Accurate and Efficient Digital Twin Construction Using Concurrent End-to-End Synchronization and Multi-Attribute Data Resampling

Pengyi Jia, *Member, IEEE*, Xianbin Wang, *Fellow, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—Accurate and efficient digital twin construction through real-time multi-attribute sensing and remote concurrent data analysis is essential in supporting complex connected industrial applications. Given the unsynchronized nature and heterogeneous sampling rates of distributed sensing processes, the varying time misalignment among different attributes will inevitably deteriorate the remote correlation analysis and digital twin construction. Furthermore, application-agnostic digital twin construction approaches could potentially involve high communication and computation overhead for comprehensive digital twin construction. In this article, a concurrent end-to-end time synchronization and multi-attribute data resampling scheme is proposed to enable accurate and efficient digital twin construction at the remote end. Specifically, digital clocks are concurrently established at the remote end, with each of them associated with a sampling rate of a unique sensing attribute. To tackle the temporal misalignment among multiple sensing attributes, raw data are accurately resampled according to the same reference frequency, with attribute-specific synchronized digital clocks providing cohesively aligned time information. An edge-centric platform is established to efficiently guide the multidimensional data processing during digital twin construction. Simulation results demonstrate that the proposed scheme can achieve more accurate and efficient digital twin construction than existing modeling methods. In the end, the digital twin-driven predictive maintenance is presented as a case study, aiming at illustrating the potential applications and benefits expected of the proposed scheme in industrial environments.**

***Index Terms*—Data resampling, digital twin, edge computing, Industrial Internet of Things (IIoT), predictive maintenance, system identification, time synchronization.**

I. Introduction

THE PAST few years have witnessed the unprecedented evolution and massive deployment of information and communication technologies (ICTs), particularly 5G, machine learning, and artificial intelligence [1], [2]. Thanks to these expeditious technological advancements, digital transformation of vertical industries (e.g., manufacturing [3] and production [4]) now becomes inevitable, leading to various emerging paradigms, like Industry 4.0 [5], digital twin [6], and Industrial Internet of Things (IIoT).

Manuscript received 29 August 2022; accepted 26 October 2022. Date of publication 9 November 2022; date of current version 7 March 2023. This work was supported in part by the NSERC Discovery Program under Grant RGPIN2018-06254; in part by the NSERC Idea to Innovation Program under Grant I2IPJ 538563-19; and in part by the Canada Research Chair Program. *(Corresponding author: Xianbin Wang.)*

Pengyi Jia and Xianbin Wang are with the Department of Electrical and Computer Engineering, Western University, London, ON N6A 5B9, Canada (e-mail: pjia7@uwo.ca; xianbin.wang@uwo.ca).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/JIOT.2022.3221012

At the core of Industry 4.0, the digital transformation of industrial entities and their associated applications relies on the effective observation, understanding, and controlling of physical industry processes in the digital domain. For this purpose, a transformative data-driven concept, referred to as digital twin, is envisioned to bridge the gap between the physical world and the digital domain. As the digitally mirrored images of physical objects and processes, digital twins can comprehensively represent the real-time conditions of their physical counterparts in the digital domain and thereby efficiently enable situation-aware critical decisions [7], [8]. To closely cohere the physical and digital domains, digital twin hinges on multifaceted observation, node-level analysis, extensive simulations, holistic coordination, and long-term optimization of the entire industrial system throughout its life cycle. As a result, the approach of digital twin construction dominates its effectiveness of further implementation [9].

Unfortunately, establishing the digital model of a physical entity/process and further utilizing it for industry process management and eventual value realization can be highly complex due to the following two main challenges.

On the one hand, constructing an effective digital twin model relies exclusively on the coherent analysis and fast interpretation of multiple correlated attributes associated with one physical object/process [10], [11]. Maintaining the precise temporal correlation among multidimensional data is of the utmost importance for correlation discovery techniques, such as the autoregressive (AR) model, least absolute shrinkage and selection operator (Lasso), and support vector machine (SVM). However, the multidimensional data collected over the Internet will be severely misaligned in the temporal domain due to local hardware limitations and network-induced varying latency. Initially, the data of different attributes are sampled neither simultaneously nor synchronously because of their diverse sampling rates, dynamic task schedules, mutual constraint hardware accessibility, and unsynchronized local clocks [12], [13]. The temporal correlation among the chaotic multidimensional data will be increasingly challenging to discover with the increment of the modeling dimension. Furthermore, during data transmission over the Internet, the best-effort delivery strategy will lead to nondeterministic issues, including uncontrollable propagation/access/queuing delays and stochastic packet losses. This will further aggravate the temporal inconsistency among multidimensional data and make the received data totally disordered at the remote end [14].

On the other hand, the low efficiency of the digital twin construction process caused by redundant data acquisition and application-agnostic attributes exploration will severely reduce the productivity of the overall industrial applications. Existing digital twin construction concepts, e.g., [15] and [16], tend to achieve detailed interpretation of the multidimensional data about each and every attribute of the physical entity while ignoring the specific application requirements. However, lacking application-centric feature mirroring will be fatal to supporting the scope of physical entities [17]. Application-agnosticism would also create an overly fine-grained digital twin at the expense of unnecessarily high communication and computational overhead. Moreover, this burden will continuously grow during the lifelong model maintenance and exponentially accumulate with the expansion of network scale [18], which may eventually hinder the normal operation of the original industrial system. Consequently, designing novel digital twin construction methodologies that can efficiently reflect the application-specific features of the physical entities becomes a must.

In most of the existing literature, digital twins are utilized as applications, with only a few studies paying attention to addressing the modeling challenges. For example, the missing LiDAR data problems during digital twin application are explored in [19], where a series of machine learning techniques are adopted to substitute the missing information. Similarly, Hu et al. [20] utilized the locality-sensitive hashing algorithm to achieve digital-twin-assisted missing traffic flow and traffic velocity data prediction. Moreover, to reduce the necessity of frequent interactions between the physical domain and virtual representatives, an integrated digital twin clone flow and smart digital twinning boarding is designed in [21], which can intelligently adapt to the development environment. Meanwhile, a decomposition digital twinning method is proposed in [22], which can reduce the digital twin construction complexity by dividing the entire process into basic components characterization and the overall digital twin integration. However, the critical issues in terms of temporal misalignment among concurrent multiple sensing attributes and application-agnostic data collection still remain unsolved, deteriorating the performance of the overall digital twin system.

Motivated by these considerations, a Concurrent end-to-end Synchronization and multi-attribute Data Resampling enabled Digital Twin (CSDR-DT) construction scheme is proposed in this article to establish digital twins for IIoT devices with enhanced accuracy and efficiency. The main contributions are summarized as follows.

1) To tackle the temporal misalignment of the multidimensional data associated with the same IIoT device, end-to-end time synchronization and data resampling are achieved for each attribute of the physical entity before digital twin creation. A virtual clock model is established at the remote processing center for each sensing attribute to predict and compensate for its local clock offsets. Moreover, the synchronized data will be further resampled according to the same reference frequency, so that the multidimensional information can be completely aligned in the temporal domain. Therefore, by concurrently processing the data of each attribute, a multidimensional digital twin with enhanced accuracy can be established.
2) A feedback-based sampling rate adjustment and attribute selection mechanism is designed aiming at improving the digital twinning efficiency. By comparing the data resampling accuracy and the application-specific requirement, an optimal local sampling rate can be determined for each attribute. Meanwhile, a Lasso-based penalized regression method is adopted during digital twin creation, which is effective in filtering unnecessary attributes during temporal correlation discovery. Local sensing components can adjust their sampling tasks accordingly to avoid unnecessary resource consumption.
3) An edge-centric digital twin construction platform is established, consisting of time synchronization, premodel data resampling, and multidimensional digital twin creation. The massive data transmitted from distributed local IIoT devices can be efficiently processed by utilizing the enhanced computational capability of edge devices. The impact of network uncertainties can also be alleviated by avoiding extremely distant communication between local devices and the remote processing center.

The remainder of this article is organized as follows. The overall system model in terms of the digital twin platform-enabled IIoT system and data transmission schedule is illuminated in Section II. Section III introduced the design of the CSDR-DT scheme in detail, including clock offset compensation, data resampling, feedback-based sampling adjustment, and digital twin creation. Simulations are carried out to demonstrate the effectiveness of the proposed method from the aspects of modeling accuracy, efficiency, and robustness in Section IV, with a case study of industrial predictive maintenance in the end. Section V concludes this article.

Notations: In this article, scalars are denoted by italic letters, while vectors are denoted by letters in bold. The main symbols used are summarized in Table I.

II. Digital Twin Enabled IIoT Systems

In this section, the proposed digital twin platform for a large-scale IIoT system is demonstrated from two aspects. First, the overall architecture of the IIoT system with the involvement of digital twin platform is illustrated, focusing on the interaction between the physical IIoT devices and the remote end during digital twin construction. Moreover, we introduce the schedule of data transmission among distributed IIoT devices and the remote end in the proposed design, including timestamp exchange and data sample sharing.

A. Overall Architecture of Digital Twin Platform

The overall architecture of a large-scale IIoT system enabled by the proposed digital twin platform is shown in Fig. 1, which

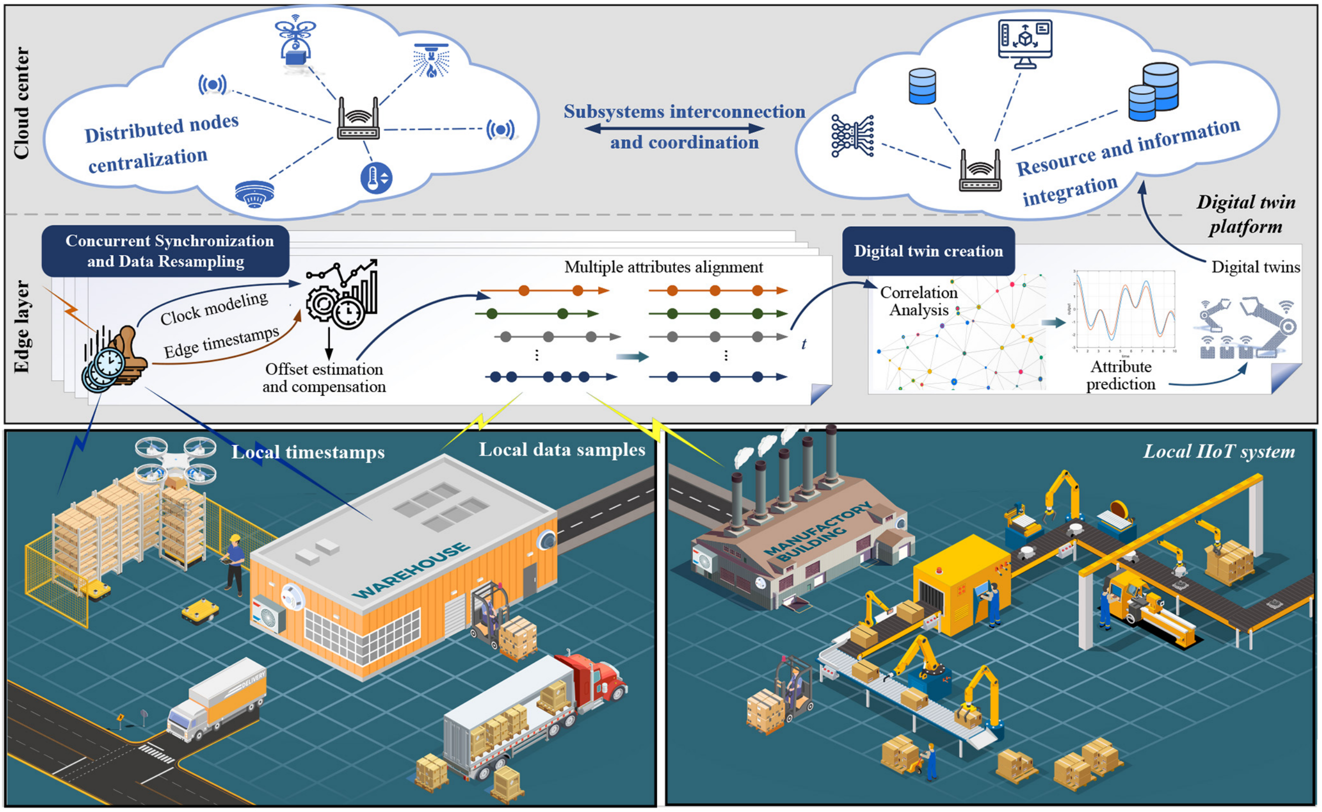

Fig. 1. Overall architecture of the digital twin enabled IIoT system, including the local IIoT system and the digital twin platform, which comprise the edge-centric data processing and cloud-based service provisioning. The construction of the digital twin platform consists of three main successive steps, namely, concurrent synchronization, multi-attribute data resampling, and digital twin creation.

TABLE I
MAJOR NOTATIONS AND DEFINITIONS

Notation	Definition
i	Sensing attribute
$\mathcal{I}$	Total number of attributes
r	Resampling instant
$\mathcal{R}$	Total number of resamples for one attribute
d	Symmetric propagation delay
α, β	Initial clock skew and offset
$\hat{\alpha}$	Relative skew between two entities
t	Absolute time
$\hat{C}$	Relative clock model
t^s, t^r	Original sampling instant and resampling instant
ϵ	Local clock error
τ	Attribute uploading frequency
s	Local sample for one attribute
$\mathcal{S}$	Total samples for one attribute
K	Optimal number of neighbors
δ^{rs}	Time difference between two instants
$\mathcal{D}^{rs}$	Euclidean distance between two instants
ω_ρ	Weighting factor in ADWKNN
f, $\hat{f}$	Local sampling rate and optimal sampling rate
Q	Ideal resampling accuracy
η	Digital twin parameters
$\mathcal{DL}$	Deterioration level

consists of two interconnected layers, namely, edge and cloud. On the one hand, critical information of the local IIoT system will be uploaded to the cloud-edge collaborative digital twin platform for edge-centric data processing. On the other hand, the correspondingly established digital twins will be utilized to provide instructive services in the remote cloud center to support various industrial applications.

In this article, we consider the application scenario of an IIoT system, which comprises heterogeneous and distributed IIoT devices ranging from sensors (e.g., for environmental monitoring and machine condition observing) to unmanned vehicles (e.g., for material delivery and photo capture). The cooperation and interconnection of the IIoT devices will enable various advanced industrial applications like predictive maintenance and networked control, which require coherent data processing with the support of digital twins. During digital twin construction, the multiple attributes and sensing timestamps relevant to the target IIoT device will be continuously uploaded to the edge device for edge-centric data processing, by following the predefined transmission schedule. After accurate digital twin creation, local actuators will execute commands upon the received information from the upper platform to support advanced industrial applications.

Edge-centric data processing is proposed to provide near-node data analysis while overcoming the limitation of local IIoT sensors in terms of lacking sufficient processing capability. The digital twin construction is organized into four successive phases, namely, virtual clock modeling, multi-attribute data resampling, feedback-based sampling adjustment, and digital twin creation. Initially, by investigating the difference between the timestamps of each sensing attribute and the ones generated at the edge device, a virtual clock model can be established for each attribute, which can be further used to predict its future offsets. Then, the time information associated with the real-time uploaded attribute samples can be accurately synchronized by compensating for the predicted offsets. The synchronized data will be further resampled according to the same sampling rate to achieve completely temporal alignment. Finally, the resampled multi-attribute data will be modeled into a digital twin for each device by adopting proper statistical analysis techniques. In addition, edge devices will generate feedback information based on the resampling and modeling performance after each training iteration. By adjusting the required attributes and corresponding sampling rates accordingly, the communication and computation overhead of local IIoT devices can be significantly reduced.

The cloud center will be responsible for gathering the established digital twins of all IIoT devices and forming potential subsystems in the digital domain by investigating their interconnections. Various services, including resource sharing, information integration, predictive decision-making, and inter-subsystem coordination, can be conveniently provisioned to guide the further operation of physical entities. Meanwhile, a long-term analysis of the digital twins established will be conducted in the cloud center, which is necessary for adapting to potential system variations and network dynamics.

B. Data Transmission Schedule

For a group of IIoT entities, an edge device will be assigned according to a series of criteria, including its physical location, processing capability, and ability to maintain a consistent clock reference. The latter one is mainly achieved by equipping temperature compensated crystal oscillator (TCXO) to drive the clock. For accurate digital twin construction, the edge device will be first responsible for establishing a virtual clock model for each sensing attribute uploaded from the IIoT devices according to the associated timestamps, and then create a digital twin model for the IIoT device by analyzing the correlation of the multiple sensing attributes. Due to the involvement of a large number of IIoT devices and the various sensing attributes, the local timestamps and sensing samples should be transmitted to the edge devices by following the predefined schedule, which can help alleviate unnecessary contention and resource wasting.

The data transmission schedule is shown in Fig. 2, where the local time for each sensing attribute will deviate from the edge device due to the unsynchronized clocks. To eliminate the impact of clock drift while minimizing the data exchange during digital twin training, timestamps, and data samples are scheduled to transmit separately. On the one hand, each IIoT sensor will periodically transmit timestamps to the edge device for intrinsic clock parameter estimation. After receiving the timestamp, an acknowledgment packet will be delivered backward, where the two pairs of timestamps can be used for the initial clock parameter estimation. On the other hand, a series of data samples for each sensing attribute will be continuously transmitted to the edge device with a predefined interval τ, where d_i indicates the random delay between two devices that affect the timeliness of local data. In this article,

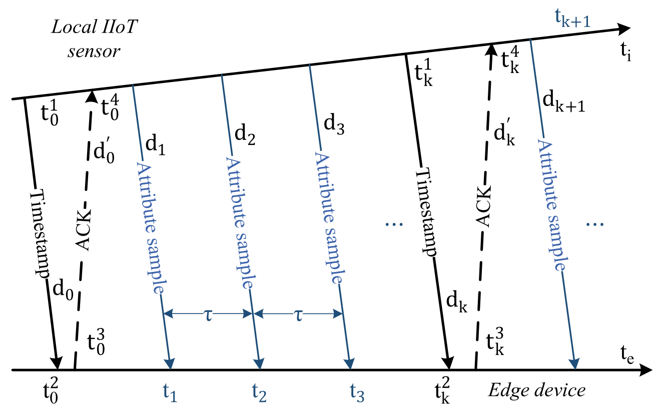

Fig. 2. Data transmission schedule between each IIoT sensor and the edge device for digital twin construction, which includes timestamp exchange and attribute sample transmission.

it is assumed that the delay between the local IIoT sensor and the edge device will follow the Gaussian distribution, i.e., $d \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ are the mean and standard deviations of the delay, respectively. Due to the influence of random network delay and stochastic packet losses, the accuracy of digital twin construction will be further deteriorated, which requires effective premodeling data processing to improve the performance.

III. Accurate Digital Twin Construction

Construction of the digital twin platform hinges on the accurate and efficient edge-centric analysis of distributed IIoT data. Details of the four successive steps of digital twin construction, including virtual clock formation, premodeling data processing, feedback-based sampling rate adjustment, and digital twin creation will be introduced in this section.

A. Virtual Clock Modeling

Multidimensional digital twin modeling relies on the cohesive processing of multiple sensing attributes of a single physical entity. The misalignment of the time information associated with different attributes will cause a significant negative impact on the modeling accuracy. As shown in Fig. 3, a two-step attribute alignment scheme is designed for premodeling data processing, including clock offset compensation and multi-attribute data resampling.

For electrical IIoT devices, an oscillator-driven clock is typically embedded to provide local time information continuously. Simply packaged crystal oscillators (SPXOs), which are widely utilized in IIoT systems to reduce implementation costs, cannot generate stable timestamps due to defective manufacturing and lack of temperature compensation techniques. The time inaccuracy of the sensing attribute i is mainly dominated by the local clock of its sensing device, which is associated with the initial clock skew α_i and clock offset β_i, while an unacceptable clock error will occur without proper time synchronization methods. Compared to the time reference t, a time-varying clock error ϵ_i will occur, written as

$$\epsilon_i(t) = \alpha_i(t)t + \beta_i - t = (\alpha_i(t) - 1)t + \beta_i. \tag{1}$$

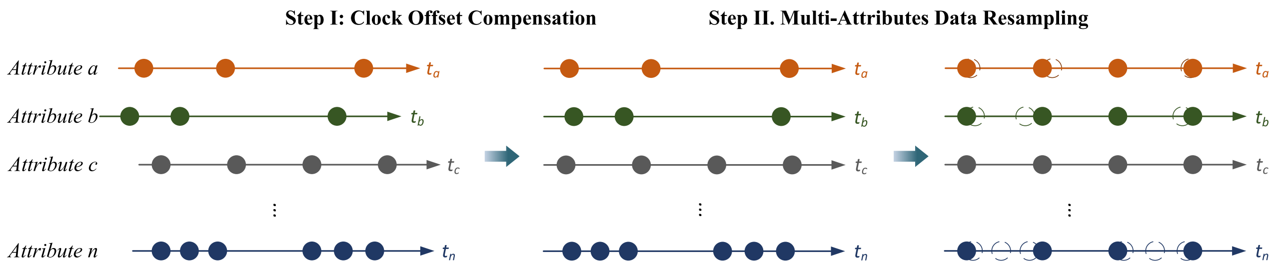

Fig. 3. Alignment of the multiple sensing attributes in the time domain enabled by clock offset compensation and multi-attribute data resampling. The temporal information will be precisely aligned after premodeling data processing.

In long-term operations, clock skew α_i of inexpensive IIoT sensors will be inconsistent with the variation of external operating conditions, e.g., ambient temperature, resulting in even more unpredictable clock inaccuracy. The lacking of temporal consistency among multiple sensing attributes caused by this clock error can severely affect the modeling accuracy of digital twins, while the traditional packet-switching-based time synchronization methods will inevitably lead to high communication overhead. Based on this observation, a model-based offset estimation scheme is designed in this section to support the digital twin construction by properly analyzing the sequential timestamps of each IIoT attribute.

As shown in Fig. 2, timestamps of each local sensor are periodically uploaded to the edge device for virtual clock modeling. The main challenge of clock modeling is to accurately estimate the clock skew and offset based on the timestamps. Similar to precision time protocol (PTP), the estimation of the initial clock offset β_i of each clock is obtained by analyzing the first two pairs of timestamps, given by

$$\hat{\beta}_i = \frac{t_0^2 - t_0^1 + t_0^3 - t_0^4}{2} + \frac{d_2 - d_1}{2} \tag{2}$$

where d_1 and d_2 are the propagation delay between the two nodes in the two successive links. By assuming the delay is symmetric, the estimated offset can be further simplified into

$$\hat{\beta}_i = \frac{t_0^2 - t_0^1 + t_0^3 - t_0^4}{2}. \tag{3}$$

Different from clock offset estimation, timestamps from the local IIoT devices are not required during the clock skew estimation, which can reduce the network overhead of clock modeling by half. As discussed in Section II-B, a series of samples for attribute i will be uploaded to the edge device for digital twin construction with a predefined interval τ_i. After obtaining each attribute sample, the edge device will record the receiving time, while ideally, the intervening period between two successive receiving instants should be identical to the predefined interval, i.e., $t_2 - t_1 = \tau_i$. However, there will be a difference between the two intervals due to the existence of the clock skew, which can be thereby estimated by

$$\hat{\alpha}_{ei}^1 = \frac{t_2 - t_1}{\tau} \tag{4}$$

where $\hat{\alpha}_{ei}^1$ denotes the first estimation of the relative skew between the edge device and the sensing attribute i. After receiving a series of samples, the estimation of the clock skew can be improved by taking the average of the historical values, given by

$$\hat{\alpha}_{ei} = \overline{\sum_{s=2}^{S_i} \frac{t_s - t_{s-1}}{\tau}} = \frac{t_{S_i} - t_1}{\tau_i(S_i - 1)} \tag{5}$$

where S_i is the overall samples collected for attribute i.

Therefore, the relative clock model of the attribute i can be written as

$$\hat{C}_i(t) = \hat{\alpha}_{ei} t + \hat{\beta}_i = \frac{t_{S_i} - t_1}{\tau_i(S_i - 1)} t + \frac{t_0^2 - t_0^1 + t_0^3 - t_0^4}{2} \tag{6}$$

which can be straightforwardly used to predict the relative clock offset of the attribute i compared to its edge device. Another two pairs of timestamps will be exchanged between the local sensor and the edge device for validation.

B. Data Synchronization and Resampling

After establishing the virtual clock model for each sensing attribute, the series of samples collected from each IIoT sensor can be compensated for according to the real-time clock errors obtained from the estimated clock parameters, given by

$$\hat{\epsilon}(t) = \left(\hat{\alpha}_{ei} - 1\right)t + \hat{\beta}_i. \tag{7}$$

The sample of attribute i at a given instant t, i.e., $s_i(t)$, can be thereby corrected as

$$\hat{s}_i(t) = s_i\left(t - \hat{\epsilon}(t)\right) \tag{8}$$

where the temporal misalignment caused by clock errors can be eliminated if accurate clock modeling is achieved.

However, due to the limitation of IIoT devices in terms of their heterogeneous sampling rates and dynamic task schedules, the sampling instants for each attribute about the same device will be nonsimultaneously, which can lead to temporal misalignment among the multi-attribute samples. This incoherence will inevitably cause modeling errors if the raw data are used for digital twin training without proper processing. To address this issue, an adaptive distance-weighted K-nearest neighbors (ADWKNNs) algorithm is proposed for multi-attribute data resampling, as illustrated in Algorithm 1.

The main difference between the proposed ADWKNN algorithm and the traditional distance-weighted K-nearest neighbors (DWKNNs) [23] lies in the adaptive selection of the K nearest neighbors according to the multi-attribute samples. More specifically, ADWKNN comprises two main steps, namely, optimal nearest neighbors selection and adaptive

Algorithm 1: Find Optimal K Using Adaptive Distance-Weighted KNN Data Resampling for Each Resample

Input: Synchronized samples $\hat{s}_{i,j}$, sampling instants t_s, resampling instants t_r
Output: Resampled data $\check{s}_i^j$

for $j = 1 : \mathcal{J}$ **do**
  for $s = 1 : \mathcal{S}$ **do**
    Calculate time difference $\boldsymbol{\delta}^{rs}_{\boldsymbol{m}}$ based on Eq. (9)
    Obtain Euclidean distance $\boldsymbol{\mathcal{D}}^{rs}_{\boldsymbol{m}}$ according to Eq. (10)
  Obtain median of Euclidean distance $\tilde{\mathcal{D}}^{rs}$
  for $s = 1 : \mathcal{S}$ **do**
    if $\mathcal{D}^{rs}_m \geq \tilde{\mathcal{D}}^{rs}$ **then**
      Remove $\mathcal{D}^{rs}_m$ from $\boldsymbol{\mathcal{D}}^{rs}$
  Sort $\boldsymbol{\mathcal{D}}^{rs}$ in an ascending order
  List $\boldsymbol{\delta}^{rs}$ by following the order of $\boldsymbol{\mathcal{D}}^{rs}$
  Set initial number of neighbors $K_{i,j} = 2$
  Sum the nearest two instants $\hat{\delta}^{\min} = |\delta_1^{rs} + \delta_2^{rs}|$
  for $s = 2 : \#\boldsymbol{\mathcal{D}}^{rs}$ **do**
    Accumulate differences $\hat{\delta}(s)$ based on Eq. (11)
    if $\hat{\delta}(s) \leq \hat{\delta}^{\min}$ **then**
      Replace minimal value $\hat{\delta}^{\min} = \hat{\delta}(s)$
      Update proper neighbors $K_{i,j} = s$
  Sort the $K_{i,j}$ samples in an ascending order
  for $k = 1 : K_{i,j}$ **do**
    Calculate the k^{th} weight according to Eq. (13)
  Obtain weighted resampling $\check{s}_{i,j}$ based on Eq. (14)

distance-weighted resampling calculation. Aiming to select the optimal nearest neighbors of the attribute i, a series of original data samples with the corresponding sampling instants should be recorded at the edge device. The desired resampling instants will be selected as the reference so that the samples of multiple sensing attributes can be fully aligned with each other in the time domain. To select the optimal $K_{i,j}$ at the jth resampling instant of the attribute i, denoted as $t^r_{i,j}$, its time difference $\boldsymbol{\delta}^{rs}_{\boldsymbol{i,j}}$ and Euclidean distance $\boldsymbol{\mathcal{D}}^{rs}_{\boldsymbol{i,j}}$ compared to the original recorded data samples are calculated, respectively, given by

$$\boldsymbol{\delta}^{rs}_{\boldsymbol{i,j}} = t^r_{i,j} - \boldsymbol{t}^s_{\boldsymbol{i}} \tag{9}$$

and

$$\boldsymbol{\mathcal{D}}^{rs}_{\boldsymbol{i,j}} = \sqrt{\left(t^r_{i,j} - \boldsymbol{t}^s_{\boldsymbol{i}}\right)^2} \tag{10}$$

where $\boldsymbol{t}^s_{\boldsymbol{i}}$ is a vector containing the original data sampling instants about the same attribute i.

Due to the randomness of local samples, some of the original samples would be excessively distant from the desired resampling instant, which can lead to inaccurate resampling results. To filter uncorrelated faraway data samples, the median of the distance vector $\boldsymbol{\mathcal{D}}^{rs}_{\boldsymbol{j}}$ is calculated as a threshold, which can help to discover excessive elements. Any original data samples with a distance greater than the median will be removed from the candidate neighbor set, while the distance of the remaining samples $\boldsymbol{\mathcal{D}}^{rs}_{\boldsymbol{j}}$ will be sorted in ascending order so that the sample with the strongest correlation will be listed as the first element.

The selection of $K_{i,j}$ aims at minimizing the imbalance of the impact from original data samples during data interpolation. In other words, simply considering more unidirectional neighbors (i.e., only before or after) will not be beneficial to the final resampling accuracy. Motivated by this observation, the accumulated difference $\hat{\delta}(s)$ of each original sample s is calculated, given by

$$\hat{\delta}(s) = \sum_{l=1}^{s} \delta_j^{rs}(l) \ \forall s = \{2, 3, \ldots, S\} \tag{11}$$

where S is the number of samples after initial filtering, while at least two of the samples should be selected to meet the basic requirement, i.e., $S \geq 2$. By observing the number of samples leading to the minimized accumulated difference $\hat{\delta}^{\min}$, the optimal $K_{i,j}$ can be obtained as

$$K_{i,j} = \left\{ s \middle| \hat{\delta}(s) = \hat{\delta}^{\min}, 2 \leq s \leq S \right\}. \tag{12}$$

After obtaining the optimal $K_{i,j}$ for each desired resampling instant t^r_j, a distance-weighting factor for each original data can be thereby calculated. The weight ω_ρ associated with the ρth closest neighbor is defined as

$$\omega_\rho = \begin{cases} 1, & \text{if } \mathcal{D}^{rs}_{K_{i,j}} = \mathcal{D}^{rs}_1 \\ \frac{\mathcal{D}^{rs}_{K_{i,j}} - \mathcal{D}^{rs}_\rho}{\mathcal{D}^{rs}_{K_{i,j}} - \mathcal{D}^{rs}_1}, & \text{otherwise.} \end{cases} \tag{13}$$

Based on the obtained weighting factors, the final resampling result of the jth resampling instant for attribute i can be calculated by taking the weighted average of the $K_{i,j}$ neighboring samples, given by

$$s^r_{i,j} = \frac{\sum_{k=1}^{K_{i,j}} \omega_k \hat{s}^k_{i,j}}{\sum_{k=1}^{K_{i,j}} \omega_k} \tag{14}$$

where $\hat{s}^k_{i,j}$ is the data sample belongs to the kth nearest neighbor of the resampling instant $t^r_{i,j}$ after clock compensation according to (8).

C. Feedback-Based Sampling Adjustment

Data resampling provides an opportunity for heterogeneous sensors to adjust their sampling rate according to the resampling performance. In this section, a feedback-based sampling rate adjustment mechanism is designed at the edge device, aiming at maintaining the application-specific processing accuracy while minimizing the network overhead during data uploading.

The main purpose of sampling adjustment is to maximize the data efficiency for resampling. Typically, the initial resampling accuracy will be excessive or insufficient compared to the application-specific requirement due to data redundancy or low data rate, respectively. As a direct result, either unnecessary resource wasting or nonideal data modeling accuracy will be expected during digital twin construction. To address these issues, the edge device will responsible for estimating the initial data resampling accuracy by conducting cross-validation based on the attribute samples collected. Attribute with an exceeding or insufficient data resampling accuracy compared to the predefined requirement will be asked to adjust its local sensing rate accordingly. Furthermore, different optimization techniques, such as golden-section search and Ternary search, can be utilized to determine the optimal local sampling rate $\hat{f}$ of each attribute efficiently in meeting the resampling accuracy Q_i, which can be generalized as

$$\hat{f}_i = \underset{f_i}{\operatorname{argmin}}\, \Psi(f_i) - Q_i \tag{15}$$

where f_i is the local sampling rate for attribute i, while $\Psi(f_i)$ is the data resampling accuracy obtained based on the cross-validation.

In addition, the attribute selection for digital twin modeling should also be application driven since not all attributes collected from the local device will be useful. Therefore, it is necessary to only upload correlated information for data resampling and digital twin creation. A penalized-regression-enabled digital twin creation method is introduced in the next section to filter the unnecessary attributes during digital twin modeling. By recording the filtered information at the edge device, local IIoT sensors can further adjust the information to be uploaded, which can significantly help to reduce the network resource consumption and successive digital twin modeling complexity.

D. Digital Twin Creation

Based on the resampled data, digital twins can be established at the edge device by investigating the temporal relationships among the multiple sensing attributes. The digital twins can be identified and modeled by a series of statistical tools, including Tikhonov regularization [24], Lasso [25], and sparse identification of nonlinear dynamics (SINDy) [26], based on the nature of the physical IIoT device to be modeled and the type of data samples. Some selection criteria include the sparsity of the data, the number of attributes, and the linearity of the relations.

To give some general ideas for digital twin creation, Lasso is selected in this section as an example after careful data processing. As a mature system identification approach, Lasso can achieve good performance for sparse data with multicollinearity. Moreover, Lasso can help to filter uncorrelated information from a large number of attributes to reduce the model complexity for comprehensive digital twin modeling. More specifically, the goal of Lasso in the proposed CSDR-DT scheme is to solve the optimization problem defined as

$$\hat{\boldsymbol{\eta}} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}}\, L(\check{\boldsymbol{s}}) \tag{16}$$

where $L(\check{s})$ is the loss function to be minimized, given by

$$L(\check{\boldsymbol{s}}) = \sum_{r=1}^{\mathcal{R}} \left\| \check{s}_r^{\text{out}} - \sum_{i=1}^{\mathcal{I}} \check{s}_r^{\text{in}} \eta_i \right\|^2 + \lambda \sum_{i=1}^{\mathcal{I}} |\eta_i| \tag{17}$$

where $\mathcal{I}$ is the total number of recorded attributes to be modeled and $\mathcal{J}$ is the total number of resampled data for each attribute, respectively. In the loss function $\check{s}_j^{\text{in}}$ and $\check{s}_j^{\text{out}}$ are the input attribute and output attribute obtained after data resampling, given in (14). η_i is the digital twin parameter for each attribute to be solved, while a penalized term defined by a nonnegative regularization parameter λ is added at the end to avoid overfitting issues during optimization. Therefore, by properly conducting the premodeling data resampling for each involved IIoT device and minimizing the loss function given in (17), the parameters of the digital twin can be identified, and replication of the model can be established accordingly in the digital domain.

As the primary goal of this article is to introduce the proposed premodeling data processing, including end-to-end time synchronization and multi-attribute data resampling, the detailed procedures of the statistical digital twin modeling are beyond the main scope of this study. Therefore, the final establishment of digital twins through Lasso will be omitted to maintain our emphasis. More details about Lasso-based model creation can be found in [25], [27], [28], and [29].

IV. Performance Evaluation

In this section, a series of simulations are carried out to evaluate the performance of the CSDR-DT scheme in terms of premodeling data processing, digital twin construction, and the ability to maintain performance under various network conditions. Moreover, a case study about predictive maintenance in IIoT systems is given as a potential application scenario for the proposed CSDR-DT platform.

A. Premodeling Data Processing

In this simulation, an IIoT system with a wide variety of industrial infrastructures and machines is simulated in MATLAB, where 20 sensors generating their unique sensing attributes are specifically considered. Initially, the clock at each device is approximately modeled as a linear function, where the initial offset β_i is a random constant ranging at a submillisecond level. In contrast, clock skew α_i will be more diverse and time-varying (typically from 1 to 100 ppm) due to different manufacturing quality and temperature sensitivity. The modeling of each clock is achieved by analyzing the sequential timestamps from the IIoT device according to (6), based on which the percentage modeling error $e_i(t)$ can be obtained, given by

$$e_i(t) = \frac{\hat{C}_i - C_i(t)}{C_i(t)}. \tag{18}$$

The simulation results of four random attributes in terms of clock modeling error with respect to training times are given as an example in Fig. 4, where it can be observed that more available timestamps and training iterations can help to reduce the clock modeling error. With less than ten training iterations, a slight modeling error around 3% is expected for each attribute, which can help to generate accurate time information for data synchronization in a given period. It is worth noting that, due to the time-varying clock skew for each device, the clock model will gradually deviate from the ground truth, especially for devices with an inexpensive oscillator. This issue can be straightforwardly solved by continuously updating the clock model based on the newly arrived timestamps, which are associated with the data samples used for digital twin creation. A long-term observation regarding the clock

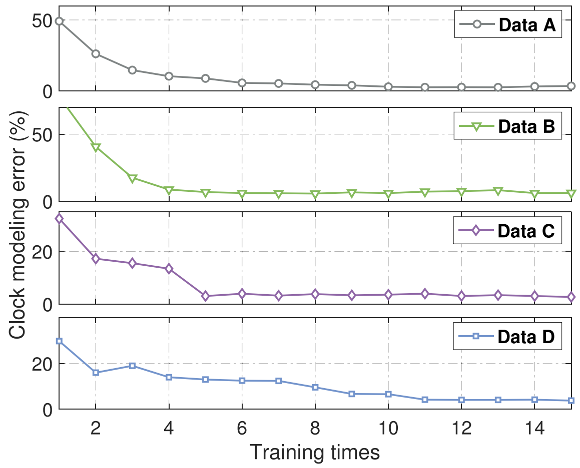

Fig. 4. Accuracy of the clock modeling for four different attributes. Accurate modeling will be achieved after less than ten training iterations.

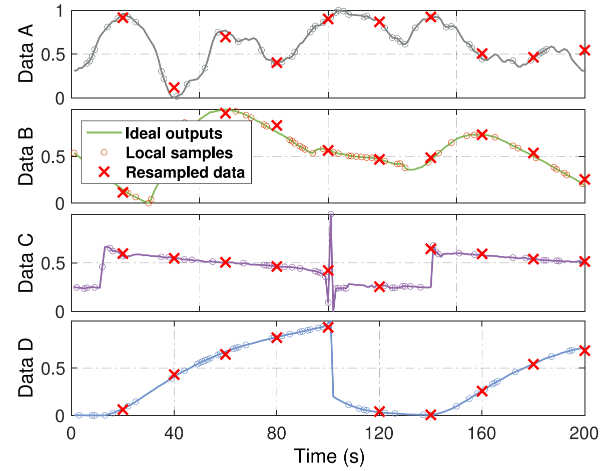

Fig. 5. Temporal consistency of four attributes collected from different sensing processes. The data will be precisely aligned after properly resampling with the proposed CSDR-DT approach.

TABLE II
100-H OBSERVATION OF THE CLOCK MODELING ERROR (s) FOR THE FOUR SELECTED ATTRIBUTES AFTER INITIAL TRAINING

	Initial error	Maximum error	Averaged error
Attribute 1	3.04×10^{-6}	6.95×10^{-4}	2.59×10^{-4}
Attribute 2	2.35×10^{-7}	1.21×10^{-5}	6.11×10^{-6}
Attribute 3	1.03×10^{-6}	6.01×10^{-5}	3.05×10^{-5}
Attribute 4	2.29×10^{-5}	6.50×10^{-3}	6.42×10^{-4}

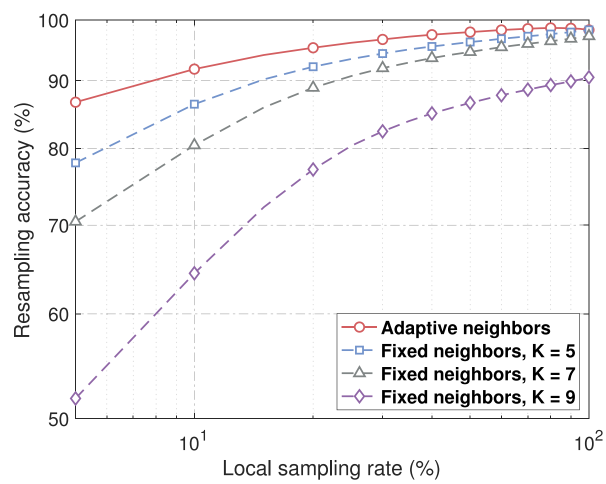

Fig. 6. Proposed ADWKNN algorithm can achieve more accurate resampling performance compared to the traditional DWKNN with a fixed number of neighbors.

modeling error is given in Table II, where the four selected clocks can maintain good accuracy by infrequently estimating the clock parameters (once per hour). The averaged modeling error can be limited to a submillisecond level throughout the entire digital twin construction process, while no additional timestamps that lead to unnecessary network overhead will be induced. Therefore, accurate time information can be maintained with significantly reduced two-way timestamp exchange by adopting the proposed end-to-end time synchronization scheme.

Moreover, the ADWKNN algorithm is conducted to resample distributed IIoT data after its calibration according to the estimated real-time clock offset given by (7). Fig. 5 shows the normalized data from 4 selected attributes, which includes the ambient operating temperature, temperature of a motor, operating velocity of a free-running rotor, and condition of a controlled motor. The raw data (i.e., small circles on each curve) are inconsistently sampled from two perspectives. On the one hand, the data at each device is sampled irregularly due to its dynamic task schedule. The availability and frequency of the samples cannot be guaranteed as a result. On the other hand, the data are sampled nonsimultaneously, which will be critical for temporal correlation analysis for the multiple sensing attributes.

After adopting the ADWKNN algorithm with an adaptive K selection for each series of data, the original data are fully resampled, which is shown as the large red crosses at each curve with the desired interval. It can be observed that the resampled data can closely fit the original data curve, meaning that significant data resampling accuracy is achieved, especially for the areas with smoother curves and considerable samples. In addition, as shown in Fig. 6, by involving more available data samples for each attribute, the resampling accuracy can be further improved. However, the improvement will converge to its limitation after around 60% of the overall data, meaning that the proposed ADWKNN algorithm does not hinge on excessive data for accurate resampling. The performance of the premodeling data processing will be still extremely reliable even if the available data and sampling rates cannot be guaranteed.

Furthermore, to demonstrate the effectiveness of the adaptive neighbor selection compared to traditional schemes where a fixed number of neighbors are selected, resampling accuracy with different scenarios is compared in Fig. 6. Based on the averaged resampling error from the 20 devices, it can be observed that the proposed adaptive K selection can always achieve a smaller error compared to each fixed K. The reason behind this result is the randomness of the available samples.

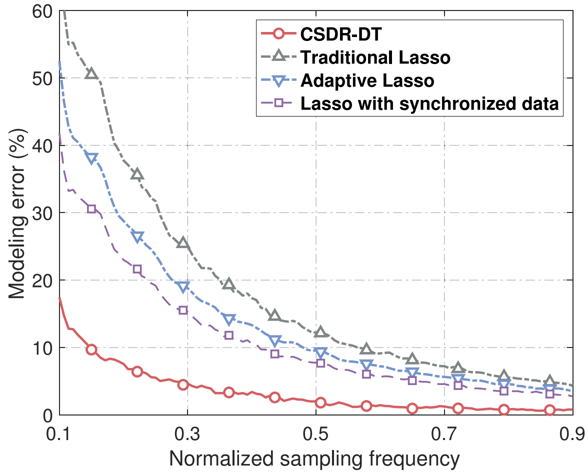

Fig. 7. Digital twin accuracy with different modeling approaches, including CSDR-DT, traditional Lasso, synchronized Lasso, and adaptive Lasso. The modeling accuracy is significantly improved by adopting the CSDR-DT scheme due to the enhanced temporal consistency among multiple attributes.

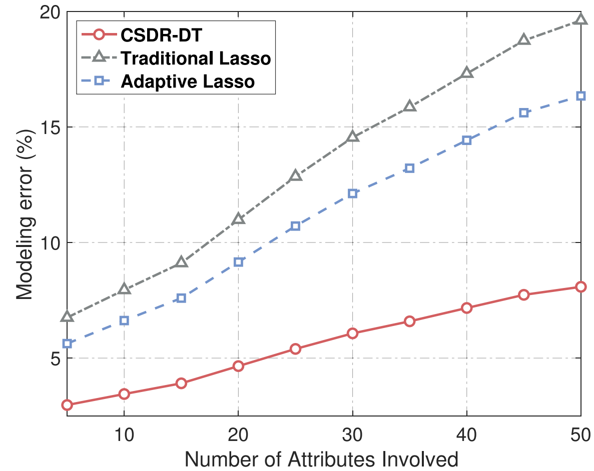

Fig. 8. Digital twin modeling accuracy with different attribute dimensions. With more attributes involved, the correlation analysis will be degraded, especially when the data are not accurately aligned.

A fixed K cannot provide identical benefits for each attribute due to their different sampling situations. In other words, a larger K might be more appropriate for attributes with dense data but will lead to degraded resampling results for other devices with sparser data. The dynamic of sampling conditions will inevitably exacerbate this issue. In contrast, by selecting an adaptive and customized K at each resampling instant according to its neighboring samples, the problem of sampling randomness can be alleviated. There might be an additional computational burden induced from the selection of K, but the calculation is straightforward, and its computational complexity is much lower than other advanced learning-based methods, e.g., [30]. Meanwhile, the involved computational overhead can be considered negligible since the data resampling is achieved in an offline mode at the edge device with enhanced processing capabilities.

B. Modeling Accuracy and Efficiency

Accurate digital twin construction serves as the backbone to support successive industrial applications, while the efficiency during digital twin modeling affects the versatility of the platform due to the limited resources in IIoT systems. Based on this observation, the modeling accuracy and efficiency of the proposed CSDR-DT scheme are, respectively, evaluated in this section. Two different modeling approaches, including traditional Lasso [25] and adaptive Lasso [31], which had been widely adopted in various areas for different modeling proposes (e.g., [27] and [28], and [32], [33], respectively), are selected as the benchmarks in the simulation.

On the one hand, the comparison of digital twin modeling accuracy is shown in Fig. 7, where one more modeling scenario, synchronized Lasso (i.e., traditional Lasso with synchronized data), is also selected. The sampling rate is normalized based on the minimum and maximum of the available samples in this simulation, where it can be observed that a higher sampling frequency can continuously bring an increment to digital twin modeling accuracy. Moreover, the accuracy of the proposed CSDR-DT scheme is always higher than directly adopting the Lasso algorithms, especially when no synchronization strategy is adopted. Although adaptive Lasso can achieve higher accuracy than traditional Lasso, the lack of temporal consistency among the multiple attributes still limits its performance. Additionally, the difference between scenarios with and without synchronization is relatively small, but accurate time synchronization is indispensable in the proposed scheme since it not only directly increases the modeling accuracy but also benefits the resampling performance. With both time synchronization and data resampling mechanisms, a significant improvement in the modeling accuracy (more than 50%) is expected.

Furthermore, the impact of the system scale is also considered to evaluate the proposed scheme. A larger system scale will indicate a higher attribute dimension, with more data involved during data processing and modeling. However, not all attributes collected will be closely relevant to the digital twin modeling. As shown in Fig. 8, different modeling scales ranging from 5 to 50 attributes are considered during the simulation, where a larger attribute dimension will lead to lower modeling accuracy. Although Lasso methods are effective in filtering uncorrelated attributes and shrinking the modeling dimension, the unsolved temporal misalignment will deteriorate the modeling performance for both traditional and adaptive Lasso. In contrast, CSDR-DT is less sensitive to the attribute dimension, which can improve the digital twin modeling behavior for all attribute dimensions with more than doubled accuracy.

On the other hand, the induced network overhead during digital twin construction is used to evaluate the efficiency of the proposed CSDR-DT scheme. As demonstrated in Fig. 9, higher network overhead will be expected with lower digital twin modeling error, leading to the necessary tradeoff between accuracy and efficiency. However, the induced network overhead of the proposed CSDR-DT scheme is always lower than both the traditional Lasso and adaptive Lasso algorithms, especially when the required modeling accuracy is very high. In

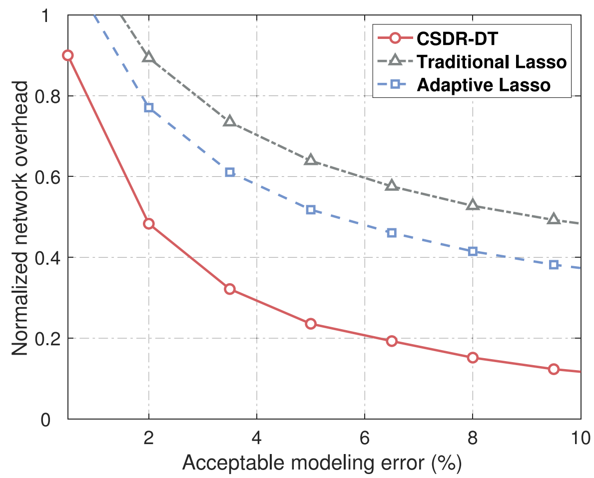

Fig. 9. Digital twin modeling efficiency of the proposed CSDR-DT method compared to synchronized Lasso and adaptive Lasso. The CSDR-DT-based approach will pose a lower overhead for different modeling requirements.

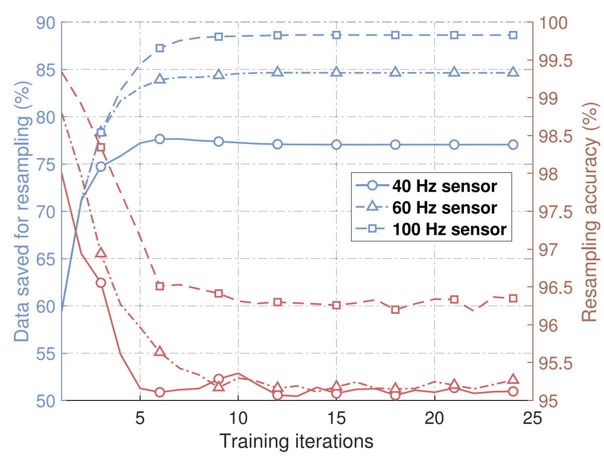

Fig. 10. Resource consumption and accuracy of data resampling for different attributes collected from sensors with diverse sampling capabilities.

addition, it can be observed that the highest modeling accuracy cannot be achieved by the methods without accurate temporal alignment, no matter how many attribute data are sampled and transmitted. Therefore, by adopting the proposed CSDR-DT scheme, a higher modeling accuracy is achievable, while more network resources can also be saved to support other critical industrial applications.

In addition, the network overhead can be further reduced by adopting the proposed feedback-based sampling adjustment mechanism. As shown in Fig. 10, with more training iterations, the resources used for digital twin creation will be gradually reduced due to the adaptation of local sampling rates. Although the training accuracy is slightly reduced as a negative impact, the application-specific requirement, which is set to be 95%, can always be achieved. Moreover, for sensors with higher sampling rates, more resources are expected to be saved due to their increasingly redundant local samples. This is also validated in Fig. 11, where it can be observed that the percentage of data used to maintain modeling accuracy in sensors with higher sampling rates will be the least. Therefore, it can be concluded that the proposed feedback-based sampling rate

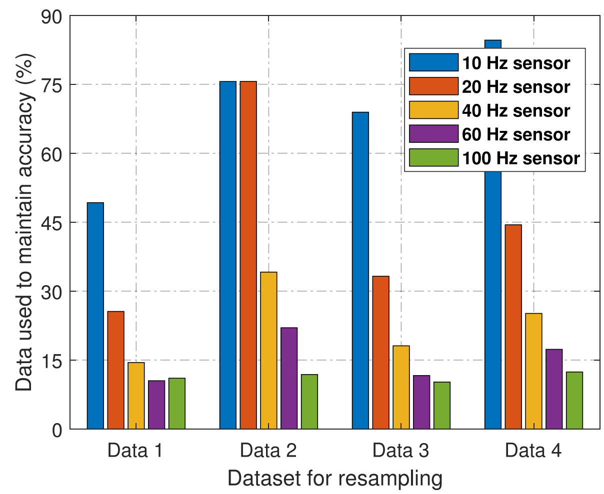

Fig. 11. Change of samplings required and accuracy achieved after adopting the feedback-based sampling adjustment.

adjustment mechanism can significantly reduce the network overhead during digital twin creation, especially for sensors with higher sampling capabilities.

C. Modeling Robustness

Due to the involvement of a large number of heterogeneous industrial machines and nonunified communication standards in complicated IIoT systems, non-negligible random network delays and stochastic packet losses will be ubiquitously associated with the frequent timestamp and data exchange during digital twin construction. Therefore, it is necessary to evaluate the digital twin modeling performance under various network conditions. Similarly, traditional Lasso-based modeling and the proposed CSDR-DT scheme will be compared, aiming at demonstrating the improvement by adopting accurate premodeling data processing.

More specifically, as mentioned in Section II-B, the network latency during digital twin construction is assumed to follow the Gaussian distribution, where both the mean (μ) and standard deviation (σ^2) will affect the modeling performance. The modeling accuracy of the proposed CSDR-DT scheme is shown as a contour diagram in Fig. 12, where a larger network latency will always lead to a more significant modeling error. Moreover, a larger σ^2 can affect the modeling accuracy more severely compared to μ, given the reason that σ^2 will result in more randomness in the data during model establishment. The reduced timeliness of distributed data can further deteriorate the temporal consistency among the devices, which is fatal for digital twin construction. However, by adopting CSDR-DT data processing, the modeling accuracy can be greatly improved, shown in Fig. 13, especially for the scenario with larger σ. Moreover, it can be observed that the modeling accuracy will be maintained within an acceptable range for extreme network conditions compared to the traditional modeling method.

Besides network latency, the stochastic packet loss is another important reason leading to increased randomness of the data samples. The amount of packet losses is condition-dependent,

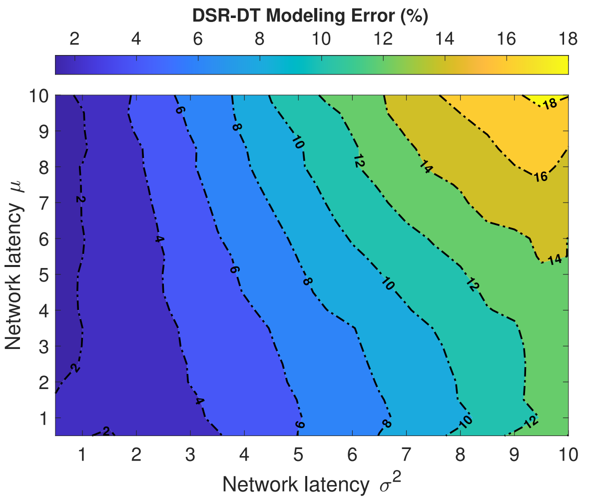

Fig. 12. Modeling accuracy of CSDR-DT with different network conditions. With both the increment of mean and deviation of the network delay, a small but increasing modeling error will be observed.

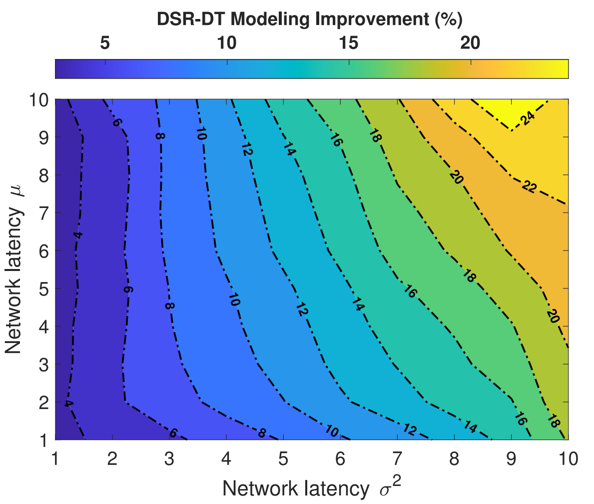

Fig. 13. Compared to the Lasso modeling with raw data, a significant improvement is expected, especially for worse network conditions.

which is highly correlated to the number of devices, network access mechanisms, and channel quality. As shown in Fig. 14, a wide range of packet loss rates from 5% to 30% is considered, where the latter scenarios can be considered beyond normal conditions. It is notable that the modeling error of the CSDR-DT scheme remains at a relatively low level, while packet losses will have a more significant impact on the modeling performance if Lasso and adaptive Lasso algorithms are directly adopted. Furthermore, the orange bars show the modeling improvement compared to the traditional Lasso. Based on Figs. 12 and 13, it can be concluded that the proposed CSDR-DT scheme can improve the digital twin construction performance under different network conditions, which can meet the practical requirements of industrial applications.

D. Case Study: Predictive Maintenance

To demonstrate the effectiveness of the proposed CSDR-DT scheme and introduce some potential application scenarios

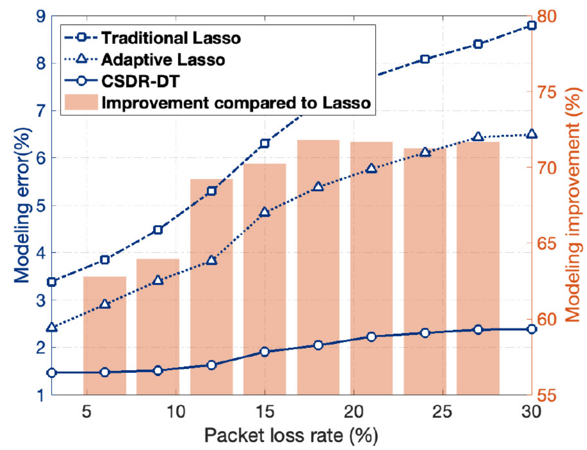

Fig. 14. With different packet loss rates, modeling accuracy can be improved for more than 50% by adopting the proposed CSDR-DT method.

in IIoT systems, a case study about predictive maintenance is demonstrated in this section. More specifically, predictive maintenance serves as a very promising technique in Industrial 4.0, which can achieve proactive machine condition guaranteeing, with minimized interruption of the normal operation [34], [35], [36]. In this simulation, three different maintenance approaches are considered and compared, namely:

1) *corrective maintenance* is carried out after detecting anomalies. This method is used to restore normal operating conditions and only repairs upon fault detection.
2) *periodic maintenance* will be conducted according to the predefined schedule, aiming to minimize the impact of the time-induced errors. However, abnormal operations and external anomalies will be very challenging for this type of maintenance to deal with.
3) *CSDR-DT-based predictive maintenance*, which is carried out at predetermined intervals or according to prescribed criteria, aiming to reduce the failure risk or performance degradation of the equipment. The maintenance cycles are planned according to the operation schedule and requirements of the IIoT devices.

Moreover, two different kinds of machine failure are considered, including spontaneous deteriorations (e.g., aging, fatigue, and wear) and external accidents (e.g., abnormal operations and unexpected materials). The deterioration level of the target machine is modeled as a time-dependent function according to [37], given by

$$\mathcal{DL}_n(t) = \left(\frac{a_n}{a_n - \xi t}\right)^{b_n} \tag{19}$$

where a_i and b_i are the intrinsic deterioration factors of the device i, while ξ is the imaginary unit to avoid confusion.

The proposed CSDR-DT-based predictive maintenance is achieved from two perspectives, as shown in Fig. 15. On the one hand, the continuously generated system states will be collected at the edge devices, which will be responsible for the condition check. Any newly generated system states that significantly violate the historical feature will be reported as abnormal inputs, which can help the target machine to avoid extremely accidental events. In this case, the previously established digital twin will generate reliable system states

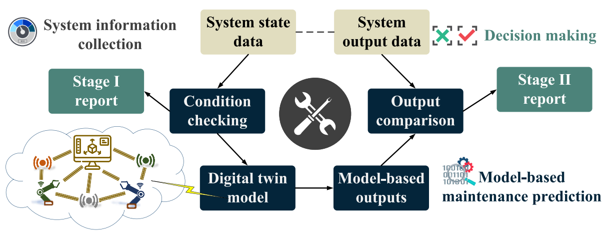

Fig. 15. CSDR-DT-enabled predictive maintenance flowchart comprises input abnormal filtering and output data evaluation.

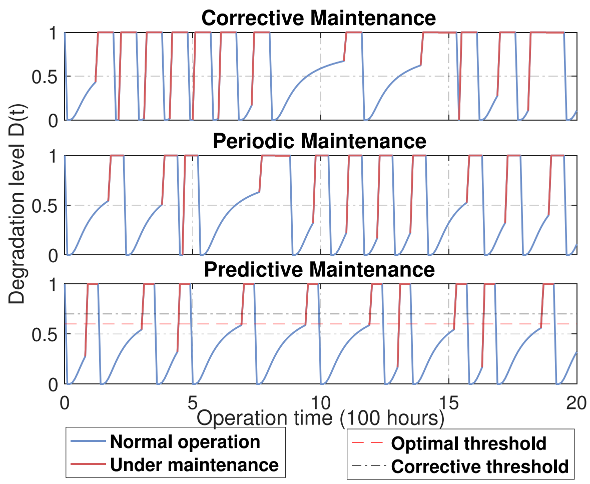

Fig. 16. Deterioration of the target device under different maintenance strategies for 100-h observation. With CSDR-DT-enabled predictive maintenance, the degradation level will be much slighter and more predictable.

to maintain normal. On the other hand, the newly generated system outputs will be compared with the ones from the digital twin, while outliers will be reported as well. By adopting this two-stage abnormal report, accidental events can be alleviated by properly scheduling timely maintenance.

The normalized deterioration levels of the target devices after adopting these three maintenance approaches are shown in Fig. 16, where the devices will shut down if no maintenance is adopted correctly (i.e., when the line goes to 1). Periodic maintenance can outperform corrective one by avoiding shutdowns of the machines. However, due to the existence of accidental events, the performance of these two maintenance strategies will be degraded as they cannot predict and avoid accidental errors. In contrast, predictive maintenance enabled by the CSDR-DT can achieve better performance by observing both the input and output of the system. Thus, the deterioration level and the operation time can be significantly enhanced.

Furthermore, remaining useful life (RUL) is commonly used in traditional studies [37], [38] to describe the performance of maintenance, which shows the time that the machine is expected to operate normally. However, the use of RUL is not always sufficient to show the effectiveness of predictive maintenance since accidental events, which dominate industrial breakdowns, will not be demonstrated by using RUL. As a result, we will use the operation rate in this case study to illustrate the healthy of the target devices and evaluate the performance of the digital twin platform by observing the operating rate under different operating periods and accidental

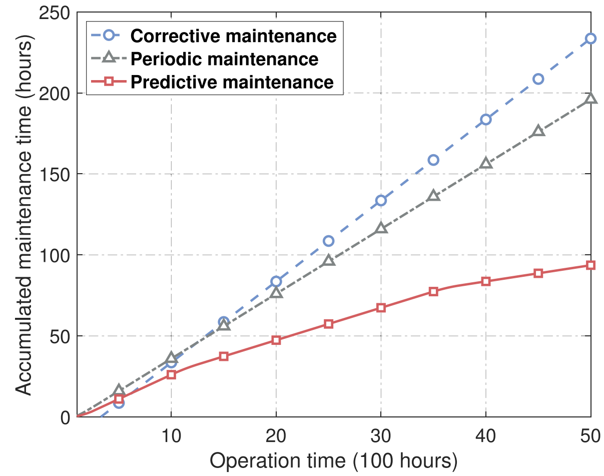

Fig. 17. Maintenance time required for the three different maintenance strategies. CSDR-DT-enabled predictive maintenance will consume the least amount of time to maintain the functionality of machines.

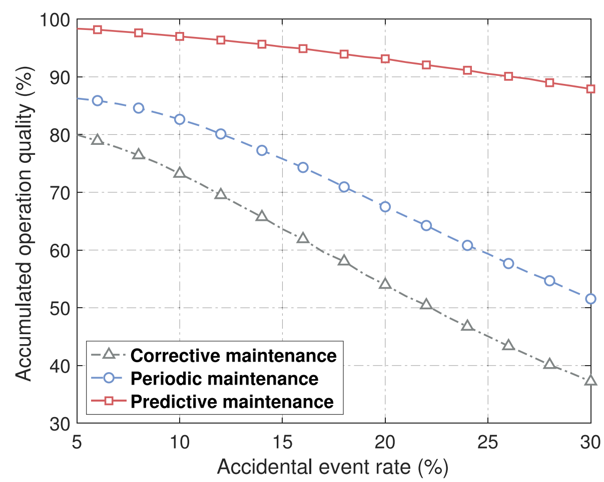

Fig. 18. Overall operation quality of the target devices by adopting the three maintenance strategies, where CSDR-DT-enabled predictive maintenance can help the machine maintain a higher operation quality.

events rates. Thanks to the accurate prediction, the accumulated maintenance time and overall operation quality of the target devices can be enhanced, as shown in Figs. 17 and 18, respectively. As the digital twin of the physical system will be updated based on the newly gathered information, the prediction accuracy can be further improved for long-term operation. It is clear that the proposed CSDR-DT-enabled predictive maintenance can achieve the least shutdown time compared to the other two strategies. Moreover, with different accidental event rates, the CSDR-DT-based approach can always overweight other maintenance methods. Therefore, it can be concluded that the proposed CSDR-DT scheme can help to provide a more accurate and timely data analysis, while the corresponding decision making can improve industrial applications significantly.

V. Conclusion

In this article, a concurrent end-to-end synchronization and multi-attribute data resampling scheme has been designed to achieve accurate and efficient digital twin construction for IIoT applications. An edge-centric digital twin platform has been formed for multidimensional data collection, premodeling processing, and digital twin creation. To tackle the temporal misalignment among the multiple attributes, the raw data will be resampled according to the same reference frequency, with the help of virtually digital twined clocks providing cohesively aligned time information. A feedback-based sampling rate adjustment and attribute selection mechanism has been designed afterward to optimize the local samples regarding the application-specific requirements. Simulation results have shown that the proposed CSDR-DT scheme can achieve more accurate, efficient, and robust digital twin construction than existing modeling methods. In the end, digital-twin-driven predictive maintenance has been given as a case study, which illustrated the potential applications and benefits of the proposed CSDR-DT in industrial environments. In the future, we will explore the feasibility of utilizing machine learning-based methods for achieving accurate data resampling. Moreover, extending the proposed digital twin platform to larger scale IIoT systems will be a promising research direction in supporting advanced collaborative applications.

REFERENCES

[1] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, “Toward the 6G network era: Opportunities and challenges,” *IT Prof.*, vol. 22, no. 1, pp. 34–38, Jan./Feb. 2020.

[2] D. C. Nguyen et al., “6G Internet of Things: A comprehensive survey,” *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[3] Q. Qi and F. Tao, “Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison,” *IEEE Access*, vol. 6, pp. 3585–3593, 2018.

[4] N. Dall’Ora, K. Alamin, E. Fraccaroli, M. Poncino, D. Quaglia, and S. Vinco, “Digital transformation of a production line: Network design, online data collection and energy monitoring,” *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 1, pp. 46–59, Jan./Mar. 2022.

[5] A. W. Colombo, S. Karnouskos, O. Kaynak, Y. Shi, and S. Yin, “Industrial Cyberphysical systems: A backbone of the fourth industrial revolution,” *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 6–16, Mar. 2017.

[6] L.-H. Lee et al., “All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda,” 2021, *arXiv:2110.05352*.

[7] R. Minerva, G. M. Lee, and N. Crespi, “Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models,” *Proc. IEEE*, vol. 108, no. 10, pp. 1785–1824, Oct. 2020.

[8] P. Jia, X. Wang, and X. Shen, “Digital-twin-enabled intelligent distributed clock synchronization in industrial IoT systems,” *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4548–4559, Mar. 2021.

[9] B. R. Barricelli, E. Casiraghi, and D. Fogli, “A survey on digital twin: Definitions, characteristics, applications, and design implications,” *IEEE Access*, vol. 7, pp. 167653–167671, 2019.

[10] M. Wollschlaeger, T. Sauter, and J. Jasperneite, “The future of industrial communication: Automation networks in the era of the Internet of Things and industry 4.0,” *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 17–27, Mar. 2017.

[11] 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.

[12] H. E. van Staden and R. N. Boute, “The effect of multi-sensor data on condition-based maintenance policies,” *Eur. J. Oper. Res.*, vol. 290, no. 2, pp. 585–600, 2021.

[13] P. Jia, X. Wang, and K. Zheng, “Distributed clock synchronization based on intelligent clustering in local area industrial IoT systems,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3697–3707, Jun. 2020.

[14] A. Sari, A. Lekidis, and I. Butun, “Industrial networks and IIoT: Now and future trends,” in *Industrial IoT*. Cham, Switzerland: Springer, 2020, pp. 3–55.

[15] H. Darvishi, D. Ciuonzo, E. R. Eide, and P. S. Rossi, “Sensor-fault detection, isolation and accommodation for digital twins via modular data-driven architecture,” *IEEE Sensors J.*, vol. 21, no. 4, pp. 4827–4838, Feb. 2021.

[16] G. N. Schroeder, C. Steinmetz, R. N. Rodrigues, R. V. B. Henriques, A. Rettberg, and C. E. Pereira, “A methodology for digital twin modeling and deployment for industry 4.0,” *Proc. IEEE*, vol. 109, no. 4, pp. 556–567, Apr. 2021.

[17] S. R. Newrzella, D. W. Franklin, and S. Haider, “5-dimension cross-industry digital twin applications model and analysis of digital twin classification terms and models,” *IEEE Access*, vol. 9, pp. 131306–131321, 2021.

[18] Y. Wu, Y. Li, S. He, and Y. Guan, “Sampled-data synchronization of network systems in industrial manufacture,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 9, pp. 3210–3219, 2020.

[19] A. Gellert, D. Sarbu, S.-A. Precup, A. Matei, D. Circa, and C.-B. Zamfirescu, “Estimation of missing LiDAR data for accurate AGV localization,” *IEEE Access*, vol. 10, pp. 68416–68428, 2022.

[20] C. Hu et al., “Digital twin-assisted real-time traffic data prediction method for 5G-enabled Internet of Vehicles,” *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2811–2819, Apr. 2022.

[21] Y. Yi, Y. Yan, X. Liu, Z. Ni, J. Feng, and J. Liu, “Digital twin-based smart assembly process design and application framework for complex products and its case study,” *J. Manuf. Syst.*, vol. 58, pp. 94–107, Jan. 2021.

[22] Q. Liu et al., “Digital twin-based designing of the configuration, motion, control, and optimization model of a flow-type smart manufacturing system,” *J. Manuf. Syst.*, vol. 58, pp. 52–64, Jan. 2021.

[23] S. A. Dudani, “The distance-weighted k-nearest-Neighbor rule,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. SMC-6, no. 4, pp. 325–327, Apr. 1976.

[24] H. Liu, L. Yan, Y. Chang, H. Fang, and T. Zhang, “Spectral deconvolution and feature extraction with robust adaptive Tikhonov regularization,” *IEEE Trans. Instrum. Meas.*, vol. 62, no. 2, pp. 315–327, Feb. 2013.

[25] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Royal Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[26] E. Kaiser, J. N. Kutz, and S. L. Brunton, “Sparse identification of nonlinear dynamics for model predictive control in the low-data limit,” *Proc. Royal Soc. A*, vol. 474, no. 2219, 2018, Art. no. 20180335.

[27] N. Tang, S. Mao, Y. Wang, and R. M. Nelms, “Solar power generation forecasting with a LASSO-based approach,” *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1090–1099, Apr. 2018.

[28] W. Wu, C. Tan, S. Zhang, and F. Dong, “Sparse local fisher discriminant analysis for gas-water two-phase flow status monitoring with multisensor signals,” *IEEE Trans. Ind. Informat.*, early access, Jun. 21, 2022, doi: 10.1109/TII.2022.3185077.

[29] Y. Plan and R. Vershynin, “The generalized lasso with non-linear observations,” *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1528–1537, Mar. 2016.

[30] S. Mandelli, V. Lipari, P. Bestagini, and S. Tubaro, “Interpolation and denoising of seismic data using convolutional neural networks,” 2019, *arXiv:1901.07927*.

[31] H. Zou, “The adaptive lasso and its oracle properties,” *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.

[32] J. Li, Q. Lu, and Y. Wen, “Multi-kernel linear mixed model with adaptive lasso for prediction analysis on high-dimensional multi-omics data,” *Bioinformatics*, vol. 36, no. 6, pp. 1785–1794, 2020.

[33] C. Wang, Q. Li, X. Song, and X. Dong, “Bayesian adaptive lasso for additive hazard regression with current status data,” *Stat. Med.*, vol. 38, no. 20, pp. 3703–3718, 2019.

[34] T. Hafeez, L. Xu, and G. Mcardle, “Edge intelligence for data handling and predictive maintenance in IIoT,” *IEEE Access*, vol. 9, pp. 49355–49371, 2021.

[35] P. Wen, Y. Li, S. Chen, and S. Zhao, “Remaining useful life prediction of IIoT-enabled complex industrial systems with hybrid fusion of multiple information sources,” *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9045–9058, Jun. 2021.

[36] M. Pech, J. Vrchota, and J. Bednář, “Predictive maintenance and intelligent sensors in smart factory,” *Sensors*, vol. 21, no. 4, p. 1470, 2021.

[37] Q. Wang, S. Bu, and Z. He, “Achieving predictive and proactive maintenance for high-speed railway power equipment with LSTM-RNN,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6509–6517, Oct. 2020.

[38] C. Chen, N. Lu, B. Jiang, and C. Wang, “A risk-averse remaining useful life estimation for predictive maintenance,” *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 2, pp. 412–422, Feb. 2021.

Pengyi Jia (Member, IEEE) received the M.Eng. and Ph.D. degrees from the Department of Electrical and Computer Engineering, Western University, London, ON, Canada, in 2016 and 2021, respectively.

He is currently a Postdoctoral Associate with Western University. His research interests include intelligent network synchronization, digital twin, and machine learning, as well as their applications in vertical IoT systems, cellular networks, and advanced manufacturing. One focus of his recent research is to develop accurate and efficient digital twin paradigms by investigating the intrinsic temporal correlations behind the massive sampling data to support network optimization and application-oriented service provisioning.

Xianbin Wang (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2001.

He is a Professor and a Tier-1 Canada Research Chair with Western University, London, ON, Canada. Prior to joining Western University, he was a Research Scientist/Senior Research Scientist with the Communications Research Centre, Ottawa, ON, Canada, from July 2002 to December 2007. He was a System Designer with STMicroelectronics, Geneva, Switzerland, from January 2001 to July 2002. He has over 500 highly cited journal and conference papers, in addition to 30 granted and pending patents and several standard contributions. His current research interests include 5G/6G technologies, Internet of Things, communications security, machine learning, and intelligent communications.

Dr. Wang has received many prestigious awards and recognitions, including the IEEE Canada R.A. Fessenden Award, the Canada Research Chair, the Engineering Research Excellence Award at Western University, the Canadian Federal Government Public Service Award, and the Ontario Early Researcher Award and six IEEE Best Paper Awards. He currently serves/has served as an Editor-in-Chief, an Associate Editor-in-Chief, and an Editor/Associate Editor for over ten journals. He was involved in many IEEE conferences, including GLOBECOM, ICC, VTC, PIMRC, WCNC, CCECE, and CWIT, in different roles, such as the general chair, the symposium chair, the tutorial instructor, the track chair, the session chair, and the TPC co-chair and keynote speaker. He has been nominated as an IEEE Distinguished Lecturer several times during the last ten years. He was the Chair of the IEEE ComSoc Signal Processing and Computing for Communications Technical Committee and is currently serving as the Central Area Chair for IEEE Canada. He is an IEEE Distinguished Lecturer and a Fellow of the Canadian Academy of Engineering and Engineering Institute of Canada.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990.

He is a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research focuses on network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular ad hoc and sensor networks.

Dr. Shen received the Canadian Award for Telecommunications Research from the Canadian Society of Information Theory in 2021, the R.A. Fessenden Award in 2019 from IEEE, Canada, the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, the Joseph LoCicero Award in 2015 and the Education Award in 2017 from the IEEE Communications Society, and the Technical Recognition Award from Wireless Communications Technical Committee in 2019 and AHSN Technical Committee in 2013. He has also received the Excellent Graduate Supervision Award in 2006 from the University of Waterloo and the Premier's Research Excellence Award in 2003 from the Province of Ontario, Canada. He served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom'16, IEEE Infocom's14, IEEE VTC'10 Fall, and IEEE Globecom'07, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications. He is the President of the IEEE Communications Society. He was the Vice President for Technical & Educational Activities, and Publications, the Member-at-Large on the Board of Governors, the Chair of the Distinguished Lecturer Selection Committee, and a member of the IEEE Fellow Selection Committee of the ComSoc. He served as the Editor-in-Chief for the IEEE INTERNET OF THINGS JOURNAL, IEEE NETWORK, and *IET Communications*. He is a registered Professional Engineer of Ontario, a Chinese Academy of Engineering Foreign Member, a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society, and a Fellow of the Engineering Institute of Canada, Canadian Academy of Engineering, and Royal Society of Canada.