnosis method of industrial bearings using multi-source information," *Appl. Sci.*, vol. 13, no. 2, p. 933, Jan. 2023.

[93] X. Zhou, N. Zhai, S. Li, and H. Shi, "Time series prediction method of industrial process with limited data based on transfer learning," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6872–6882, May 2023.

[94] K. I. Wang, X. Zhou, W. Liang, Z. Yan, and J. She, "Federated transfer learning based cross-domain prediction for smart manufacturing," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 4088–4096, Jun. 2022.

[95] H. Zhang, Z. Tang, Y. Xie, H. Yuan, Q. Chen, and W. Gui, "Siamese time series and difference networks for performance monitoring in the froth flotation process," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2539–2549, Apr. 2022.

[96] T. Mezair, Y. Djenouri, A. Belhadi, G. Srivastava, and J. C.-W. Lin, "A sustainable deep learning framework for fault detection in 6G Industry 4.0 heterogeneous data environments," *Comput. Commun.*, vol. 187, pp. 164–171, Apr. 2022.

[97] S. Shao, R. Yan, Y. Lu, P. Wang, and R. X. Gao, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2658–2669, Jun. 2020.

[98] S. Deldari, H. Xue, A. Saeed, D. V. Smith, and F. D. Salim, "COCOA: Cross modality contrastive learning for sensor data," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 3, pp. 1–28, Sep. 2022.

[99] Y. Yao, G. Gui, S. Yang, and S. Zhang, "A recursive denoising learning for gear fault diagnosis based on acoustic signal in real industrial noise condition," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2021.

[100] F. Yan, C. Yang, and X. Zhang, "DSTED: A denoising spatial–temporal encoder–decoder framework for multistep prediction of burn-through point in sintering process," *IEEE Trans. Ind. Electron.*, vol. 69, no. 10, pp. 10735–10744, Oct. 2022.

[101] Y. Xu, X. Yan, B. Sun, J. Zhai, and Z. Liu, "Multireceptive field denoising residual convolutional networks for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 69, no. 11, pp. 11686–11696, Nov. 2022.

[102] S. Chang, C. Zhao, and K. Li, "Consistent-contrastive network with temporality-awareness for robust-to-anomaly industrial soft sensor," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[103] Z. Yao and C. Zhao, "FIGAN: A missing industrial data imputation method customized for soft sensor application," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 3712–3722, Oct. 2022.

[104] Z. Pan, Y. Wang, K. Wang, H. Chen, C. Yang, and W. Gui, "Imputation of missing values in time series using an adaptive-learned median-filled deep autoencoder," *IEEE Trans. Cybern.*, vol. 53, no. 2, pp. 695–706, Feb. 2023.

[105] S. Gao, S. Qiu, Z. Ma, R. Tian, and Y. Liu, "SVAE-WGAN-based soft sensor data supplement method for process industry," *IEEE Sensors J.*, vol. 22, no. 1, pp. 601–610, Jan. 2022.

[106] S. Kang, "Product failure prediction with missing data using graph neural networks," *Neural Comput. Appl.*, vol. 33, no. 12, pp. 7225–7234, Jun. 2021.

[107] M. Kavianpour, A. Ramezani, and M. T. H. Beheshti, "A class alignment method based on graph convolution neural network for bearing fault diagnosis in presence of missing data and changing working conditions," *Measurement*, vol. 199, Aug. 2022, Art. no. 111536.

[108] J. Wang, K. Wang, Y. Wang, Z. Huang, and R. Xue, "Deep Boltzmann machine based condition prediction for smart manufacturing," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 3, pp. 851–861, Mar. 2019.

[109] L. Yao and Z. Ge, "Cooperative deep dynamic feature extraction and variable time-delay estimation for industrial quality prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 3782–3792, Jun. 2021.

[110] K. Liu, Y. Shang, Q. Ouyang, and W. D. Widanage, "A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3170–3180, Apr. 2021.

[111] L. Ren, X. Cheng, X. Wang, J. Cui, and L. Zhang, "Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction," *Future Gener. Comput. Syst.*, vol. 94, pp. 601–609, May 2019.

[112] W. Zhang, D. Yang, Y. Xu, X. Huang, J. Zhang, and M. Gidlund, "DeepHealth: A self-attention based method for instant intelligent predictive maintenance in industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5461–5473, Aug. 2021.

[113] C. Yang, K. Zhou, and J. Liu, "SuperGraph: Spatial–temporal graph-based feature extraction for rotating machinery diagnosis," *IEEE Trans. Ind. Electron.*, vol. 69, no. 4, pp. 4167–4176, Apr. 2022.

[114] L. Ren, Y. Liu, D. Huang, K. Huang, and C. Yang, "MCTAN: A novel multichannel temporal attention-based network for industrial health indicator prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, early acces, Jan. 10, 2022, doi: 10.1109/TNNLS.2021.3136768.

[115] K. Jang, S. Hong, M. Kim, J. Na, and I. Moon, "Adversarial autoencoder based feature learning for fault detection in industrial processes," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 827–834, Feb. 2022.

[116] Z. Geng, Z. Chen, Q. Meng, and Y. Han, "Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1521–1529, Mar. 2022.

[117] H. Chen, R. Liu, Z. Xie, Q. Hu, J. Dai, and J. Zhai, "Majorities help minorities: Hierarchical structure guided transfer learning for few-shot fault recognition," *Pattern Recognit.*, vol. 123, Mar. 2022, Art. no. 108383.

[118] X. Cao, Y. Wang, B. Chen, and N. Zeng, "Domain-adaptive intelligence for fault diagnosis based on deep transfer learning from scientific test rigs to industrial applications," *Neural Comput. Appl.*, vol. 33, no. 9, pp. 4483–4499, May 2021.

[119] X. Zhou, Y. Hu, J. Wu, W. Liang, J. Ma, and Q. Jin, "Distribution bias aware collaborative generative adversarial network for imbalanced deep learning in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 570–580, Jan. 2023.

[120] G. Yang, Y. Zhong, L. Yang, H. Tao, J. Li, and R. Du, "Fault diagnosis of harmonic drive with imbalanced data using generative adversarial network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[121] J. Kuang, G. Xu, T. Tao, and Q. Wu, "Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[122] W. Mao, J. He, and M. J. Zuo, "Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1594–1608, Apr. 2020.

[123] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, "Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8702–8712, Nov. 2020.

[124] F. Kong, J. Li, B. Jiang, H. Wang, and H. Song, "Integrated generative model for industrial anomaly detection via bidirectional LSTM and attention mechanism," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 541–550, Jan. 2023.

[125] M. Abdel-Basset, N. Moustafa, and H. Hawash, "Privacy-preserved generative network for trustworthy anomaly detection in smart grids: A federated semisupervised approach," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 995–1005, Jan. 2023.

[126] C. Zhao and W. Shen, "Adversarial mutual information-guided single domain generalization network for intelligent fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2909–2918, Mar. 2023.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[127] P. Peng, J. Lu, T. Xie, S. Tao, H. Wang, and H. Zhang, "Open-set fault diagnosis via supervised contrastive learning with negative out-of-distribution data augmentation," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2463–2473, Mar. 2023.

[128] L. Ren, T. Mo, and X. Cheng, "Meta-learning based domain generalization framework for fault diagnosis with gradient aligning and semantic matching," *IEEE Trans. Ind. Informat.*, early access, Apr. 3, 2023, doi: 10.1109/TII.2023.3264111.

[129] S. Sun, T. Wang, H. Yang, and F. Chu, "An environmentally adaptive and contrastive representation learning method for condition monitoring of industrial assets," *IEEE Trans. Cybern.*, early access, Oct. 11, 2022, doi: 10.1109/TCYB.2022.3209707.

[130] J. Li, C. Shen, L. Kong, D. Wang, M. Xia, and Z. Zhu, "A new adversarial domain generalization network based on class boundary feature detection for bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.

[131] B. Yang et al., "A joint energy and latency framework for transfer learning over 5G industrial edge networks," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 531–541, Jan. 2022.

[132] J. Wang, P. Zheng, and J. Zhang, "Big data analytics for cycle time related feature selection in the semiconductor wafer fabrication system," *Comput. Ind. Eng.*, vol. 143, May 2020, Art. no. 106362.

[133] G. S. Chadha, M. Krishnamoorthy, and A. Schwung, "Time series based fault detection in industrial processes using convolutional neural networks," in *Proc. 45th Annu. Conf. IEEE Ind. Electron. Soc.*, vol. 1, Oct. 2019, pp. 173–178.

[134] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Eng. Appl. Artif. Intell.*, vol. 92, Jun. 2020, Art. no. 103678.

[135] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021.

[136] T. Zonta, C. A. Da Costa, R. D. R. Righi, M. J. De Lima, E. S. Da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Comput. Ind. Eng.*, vol. 150, Dec. 2020, Art. no. 106889.

[137] T. P. Carvalho, F. A. Soares, R. Vita, R. D. P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Comput. Ind. Eng.*, vol. 137, Nov. 2019, Art. no. 106024.

[138] L. Ren, Y. Liu, X. Wang, J. Lu, and M. J. Deen, "Cloud-edge-based lightweight temporal convolutional networks for remaining useful life prediction in IIoT," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12578–12587, Aug. 2021.

[139] J. Chen, H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Rel. Eng. Syst. Saf.*, vol. 185, pp. 372–382, May 2019.

[140] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3478–3487, May 2021.

[141] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2019.

[142] L. Wen, X. Li, and L. Gao, "A new reinforcement learning based learning rate scheduler for convolutional neural network in fault classification," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12890–12900, Dec. 2021.

[143] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1310–1320, Jun. 2017.

[144] Y.-N. Sun, W. Qin, H.-W. Xu, R.-Z. Tan, Z.-L. Zhang, and W.-T. Shi, "A multiphase information fusion strategy for data-driven quality prediction of industrial batch processes," *Inf. Sci.*, vol. 608, pp. 81–95, Aug. 2022.

[145] Y. Liu, C. Yang, K. Liu, B. Chen, and Y. Yao, "Domain adaptation transfer learning soft sensor for product quality prediction," *Chemometric Intell. Lab. Syst.*, vol. 192, Sep. 2019, Art. no. 103813.

[146] J. Zhang, Y. Jiang, H. Luo, and S. Yin, "Prediction of material removal rate in chemical mechanical polishing via residual convolutional neural network," *Control Eng. Pract.*, vol. 107, Feb. 2021, Art. no. 104673.

[147] S. Bag, L. C. Wood, L. Xu, P. Dhamija, and Y. Kayikci, "Big data analytics as an operational excellence approach to enhance sustainable supply chain performance," *Resour., Conservation Recycling*, vol. 153, Feb. 2020, Art. no. 104559.

[148] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, May 2020, Art. no. 106435.

[149] G. Baryannis, S. Dani, and G. Antoniou, "Predicting supply chain risks using machine learning: The trade-off between performance and interpretability," *Future Gener. Comput. Syst.*, vol. 101, pp. 993–1004, Dec. 2019.

[150] C. Lin, D. Deng, Y. Chih, and H. Chiu, "Smart manufacturing scheduling with edge computing using multiclass deep Q network," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4276–4284, Jul. 2019.

[151] D. A. Rossit, F. Tohmé, and M. Frutos, "Production planning and scheduling in cyber-physical production systems: A review," *Int. J. Comput. Integr. Manuf.*, vol. 32, nos. 4–5, pp. 385–395, May 2019.

[152] D. Ivanov, A. Dolgui, and B. Sokolov, "The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics," *Int. J. Prod. Res.*, vol. 57, no. 3, pp. 829–846, Feb. 2019.

[153] C. Morariu, O. Morariu, S. Raileanu, and T. Borangiu, "Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems," *Comput. Ind.*, vol. 120, Sep. 2020, Art. no. 103244.

[154] F. R. Lima-Junior and L. C. R. Carpinetti, "Predicting supply chain performance based on SCOR metrics and multilayer perceptron neural networks," *Int. J. Prod. Econ.*, vol. 212, pp. 19–38, Jun. 2019.

[155] D. Shi, W. Fan, Y. Xiao, T. Lin, and C. Xing, "Intelligent scheduling of discrete automated production line via deep reinforcement learning," *Int. J. Prod. Res.*, vol. 58, no. 11, pp. 3362–3380, Jun. 2020.

[156] Z. Tong, F. Ye, B. Liu, J. Cai, and J. Mei, "DDQN-TS: A novel bi-objective intelligent scheduling algorithm in the cloud environment," *Neurocomputing*, vol. 455, pp. 419–430, Sep. 2021.

[157] D. Syed, A. Zainab, A. Ghrayeb, S. S. Refaat, H. Abu-Rub, and O. Bouhali, "Smart grid big data analytics: Survey of technologies, techniques, and applications," *IEEE Access*, vol. 9, pp. 59564–59585, 2021.

[158] B. Dhupia, M. U. Rani, and A. Alameen, "The role of big data analytics in smart grid management," in *Emerging Research in Data Engineering Systems and Computer Communications*. Singapore: Springer, 2020, pp. 403–412.

[159] Z. Alavikia and M. Shabro, "A comprehensive layered approach for implementing Internet of Things-enabled smart grid: A survey," *Digit. Commun. Netw.*, vol. 8, no. 3, pp. 388–410, Jun. 2022.

[160] W. Zheng and G. Chen, "An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13902–13914, Dec. 2022.

[161] Y. Jia, X. Lyu, P. Xie, Z. Xu, and M. Chen, "A novel retrospect-inspired regime for microgrid real-time energy scheduling with heterogeneous sources," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 4614–4625, Nov. 2020.

[162] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[163] D. Wang, G. Zhao, H. Chen, Z. Liu, L. Deng, and G. Li, "Nonlinear tensor train format for deep neural network compression," *Neural Netw.*, vol. 144, pp. 320–333, Dec. 2021.

[164] Y. Gil, J.-H. Park, J. Baek, and S. Han, "Quantization-aware pruning criterion for industrial applications," *IEEE Trans. Ind. Electron.*, vol. 69, no. 3, pp. 3203–3213, Mar. 2022.

[165] Q. Xu, Z. Chen, K. Wu, C. Wang, M. Wu, and X. Li, "KDNet-RUL: A knowledge distillation framework to compress deep neural networks for machine remaining useful life prediction," *IEEE Trans. Ind. Electron.*, vol. 69, no. 2, pp. 2022–2032, Feb. 2022.

[166] W. Fang, F. Xue, Y. Ding, N. Xiong, and V. C. M. Leung, "EdgeKE: An on-demand deep learning IoT system for cognitive big data on industrial edge devices," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6144–6152, Sep. 2021.

[167] C. Qiu, X. Wang, H. Yao, J. Du, F. R. Yu, and S. Guo, "Networking integrated cloud-edge-end in IoT: A blockchain-assisted collective Q-learning approach," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12694–12704, Aug. 2021.

[168] A. H. Sodhro, S. Pirbhulal, and V. H. C. de Albuquerque, "Artificial intelligence-driven mechanism for edge computing-based industrial applications," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4235–4243, Jul. 2019.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[169] Y. Laili, F. Guo, L. Ren, X. Li, Y. Li, and L. Zhang, "Parallel scheduling of large-scale tasks for industrial cloud-edge collaboration," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 3231–3242, Feb. 2023.

[170] A. M. Ghosh and K. Grolinger, "Edge-cloud computing for Internet of Things data analytics: Embedding intelligence in the edge with deep learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2191–2200, Mar. 2021.

[171] Y. Wu, "Cloud-edge orchestration for the Internet of Things: Architecture and AI-powered data processing," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12792–12805, Aug. 2021.

[172] A. Ghosh, A. Mukherjee, and S. Misra, "SEGA: Secured edge gateway microservices architecture for IIoT-based machine monitoring," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1949–1956, Mar. 2022.

[173] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, "An empirical investigation of catastrophic forgetting in gradient-based neural networks," 2013, *arXiv:1312.6211*.

[174] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.

[175] M. De Lange et al., "A continual learning survey: Defying forgetting in classification tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3366–3385, Jul. 2022.

[176] S. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "ICaRL: Incremental classifier and representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5533–5542.

[177] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[178] L. Floridi, "Establishing the rules for building trustworthy AI," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 261–262, May 2019.

[179] S. Mohseni, N. Zarei, and E. D. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable AI systems," *ACM Trans. Interact. Intell. Syst.*, vol. 11, nos. 3–4, pp. 1–45, Dec. 2021.

[180] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.

[181] R. Bommasani et al., "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.

[182] S. Yuan et al., "A roadmap for big model," 2022, *arXiv:2203.14101*.

[183] T. B. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.

[184] W. Fedus, B. Zoph, and N. Shazeer, "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," *J. Mach. Learn. Res.*, vol. 23, no. 1, pp. 5232–5270, 2021.

[185] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, Aug. 2021.

[186] J. W. Song, Y. I. Park, J.-J. Hong, S.-G. Kim, and S.-J. Kang, "Attention-based bidirectional LSTM-CNN model for remaining useful life estimation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Daegu, South Korea, 2021, pp. 1–5, doi: 10.1109/ISCAS51556.2021.9401572.

Lei Ren (Member, IEEE) received the Ph.D. degree in computer science from the Institute of Software, Chinese Academy of Sciences, Beijing, China, in 2009.

He is currently a Professor with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, the Zhongguancun Laboratory, Beijing, and the State Key Laboratory of Intelligent Manufacturing System Technology, Beijing. His research interests include neural networks and deep learning, time-series analysis, and industrial artificial intelligence applications.

Dr. Ren serves as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and other international journals.

Zidi Jia (Graduate Student Member, IEEE) received the bachelor's and master's degrees from the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree in pattern recognition and intelligent systems with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China.

His research interests include deep learning, industrial artificial intelligence, the Industrial Internet of Things, and industrial process monitoring.

Yuanjun Laili (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, in 2009, 2012, and 2015, respectively.

She is currently an Associate Professor with the School of Automation Science and Electrical Engineering, Beihang University, the Zhongguancun Laboratory, Beijing, and the State Key Laboratory of Intelligent Manufacturing System Technology, Beijing. She is the Chief Scientist with the National Key Research and Development Program of China. Her research interests include the area of intelligent optimization, modeling, and simulation of manufacturing systems.

Dr. Laili is a member of the Society for Modeling and Simulation International (SCS). She has received the Young Talent Lift Project supported by the China Association for Science and Technology and the Young Simulation Scientist Award from SCS. She serves as an Associate Editor for the *International Journal of Modeling, Simulation, and Scientific Computing* and *Cogent Engineering*.

Di Huang (Senior Member, IEEE) received the B.S. and M.S. degrees in computer science from Beihang University, Beijing, China, in 2005 and 2008, respectively, and the Ph.D. degree in computer science from the École Centrale de Lyon, Lyon, France, in 2011.

He joined as a Faculty Member with the Laboratory of Intelligent Recognition and Image Processing, School of Computer Science and Engineering, Beihang University, where he is currently a Professor. His research interests include biometrics, 2-D/3-D face analysis, image/video processing, and pattern recognition.