

# An End-to-End Multisource Information Fusion Framework for f-CaO Content Soft Sensing in Cement Clinker Burning Process

Yuchen Zhao<sup>ID</sup>, *Student Member, IEEE*, Chunjie Yang<sup>ID</sup>, *Senior Member, IEEE*, Hang Xiao<sup>ID</sup>, Yaoyao Bao<sup>ID</sup>, Siwei Lou<sup>ID</sup>, *Student Member, IEEE*, Jiayun Mao<sup>ID</sup>, and Huanyu Liao<sup>ID</sup>

**Abstract**—In the cement clinker burning process, the soft sensing of free calcium oxide (f-CaO) content has been a challenging task due to the dynamic time delay between f-CaO content and process variables, the different time scale between process variables and f-CaO content, and the strong nonlinearity of the process data. With the development of data-driven modeling techniques, numerous soft sensing methods for f-CaO content based on process data have emerged. However, under this circumstance, the monotony of soft sensor input may become a bottleneck that limits further improvement of the f-CaO prediction performance. To address this issue, this article proposes a novel end-to-end multisource information fusion framework (MSIFF) for soft sensing the f-CaO content within the cement clinker. The MSIFF takes process data and flame images as inputs and utilizes mechanistic knowledge by generating mechanistic features from process data using a first-principle rotary kiln model. The explainable dynamic features are extracted from the matched process data and flame image sequences with a multisource dynamic feature extraction network (MSDFE), which further participates in the end-to-end modeling of f-CaO content together with the mechanistic features. The proposed MSIFF method is validated on a real cement production line. While providing valuable operating information for the cement clinker burning process, the MSIFF exhibits an improved f-CaO soft sensing performance compared to existing f-CaO estimation methods.

**Index Terms**—Dynamic feature extracting, end-to-end modeling, first-principle modeling, free calcium oxide (f-CaO) soft sensing, multisource information fusion, semisupervised learning.

## I. INTRODUCTION

THE manufacturing of cement can be mainly divided into three processes: raw material grinding, clinker burning, and cement finishing grinding. Among them, the clinker burning process, which mainly takes place inside

Manuscript received 31 May 2023; revised 16 July 2023; accepted 27 July 2023. Date of publication 14 August 2023; date of current version 28 August 2023. This work was supported by the Alibaba-Zhejiang University Joint Institute of Frontier Technologies (AZFT). The Associate Editor coordinating the review process was Dr. Rosenda Valdes Arencibia. (*Corresponding author: Chunjie Yang*)

Yuchen Zhao, Chunjie Yang, Hang Xiao, and Siwei Lou are with the State Key Laboratory of Industrial Control Technology (SKLICT), College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: zhaoyuchen@zju.edu.cn; cjiang999@zju.edu.cn; xiaohang01@zju.edu.cn; swlou@zju.edu.cn).

Yaoyao Bao, Jiayun Mao, and Huanyu Liao are with the Alibaba Cloud, Alibaba Group, Hangzhou 310030, China (e-mail: yaoyao.byy@alibaba-inc.com; jiayun.mjy@alibaba-inc.com; huanyu.lhy@alibaba-inc.com). Digital Object Identifier 10.1109/TIM.2023.3304693

the precalciner and rotary kiln, transforms the raw material into cement clinker by heating and calcinating. This process plays a vital role in determining the cement's quality, further impacting the concrete's strength and stability. The cement quality is usually measured by the amount of free calcium oxide (f-CaO) content, an unstable compound present in the clinker. A high level of f-CaO content indicates that the cement may not be stable enough, while a level of f-CaO content too low suggests that the clinker has been over-calcinated, resulting in greater energy consumption [1]. Therefore, the measurement of f-CaO content is necessary for cement production.

Fig. 1 exhibits a flow diagram of the typical clinker burning process. The process begins with preheating the powdered raw material in the preheaters. Then, the preheated raw material falls into the precalciner, where 90% of the calcium carbonate is decomposed at a relatively high temperature. After that, the raw material enters the rotary kiln, the core equipment in the cement clinker burning system for an overall calcination to convert raw material into cement clinker.

During the clinker calcination, the majority of calcium oxide (CaO) combines with other oxides to form stable compounds within the rotary kiln during the calcination process. However, a small portion of CaO remains uncombined, resulting in the formation of f-CaO. Based on the temperature and clinker form, the areas within the rotary kiln can be divided into different zones as the calcining zone, transition zone, burning zone, and discharging zone. The temperature of the burning zone, which is annotated in Fig. 1, indicates the degree of clinker reaction within the rotary kiln and therefore provides an important reference for estimating f-CaO content within the clinker. Following the calcination process, the grate cooler is used to cool the discharged cement clinker while recovering heat through the heat exchange between the clinker and airflow. In the end, the hot air is sent back to the rotary kiln and precalciner to aid the calcinating process, and these types of air are, respectively, called secondary wind and tertiary wind.

The traditional method for obtaining f-CaO content involves sampling cooled clinker and performing offline chemical analysis in a laboratory, which takes more than an hour before the analysis results can be reported. Although automation equipment can speed up this process, the time lag caused by chemical analysis still prevents real-time control and

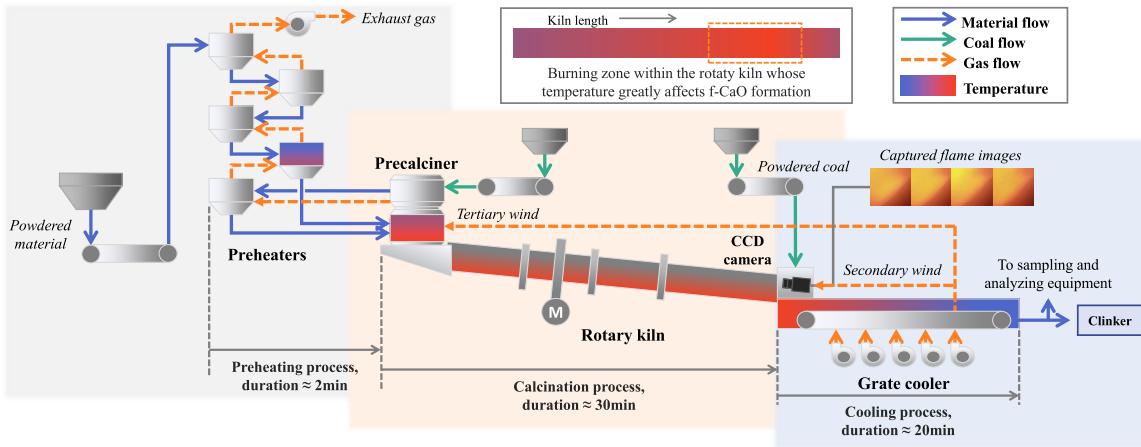


Fig. 1. Flow diagram of the cement clinker burning process.

optimization of the clinker-burning process based on f-CaO content. Therefore, the need for developing methods for real-time f-CaO soft sensing arises. Due to the difficulty in expressing the mechanistic relationship between various process variables and industrial quality indexes, such as ferrous oxide (FeO) [2], [3], silicon [4], and f-CaO content [5], data-driven modeling methods have been widely applied in the field of quality index prediction. However, the application of data-driven methods generally presents three challenges in f-CaO soft sensing.

First, in contrast to gases and liquids in process industries, the solid material in the cement clinker burning system requires a much longer time to pass through each process, especially the calcination process in the rotary kiln and the cooling process within the grate cooler. As indicated in Fig. 1, the material often takes about 30 min to pass from the inlet to the outlet of the rotary kiln, and about 20 min to pass through the grate cooler. This phenomenon results in time delays between the change in f-CaO content and the measured/controlled parameters of the devices, defined as the process variables within the clinker burning system. Moreover, the time delays are affected by the variation in process variables of the rotary kiln and grate cooler, making it difficult to match the correct input and output for f-CaO soft sensing models.

Second, the analysis of f-CaO content is performed on an hourly basis, but the other process variables are updated with an interval of 1 min or less in the distributed control system (DCS). This results in the f-CaO content and process variable data existing on different time scales, which in turn leaves a significant amount of data unlabeled.

Finally, the reactions during the clinker burning process are complex and interdependent, engendering a highly nonlinear relationship between f-CaO content and process variables, and this is also a common phenomenon in most industrial quality index soft sensing tasks [6]. Without diverse information input and appropriate information extraction methods, obtaining satisfactory f-CaO content estimation results can be challenging.

In this article, an end-to-end multisource information fusion framework (MSIFF) is proposed for the soft sensing of

f-CaO content in cement clinker using process data, burning mechanistic knowledge, and flame images. Specifically, MSIFF aims to address time-varying delays, multiple time scales, and strong nonlinearity that exists in f-CaO soft sensing tasks while providing valuable operating information on the cement clinker burning process. In MSIFF, preprocessed images and process data are dynamically matched in fixed-length sequences using the material's motion mechanism. Subsequently, mechanistic features are generated from the process data by a real-time rotary kiln temperature calculation model. Finally, the multisource data sequences, together with the mechanistic features and the mean value features of the process variables, are input into a semisupervised multisource dynamic feature extraction network (MSDFE) to extract dynamic features and at the same time estimate the f-CaO content in cement clinker.

The main contributions are summarized below.

- 1) A first-principle model is developed to calculate the temperature within the cement rotary kiln as well as other significant mechanistic features using available process variables. Compared with existing rotary kiln modeling methods, this model provides a second-scale real-time estimating way of the rotary kiln temperature and, at the same time, utilizes the mechanistic information for downstream f-CaO soft sensing tasks.
- 2) To make full use of unlabeled multisource data, a novel autoencoder that combines the advantages of long short-term memory (LSTM) and CNN networks is proposed to extract explainable dynamic features from minute-scale multisource data sequences, which further enriches the soft sensing model's information input.
- 3) Based on aforementioned information extraction methods, an end-to-end framework MSIFF is established to measure f-CaO content by fusing multisource information. The MSIFF takes the multisource data as the input and directly outputs f-CaO estimation, addressing problems from training different models separately. The experiments indicate that the MSIFF's performance surpasses the existing f-CaO soft sensing methods, which rely on monotony process data inputs.

The rest of this article is organized as follows. Section II presents the related works associated with our methods. Section III briefly introduces the preliminaries of this work. Section IV presents the details of the proposed MSIFF. In Section V, the MSIFF is verified in a real cement production line. Finally, conclusions are made in Section VI.

## II. RELATED WORKS

### A. Progress of f-CaO Content Soft Sensing Methods

In early times, f-CaO modeling used methods such as principal component regression (PCR), partial least-squares (PLSs) estimation [7], and backpropagation neural networks (BP-NNs) [8] to address the nonlinearity between process variables and f-CaO content. To improve soft sensing performance, an ensemble of support vector machine (ESVM) regression models was introduced [9]. This approach used clustering to classify operating modes and build SVM regression models. Time delays were accounted for using mechanistic methods to align the model's input and output with a fixed time length.

However, these early works did not fully utilize unlabeled data until advanced computational and deep-learning techniques were adopted. For instance, a Bayesian-optimized light gradient boosting machine (BO-LightGBM) was proposed to estimate f-CaO content using the LightGBM histogram algorithm to condense time-varying delay features from minute-level multivariate time series [10]. In another work, a multivariate time-series analysis and convolutional neural network (MVTS-CNN) was established to extract key features from prematched multivariate time series [11]. These methods provided a preliminary approach to incorporating unlabeled data into f-CaO modeling.

In recent years, semisupervised learning has gained increasing attention as a strategy to address the class imbalance between labeled and unlabeled data [12], [13]. In f-CaO soft sensing tasks, semisupervised learning has been used to promote the efficiency of unlabeled data utilization [5], [14], [15]. In the work [5], a dynamic feature extraction (DFE) semisupervised network has been proposed for objective-driven DFE using an LSTM autoencoder and a multilayer perceptron (MLP) for supervised training. Wasserstein generative adversarial networks (WGANs) have also been employed for data augmentation to extend the dataset for f-CaO content modeling [14], [15].

The development of data-driven modeling methods has facilitated progress in the performance of f-CaO soft sensors, albeit with some remaining issues, such as the problem of explainability. Since the existing works do not provide an explanation for the intermediate outcomes of their approaches, like the latent feature extracted by the models that could aid in monitoring the clinker burning process, the model interpretability can be an issue lying to be solved.

### B. Mechanistic Modeling for f-CaO Soft Sensing

Besides model interpretability, another issue in f-CaO soft sensing is the monotony source of the model input, as the

strategies presented above rely exclusively on the information provided by process data as input to the model. This can become a bottleneck in further improving the performance of the f-CaO soft sensors. In fact, the model's performance can be further elevated by introducing the process mechanism and domain expert knowledge by enhancing the input data [16], modifying the model's loss function [17], or integrating the output of both the mechanistic and data-driven models [18].

As previously mentioned, the temperature inside the rotary kiln has a significant impact on the f-CaO content in cement clinker. However, due to the rotary kiln's structure and harsh operating environment, the kiln temperature cannot be obtained directly. Therefore, various estimation methods were proposed, including multidimensional numerical simulation [19], [20] or 1-D heat transfer models [21], [22]. However, the computation process of these methods is relatively complex and cannot provide real-time responses at the scale of seconds. Accordingly, there emerges a need to develop a real-time calculation method for the clinker temperature, so that it can be used in the soft sensing of f-CaO content in cement clinker.

### C. Flame IFE for f-CaO Modeling

Flame images obtained at the head of the rotary kiln, which are initially used for kiln state monitoring, can also help better predict the f-CaO content. In the early stages, there was an attempt to use multivariate image analysis (MIA), principal component analysis (PCA), and scale-invariant feature transform (SIFT) operators to extract features from flame images and then reduce their dimensionality using a kernel partial least-squares (KPLSs) algorithm. The reduced image features were then fed into random vector functional-link (RVFL) networks [23] or a decorrelated neural-net ensemble (DNNE) [24] for f-CaO prediction. While these image feature extraction (IFE) techniques are somewhat effective, considering that the flame pattern changes with the clinker burning system's operating conditions, the parameters of these methods need to be constantly adjusted, which could bring much inconvenience to the plant operators.

With the recent advances in neural network-based image-processing techniques, automatically learning important features the flame image contains has become possible [25], [26]. In FeO content modeling, a multisource information fusion autoformer (MIF-Autoformer) was proposed to fuse image information with process data by training a squeeze-and-excitation-module-based residual network (SE-ResNet) for IFE [3]. This approach has been successful in the domain of industrial multisource data fusion. However, it should be noticed that the IFE and FeO content soft sensing are carried out using separate models, which may increase the approach's overall training complexity. Furthermore, the distinct evaluation metrics for different models may deviate from the training direction of the models and the approach's ultimate goal. To overcome the aforementioned problems, an end-to-end modeling method would be highly beneficial.

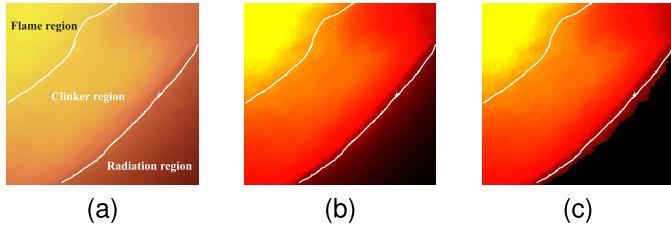


Fig. 2. Comparison of the flame image before and after preprocessing. (a) Original flame image and area division. (b) Flame image after image enhancement. (c) Flame image after area segmentation.

### III. PRELIMINARIES

#### A. Multisource Data Acquisition

The cement clinker burning process data is collected through sensors deployed at various locations in the system. Due to the complexity of the f-CaO formation mechanism, interactions between process variables, data noise, and time-varying delays, there is a high degree of nonlinearity between f-CaO content and process variables. Additionally, flame images are captured using a charge-coupled device (CCD) industrial camera as shown in Fig. 1. To avoid data redundancy, the acquired multisource data is downsampled at the interval of 1 min. The f-CaO content labels are obtained by mixing and sampling the cement clinker discharged every hour, which is then analyzed using automatic testing equipment. Accordingly, each batch of clinker samples contains information about the burning process over the past whole hour.

#### B. Flame Image Preprocessing

The captured flame images contain certain noises, which are mainly caused by the flying ash and dust emitted during the calcination process. These noises have a significant impact on flame image information extraction and further affect the effectiveness of downstream tasks [27]. Based on existing research [28], [29] and expert knowledge, the rotary kiln flame image is divided into three regions: the radiation region, the clinker region, and the flame region, as shown in Fig. 2(a). The radiation region is mainly generated by visible light scattering in the flying ash and contains much redundant information. The clinker region mainly indicates the long-term state of calcinating, while the flame region reflects a more dynamic combustion status, and both regions contain the main information of the clinker burning process. To enhance important regions while removing excessive ones, image transformation and segmentation methods are used to preprocess the flame images. First of all, the flame image is filtered using the median filtering method to reduce particle noises. Next, a transformation equation is employed to enhance the image

$$f(x, y, c) = sI(x, y, c)^\gamma. \quad (1)$$

In this equation,  $I(x, y, c)$  denotes the scaled pixel intensity at the location  $(x, y, c)$ , where  $x$ ,  $y$ , and  $c$  represent the horizontal, vertical position, and the channel of this pixel, respectively. The  $s$  and  $\gamma$  denote the linear and nonlinear transformation coefficients, which alter the pixel intensity's distribution. For image enhancement,  $s$  is set above 1, while

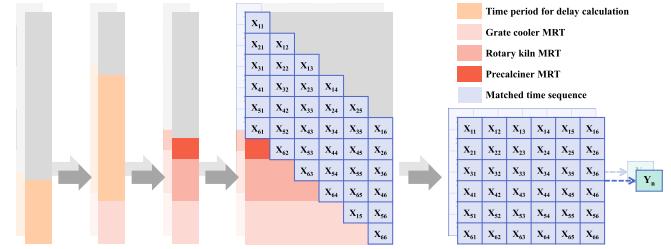


Fig. 3. Time sequence matching procedure.

$\gamma$  is below 1. Fig. 2(b) shows the enhanced flame image after image transformation, and it can be noticed that most of the radiation region is removed. To further clean up the radiation region, the image's gray level  $I(x, y)$  is calculated according to [30]

$$I(x, y) = 0.30R(x, y) + 0.59G(x, y) + 0.11B(x, y) \quad (2)$$

where  $R(x, y)$ ,  $G(x, y)$ , and  $B(x, y)$  denote the red, green, and blue components of the pixel at  $(x, y)$ . Finally, threshold segmentation is performed based on the image's gray level

$$f(x, y, c) = \begin{cases} I(x, y, c), & \text{if } I(x, y) \geq \text{th} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Here, th stands for the threshold to distinguish radiation and clinker region. The flame image after area segmentation is shown in Fig. 2(c). Compared with Fig. 2(a), the preprocessed flame image contains more prominent features, which facilitates the subsequent feature extraction tasks.

#### C. Time Sequence Matching

To address the time-varying delays between f-CaO content and process variables, a time sequence matching approach is proposed to dynamically match multisource data sequences with corresponding f-CaO content. First, to determine the time delay caused by the rotary kiln, the average material bed height  $h$  is calculated using Kramer's and Croockewit's methods [31]

$$h = \frac{3 \tan \phi}{4\pi n} V_s [R_i^2 - (h - R_i)^2]^{-3/2} - \frac{\tan \beta}{\cos \phi} \quad (4)$$

where  $R_i$  denotes the inner radius of the rotary kiln, m,  $\beta$ , and  $\phi$  are the rotary kiln's tilt angle, rad, and repose angle, rad, respectively. These parameters usually maintain stability during the burning process. In addition,  $V_s$  stands for the volume flow of material ( $\text{m}^3/\text{min}$ ),  $n$  stands for the kiln speed ( $\text{r}/\text{min}$ ), and  $\theta$  is the central angle of the material bed (rad)

$$\theta = 2\arccos((R_i - h)/R_i). \quad (5)$$

After obtaining average bed height  $h$  (m), the residence time  $\tau_k$  of the material inside rotary kiln (min) can be expressed as follows:

$$\tau_k = \frac{\rho_s R_i^2 (\theta - \sin \theta) L_k}{2M_s} \quad (6)$$

where  $\rho_s$  denotes the material density ( $\text{kg}/\text{m}^3$ ),  $M_s$  is the material's mass flow ( $\text{kg}/\text{min}$ ), and  $L_k$  represents the length of the rotary kiln (m). For time delay in the grate cooler, a simple empirical expression is established between the

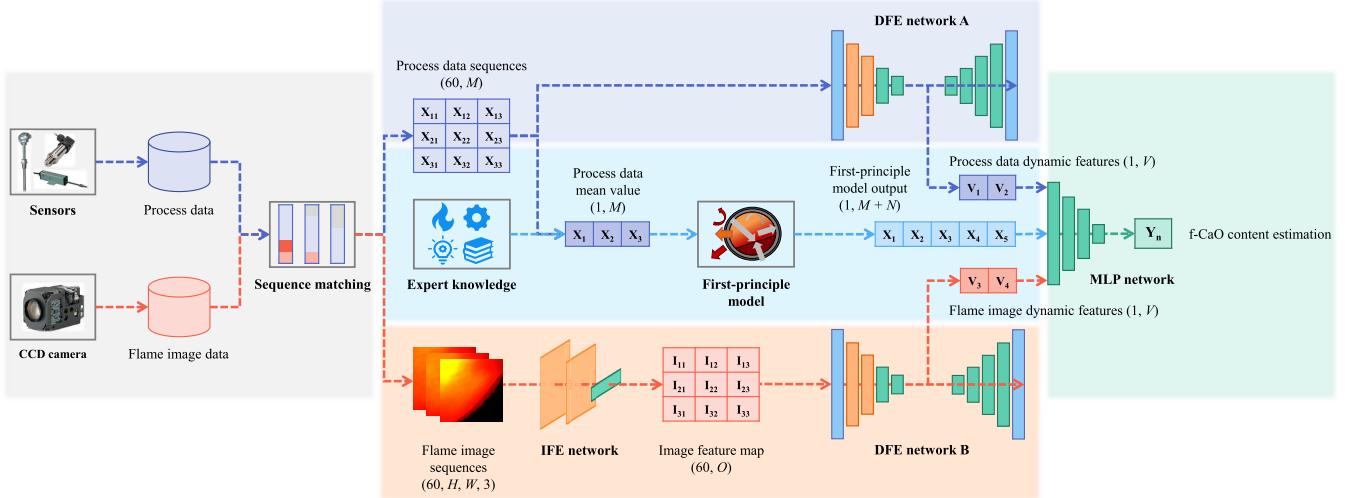


Fig. 4. Structure and process flow of the MSIFF.

material retention time  $\tau_c$  (min) and the grate cooler speed  $U_c$  (r/min)

$$\tau_c = \frac{(\bar{U}_c \bar{\tau}_c)}{U_c} \quad (7)$$

where  $\bar{U}_c$  and  $\bar{\tau}_c$  denote the cooler's rated speed (r/min) and the retention time under the rated speed (min), respectively. Based on the above time lag calculation method, the timestamps of different process variables are matched with the sampling time of the same batch of clinker through the following steps.

- 1) Starting with the clinker sampling time, the average grate cooler speed within the past half hour is calculated to obtain the clinker's mean residence time (MRT) in the grate cooler, which is used to determine when the clinker enters the grate cooler.
- 2) Taking the grate cooler inlet time as the starting point, the time when the material enters the rotary kiln is determined by calculating the clinker's MRT in the rotary kiln during the past hour.
- 3) To consider the movement of the raw material inside the precalciner, the precalciner inlet time is set 1 min ahead of the rotary kiln inlet time.

For f-CaO soft sensing, process data sequences, as well as flame image series are extracted starting from the matched timestamps. The length of time sequences is determined to 1 h to maximize the information contained in the time sequences for each batch of samples. The overall procedure of time sequence matching is illustrated in Fig. 3.

#### IV. END-TO-END MSIFF FOR f-CaO CONTENT SOFT SENSING

Fig. 4 shows the structure and process flow of the MSIFF. First, process data and flame images from databases are segmented into time sequences using the above-mentioned sequence matching method. Following this, key features are extracted from both types of sequences. For process data sequences, a first-principle model is established to generate mechanistic features that reflect the average status of the

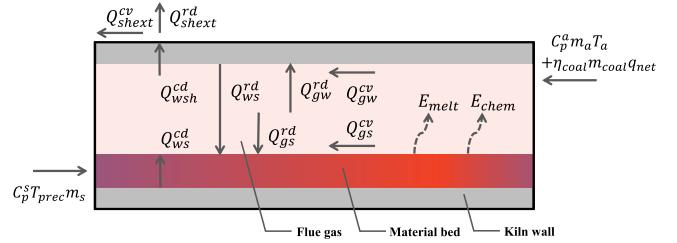


Fig. 5. Heat balance of the rotary kiln.

rotary kiln burning process during the past hour. In addition, a DFE network is developed to extract dynamic features from process data sequences. For flame image sequences that have a different structure from process data sequences, an IFE network is built to preliminarily generate multivariate image feature series from the flame image sequences. The IFE network is then connected with a second DFE network to form an end-to-end image DFE network, outputting image dynamic features. Finally, mechanistic features, dynamic features, and mean level of process data are concatenated and input into an MLP network to estimate the f-CaO content. The MSIFF's components are further detailed below.

##### A. First-Principle Model of Cement Rotary Kiln

The first-principle model takes data of  $M$  process variables as the input and output  $N$  mechanistic features as well as the mean value of process variables. The modeling of the cement rotary kiln mainly includes three steps: To begin with, the structural parameters, like the material bed height and central angle, are evaluated to calculate various heat transfer areas. Then, a macroscopic mechanistic model is constructed by integrating the energy balance among heat transfer, physical transformation, and chemical reactions within the rotary kiln, which is illustrated in Fig. 5. Finally, significant mechanistic features, such as the material bed temperature, are obtained through an iterative computation approach. Since the purpose of modeling rotary kiln is to obtain the average condition of clinker burning over the hour, the established model is steady-state, neglecting time-varying processes to achieve

TABLE I  
NOMENCLATURE FOR FIRST-PRINCIPLE MODELING

Symbol	Description, unit	Symbol	Description, unit
$A_{cws}$	Conductive heat transfer area (HTA) between inner wall and material, $\text{m}^2$	$T_{ext}$	External environment temperature, K
$A_g$	Flue gas cross-sectional area, $\text{m}^2$	$T_g$	Flue gas temperature, K
$A_{gs}$	HTA between flue gas and material, $\text{m}^2$	$T_{prec}$	Precalciner outlet temperature, K
$A_{gw}$	HTA between flue gas and inner wall, $\text{m}^2$	$T_s$	Material temperature, K
$A_{rsw}$	Radioactive HTA from material to inner wall, $\text{m}^2$	$T_{sh}$	Shell temperature, K
$A_{rws}$	Radioactive HTA from inner wall to material, $\text{m}^2$	$T_w$	Inner wall temperature, K
$A_{sh}$	HTA between shell and environment, $\text{m}^2$	$e_{melt}$	Latent melting heat, $\text{kJ/kg}$
$C_p^a$	Secondary wind specific heat capacity, $\text{kJ}/(\text{kg} \cdot \text{K})$	$h_{gs}$	Convective HT coefficient between flue gas and material, $\text{W}/(\text{m}^2 \cdot \text{K})$
$C_p^s$	Material specific heat capacity, $\text{kJ}/(\text{kg} \cdot \text{K})$	$h_{gw}$	Convective HT coefficient between flue gas and inner wall, $\text{W}/(\text{m}^2 \cdot \text{K})$
$D_i$	Kiln inner diameter, m	$h_{ncv}$	Natural convective HT coefficient of shell, $\text{W}/(\text{m}^2 \cdot \text{K})$
$D_o$	Kiln outer diameter, m	$k_{comp}$	Composite conductivity between inner wall and shell, $\text{W}/(\text{m}^2 \cdot \text{K})$
$E_{chem}$	Chemical reaction heat, kJ	$k_{ws}$	Conductivity between inner wall and material, $\text{W}/(\text{m}^2 \cdot \text{K})$
$E_{melt}$	Material melting heat, kJ	$m_a$	Secondary wind inlet mass, kg
$\Delta H_i^*$	$i$ -th chemical component reaction heat, $\text{kJ/mol}$	$m_{coal}$	Coal feeding mass, kg
$Q_{ws}^{cd}$	Conductive heat transfer (HT) between inner wall and material, kJ	$m_s$	Material feeding mass, kg
$Q_{wsh}^{cd}$	Conductive HT between inner wall and shell, kJ	$q_{net}$	Low heating value of feeding coal, $\text{kJ/kg}$
$Q_{gs}^{cv}$	Convective HT between flue gas and material, kJ	$r_i$	number of moles participated in the reaction, mol
$Q_{gw}^{cv}$	Convective HT between flue gas and inner wall, kJ	$u_g$	Flue gas velocity (time dimension removed), m
$Q_{shext}^{cv}$	Convective HT between shell and environment, kJ	$\alpha_g$	Flue gas absorptivity, /
$Q_{gs}^{rd}$	Radioactive HT from flue gas to material, kJ	$\epsilon_g$	Flue gas emissivity, /
$Q_{gw}^{rd}$	Radioactive HT from flue gas to inner wall, kJ	$\epsilon_s$	Material emissivity, /
$Q_{shext}^{rd}$	Radioactive HT from shell to environment, kJ	$\epsilon_{sh}$	Shell emissivity, /
$Q_{ws}^{rd}$	Radioactive HT from inner wall to material, kJ	$\epsilon_w$	Inner wall emissivity, /
$T_L$	Material liquidus temperature, K	$\eta_{coal}$	Coal combustion rate, /
$T_S$	Material solidus temperature, K	$\rho_g$	Flue gas density, $\text{kg}/\text{m}^3$
$T_a$	Secondary air temperature, K	$\sigma$	Stefan-Boltzmann constant, /

better real-time performance. To improve the clarity of the following mechanistic equations, a nomenclature for first-principle modeling is provided in Table I.

The heat transfer within the cement rotary kiln can be mainly divided into three categories: conduction, convection, and radiation. Conduction refers to the heat transfer through direct contact between the kiln's inner wall and material bed as well as the kiln shell, which can be expressed as follows:

$$Q_{ws}^{cd} = k_{ws} A_{cws} (T_w - T_s) \quad (8)$$

$$Q_{wsh}^{cd} = \frac{2\pi k_{comp} (T_w - T_{sh})}{\ln\left(\frac{D_o}{D_i}\right)}. \quad (9)$$

Convection includes forced convection and natural convection, where forced convection represents the transfer of heat through the movement of the flue gas driven by the inlet secondary wind, and natural convection is caused by the temperature gradient between the kiln shell and external environment

$$Q_{gs}^{cv} = h_{gs} A_{gs} (T_g - T_s) \quad (10)$$

$$Q_{gw}^{cv} = h_{gw} A_{gw} (T_g - T_w) \quad (11)$$

$$Q_{shext}^{cv} = h_{ncv} A_{sh} (T_{sh} - T_{ext}). \quad (12)$$

Finally, radiation represents the heat transfer by electromagnetic waves, such as light or infrared radiation. It takes place

not only between the inner wall and the material bed, but also between the flue gas and the other parts inside the rotary kiln

$$Q_{gs}^{rd} = \sigma (\epsilon_s + 1) A_{gs} \frac{\epsilon_g T_g^4 - \alpha_g T_s^4}{2} \quad (13)$$

$$Q_{gw}^{rd} = \sigma (\epsilon_w + 1) A_{gw} \frac{\epsilon_g T_g^4 - \alpha_g T_w^4}{2} \quad (14)$$

$$Q_{ws}^{rd} = \frac{\sigma (T_w^4 - T_s^4)}{\frac{(1-\epsilon_w)}{\epsilon_w} A_{rsw} + \frac{1}{A_{rws}} + \frac{(1-\epsilon_s)}{\epsilon_s} A_{rws}} \quad (15)$$

$$Q_{shext}^{rd} = \epsilon_{sh} A_{sh} \sigma (T_{sh}^4 - T_{ext}^4). \quad (16)$$

Furthermore, the physical and chemical reaction of the material can be described by referring to [21]

$$E_{melt} = m_s e_{melt} \left( \max \left[ 0, \frac{T_s - T_S}{T_L - T_S} \right] \right) \quad (17)$$

$$E_{chem} = m_s \sum_{i=1}^n \Delta H_i^* r_i. \quad (18)$$

To obtain the temperature within the rotary kiln and other mechanistic features, the following equations are established using mass and energy balances. In terms of mass balance, the balance between the mass of air entering and leaving the rotary kiln is studied, and the equation is given as follows:

$$m_a = A_g \rho_g u_g - \eta_{coal} m_{coal}. \quad (19)$$

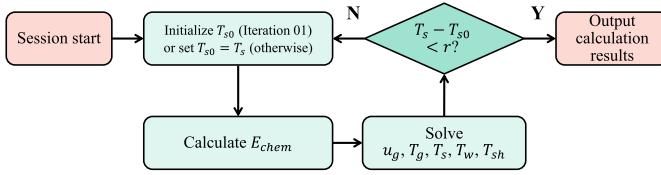


Fig. 6. First-principle model solution process.

In the above equation,  $u_g$  is an unknown variable to be solved in the equation system. Moreover, the energy balance of the rotary kiln can be expressed using the following equations:

$$m_a c_p^a T_a + \eta_{\text{coal}} m_{\text{coal}} q_{\text{net}} = Q_{\text{gs}}^{\text{cv}} + Q_{\text{gw}}^{\text{cv}} + Q_{\text{gs}}^{\text{rd}} + Q_{\text{gw}}^{\text{rd}} \quad (20)$$

$$C_p^s (T_s - T_{\text{prec}}) m_s + E_{\text{melt}} + E_{\text{chem}} = Q_{\text{gs}}^{\text{cv}} + Q_{\text{gs}}^{\text{rd}} + Q_{\text{ws}}^{\text{cd}} + Q_{\text{ws}}^{\text{rd}} \quad (21)$$

$$Q_{\text{gw}}^{\text{rd}} + Q_{\text{gw}}^{\text{cv}} = Q_{\text{ws}}^{\text{cd}} + Q_{\text{ws}}^{\text{rd}} + Q_{\text{wsh}}^{\text{cd}} \quad (22)$$

$$Q_{\text{wsh}}^{\text{cd}} = Q_{\text{shext}}^{\text{cv}} + Q_{\text{shext}}^{\text{rd}}. \quad (23)$$

The solution process of the model is shown in Fig. 6. According to the chemical reaction mechanism, the material temperature  $T_s$  and chemical reaction heat  $E_{\text{chem}}$  are mutually dependent, and solving the  $T_s$  and  $E_{\text{chem}}$  together is usually infeasible. Therefore, an iterative computation strategy is adopted. First, an initial value of material temperature  $T_{s0}$  is set, which is 1600 K in this work, and  $E_{\text{chem}}$  of the material is preliminarily solved. Then, by utilizing (19)–(23), the equation system is established to solve the rotary kiln temperature. Next,  $E_{\text{chem}}$  is evaluated one more time using the obtained rotary kiln temperature, and the rotary kiln temperature is solved again. This process is repeated iteratively until the difference between  $T_{s0}$  and solved  $T_s$  is within an acceptable range  $r$ . The calculation result of the last iteration is regarded as the output of the first-principle model.

### B. Multisource Dynamic Feature Extraction

DFE utilizes unlabeled minute-level data to produce hour-level dynamic information, which plays an important role in building the f-CaO soft sensors. In MSIFF, dynamic features are extracted from both process data and flame image sequences. Since the structure of flame images is different from process data, based on the tradeoff between model complexity and flame image characteristics, a 2-D convolution IFE network is developed to preliminarily transform  $(60, H, W, 3)$  flame image sequences into  $(60, O)$  multivariate image feature series, as shown in Fig. 7. The IFE network is mainly consisted of two basic convolution blocks, which are formed with convolution, batch normalization (BN), activation (ReLU), and pooling layers. The convolution layer is the main component of the IFE network, whose main function is to extract features from input images by performing sliding window convolutions with a trainable 2-D convolution kernel

$$\mathbf{M}_{i,j,k_{\text{out}}} = \sum_{m=1}^M \sum_{n=1}^N \sum_{k_{\text{in}}=1}^{k_{\text{in}}} \mathbf{C}_{m,n,k_{\text{in}},k_{\text{out}}} \mathbf{I}_{i+m, j+n, k_{\text{in}}} \quad (24)$$

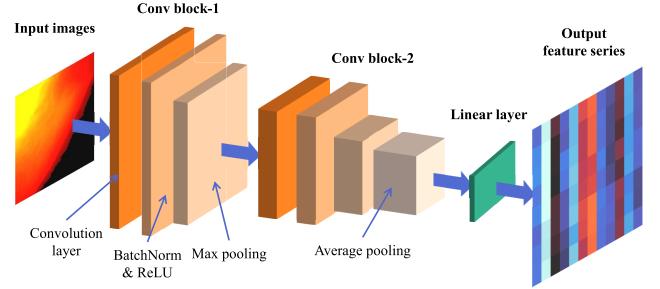


Fig. 7. IFE network for flame IFE.

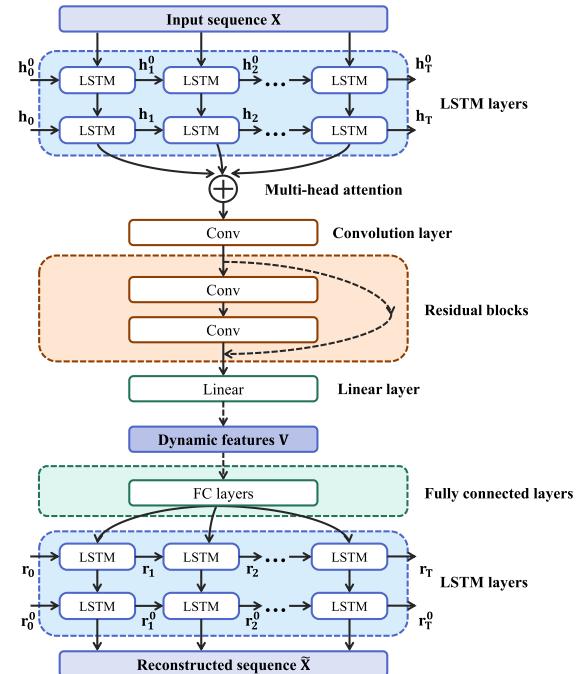


Fig. 8. Architecture of the proposed DFE network.

where  $\mathbf{I}$  represents the input image,  $\mathbf{M}$  represents the output feature map,  $\mathbf{C}$  denotes the convolution kernel,  $(i, j)$  is the pixel position on the output feature map,  $(m, n)$  is the position on the convolution kernel, and  $k_{\text{in}}$  and  $k_{\text{out}}$  are the channels of the input image and output feature map, respectively. The normalization layer and activation layer are used to accelerate the convergence while improving the stability of the network, and the pooling layer is employed to downsample the feature maps. After being processed by two basic convolution blocks, the feature maps are flattened and passed through a linear layer to produce multivariate image feature series, which have the same dimension as process data sequences.

After that, a DFE network is proposed to extract dynamic features from the multisource data sequences, whose architecture is illustrated in Fig. 8. The DFE network takes the  $(60, P)$  data sequence as the input, where  $P$  is the feature dimension of the input sequence, and outputs dynamic feature with the shape  $(1, V)$ .

First, the data sequence is processed by the LSTM layers through the functions within each LSTM unit: the input gate  $\mathbf{i}_t$ , forget gate  $\mathbf{f}_t$ , output gate  $\mathbf{o}_t$ , candidate cell gate  $\tilde{\mathbf{c}}_t$ , cell state  $\mathbf{c}_t$ , and hidden state  $\mathbf{h}_t$ . The working principle of each

LSTM unit can be represented below

$$\begin{cases} \mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \\ \mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \\ \mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \\ \tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \\ \mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \\ \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \end{cases} \quad (25)$$

where  $\mathbf{h}_{t-1}$  and  $\mathbf{c}_{t-1}$  are the previous hidden state and cell state of the LSTM cell, respectively,  $\mathbf{W}_i$ ,  $\mathbf{b}_i$ ,  $\mathbf{W}_f$ ,  $\mathbf{b}_f$ ,  $\mathbf{W}_o$ ,  $\mathbf{b}_o$ ,  $\mathbf{W}_c$ , and  $\mathbf{b}_c$  are the weights and biases of each function,  $\sigma$  denotes the Sigmoid activation function, and  $\odot$  denotes the Hadamard product [5]. In this work, the hidden states outputted from the last LSTM layer at each time step of the sequence are regarded as the dynamic information of the input sequence. Therefore, the shape of the LSTM layers' output is  $(60, V)$ , where  $V$  stands for the dimension of the compressed dynamic features. The mapped dynamic information series are then weighted through the multihead self-attention mechanism to focus on specific parts of the series that are potentially important to downstream tasks

$$f(\mathbf{X}) = \mathbf{W}_{\text{att}} \cdot \text{concat}[\text{head}_1, \dots, \text{head}_h] \quad (26)$$

where  $\mathbf{W}_{\text{att}}$  denotes a trainable weight and  $\text{head}_i$  is the weighted vector outputted by a single attention head

$$\text{head}_i = \text{softmax}\left(\frac{\mathbf{X}\mathbf{X}^T}{\sqrt{\mathbf{d}_k}}\right)\mathbf{X} \quad (27)$$

where  $\mathbf{d}_k$  is the dimension of the input sequence  $\mathbf{X}$ . After applying the attention mechanism, the dynamic information is further condensed using a residual convolution network (ResNet) [32] that is mainly composed of a convolution layer followed by a series of residual blocks and ends with a linear layer. In this network, each dimension of the series input is treated as a channel [11], and a set of single-dimensional convolution kernels are applied to process the input series. Besides, residual connections are introduced to address the vanishing gradient problem during the application of a deeper network. As shown in Fig. 9, the residual blocks are formed with two convolution layers connected by normalization and activation layers. Given the input  $\mathbf{x}$ , the residual block output  $\mathbf{U}(\mathbf{x})$  is

$$\mathbf{U}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) + \mathbf{x} \quad (28)$$

where  $\mathbf{F}(\mathbf{x})$  represents the residual mapping, which can be expressed as  $\mathbf{F}(\mathbf{x}) = \mathbf{U}(\mathbf{x}) - \mathbf{x}$ . To output  $\mathbf{U}(\mathbf{x})$ ,  $\mathbf{x}$  is downsampled before adding with  $\mathbf{F}(\mathbf{x})$ , and the two-convolution-layer form is adopted to prevent the residual connection from degrading to a linear layer [3]. The output of the DFE network is a  $(1, V)$  array that contains the further compressed dynamic information, which is taken as the dynamic feature of the input  $(60, P)$  sequence. Finally, the input sequence is reconstructed by a group of fully connected layers and symmetric LSTM layers from the extracted dynamic features.

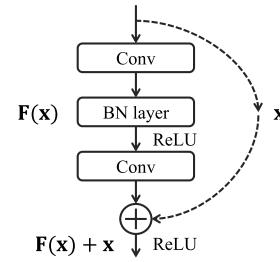


Fig. 9. Structure of the residual block.

### C. Multisource Information Fusion and f-CaO Modeling

To fuse the information provided by the process data and image sequences as well as mechanistic knowledge and, at the same time, utilize the multisource information for f-CaO soft sensing, the DFE networks, along with the first-principle model, IFE network, and an MLP network form the MSIFF. Before implementation, the MSIFF needs to be trained beforehand. During the training, the IFE network, the DFE networks, and an MLP network are connected together to form the MSDFE network. The training of the network adopts a semisupervised learning paradigm, and its loss function is defined as follows:

$$\begin{aligned} \mathbf{L} = & \frac{\lambda_1}{B \times T \times M} \sum_{m=1}^B \sum_{t=1}^T \sum_{n=1}^M (\mathbf{X}_{t,n}^m - \tilde{\mathbf{X}}_{t,n}^m)^2 \\ & + \frac{\lambda_2}{B \times T \times O} \sum_{m=1}^B \sum_{t=1}^T \sum_{n=1}^O (\mathbf{I}_{t,n}^m - \tilde{\mathbf{I}}_{t,n}^m)^2 \\ & + \frac{\lambda_3}{B} \sum_{m=1}^B (\mathbf{Y}^m - \hat{\mathbf{Y}}^m)^2. \end{aligned} \quad (29)$$

In this loss function,  $B$  represents the batch size used during training, and  $T$  represents the length of the process data and image feature sequences, which is set to 60 based on the previous design. The variables  $\mathbf{X}_{t,n}^m$ ,  $\mathbf{I}_{t,n}^m$ ,  $\tilde{\mathbf{X}}_{t,n}^m$ , and  $\tilde{\mathbf{I}}_{t,n}^m$  refer to the input and reconstructed process data and image feature values for the  $m$ th batch, at time  $t$  for feature  $n$ , respectively. Besides,  $\mathbf{Y}^m$  and  $\hat{\mathbf{Y}}^m$  represent the true and estimation values of the f-CaO content, respectively. Finally,  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  denote the regularization parameters. The network is then trained with the Adam optimizer [33] to minimize this loss function.

## V. EXPERIMENTS AND RESULTS

### A. Experiment Configurations

The process data and flame images for the experiment are obtained from a real cement production line. Within the duration of 2 months, 1003 clinker samples are analyzed, and the process data and flame image sequences are matched to correspond with the f-CaO content. For process data, 28 process variables that could impact the f-CaO content are selected, including secondary and tertiary air properties, measurable material temperatures, and various controllable parameters of the cement clinker burning system. The preprocessing methods mentioned above are used for image enhancement and segmentation for flame images. After that, the 1003 data points are divided into 602 training data, 201 validation data, and 200 testing data in chronological

TABLE II  
PARAMETERS FOR TRAINING THE MSIFF

Components	Architectures	Hyper-parameters
IFE network	Conv $\times 2$ , FC $\times 1$	$O=12$ , $V=4$ , $\lambda_1=0.3$ ,
DFE networks	(LSTM $\times 3$ , Conv, Residual block $\times 3$ , FC $\times 1$ ) +(FC $\times 4$ , LSTM $\times 3$ )	$\lambda_2=0.7$ , $\lambda_3=1$ , Batch size $bs=24$ , Epoch $ep=100$ , Learning rate $lr=0.001$ , Dropout rate $dr=0.3$
MLP network	FC $\times 6$ with dropouts	

order. The scale of our dataset is close to the datasets utilized in previous f-CaO prediction studies [5], [9], [11] to ensure the comparability and consistency of our work and represent a more common dataset size in f-CaO content modeling. Then, the MSIFF is developed using the PyTorch framework with detailed parameters listed in Table II. All the methods are trained and tested on a computer with an Apple M1 Pro silicon, 16GB RAM.

Additionally, the PLS [7], BP-NN [8], ESVM [9], MVTS-CNN [11], and DFETM [5] methods are replicated to compare with MSIFF on f-CaO soft sensing task. The PLS, BP-NN, and ESVM methods use the hour average value of the process data to model the f-CaO content, which is the traditional approach for f-CaO soft sensing. The MVTS-CNN takes the process data sequences as the model input, while the DFETM utilizes both the hour average value and data sequences of the process variable to develop the f-CaO soft sensor.

### B. First-Principle Modeling Results

Before training the MSDFE network, the hour average of the process variables is calculated and input into the first-principle model to obtain the average levels of the rotary kiln temperature and other mechanistic features. To assess the reliability of the model's output, the energy input and output of the rotary kiln are calculated, and the energy balance error of the rotary kiln is then used as the criterion for modeling accuracy since the true value of the rotary kiln temperature is unavailable. The energy input of the rotary kiln consists of the heat carried by the material input from the precalciner and the secondary air entering the inlet, as well as the heat released during coal combustion. On the other hand, the energy output includes the heat carried by the output clinker and the heat discharged from the kiln tail gas. Accordingly, the energy balance error  $e_k$  is defined as follows:

$$\begin{aligned} E_{\text{in}} &= m_a C_p^a T_a + \eta_{\text{coal}} m_{\text{coal}} q_{\text{net}} + C_p^s T_{\text{prec}} m_s \\ E_{\text{out}} &= C_p^g T_{\text{tail}} (m_a + \eta_{\text{coal}} m_{\text{coal}}) - C_p^s T_s m_s \\ e_k &= \left| \frac{E_{\text{in}} - E_{\text{out}}}{E_{\text{out}}} \right| \times 100\%. \end{aligned} \quad (30)$$

In the above equations,  $T_{\text{tail}}$  represents the temperature of the tail gas, which is measured by a thermal coupe. This temperature is not involved in the rotary kiln modeling but is only used when measuring  $e_k$  to ensure the fairness of the model evaluation. To protect data privacy, the displayed temperature data is subtracted from its absolute value to generate a data trend. The trend of the calculated clinker

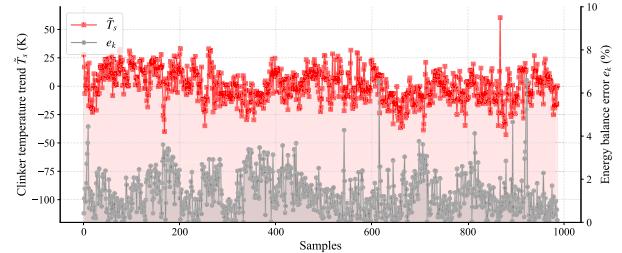


Fig. 10. First-principle modeling results of clinker temperature trend  $\tilde{T}_s$  and the energy balance error  $e_k$ .

temperature  $\tilde{T}_s$  and the error  $e_k$  is shown in Fig. 10. Generally, an energy balance error below 5% is acceptable in industrial device modeling. In this experiment, the 1003 measurement points' average  $e_k$  is 1.31%, indicating a good energy balance accuracy of the proposed first-principle model.

### C. Process Data DFE Results

To explain the dynamic features extracted by the MSDFE network from process data sequences,  $V$  dynamic features are summarized into one using a PCA method, which allows for the identification of the system's operating condition by obtaining the most representative portion of the dynamic features. After that, each PCA output value is matched with its corresponding process data sequence. The process data sequences are sorted in ascending order based on the PCA output values, reflecting the properties of the process data sequences under different operating conditions.

The visualization of the data trend sequences for the rotary kiln's coal feeding rate, the clinker burning system's SNCR ammonia flow rate, and the kiln speed under three operating conditions are presented in Fig. 11. The coal feeding rate directly impacts the amount of energy input of the rotary kiln and affects the burning status at the head of the rotary kiln. The ammonia is used to remove the nitrogen oxides ( $\text{NO}_x$ ) generated during the calcination process and is influenced by multiple parameters of the kiln tail state that reflect the state of the tail of the rotary kiln. The kiln speed affects the motion state of the material, which can reveal the physical changes of the material within the rotary kiln.

As shown in Fig. 11, under different operating conditions identified by the MSDFE network, the shapes of the three process variable sequences vary significantly, indicating obvious differences in dynamic information. In operating condition 1, the coal feeding rate is relatively stable, and the system mainly concentrates on altering the kiln speed. In operating condition 3, the coal feeding rate changes greatly within an hour, resulting in an unstable burning status, which further leads to the frequent adjustment of the ammonia flow rate while reducing the kiln speed fluctuation. Operating condition 2, whose PCA output value is the average of conditions 1 and 3, exhibits an intermediate state between the other two operating conditions. The experimental results demonstrate that the MSDFE network can effectively extract dynamic information from process data sequences, providing valuable information about the clinker burning system's operating conditions for the f-CaO soft sensing task.

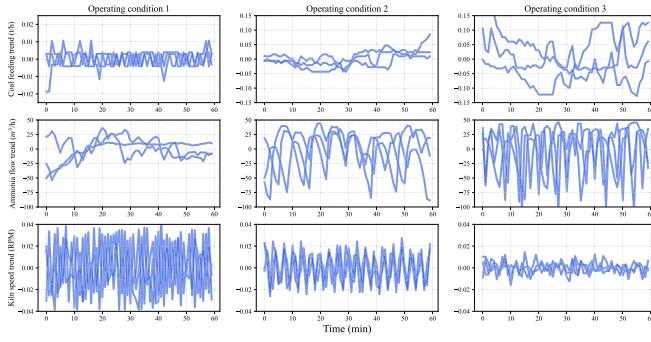


Fig. 11. Visualization of the coal feeding rate, SNCR ammonia flow rate, and kiln speed data trend sequences under different operating conditions.

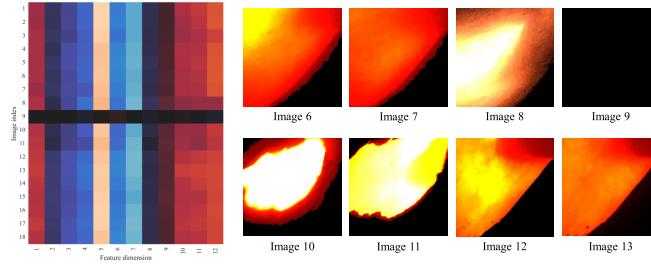


Fig. 12. Flame IFE results.

#### D. Flame Image DFE Results

Considering the end-to-end process of dynamic extraction from flame image sequences, the effectiveness of the IFE network is first analyzed. To this end, a typical flame image sequence with a length of 18 is input into the IFE network, which extracts an image feature series with a feature dimension of 12 from the input image sequence. For ease of visualization and observation, the feature series are processed into a heatmap, as illustrated in Fig. 12. By comparing the extracted features with the corresponding preprocessed images, it can be noted that the differences in flame image patterns are clearly reflected in the heatmap. This indicates that the IFE network can effectively extract important features from the flame images.

Based on the results of flame IFE, an MSDFE network is employed to further extract dynamic features from 1003 flame image sequences. To validate the effectiveness of extracted flame image dynamic features, the Spearman's correlation coefficients  $R$  between the dynamic features, and f-CaO content are calculated using the following expression:

$$R = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (31)$$

where  $x_i$  and  $y_i$  stand for the  $i$ th image dynamic feature value and f-CaO content value, respectively, and  $\bar{x}$  and  $\bar{y}$  are their mean values. When the number of dynamic features  $V$  is determined as 4, the  $R$  between each dynamic feature and f-CaO content are  $-0.19$ ,  $-0.24$ ,  $0.26$ , and  $0.21$ , respectively, demonstrating the significance of flame image information for soft sensing the f-CaO content.

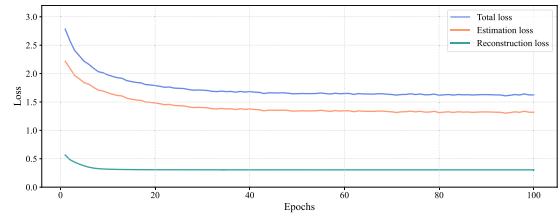


Fig. 13. Loss functions during the training process.

#### E. f-CaO Content Soft Sensing Results

To validate the performance of the proposed method in f-CaO soft sensing, the MSIFF is constructed based on the parameters listed in Table II and is trained on 602 training data and evaluated on 201 validation data. The loss functions during the training process are shown in Fig. 13. Specifically, the reconstruction loss and estimation loss are exhibited separately to better illustrate the semisupervised training process. As the training epoch increases, the level of the total loss continuously decreases and eventually converges.

After that, the MSIFF together with other replicated comparison methods is tested on 200 testing data. To quantitatively evaluate the methods' performance, the correlation coefficient  $R$  is used to assess the overall correctness of the f-CaO estimation, and the mean absolute error (MAE) and root mean square error (RMSE) are introduced to evaluate the absolute error between the true values and the estimated values. Furthermore, the metric direction precision (DP) is defined to judge the correctness of the f-CaO prediction direction between the adjacent two samples. In some downstream tasks, such as control optimization, whether the soft sensing methods can effectively determine the direction of performance indicator changes is a more important factor than whether they can correctly predict the absolute value. The calculation methods for MAE, RMSE, and DP are defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (32)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (33)$$

$$\text{DP} = (N_c/N) \times 100\%. \quad (34)$$

Here,  $y_i$  and  $\hat{y}_i$  are the  $i$ th clinker sample's real f-CaO content value and its corresponding estimated value, respectively.  $N$  denotes the total number of the clinker sample, while  $N_c$  represents the number of estimation results that correctly predict the change direction of the f-CaO content.

To avoid contingency caused by neural network training, the BP-NN, MVTS-CNN, DFETM, and MSIFF models are trained and tested 20 times, and the average performance metrics were used to evaluate their effectiveness. The test results are listed in Table III and plotted in Fig. 14. What is more, the RMSE distribution of the four neural-network-based methods is displayed in Fig. 15 using boxplots.

Among all methods, the PLS's performance is the worst due to its limited ability to handle nonlinear relationships between f-CaO content and process variables. Despite careful

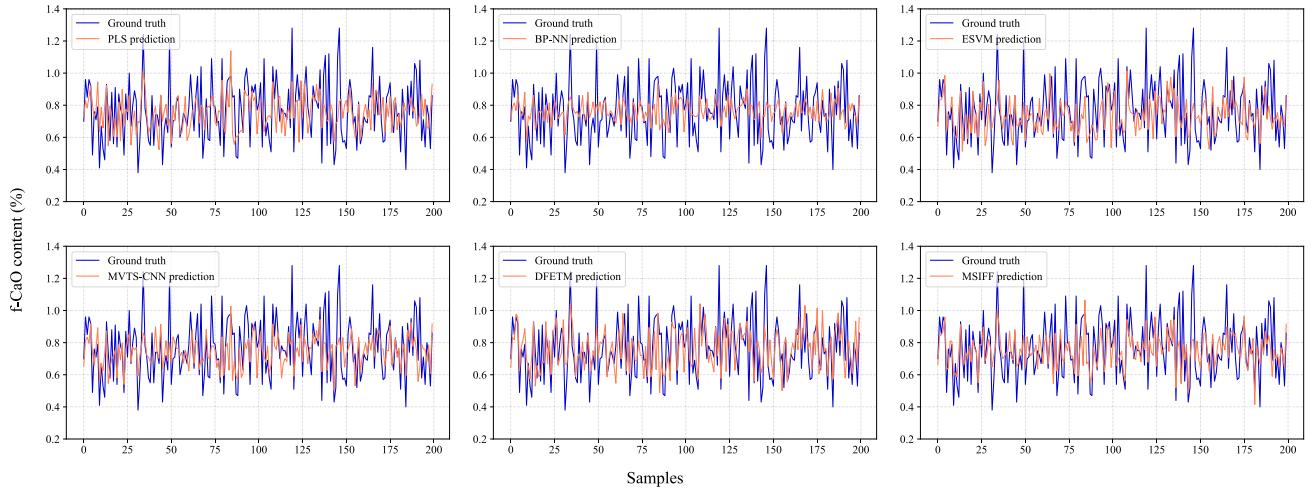


Fig. 14. f-CaO soft sensing results by different methods.

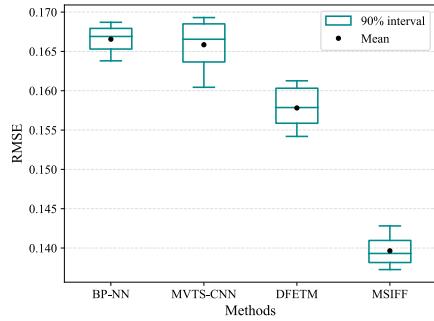


Fig. 15. RMSE distribution of the neural-network-based methods.

TABLE III  
PERFORMANCES OF DIFFERENT METHODS ON THE TESTING DATA

Methods	MAE	RMSE	R	DP
PLS	0.1360	0.1705	0.4235	69.34%
BP-NN	0.1335	0.1666	0.4784	69.72%
ESVM	0.1249	0.1584	0.5976	69.84%
MVTS-CNN	0.1308	0.1658	0.4932	68.64%
DFETM	0.1252	0.1578	0.5566	72.31%
<b>MSIFF</b>	<b>0.1089</b>	<b>0.1396</b>	<b>0.7167</b>	<b>74.15%</b>

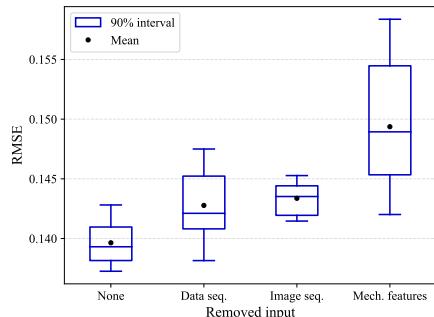


Fig. 16. RMSE distribution of the ablation experiment results.

hyperparameter tuning, the BP-NN model still exhibits underfitting during the training and prediction process, which is mainly caused by the limitation of dataset size and model complexity issues. The ESVM method achieves relatively ideal soft sensing performance by classifying operating conditions before training and utilizing a kernel function that has a

TABLE IV  
ABLATION EXPERIMENT RESULTS

Removed input	MAE	RMSE	R	DP
None	0.1089	0.1396	0.7167	74.15%
Data sequences	0.1103	0.1428	0.7172	73.59%
Image sequences	0.1105	0.1434	0.7171	73.29%
Mechanistic features	0.1153	0.1494	0.7169	72.51%

certain ability to handle nonlinear relationships. In general, the MVTS-CNN method slightly outperforms the BP-NN method. Nevertheless, as shown in Fig. 15, the network's stability is relatively weak, since the minute-level data sequences may contain a relatively high level of noise. This indicates that using process data sequences alone is not suitable in cases of high minute-level noise-to-signal ratios. The DFETM method, which integrates time averages and process data sequences for f-CaO modeling and adopts a semisupervised learning paradigm, achieves relatively satisfying results. Eventually, the MSIFF method proposed by us has significantly surpassed all the comparison methods in every aspect, while guaranteeing its network stability. This is attributed not only to the MSIFF's utilization of multisource information, but also to the end-to-end design of MSIFF's information fusion process, which is not possessed by other existing methods.

#### F. Ablation Study

To investigate the impact of each component of the multisource information on MSIFF's f-CaO soft sensing performance, the ablation experiment is conducted as follows. Considering that besides the process variable's hour average value, the additional multisource information is consisted of process data sequences, flame image sequences, and mechanistic features, each one of the information inputs is removed in turn, and the structure as well as parameters of MSIFF are adjusted accordingly. Then, the MSIFF is retrained and tested on the testing data. The experimental results are presented in Table IV, and the RMSE distribution of each circumstance is demonstrated in Fig. 16.

Generally, the MSIFF still outperforms the comparison methods when a part of the input information is absent. However, the performance degradation varies under different circumstances. When the process data sequences are removed, the performance and stability of MSIFF in predicting the f-CaO content decrease. This is because the dynamic features contained in process data sequences are absent during the MSIFF training and decision-making processes, rendering MSIFF unable to acquire information regarding the operating conditions of the clinker burning system. The importance of operating condition identification has already been demonstrated in the experimental result of the ESVM method.

Additionally, when the flame image sequences are removed, the prediction performance of MSIFF also decreases to a certain extent, while the network stability is slightly improved. This is due to the reduction in network complexity caused by the removal of the IFE network used for image processing.

As shown in Fig. 16, the removal of mechanistic features has the greatest impact on the prediction performance of MSIFF. This suggests that the exploration of mechanistic knowledge and the combination of mechanism-driven and data-driven modeling are crucial for improving the performance of the f-CaO soft measurement. In conclusion, the ablation experiment further demonstrates that each part of the multisource information contributes to the performance of f-CaO soft sensing, which is valuable and indispensable.

## VI. CONCLUSION

This article proposes a novel end-to-end MSIFF, for soft sensing the f-CaO content in the cement clinker burning process. The information provided by the burning mechanistic knowledge, matched process data, and flame image sequences within the rotary kiln is effectively explained and utilized by the MSIFF through a first-principle model and an MSDFE network. The MSIFF's performance in cement clinker f-CaO soft sensing is further verified by the data obtained from a real cement production line, which surpasses all existing f-CaO prediction methods. Moreover, the work presented in this article also suggests that the application of mechanistic knowledge and multimode data is offering promising research opportunities in industrial quality index soft sensing.

## REFERENCES

- [1] W. Zhang, A. Maleki, M. G. Khajeh, Y. Zhang, S. M. Mortazavi, and A. Vasel-Be-Hagh, "A novel framework for integrated energy optimization of a cement plant: An industrial case study," *Sustain. Energy Technol. Assessments*, vol. 35, pp. 245–256, Oct. 2019.
- [2] F. Yan, X. Zhang, C. Yang, B. Hu, W. Qian, and Z. Song, "Data-driven modelling methods in sintering process: Current research status and perspectives," *Can. J. Chem. Eng.*, vol. 101, no. 8, pp. 4506–4522, Aug. 2023.
- [3] C. Yang, C. Yang, X. Zhang, and J. Zhang, "Multi-source information fusion for autoformer: Soft sensor modeling of FeO content in iron ore sintering process," *IEEE Trans. Ind. Informat.*, early access, Feb. 23, 2023, doi: [10.1109/TII.2023.3248059](https://doi.org/10.1109/TII.2023.3248059).
- [4] L. Kong, C. Yang, S. Lou, Y. Cai, X. Huang, and M. Sun, "Collaborative extraction of intervariable coupling relationships and dynamics for prediction of silicon content in blast furnaces," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.
- [5] L. Yao et al., "Virtual sensing f-CaO content of cement clinker based on incremental deep dynamic features extracting and transferring model," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [6] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, Apr. 2009.
- [7] B. Lin, B. Recke, J. K. H. Knudsen, and S. B. Jørgensen, "A systematic approach for soft sensor development," *Comput. Chem. Eng.*, vol. 31, nos. 5–6, pp. 419–425, May 2007.
- [8] A. K. Pani, V. K. Vadlamudi, and H. K. Mohanta, "Development and comparison of neural network based soft sensors for online estimation of cement clinker quality," *ISA Trans.*, vol. 52, no. 1, pp. 19–29, Jan. 2013.
- [9] X. Liu, J. Jin, W. Wu, and F. Herz, "A novel support vector machine ensemble model for estimation of free lime content in cement clinkers," *ISA Trans.*, vol. 99, pp. 479–487, Apr. 2020.
- [10] X. Hao, Z. Zhang, Q. Xu, G. Huang, and K. Wang, "Prediction of f-CaO content in cement clinker: A novel prediction method based on LightGBM and Bayesian optimization," *Chemometric Intell. Lab. Syst.*, vol. 220, Jan. 2022, Art. no. 104461.
- [11] Y. Zhao, B. Ding, Y. Zhang, L. Yang, and X. Hao, "Online cement clinker quality monitoring: A soft sensor model based on multivariate time series analysis and CNN," *ISA Trans.*, vol. 117, pp. 180–195, Nov. 2021.
- [12] B. Shen and Z. Ge, "Supervised nonlinear dynamic system for soft sensor application aided by variational auto-encoder," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6132–6142, Sep. 2020.
- [13] Q. Sun and Z. Ge, "A survey on deep learning for data-driven soft sensors," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 5853–5866, Sep. 2021.
- [14] X. Hao, L. Liu, G. Huang, Y. Zhang, Y. Zhang, and H. Dang, "R-WGAN-based multitemplescale enhancement method for predicting f-CaO cement clinker," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.
- [15] X. Hao et al., "SSP-WGAN-based data enhancement and prediction method for cement clinker f-CaO," *IEEE Sensors J.*, vol. 22, no. 23, pp. 22741–22753, Dec. 2022.
- [16] Y. Ni, J. Xu, C. Zhu, and L. Pei, "Accurate residual capacity estimation of retired LiFePO<sub>4</sub> batteries based on mechanism and data-driven model," *Appl. Energy*, vol. 305, Jan. 2022, Art. no. 117922.
- [17] 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.
- [18] J. Yang, T. Chai, C. Luo, and W. Yu, "Intelligent demand forecasting of smelting process using data-driven and mechanism model," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9745–9755, Dec. 2019.
- [19] A. A. Boateng and P. V. Barr, "A thermal model for the rotary kiln including heat transfer within the bed," *Int. J. Heat Mass Transf.*, vol. 39, no. 10, pp. 2131–2147, Jul. 1996.
- [20] H. F. Elattar, E. Specht, A. Fouad, and A. S. Bin-Mahfouz, "CFD modeling using PDF approach for investigating the flame length in rotary kilns," *Heat Mass Transf.*, vol. 52, no. 12, pp. 2635–2648, Dec. 2016.
- [21] K. S. Mujumdar and V. V. Ranade, "Simulation of rotary cement kilns using a one-dimensional model," *Chem. Eng. Res. Des.*, vol. 84, no. 3, pp. 165–177, Mar. 2006.
- [22] T. Hanein, F. P. Glasser, and M. N. Bannerman, "One-dimensional steady-state thermal model for rotary kilns used in the manufacture of cement," *Adv. Appl. Ceram.*, vol. 116, no. 4, pp. 207–215, May 2017.
- [23] W. Li, D. Wang, and T. Chai, "Multisource data ensemble modeling for clinker free lime content estimate in rotary kiln sintering processes," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 2, pp. 303–314, Feb. 2015.
- [24] W. Li, D. Wang, X. Zhou, and T. Chai, "An improved multi-source based soft sensor for measuring cement free lime content," *Inf. Sci.*, vol. 323, pp. 94–105, Dec. 2015.
- [25] Z. Wang, C. Song, and T. Chen, "Deep learning based monitoring of furnace combustion state and measurement of heat release rate," *Energy*, vol. 131, pp. 106–112, Jul. 2017.
- [26] Z. Han, J. Li, B. Zhang, M. M. Hossain, and C. Xu, "Prediction of combustion state through a semi-supervised learning model and flame imaging," *Fuel*, vol. 289, Apr. 2021, Art. no. 119745.
- [27] H. Chen, X. Zhang, P. Hong, H. Hu, and X. Yin, "Recognition of the temperature condition of a rotary kiln using dynamic features of a series of blurry flame images," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 148–157, Feb. 2016.

- [28] G. Szatvanyi, C. Duchesne, and G. Bartolacci, "Multivariate image analysis of flames for product quality and combustion control in rotary kilns," *Ind. Eng. Chem. Res.*, vol. 45, no. 13, pp. 4706–4715, Jun. 2006.
- [29] R. Zhang, S. Lu, H. Yu, and X. Wang, "Recognition method of cement rotary kiln burning state based on Otsu-Kmeans flame image segmentation and SVM," *Optik*, vol. 243, Oct. 2021, Art. no. 167418.
- [30] H. Bae, S. Kim, B.-H. Wang, M. H. Lee, and F. Harashima, "Flame detection for the steam boiler using neural networks and image information in the Ulsan steam power generation plant," *IEEE Trans. Ind. Electron.*, vol. 53, no. 1, pp. 338–348, Feb. 2006.
- [31] H. Kramers and P. Croockewit, "The passage of granular solids through inclined rotary kilns," *Chem. Eng. Sci.*, vol. 1, no. 6, pp. 259–265, 1952.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.



**Yuchen Zhao** (Student Member, IEEE) received the B.Eng. degree in thermal engineering from the College of Energy Engineering, Zhejiang University, Hangzhou, China, in 2022, where he is currently pursuing the Ph.D. degree in control science and engineering with the State Key Laboratory of Industrial Control Technology (SKLICT), College of Control Science and Engineering.

His research interests are in multisource information fusion approaches to process industry modeling.



**Chunjie Yang** (Senior Member, IEEE) received the B.Eng. degree in machine design, the M.Eng. degree in fluid transmission and control, and the Ph.D. degree in industrial automation from Zhejiang University, Hangzhou, China, in 1992, 1995, and 1998, respectively.

He is currently a Professor with the College of Control Science and Engineering and a Qiushi Distinguished Professor with Zhejiang University. His research interests include modeling and control for industrial processes, smart energy, and intelligent manufacturing.



**Hang Xiao** received the B.Eng. degree in automation from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2023, where he is currently pursuing the master's degree in control science and engineering with the State Key Laboratory of Industrial Control Technology (SKLICT), College of Control Science and Engineering.

His research interests include data-driven modeling, process data analysis, and their applications in the process industry.



**Yaoyao Bao** received the Ph.D. degree from the East China University of Science and Technology, Shanghai, China, in 2021.

He is currently an Engineer with Alibaba Cloud, Hangzhou, China. His research interests include industrial process modeling and intelligent decision.



**Siwei Lou** (Student Member, IEEE) received the B.Eng. degree in automation and the M.Sc. degree in control science and engineering from the Faculty of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou, China, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree in control science and engineering with the State Key Laboratory of Industrial Control Technology (SKLICT), College of Control Science and Engineering, Zhejiang University, Hangzhou.

Over the past years, he and his colleagues have perfected the modeling and control with Liuzhou Ironmaking Plant, Liuzhou, China. His research interests include statistical learning, deep learning, fault detection, and process control.



**Jiayun Mao** received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2016.

He is currently an Engineer with Alibaba Cloud, Hangzhou. His research interests include intelligent decision, time-series prediction, and soft-sensing.



**Huanyu Liao** received the M.Sc. degree from Central South University, Changsha, China, in 2010.

He is currently an Algorithm Team Leader of Industrial Brain with Alibaba Cloud, Hangzhou, China. His research interests include industrial intelligent manufacturing platforms, intelligent control, and optimization algorithm.