



Original Paper

A multiscale adaptive framework based on convolutional neural network: Application to fluid catalytic cracking product yield prediction



Nan Liu^b, Chun-Meng Zhu^{a, b, 1}, Meng-Xuan Zhang^b, Xing-Ying Lan^{b, *}

^a College of Artificial Intelligence, China University of Petroleum (Beijing), Beijing, 102249, China

^b State Key Laboratory of Heavy Oil Processing, China University of Petroleum (Beijing), Beijing, 102249, China

ARTICLE INFO

Article history:

Received 6 October 2023

Received in revised form

21 December 2023

Accepted 15 January 2024

Available online 22 January 2024

Edited by Min Li

Keywords:

Fluid catalytic cracking

Product yield

Data-driven modeling

Multiscale prediction

Data decomposition

Convolution neural network

ABSTRACT

Since chemical processes are highly non-linear and multiscale, it is vital to deeply mine the multiscale coupling relationships embedded in the massive process data for the prediction and anomaly tracing of crucial process parameters and production indicators. While the integrated method of adaptive signal decomposition combined with time series models could effectively predict process variables, it does have limitations in capturing the high-frequency detail of the operation state when applied to complex chemical processes. In light of this, a novel Multiscale Multi-radius Multi-step Convolutional Neural Network (MsrtNet) is proposed for mining spatiotemporal multiscale information. First, the industrial data from the Fluid Catalytic Cracking (FCC) process decomposition using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) extract the multi-energy scale information of the feature subset. Then, convolution kernels with varying stride and padding structures are established to decouple the long-period operation process information encapsulated within the multi-energy scale data. Finally, a reconciliation network is trained to reconstruct the multiscale prediction results and obtain the final output. MsrtNet is initially assessed for its capability to untangle the spatiotemporal multiscale relationships among variables in the Tennessee Eastman Process (TEP). Subsequently, the performance of MsrtNet is evaluated in predicting product yield for a 2.80×10^6 t/a FCC unit, taking diesel and gasoline yield as examples. In conclusion, MsrtNet can decouple and effectively extract spatiotemporal multiscale information from chemical process data and achieve a approximately reduction of 30% in prediction error compared to other time-series models. Furthermore, its robustness and transferability underscore its promising potential for broader applications.

© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Some developed and developing countries are accelerating the strategic planning and layout of intelligent manufacturing, called the fourth industrial revolution (Chai and Ding, 2018; Yuan et al., 2017). In particular, Germany has taken the lead with Industry 4.0 (Lee et al., 2019). Intelligent manufacturing will transform the refining and petrochemical industry into a connected, information-driven environment, accelerating technology and industrialization

integration (Li, 2016; Li et al., 2015). With the development of sensors, wireless communication, and the Internet of Things, industrial data generated and collected throughout the product life-cycle has increased exponentially (C.Q. Li et al., 2022; Soni and Kumar, 2022; Yang et al., 2021). Companies are gradually considering big data as an essential factor of production, with industrial data becoming a significant driver in pursuing efficient, sustainable, and intelligent manufacturing practices (Sun et al., 2017). The use of artificial intelligence technology to fully exploit industrial data and refine the operational cognition of enterprises has become a new demand (Andooz et al., 2023; He et al., 2020). It adjusts system status information promptly, assists enterprises in monitoring critical locations in the production process, assists in optimizing decisions at the production site, and gradually completes the

* Corresponding author.

E-mail address: lanxy@cup.edu.cn (X.-Y. Lan).

¹ These authors contributed equally to this work and should be considered co-first corresponding authors

intelligence of the enterprise (Hong et al., 2022; Singer and Cohen, 2021; M. Zhang et al., 2023). The Fluid Catalytic Cracking (FCC) unit is the core process unit of an oil refinery, and its operational data contains information on the operational status of the processing unit, which couples multiscale information such as reaction and flow (Stratiev et al., 2023). Therefore, there is an urgent need to develop a fundamental model framework that can decouple multiscale information across scales.

Data-driven approaches are modeled by value mining of process data (W. Liu et al., 2023; Qu et al., 2023; Wu and Chau, 2010). These approaches effectively model most complex industrial processes and have a natural advantage in generality (Li et al., 2021; Z. Zhang et al., 2023; Zhang et al., 2022). For instance, using a data-driven predictive model, Byun et al. (2021) conducted a comprehensive analysis of a newly designed methanol steam reforming system's technical, environmental, and economic feasibility. Ascher et al. (2022) explored seven machine-learning approaches for biomass gasification process modeling and discovered that Gradient Boosting Regression outperformed other models. Therefore, data-driven models have gained widespread attention and application across various fields. Convolutional Neural Network (CNN) is widely utilized because it can extract feature of varying significance through sparse interactions and convolutional kernel structures (Kang et al., 2022; Rayhan Ahmed et al., 2023; Zhao et al., 2023). Y. Xie et al. (2023) detailed the application of 1D and 2D CNN for time series data prediction and improved the daily runoff prediction accuracy by integrating the non-linear fitting capability. J. Liu et al. (2023) used CNN to extract feature from non-linear and non-stationary solar irradiance time series. J. Xie et al. (2023) enhanced the accuracy of daily runoff prediction by employing a combined CNN and Long Short-Term Memory (LSTM) network to extract spatial and temporal information from battery thermal management system data. However, because of the system's related material and energy flow, capturing the certain scale feature does not effectively depict the state derivation in engineering applications. As the refinery industry data set contains multiple variables such as temperature, pressure, and level, the fluctuation trends respond differently to the same disturbance when regulated by the control system. Therefore, the only way to focus on both short-term fluctuations and long-term trends of the feature variables is through composite spatiotemporal multiscale feature extraction. In terms of performance, CNN has the fundamental capability to implement multiscale deep neural networks.

Data-driven approaches to FCC analysis highlight their potential to improve efficiency, environmental sustainability, and safety in the process industry (Yang et al., 2023). Khaldi et al. (2023) present a systematic literature review to identify critical directions for future research, including exploring ensemble methods and developing hybrid models in modeling, controlling, and optimizing the FCC unit using artificial intelligence techniques. Taira et al. (2022) proposed to investigate the performance of a Bayesian recurrent neural network-based method in detecting faults in a real FCC unit, aiming to improve process safety in modern chemical industries. Kawai et al. (2022) developed an AI hybrid reaction model that incorporates catalyst deactivation and optimization of operational conditions for residue FCC. Usman et al. (2023) implemented ensemble machine learning paradigms to improve the modeling of the yields of light olefins in crude-to-chemical conversions. Ni et al. (2021) developed an online optimization strategy for an FCC process using a case-based reasoning method based on big data technology. Yang et al. (2020) developed a hybrid predictive framework for FCC by integrating machine learning with the 8-lump kinetics model, which significantly improves the predictability of FCC performance.

Industrial data is prone to loss and inconsistency due to sensor failures, human manipulation, and system errors, which negatively impact data quality and reliability, affecting the monitoring, control, and optimization of production processes (Cui et al., 2020; Min et al., 2019; Mohammadpoor and Torabi, 2020; Zheng et al., 2023). Particularly in the industrial sector, high-quality data is essential for building effective models and can lead to problems of reduced model accuracy and reliability. (Teh et al., 2020; Yao and Zhao, 2022). Consequently, data decomposition has attracted the attention of scholars, including techniques like Short-time Fourier Transform (STFT) (C. Li et al., 2022), Wavelet and Wavelet Packet Decomposition (Lung, 2007), Empirical Mode Decomposition (EMD) (Gulay and Duru, 2020), and Local Mean Decomposition (LMD) (Liu et al., 2015). To improve the signal-to-noise ratio of the original vibration signals of dynamic bearings, Chen et al. (2023) obtained reconstructed denoised signals and time-frequency spectrograms through Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and STFT processing to fully exploit the fault characteristics of low-quality data and achieve more efficient fault diagnosis. Similarly, Peng et al. (2021) proposed a model integrating LMD, wavelet threshold denoising, and LSTM to accurately predict the load of natural gas pipeline networks, demonstrating excellent performance. Sun et al. (2023) used CEEMDAN to decompose the voltage sequence before prediction, effectively improving the accuracy of short-term performance degradation prediction for individual commercial vehicle fuel cell systems. In summary, the appropriate treatment of non-linear and non-stationary signals through spatiotemporal multiscale mining is beneficial for accurately predicting process variables in practical industrial applications.

Based on the above analysis, we propose the composite 2D convolutional structure with different padding and stride, which is the MsrtNet framework. The first step is to build the feature dataset on the target variable using Distributed Control System (DCS) and Laboratory Information Management System (LIMS) data to ensure the dataset is non-redundant, complete, and balanced. CEEMDAN is then used to decompose the feature dataset in the time-frequency domain to remove the mode component of data noise and fully exploit the spatiotemporal information implied by the industrial data stream. The decomposed mode components are then utilized in the MsrtNet to construct an accurate adaptive multiscale convolutional prediction model. This model captures the multi-physical scale and time-scale composites present in the data. The TE process validates the need for composite spatiotemporal multiscale mining, and an application case of product yield from a 2.80×10^6 t/a FCC unit in Northwest China is presented. The results show excellent prediction accuracy and robustness. It has the potential to be applied to complex industrial scenarios.

2. Preliminaries

2.1. CEEMDAN

The system's time series can analyze the dynamic feature of the FCC unit during operation. It is necessary to fully use the spatial correlation and time dependence in the original data, manifesting as waveforms in the spatial domain and displayed as time-series data in the time domain. The significant advantage of EMD is its ability to decompose the original signal into a series of Intrinsic Mode Functions (IMFs) without needing prior knowledge. In particular, without the need to define basis functions, the individual IMF reflects the local feature information of the original signal at different time scales, reducing the influence of subjectivity on the algorithm. However, the various uniformly distributed

frequency components of the EMD decomposition make IMFs power spectrum appear to have the same bandpass characteristics, making it highly susceptible to mode mixing problems. Although the EEMD method mitigates mode aliasing by introducing positive and negative Gaussian white noise, there is the problem of a certain amount of white noise remaining in the eigenmode components. CEEMDAN, as a variant of EEMD, adds Gaussian white noise to the components to be decomposed and integrates averaging for each order of IMFs (Torres et al., 2011). Consequently, CEEMDAN can effectively address the issue of mode mixing in EMD and EEMD, thereby enhancing the efficiency of the decomposition process and reducing computational costs. The CEEMDAN consists of the following steps.

- 1) The original data $x(t)$ are superimposed with N kinds of Gaussian white noise signals satisfying $N(0, 1)$ and then decomposed by EMD to obtain N decomposed $imf_1^i(t)$, and then integrated averaging is performed to obtain the first order $imf_1(t)$ and the first residual component $R_1(t)$.

$$imf_1(t) = \frac{1}{N} \sum_{i=1}^N imf_1^i(t) \quad (1)$$

$$R_1(t) = x(t) - imf_1(t) \quad (2)$$

- 2) $R_1(t)$ is used as the new original signal, and N Gaussian white noise signals satisfying $N(0, 1)$ in positive and negative pairs are added once again, afterwards EMD is performed separately to obtain N decomposed $imf_2^i(t)$, and then integrated averaging is conducted to obtain the second order $imf_2(t)$, and the second residual component $R_2(t)$.

$$imf_2(t) = \frac{1}{N} \sum_{i=1}^N imf_2^i(t) \quad (3)$$

$$R_2(t) = R_1(t) - imf_2(t) \quad (4)$$

- 3) Repeat step 2, set the ultimate count of decompositions to j . Stop the iteration when the number of residual components $R_j(t)$ poles is at most 2. Finally, the original data is decomposed into M mode components and the residual component $R_j(t)$ after the j th decomposition.

$$x(t) = \sum_{i=1}^M imf_j^i(t) + R_j(t) \quad (5)$$

2.2. CNN

A typical CNN consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer, as shown in Fig. 1. The two-dimensional output array can represent the input at some level in the spatial dimension (width and height), also known as the feature map. All possible input regions affecting the forward computation of the matrix elements are called the receptive field of the elements. When dealing with complex industrial processes, it is essential to carefully arrange convolution, pooling, and fully connected layers rationally and optimally and adjust the width and depth of the deep neural network to achieve the best possible output. To model complex industrial processes effectively,

optimizing the arrangement of convolutional, pooling, and fully connected layers and adjusting the width and depth of the deep neural network to achieve the desired output is necessary. Using the convolutional kernel structure, CNN addresses parameter inflation arising from direct connections between neurons in adjacent layers. Through this structure, CNN can preserve the positional relationship of the original data by leveraging the convolutional operation, as depicted in Fig. 2.

Incorporating a pooling layer after the convolutional layer is a common practice in the model structure. This addition enables the network to expand its receptive field in deeper layers, aggregating a broader range of feature and enhancing its ability to handle minor translations of certain feature. Additionally, it effectively reduces the computational effort required. System-level robustness mainly considers the coupling and complexity of the system units. The convolution and pooling feature are spanned into a fully connected layer to generate the output. Dropout is often used in the output of the fully connected layer, which can inject noise into the internal hidden layer. Dropout may seem random, but due to simultaneous training, one neuron may depend on certain other neurons, and randomly discarding some of them can force the neurons to learn more prominent feature, which can effectively avoid overfitting. It is worth mentioning that the Dropout method is not applicable in the testing phase of the model.

When dealing with the complex feature input, it is essential to incorporate suitable activation functions into the network structure to introduce nonlinearity and enable a nonlinear mapping between the input and output data. Commonly used activation functions include the Rectified Linear Unit (ReLU), Parametric Rectified Linear Unit (PReLU), and Hyperbolic Tangent Function (Tanh). When the input of ReLU is less than 0, the gradient is 0, and the gradient stops forward propagation from this position. When the input is greater than 0, the gradient is 1, and the gradient keeps the same value for forward propagation, which makes the concatenation of data gradients not converge to 0, but the neuron dying phenomenon occurs. When negative values are input to the neuron, the gradient is constant at 0, and the corresponding weights and biases are not updated. At the same time, PReLU improves it by multiplying the data by a smaller fixed value α when the input is less than 0, as shown in Eq. (6).

$$PReLU(x_{t,m}) = \begin{cases} x_{t,m} & \text{if } x_{t,m} > 0 \\ \alpha x_{t,m} & \text{if } x_{t,m} < 0 \end{cases} \quad (6)$$

In the two-dimensional convolution layer, the input matrix $X[n_h \times n_w]$, the kernel matrix $W[k_h \times k_w]$, the deviation scalar b , and the output matrix $Y = X \times W + b$, where W and b are learnable parameters, and the output size is:

$$(n_h - k_h + 1) \times (n_w - k_w + 1) \quad (7)$$

Compared with the fully connected layer, convolution has the feature of "weight sharing", which reduces the model's capacity and can effectively detect spatial patterns. Furthermore, it can control the output shape by filling and stepping to obtain the edge data of the input data, expand the receptive field, and reduce the information loss. It is the most essential and unique layer in the CNN. If the row and column are filled, the output shape is:

$$(n_h - k_h + p_h + 1) \times (n_w - k_w + p_w + 1) \quad (8)$$

Commonly make $p_h = k_h - 1$, $p_w = k_w - 1$, when k_h is odd, $p_h/2$ is filled on two-side. When k_h is even, $\lceil p_h/2 \rceil$ is filled on the upper side, and $\lfloor p_h/2 \rfloor$ is filled on the lower side. Generally, the convolution kernel with odd height and width is used. Hence, the number of fillings on both ends is equal, and the output size of filling the

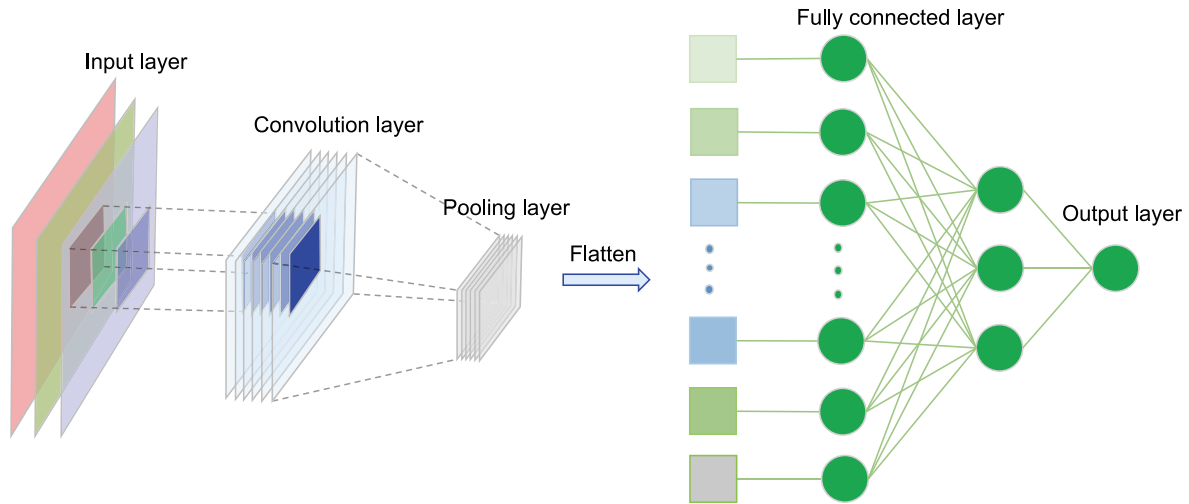


Fig. 1. Structure of the CNN.

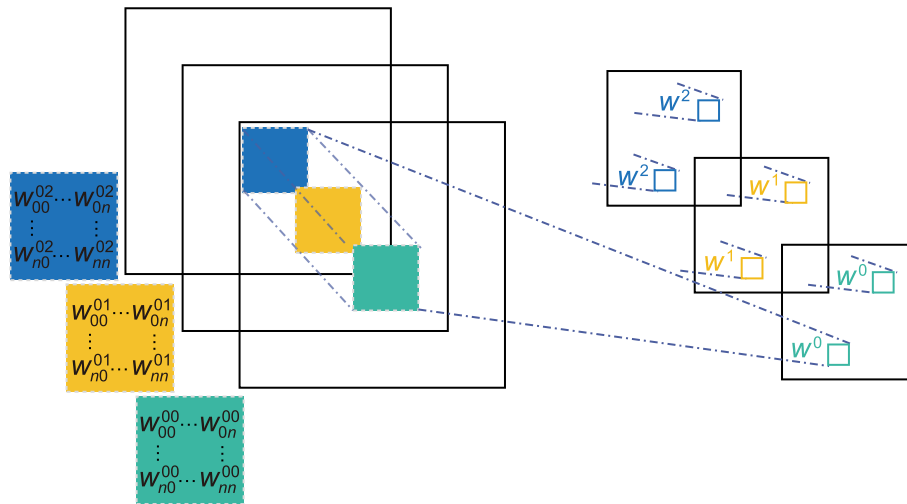


Fig. 2. Inter-correlation calculation with three input channels.

reduction is linearly related to the number of layers, which requires much calculation. The convolution kernel usually moves from left to right and top to bottom to read the data. The sliding step of the row/column is called stride. The output shape of the stride of height s_h and width s_w is shown in Eq. (9).

$$\lfloor (n_h - k_h + p_h + s_h) / s_h \rfloor \times \lfloor (n_w - k_w + p_w + s_w) / s_w \rfloor \quad (9)$$

2.3. RNN

Recurrent Neural Network (RNN) is frequently used to processing sequence data. While traditional fully connected feed-forward networks assign a separate parameter to each input feature, RNN shares the same weights over several time steps and does not need to learn the rules for each position separately. Parameter sharing enables the model to be extended and generalized to samples of different lengths of time, an advantage that is particularly important in industrial processes with time delays. The structure of the loop network without output is shown in Fig. 3, with a black square representing the delay of a single time step and

the loop after t steps of expansion represented by the function $g^{(t)}$, as shown in Eq. (10). Where f represents the transfer function using the same parameters for each time step, θ is used to parameterise f to the same value, and h is the state at the current moment.

$$p^{(t)} = g^{(t)}(x^{(t)}, x^{(t-1)}, x^{(t-2)}, \dots, x^{(2)}, x^{(1)}) \\ = f(h^{(t-1)}, x^{(t)}; \theta) \quad (10)$$

3. MsrtNet

The flow chart of the proposed MsrtNet in this paper is shown in Fig. 4. Each feature sequence is decomposed individually after a basic pre-processing of the industrial data. The high-frequency components with high noise content are removed to suppress their energy. The noise-reduced IMFs are then Min-Max normalized and used as the feature data set. Then, convolutional kernels with three sizes of receptive field, small, medium and large, are constructed to control the pace of movement respectively and perform full-domain convolution across spatiotemporal scales in

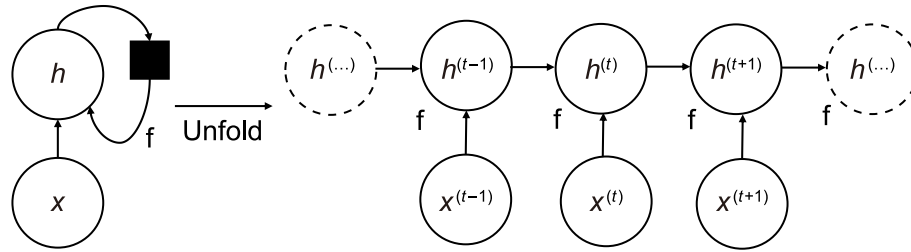


Fig. 3. RNN without output.

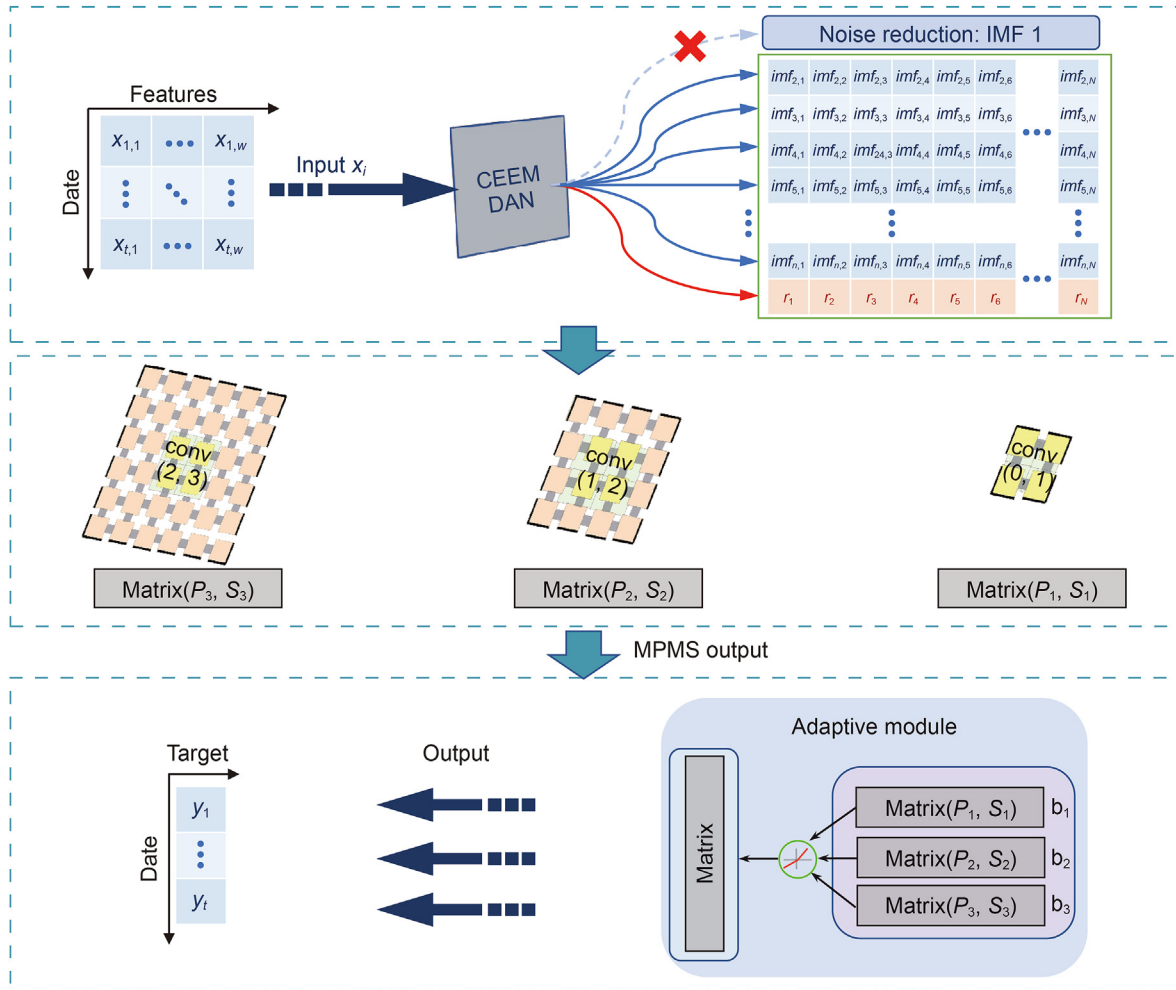


Fig. 4. Construction flowchart of MsrtNet.

the feature map. Finally, an adaptive CNN containing a 1×1 convolutional layer and a fully connected layer is used to integrate the three different spatiotemporal scales of information to obtain the final time-series prediction results for the target variable. It is worth noting that the number of channels in the network, the size and number of pooling layers, and the choice of activation functions for the convolutional and fully-connected layers will be dynamically adjusted according to the demands of the task data.

This paper considers the temporal attributes of chemical data by introducing time as an additional dimension in the dataset. Consequently, the input data is transformed into a two-dimensional matrix. Utilizing MsrtNet, both time-domain and frequency-domain feature of the data are extracted simultaneously.

Subsequently, each consecutive set of n data points is employed as either training or test sample data, resulting in a matrix with dimensions $n \times k$, where k represents the number of feature variables obtained after CEEMDAN decomposition. The number of experimental samples is determined by sliding a time window across the data at a single time interval. The flowchart detailing the data preparation process and the final flowchart for obtaining predicted values are provided visually in Fig. 5.

3.1. Multiscale-CNN module

RNN and LSTM can be extended to samples of different time lengths and generalized and have been widely applied in handling

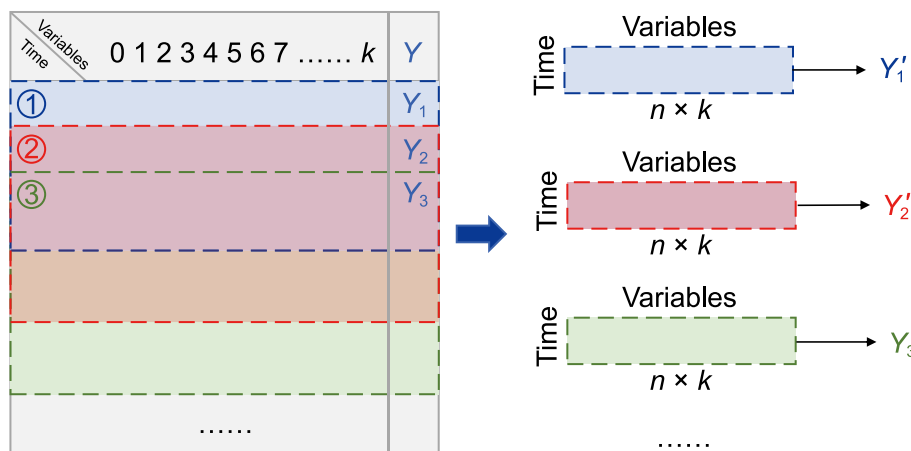


Fig. 5. Forms of dataset in MsrtNet.

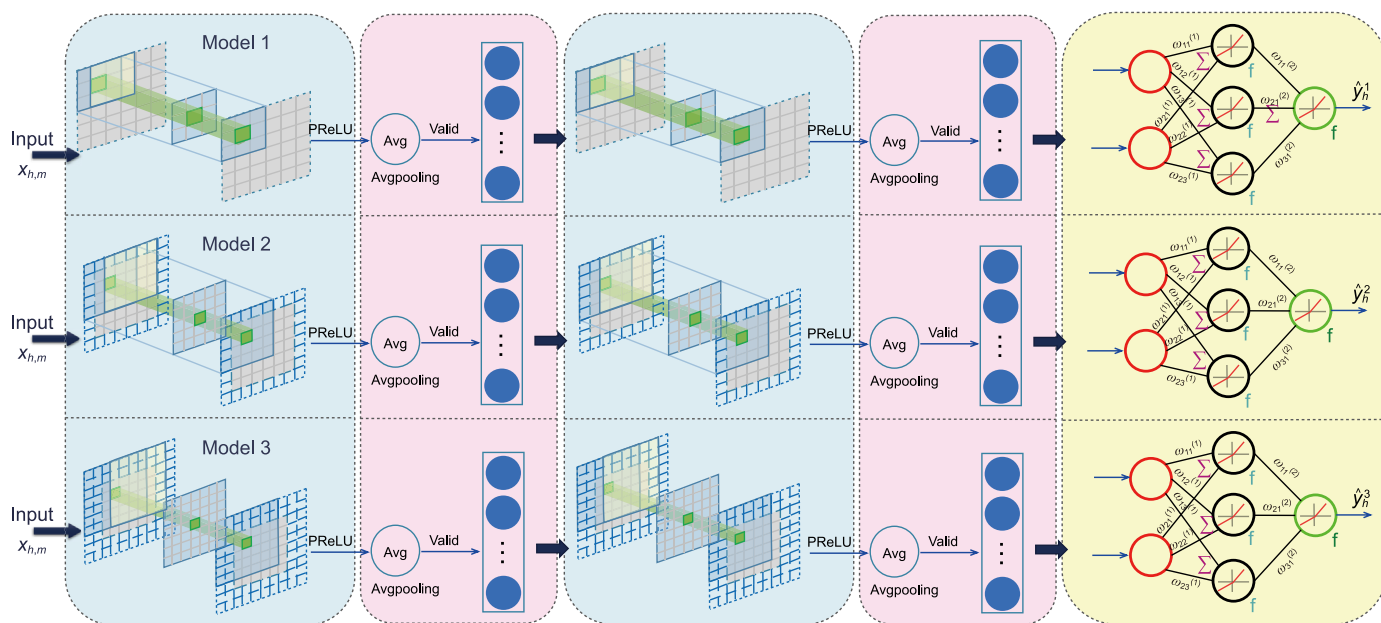


Fig. 6. Structure of Multiscale-CNN module.

industrial process time-series data. However, RNN only inputs data for one-time step of variables during the operation of the data stream, and the nonlinear hidden relationships between different variables and different time steps of data are not reflected. Chemical processes are highly nonlinear and complex processes. The reaction regeneration, distillation, and absorption stabilization units in the FCC unit are closely linked by material flow, energy flow, and information flow, and the coupling phenomenon between variables is prominent. It is necessary to capture the information between feature variables in depth to predict the product yield and determine its change trend accurately.

Most natural signals share a characteristic that the intensity decays with increasing frequency (Xu, 2020). The prediction and generalization difficulties associated with high-frequency catastrophes are challenging to alleviate by simply tuning parameters (Liu et al., 2020). In this paper, we propose a spatiotemporal multiscale CNN structure to accurately capture the oscillation trends of high-frequency signals. Taking advantage of CNN with convolutional structure in multiscale feature extraction, different

convolutional kernel structures are set to sense the full field information of the feature dataset. In particular, larger convolutional kernels can extract slowly changing components more efficiently over a broader range of time scales and more energy scales. Smaller convolutional kernels can detect the transient fluctuation behavior of time series data.

In the experiment, a two-dimensional matrix is used as input, with the two dimensions representing space and time and the individual elements being observations of different variables at different times. In the time dimension, the time step represents the length of the historical data. A series of points represents the neighboring local variables in the spatial dimension. The neighboring elements of adjacent time steps are weighted by convolution operations simultaneously perceiving more variables stitched together at different time steps and accumulating them as a new element in the feature map, which then integrates local spatiotemporal information in a higher dimension to better fit the details based on capturing the contour information of the objective function. The spatiotemporal multiscale CNN structure is shown in Fig. 6.

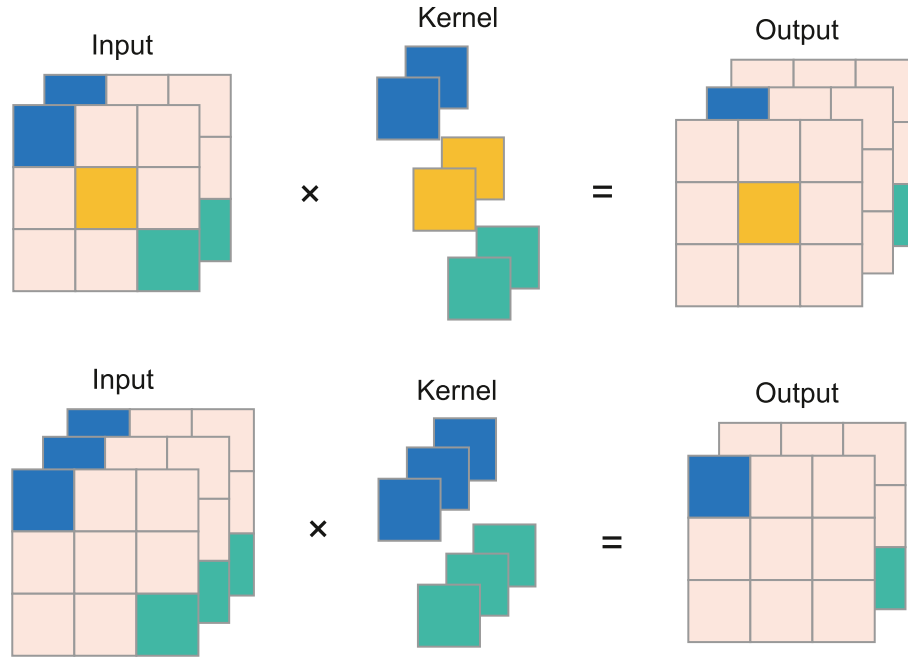


Fig. 7. Mutual correlation calculation of 1x1 convolution kernel.

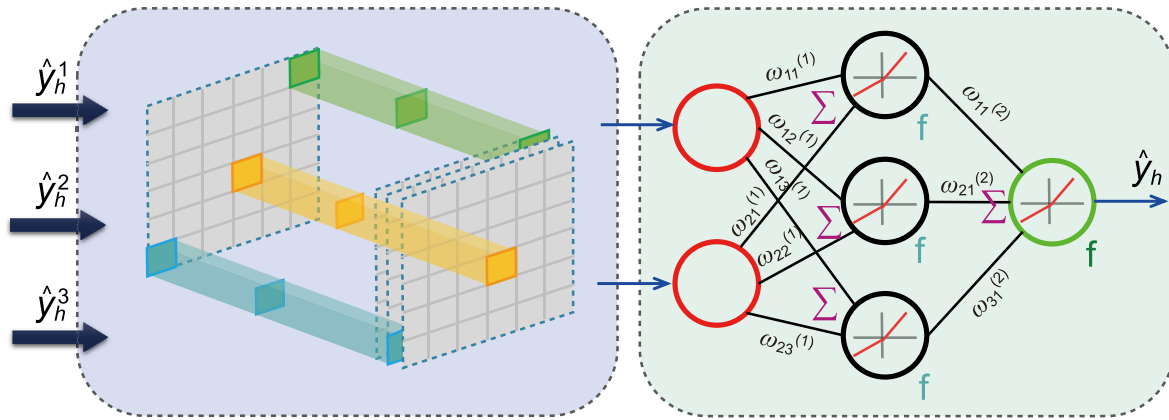


Fig. 8. Structure of Adaptive-CNN module.

More importantly, if each spatiotemporal scale can be considered a subsystem, then CEEMDAN decomposes the subsystems at each spatiotemporal scale of the original system. Although it is impossible to know precisely the manipulated variables corresponding to each subsystem analytically, both DCS and LIMS data involve variables that contribute to information at different scales to varying degree fields. Thus, CEEMDAN decomposes the highly complex, high-dimensional industrial system into several lower-dimensional, physical time and space scales implicit in the original sequence, further ensuring the convergence and robustness of the model.

The time series data $x(t)$ of time length t is decomposed j times by CEEMDAN to obtain M IMFs $imf_j^i(t)$ and a residual component $R_j(t)$, where each component is of equal length to the original data, as shown in Eq. (11).

$$x(t) = \sum_{i=1}^M imf_j^i(t) + R_j(t) \quad (11)$$

The two-dimensional matrix $\begin{bmatrix} \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,w} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{t,1} & \cdots & \mathbf{x}_{t,w} \end{bmatrix}$ with the number of feature w in the original data set $V = \{x_{t,w}, y_t\} = \begin{bmatrix} \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,w} & \mathbf{y}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{x}_{t,1} & \cdots & \mathbf{x}_{t,w} & \mathbf{y}_t \end{bmatrix}$ is decomposed by CEEMDAN. The component feature energies are extracted as signal feature to obtain a two-dimensional matrix $x(t)' = \begin{bmatrix} \mathbf{x}_{1,1} & \cdots & \mathbf{x}_{1,m} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{t,1} & \cdots & \mathbf{x}_{t,m} \end{bmatrix}$ with the number of feature m . The extracted feature energies are used as feature dataset. Assuming that the length of each period passing through the Multiscale-CNN module is h , with small, medium and large to distinguish between three convolution structures with different stride and padding structures, the shape of the output after convolution calculation is shown in Eq. (12). Three convolution structures simultaneously implement the prediction of time series data of length h . The

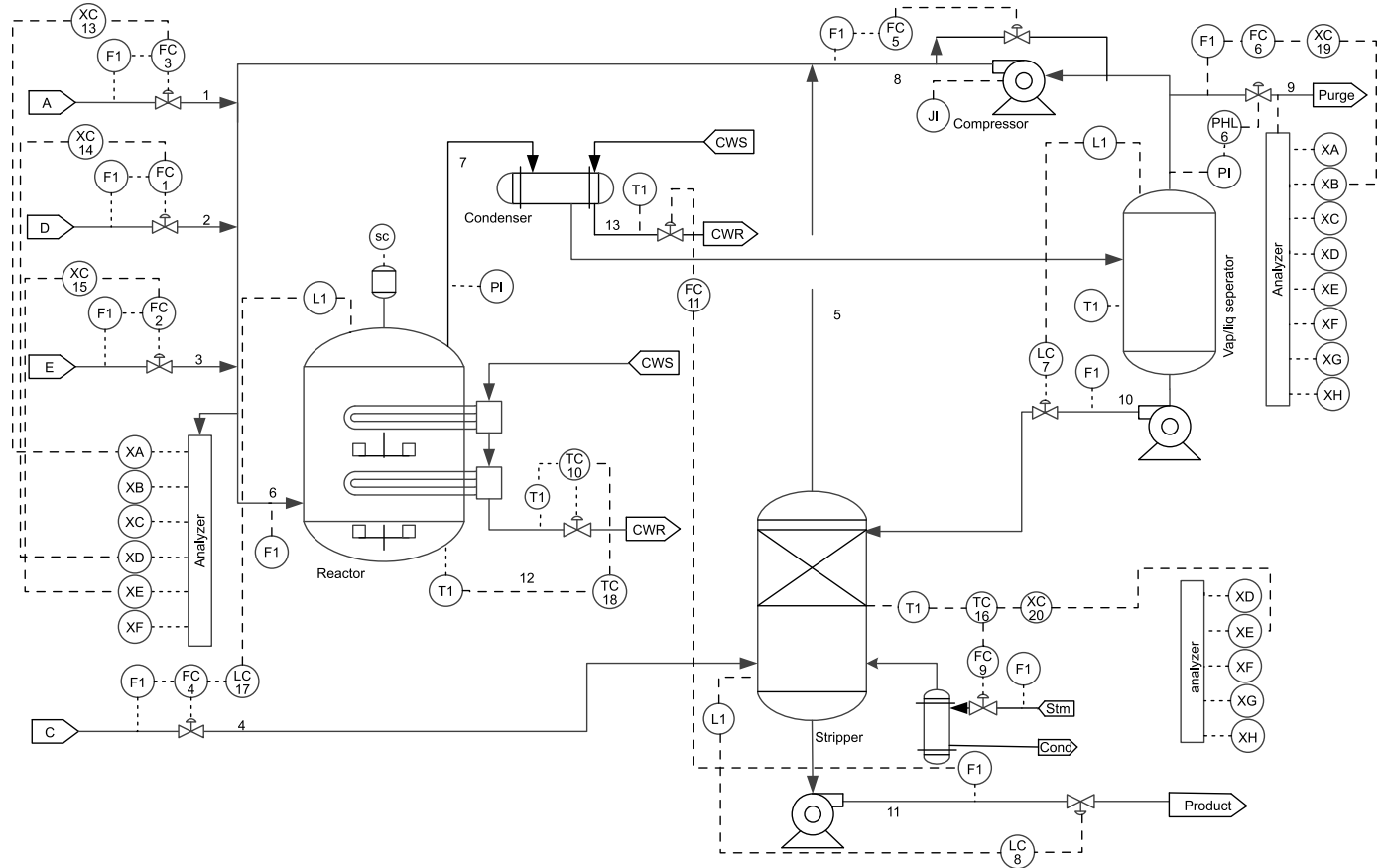


Fig. 9. Process of the TEP

outputs obtained are $\hat{y}_h^1$, $\hat{y}_h^2$ and $\hat{y}_h^3$ respectively. Among them, average pooling is used to simplify the feature map.

$$\lfloor (n_h - k_{hn} + p_{hn} + s_{hn}) / s_{hn} \rfloor \times \lfloor (n_m - k_{mn} + p_{mn} + s_{mn}) / s_{mn} \rfloor \quad (12)$$

The objective function is the central guide to the overall model optimization learning and is also known as the loss function when the objective function needs to be minimized. The objective function is used to estimate the degree of inconsistency between the predicted and true values of the model. The simplest way is using the square of Euclidean distance as the loss function, Minimizing Square Error (MSE) on training samples, as in Eq. (13). The MSE curve is smooth, continuous, and derivable, facilitating the use of the gradient descent algorithm, and as the error decreases, so does the gradient, which facilitates the convergence of the function.

$$MSE = \frac{1}{t} \sum_{h=1}^t \sum_{i=1}^h (\hat{y}_h^i - y_h^i)^2 \quad (13)$$

Adam is generally suitable for processing large-scale data. However, its learning rate can still become unstable in the later stages of training, failing to converge to values that are sufficiently good for generalization. RMSProp is a very robust optimizer that has pseudo-curvature information. It can deal with stochastic objectives nicely, making it applicable to mini-batch learning. It can deal with stochastic objectives nicely, making it applicable to mini-batch learning. Additionally, using leakage averaging to tune the

preprocessor in order of coefficients, it converges faster than the momentum algorithm.

3.2. Adaptive-CNN module

In this paper, based on subsection 3.1, an Adaptive-CNN module is constructed by setting up a 1×1 convolution layer and a fully connected layer, which, after the non-linear representation of neurons, can deeply integrate the multi-temporal and spatial scale information obtained by Multiscale-CNN module in higher dimensions to obtain the best time-series prediction results. Compared to the original measurement space, the variant feature space carries more excellent recognition capability and is more conducive to process monitoring performance. This enables MsrtNet to solve data multiscale, high-latitude, and highly non-linear problems more effectively, enhancing its data mining capabilities.

The data of $\hat{y}_h^1$, $\hat{y}_h^2$, $\hat{y}_h^3$ is up-dimensional by a 1×1 convolutional layer to increase the feature map. Each element in the output comes from the accumulation of elements in the input at the same position in height and width between different channels by weight, realizing the cross-channel combination of information, gradually abstracting low-dimensional feature to high-dimensional feature, making the feature representation of the new feature map more abundant. Indeed, using a 1×1 convolution for dimensionality reduction is similar to performing a convolution operation on the number of channels within the feature map. This process efficiently compresses the feature map and extracts feature in a two-step process, leading to an improved representation of the new

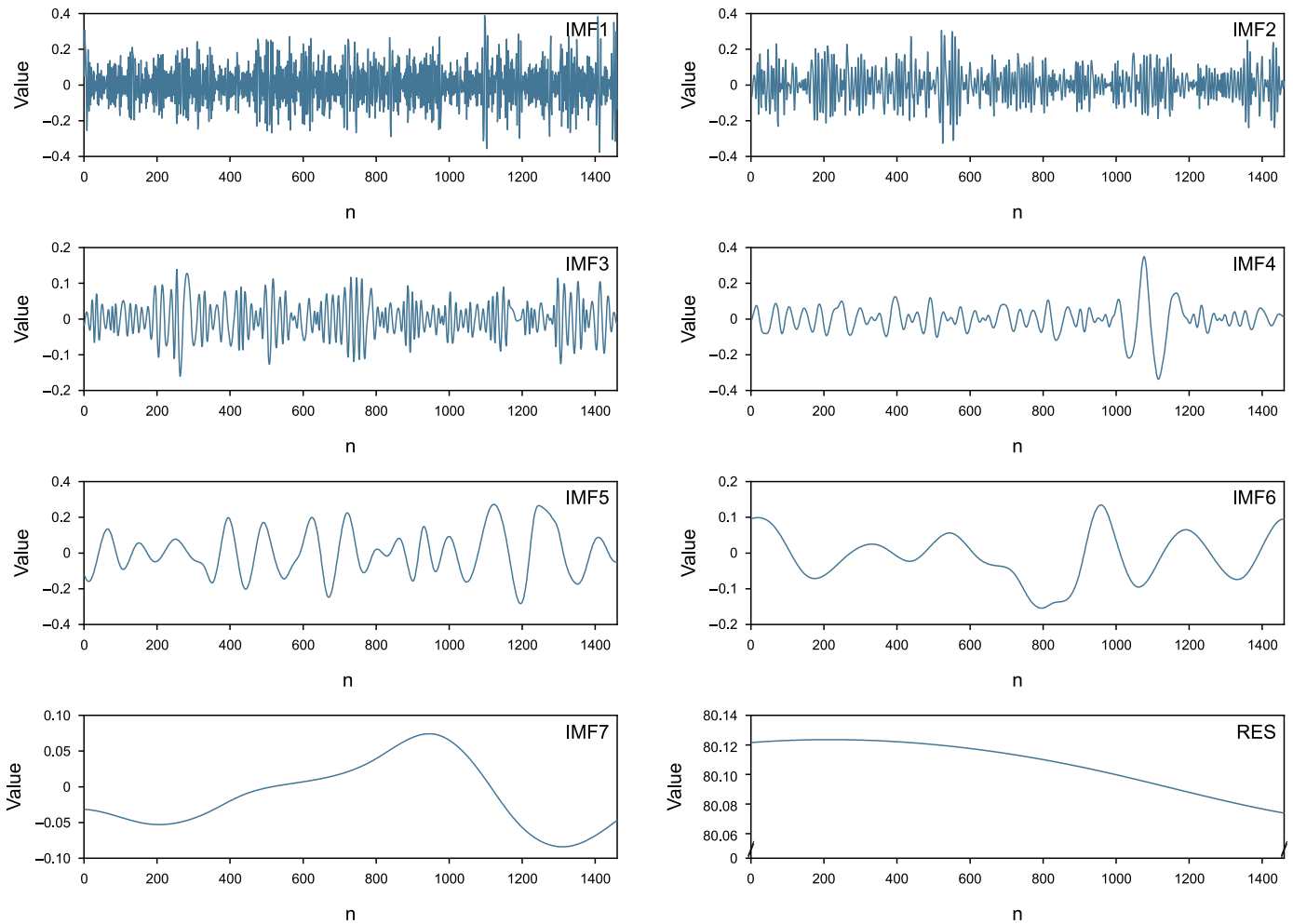


Fig. 10. Decomposition results of XMEAS11.

Table 1
Multiscale-CNN module structure and parameters.

Model	Conv1→Tanh			Conv2→Tanh			FC1→Tanh	FC2→Tanh
	Size	Padding	Strides	Size	Padding	Strides	Units	Units
Model1	5 × 5	0 × 0	1 × 1	5 × 5	0 × 0	1 × 1	120	50
Model2	5 × 5	1 × 1	2 × 2	5 × 5	1 × 1	2 × 2	120	50
Model3	5 × 5	2 × 2	3 × 3	5 × 5	2 × 2	3 × 3	120	50

Table 2
Adaptive-CNN module structure and parameters.

Model	Conv1			Conv2			FC1→Tanh	FC2
	Size	Padding	Strides	Size	Padding	Strides	Units	Units
parameter dimension	1 × 1	0 × 0 1 → 8	1 × 1	1 × 1	0 × 0 8 → 1	1 × 1	150	50

feature map. The application of a 1×1 convolution kernel for dimensionality reduction and enhancement is visually depicted in Fig. 7.

Through the fully connected layer and activation function, the two-dimensional vector of the dimensionality reduction 1×1 convolutional output is nonlinearly transformed into a one-dimensional vector, completing the feature fusion for the entire field of the feature subset. The end-to-end learning process is thus

achieved, and the time series prediction result $\hat{y}_t$ with period length h is obtained. The structure of the Adaptive CNN is shown in Fig. 8. According to the time sequence, $x(t)'$ is partitioned into $\lfloor \frac{t}{h} \rfloor$ two-dimensional matrices $\begin{bmatrix} \mathbf{y}_{1,1} & \cdots & \mathbf{y}_{1,m} \\ \vdots & \ddots & \vdots \\ \mathbf{y}_{h,1} & \cdots & \mathbf{y}_{h,m} \end{bmatrix}$, where h is the width of the matrix and m is the length of the matrix. The overall time

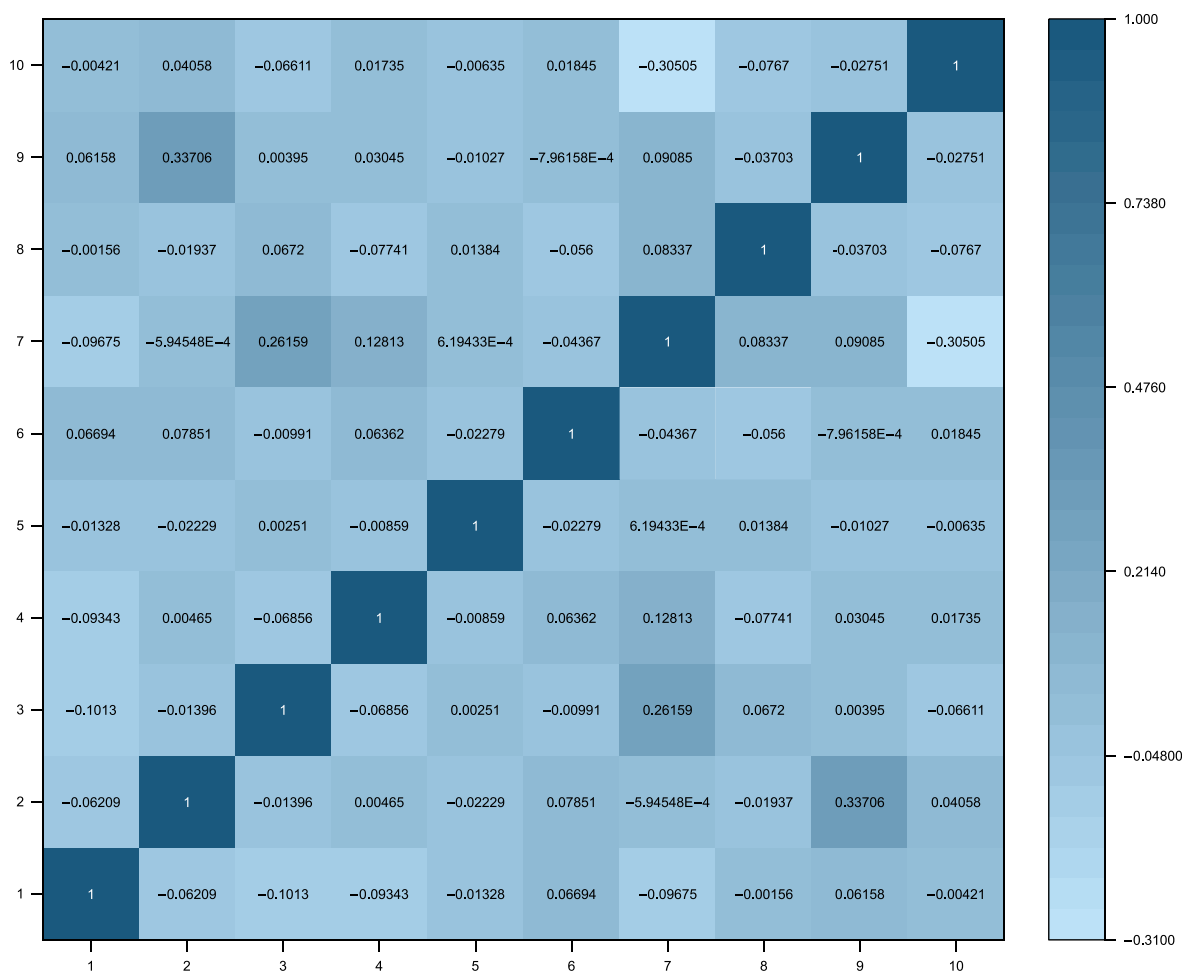


Fig. 11. PCC of the XMEAS1–10.

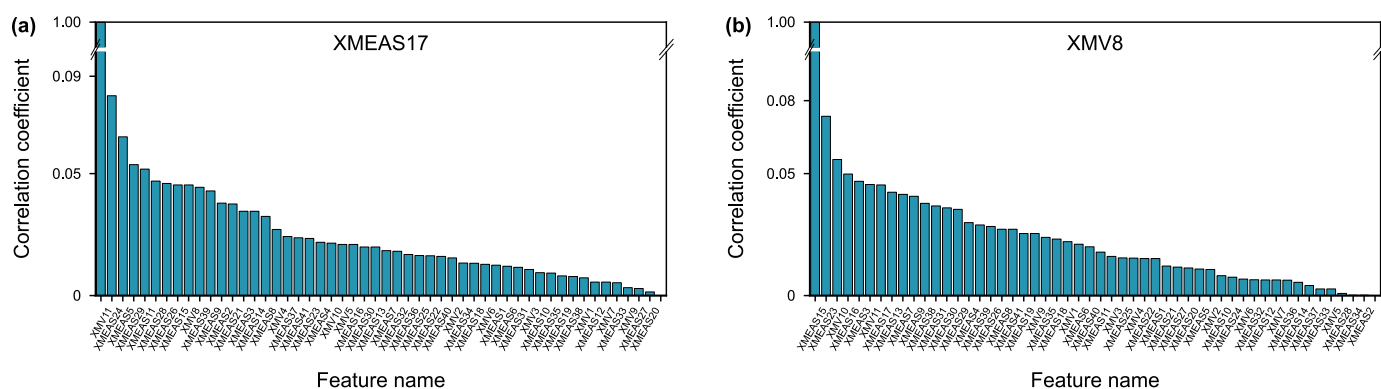


Fig. 12. Detailed presentation of the sequence weights with XMEAS17 and XMEAS8 as the target variables.

series prediction result $\hat{y}_t$ i.e. $[y_1, y_2, \dots, y_n]$ is obtained after the Multiscale-CNN and Adaptive-CNN modules.

4. Case study

4.1. Tennessee Eastman process

4.1.1. Process description

The Tennessee Eastman process (TEP), developed by Eastman Chemical Company in the USA, is based on actual chemical

processes and produces dynamic, time-varying, strongly coupled, and non-linear data. TEP consists of the reaction, separation, and recycling systems. The process flow diagram is shown in Fig. 9. TEP involves two simultaneous gas-liquid exothermic reactions, as shown in Eqs. (14) and (15), and two side reactions, as shown in Eqs. (16) and (17).



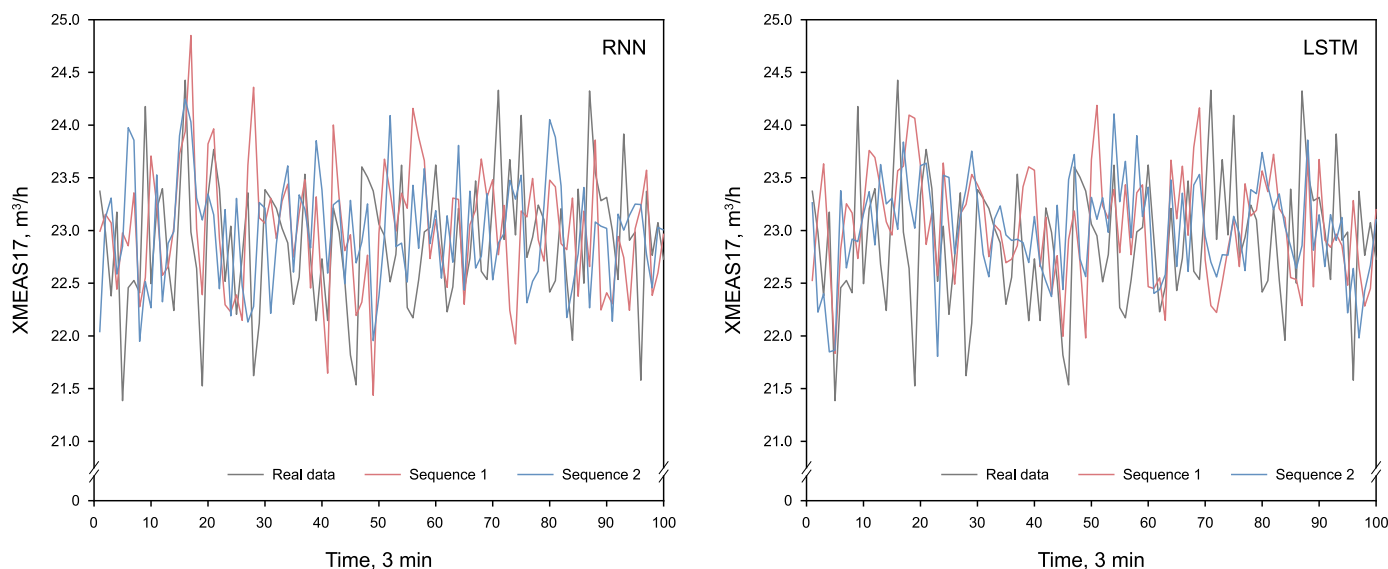


Fig. 13. Detailed prediction results of the two methods for different feature orders of XMEAS17.

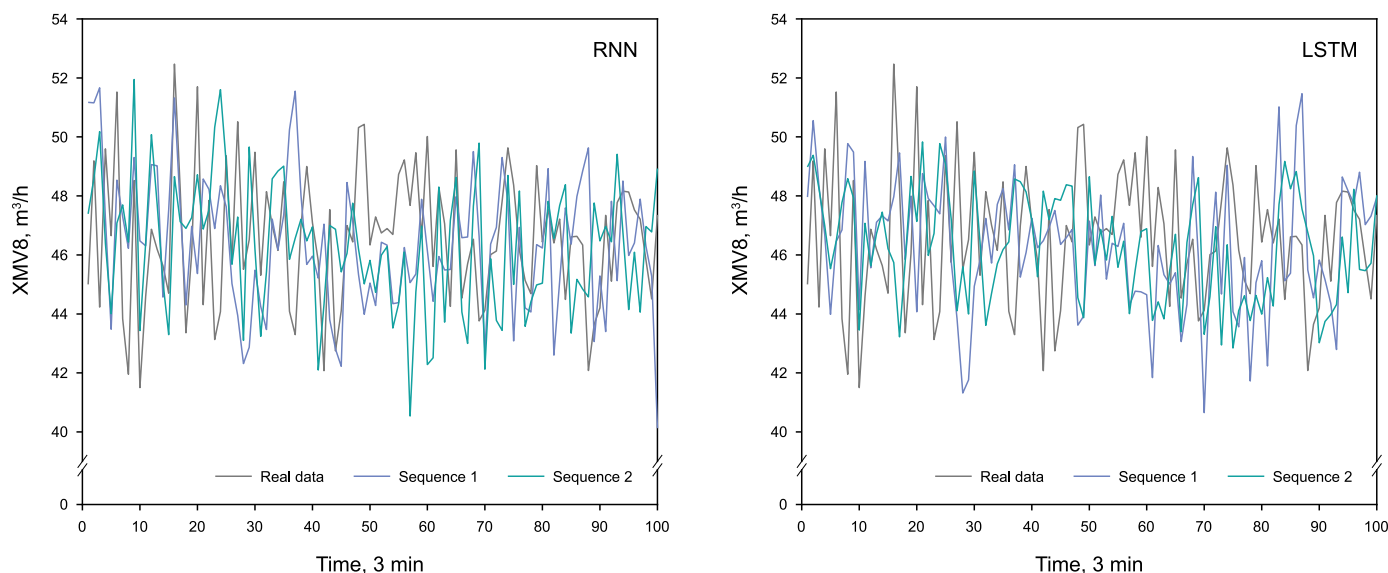


Fig. 14. Detailed prediction results of the two methods for different feature orders of XMV8.



The complete description of feature variables, including 11 operating and 41 measured variables, is provided in Tables S1–S3 of the supplementary file. XMEAS17 and XMV8 are essential indicators for monitoring product quality and are chosen as target variables to validate the performance of MsrtNet. The TEP normal condition dataset is used, with 1460 pieces of data and an interval of 3 min. One hundred items are selected in time sequence as the test dataset and the rest as the training set.

4.1.2. Model construction

The feature variables are decomposed using CEEMDAN to take

advantage and maximize the performance of MsrtNet to capture feature across spatial and temporal scales. For instance, XMEAS11 is adaptively decomposed into a series of IMFs, as shown in Fig. 10. The IMFs Oscillating around 0 are referred as the fluctuation component, and the smooth IMFs are referred as the trend component. Since the loss of precision in phase extraction with CEEMDAN is relatively tiny, it is beneficial for MsrtNet to learn the multiple coupling relationships embedded in the process data. CEEMDAN, as an advanced method applicable to industrial data processing, boasts outstanding capabilities in nonlinear analysis, noise suppression, and local feature extraction.

To ensure the rationality of the comparative experiment, the feature dataset and the hyperparameters used for the LSTM and RNN are identical. According to the actual convergence trend, the epoch is 100. In addition, the minimum batch size for each training period is 64, and the learning rate is 0.001. RMSProp optimizer is chosen to optimize the model by iteratively updating the network

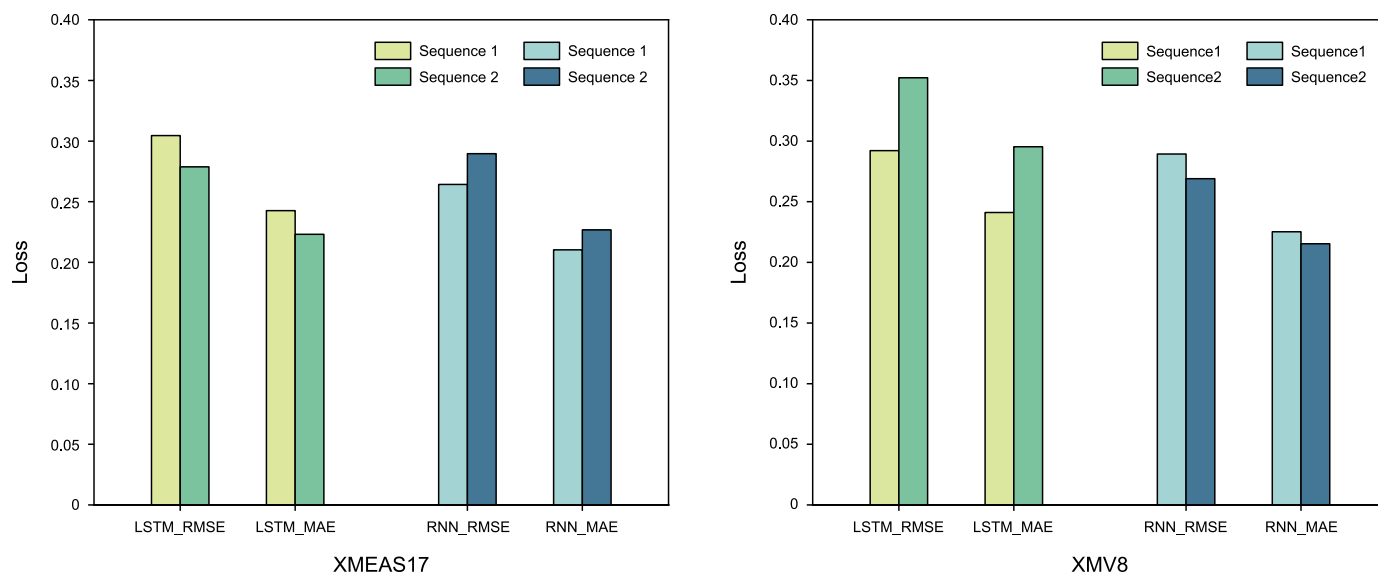


Fig. 15. Prediction loss of RNN and LSTM for XMEAS17 and XMV8 with different feature sequence.

weights based on the training set.

The parameters of the Multiscale-CNN module are shown in Table 1. The number of channels in the convolution layers is 6 and 11, respectively. The tanh activation function is used to obtain non-linear data feature. The epoch is 100, the minimum batch size for each training period is 50, the learning rate is 0.001, and the optimizer is RMSProp. The architecture of MsrtNet is structured into three levels: small (Model1), medium (Model2), and large (Model3). The large convolutional kernels extract low-frequency fluctuation behavior more efficiently over a broader range of time and energy scales. The small convolutional kernels detect high-frequency fluctuation behavior of process data. The medium convolution kernels capture the condition transition state and the fluctuation delay between different process units.

The parameters of the Adaptive-CNN module are shown in Table 2. The Adaptive-CNN module is highly flexible and adaptable, with numerous potential performance and application advantages. Firstly, by employing a filter size of 1×1 , no padding (Padding = 0×0), and a step size of 1×1 in Conv1 and Conv2 layers, the Adaptive-CNN module can efficiently extract both local and global information from the input feature, thereby preserving the essential feature of the input.

4.1.3. Results and discussion

The explication provided in Section 2.3 shows that RNN faces a limitation in assimilating information from states with varying historical lengths, attributable to their strict adherence to exact input size. Conversely, LSTM, while possessing fixed parameters, exhibits a dynamically modulated cumulative time scale in response to the input sequence. Regrettably, they do not account for the intricate, non-linear interplay within industrial process data across spatial and temporal dimensions. For this reason, we meticulously optimize the feature sequences to assess their influence on the predictive accuracy of the target variable, employing both RNN and LSTM architectures. Primarily, we compute the Pearson correlation coefficient (PCC) between the 52 variables, as depicted in Fig. S1 of the supplementary material. Notably, more pronounced absolute values at each position in the graph denote heightened correlations. For the sake of brevity, Fig. 11 exclusively presents the correlation coefficients of XMEAS 1–10.

The optimized sequence arrangement, determined by the PCC

analysis with XMEAS17 and XMV8 as the target variables, is visually depicted in Fig. 12(a)–(b). As shown in Fig. 12(b), the correlation coefficient between XMV8 and XMEAS15 attains a value of 1, signifying a highly correlated relationship. Conversely, XMEAS2, XMEAS28, XMEAS33, XMEAS34, XMEAS37, and XMV5 exhibit correlation coefficients of approximately 0 with XMV8, indicating a weaker association. In constructing the input dataset, we strategically positioned variables with pronounced correlations close to the target variables.

In Fig. 13, a comprehensive depiction of the prediction outcomes for both methods is presented, featuring distinct feature sequences of XMEAS17: sequence 1 without an optimized arrangement and sequence 2 with a meticulously curated order. At approximately the 60-min mark, the LSTM exhibits superior prediction performance with sequence 2 compared to sequence 1. Within the RNN, the prediction accuracy of sequence 1 surpasses that of sequence 2. Fig. 14 further elucidates the detailed prediction outcomes of XMV8 feature sequences for both methods. Taking the 70-min mark as an illustration, the accuracy of sequence 2 is better than that of sequence 1 in the LSTM. Conversely, in the RNN, the accuracy of sequence 1 outshines that of sequence 2. As anticipated, the prediction outcomes of feature sequences, subject to distinct permutations, manifest discernible disparities between the RNN and LSTM.

In Fig. 15, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are presented for both Sequence 1 and Sequence 2. When XMEAS17 is the target variable, it is evident that the LSTM exhibits higher predictive accuracy for Sequence 2, whereas the RNN outperforms Sequence 1. Additionally, with XMV8 as the target variable, the LSTM demonstrates superior performance for sequence 1, while the RNN excels with sequence 2. Indeed, while both RNN and LSTM exhibit competence in predicting XMV8 and XMEAS17, they yield distinct results contingent on the specific sequences and tasks.

As depicted in Fig. 16, predictions are generated for a dataset characterized by a randomized sequence order. The forecasted and observed values for XMEAS17 are compared across RNN, LSTM, and MsrtNet. Notably, in contrast to RNN and LSTM, the predictions generated by MsrtNet align more closely with the actual values. At approximately the 120-min mark, while RNN and LSTM can discern the overarching trend, MsrtNet excels in capturing finer details

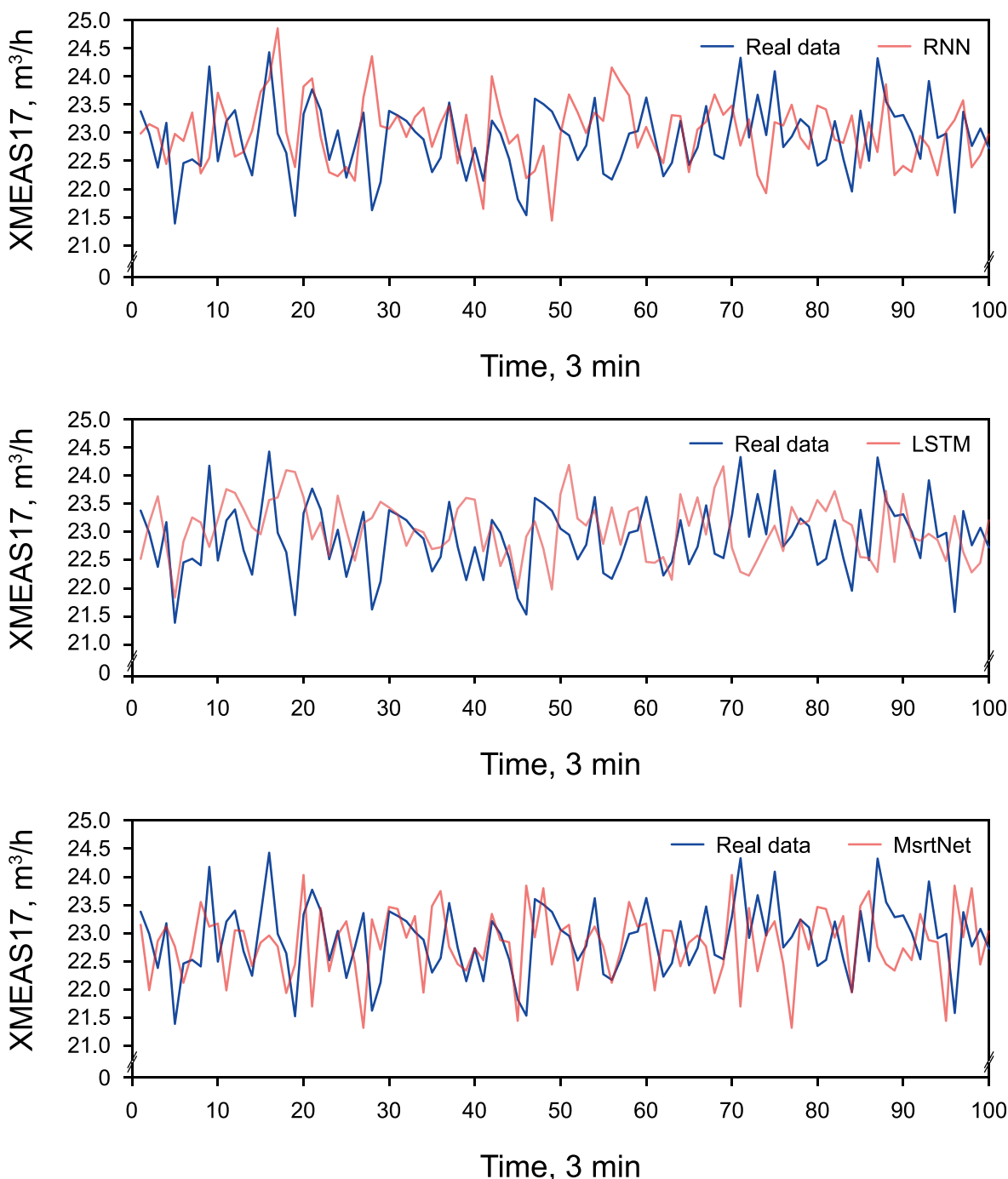


Fig. 16. Detailed comparison of predicted and true values of XMEAS17 by RNN, LSTM, and MsrtNet.

with heightened precision. This comprehensive evaluation underscores that MsrtNet is adept at effectively accommodating high-frequency intricacies while simultaneously capturing the broader trend patterns exhibited by XMEAS17.

Section 4.1 highlights the notable limitation in the time series prediction model, which does not account for the influence of temporal correlations within process data and the arrangement of feature sequences on model performance. The TE case is a compelling illustration of the substantial impact that the order of feature sequences wields on model effectiveness, emphasizing the complexity of determining an optimal sequence in industrial contexts. In response, MsrtNet employs an innovative strategy,

decomposing feature variables into a composite input encompassing multiple physical and temporal scales, accomplished through CEEMDAN. Furthermore, we employ the composite 2D convolutional structure, small, medium, and large, to identify interdependencies between a broader range of variables and aid in extracting spatiotemporal multiscale feature from the original signals. This strategy significantly enhances MsrtNet's ability to disentangle the intricate interplay inherent within complex chemical process data, which can effectively overcome the effect of variable sequence order on model performance.

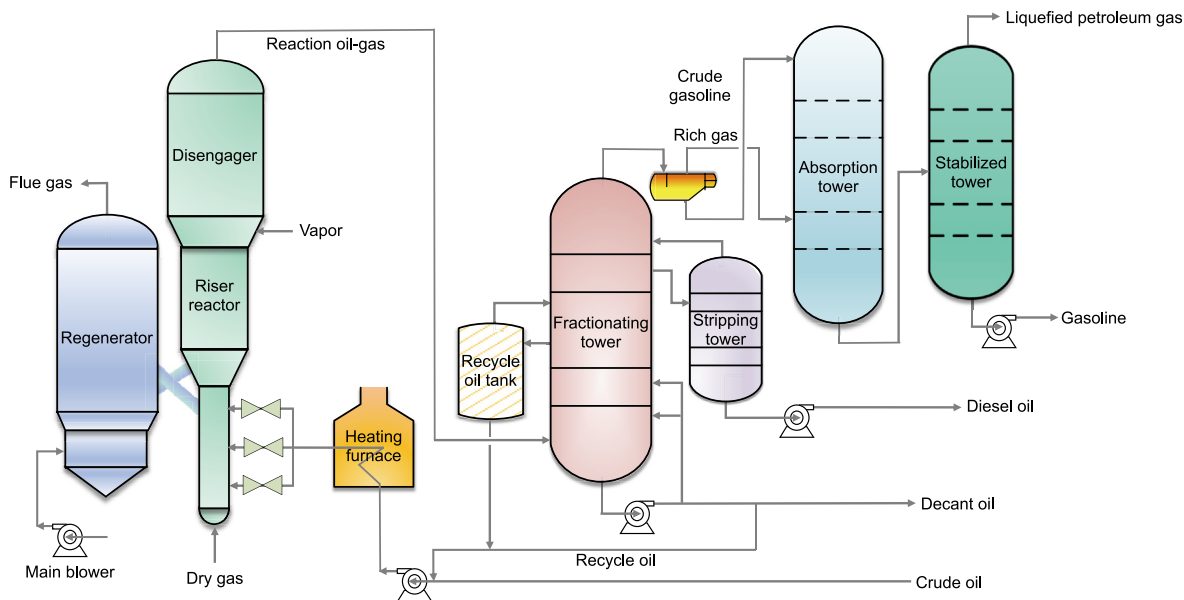


Fig. 17. Process diagram of the FCC unit.

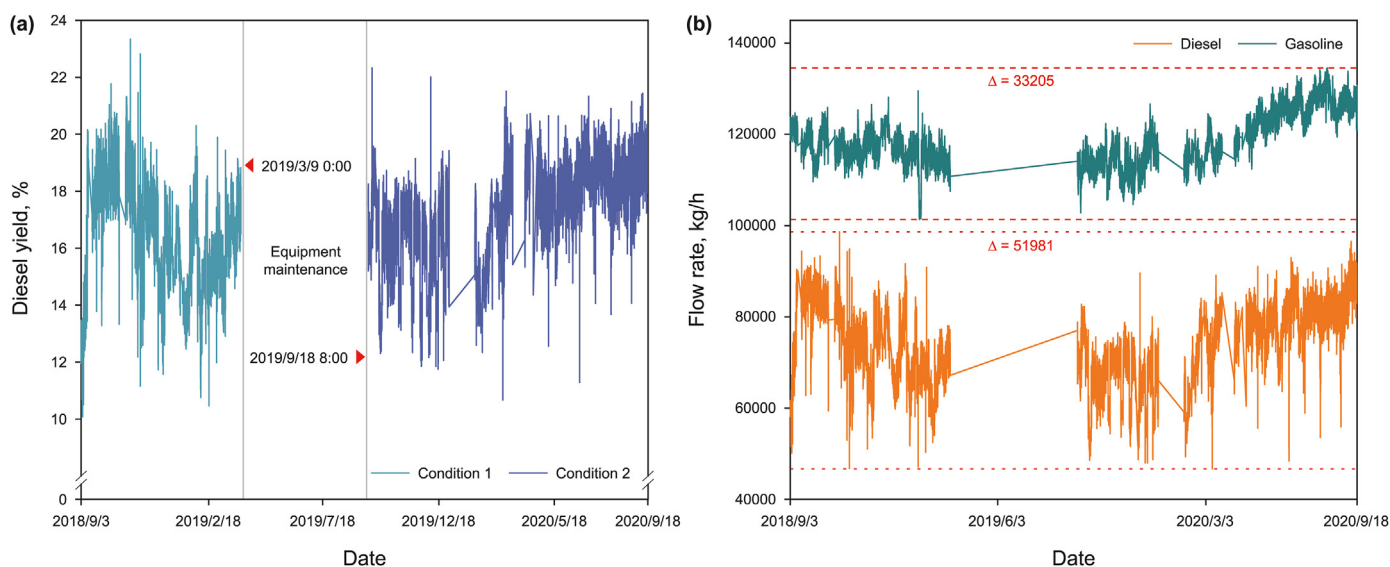


Fig. 18. Real sample values of the FCC unit.

4.2. MIP process

4.2.1. Process description

Heavy oil lightning technology is gaining increasing significance in the face of diminishing proven and exploited oil resources worldwide, coupled with a deterioration in crude oil quality. FCC has garnered significant attention among the various secondary processing and lightning techniques. The FCC unit encompasses reaction regeneration, distillation, and absorbing-stabilizing systems, as illustrated in Fig. 17. Light diesel is fed from the distillation tower to the light diesel stripper during FCC manufacture. The stripped light diesel is pumped to the diesel hydro-upgrading unit after heat exchange with the crude oil. Light diesel oil is pumped from the distillation tower to the stripping tower and pumped out by the light diesel oil pump. After heat exchange with the crude oil, one way is directly sent to the diesel oil hydro-upgrading unit; the

other is sent to the reabsorption tower as the absorber. The desorption effect of the desorption tower affects the quality of the lean gas, which makes the composition of the rich absorption oil at the bottom of the reabsorption tower significant change, and the rich absorption oil returns to the distillation tower, which will directly affect the heat balance and pressure of the fractionating tower, resulting in the fluctuation of the distillation tower performance. Moreover, the stability of installations is crucial for the refining and chemical industries in terms of safety and operational efficiency.

This study collects process data from a 2.80×10^6 t/a FCC unit in northwest China. Utilizing data from DCS and LIMS, 55 variables have been chosen as feature variables, categorized into two main groups: (1) the manipulated variable and (2) the feed property variable. Tables S4–S5 of the supplementary file provide a detailed description of these feature variables. The sampling interval is set at

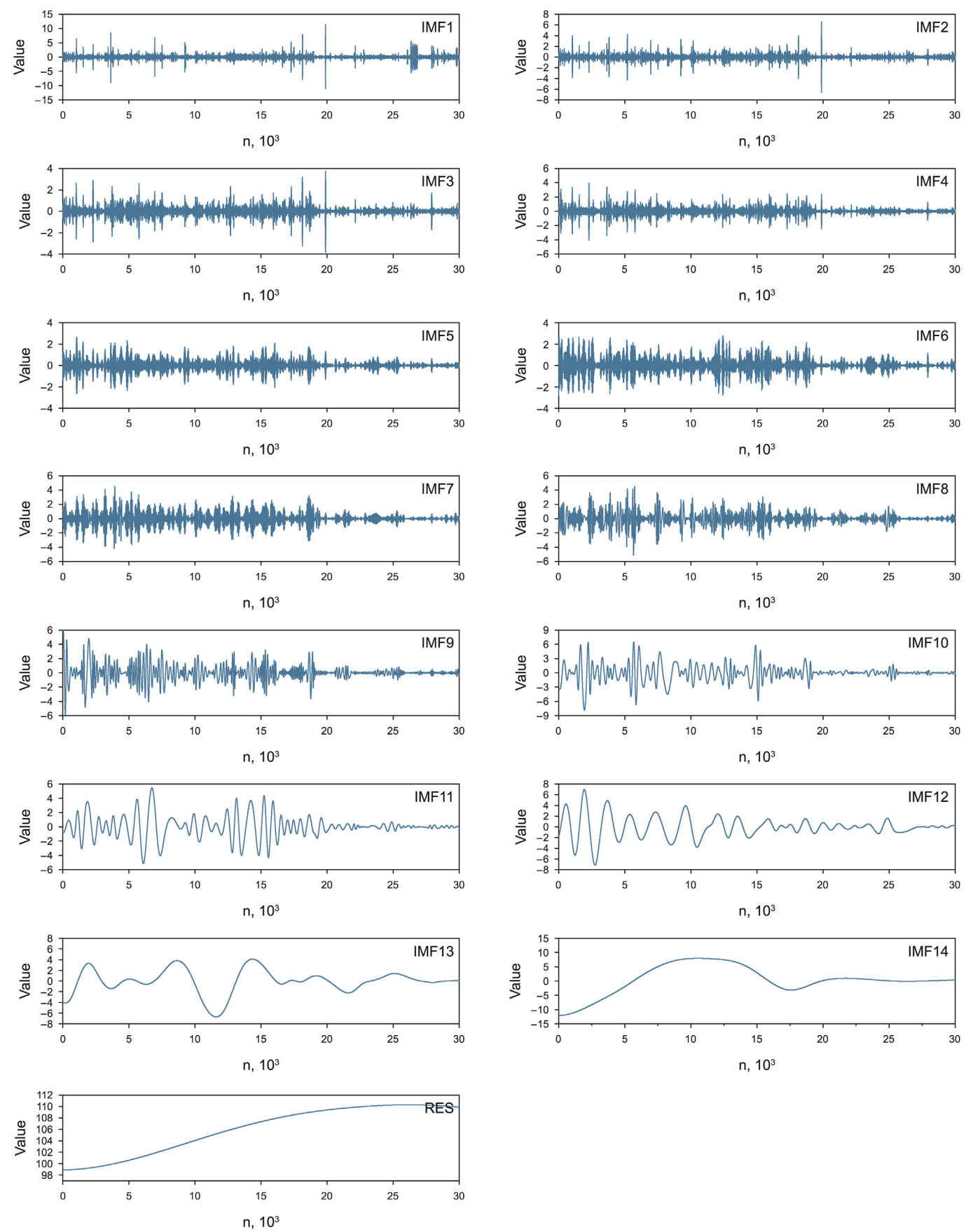


Fig. 19. Decomposition results of disengager charge.

Table 3
Multiscale-CNN module structure and parameters.

Model	Conv1 → P			Conv2 → P–M			FC1 → P–D	FC2 → P
	Size	Padding	Strides	Size	Padding	Strides	Units	Units
Model1	5 × 5	0 × 0	1 × 1	5 × 5	0 × 0	1 × 1	120	50
Model2	5 × 5	1 × 1	2 × 2	5 × 5	1 × 1	2 × 2	120	50
Model3	5 × 5	2 × 2	3 × 3	5 × 5	2 × 2	3 × 3	120	50

*P represents PReLU, M represents Mean-pooling, D represents Dropout, FC represents Fully Connected Layer.

Table 4
Adaptive-CNN module structure and parameters.

Model	Conv1 → P			Conv2 → P–M			FC1 → P	FC2 → P
	Size	Padding	Strides	Size	Padding	Strides	Units	Units
parameter dimension	1 × 1	0 × 0 1 → 16	1 × 1	1 × 1	0 × 0 16 → 1	1 × 1	150	50

*P represents PReLU, FC represents Fully Connected Layer.

Table 5
Prediction results of RNN, LSTM, 2DCNN and MsrtNet on the diesel yield of FCC.

Model	Training set		Testing set	
	RMSE	MAE	RMSE	MAE
RNN	0.110	0.096	0.176	0.135
LSTM	0.106	0.093	0.178	0.136
2DCNN	0.098	0.077	0.156	0.121
MsrtNet	0.041	0.043	0.107	0.092

Table 6
Training time of RNN, LSTM, 2DCNN and MsrtNet on the diesel yield of FCC.

Model	Training time	Unit
RNN	418	s
LSTM	1334	s
2DCNN	1474	s
MsrtNet	2678	s

20 min over 2 years, yielding approximately 30,000 data points. Data missing from the feature dataset for more than two days is defined as long-term missing and the rest as short-period missing. Short-period data that is missing is filled with data from adjacent periods to maintain the original fluctuation of the data. As shown in Fig. 18(a), between 0:00 on September 3, 2018 and 8:00 on September 18, 2020, the FCC process is considered to be in a different operating condition due to a long-period data missing in equipment maintenance. To test the robustness of MsrtNet, we refrain from processing this specific data scenario. The target variables, namely diesel and gasoline flow rates, are presented in Fig. 18(b). It is evident that both petrol and diesel flow data exhibit high-frequency variations, with the fluctuations in gasoline flow notably lower than those in diesel flow.

The prediction of diesel yield is closely related to changes in plant operating parameters, raw material properties, external environment, and other factors. When these complex factors change, the generalization performance of the data-driven model is put to a significant challenge. Resisting fluctuations in plant operating conditions is essential for the practical adaptation of MsrtNet to industrial applications. MsrtNet is used to capture the internal tangle of operating conditions and feedstock properties hidden in the dataset across time and space scales deeply. Adjusting the network structure and optimizing the hyperparameters resulted in a diesel yield prediction model with prediction accuracy and robustness, exceptionally superior in capturing fluctuation details.

4.2.2. Model construction

CEEMDAN can obtain the instantaneous frequency components of the original data. Each component represents the temporal partial feature of the periodic and no mode mixing, thus achieving a high-resolution time-frequency analysis with mining spatiotemporal multiscale information. As the frequency of each component decreases, each component contains information about the fluctuations of the original sequence. The decomposed sequences correspond to short-term fluctuation trends, medium-term excess trends, and smooth long-term trends, and the decomposed sequences have a more substantial regularity than the original sequences. IMF1 has the highest frequency and energy, generally removed as a noise component. Eventually, the feature dimension increases from 55 to 767. The disengager charge decomposition results of the IMFs are shown in Fig. 19.

Consistent with the principle of Section 4.1.2, the epoch is set to 500. In addition, the minimum batch size is set to 512 for each training period, and the learning rate is 0.001. To reduce the value of the loss function, the RMSProp optimizer is chosen to optimize the model by iteratively updating the neural network weights based on the training set. Similarly, Table 3 provides three sets of network parameters for the Multiscale-CNN module. Among them, the Dropout parameters are set to 0.5, the number of convolution layer channels is 6, 6, and 16, and the pooling layer size and strides are 2 × 2. In the training process, the epoch is 500, the minimum batch for each training period is 50, the learning rate is 0.001, and the optimizer selects RMSProp. The prediction performance is quantitatively evaluated using RMSE and MAE.

The Adaptive-CNN module utilizes two convolutional layers (Conv1 and Conv2) and two fully-connected layers (FC1 and FC2) for feature extraction and transformation, as shown in Table 4. Conv1 employs a 1 × 1 filter size, no padding (Padding = 0 × 0), and a step size 1 × 1 to preserve crucial information from the input feature. Conv2 further augments the feature dimensionality by using a 1 × 1 filter size, no padding, and a step size 1 × 1, enhancing feature expressiveness while maintaining the input information. Regarding the fully connected layers, FC1 and FC2 are 150 and 50 units, respectively. They undergo non-linear transformation through the Tanh activation function, which aids in extracting higher-order relationships between the input feature, thereby boosting the model's expressive capacity. Additionally, the model's parameter dimensions are 1–16 and 16–1, respectively, indicating high parameter efficiency and feature extraction capability.

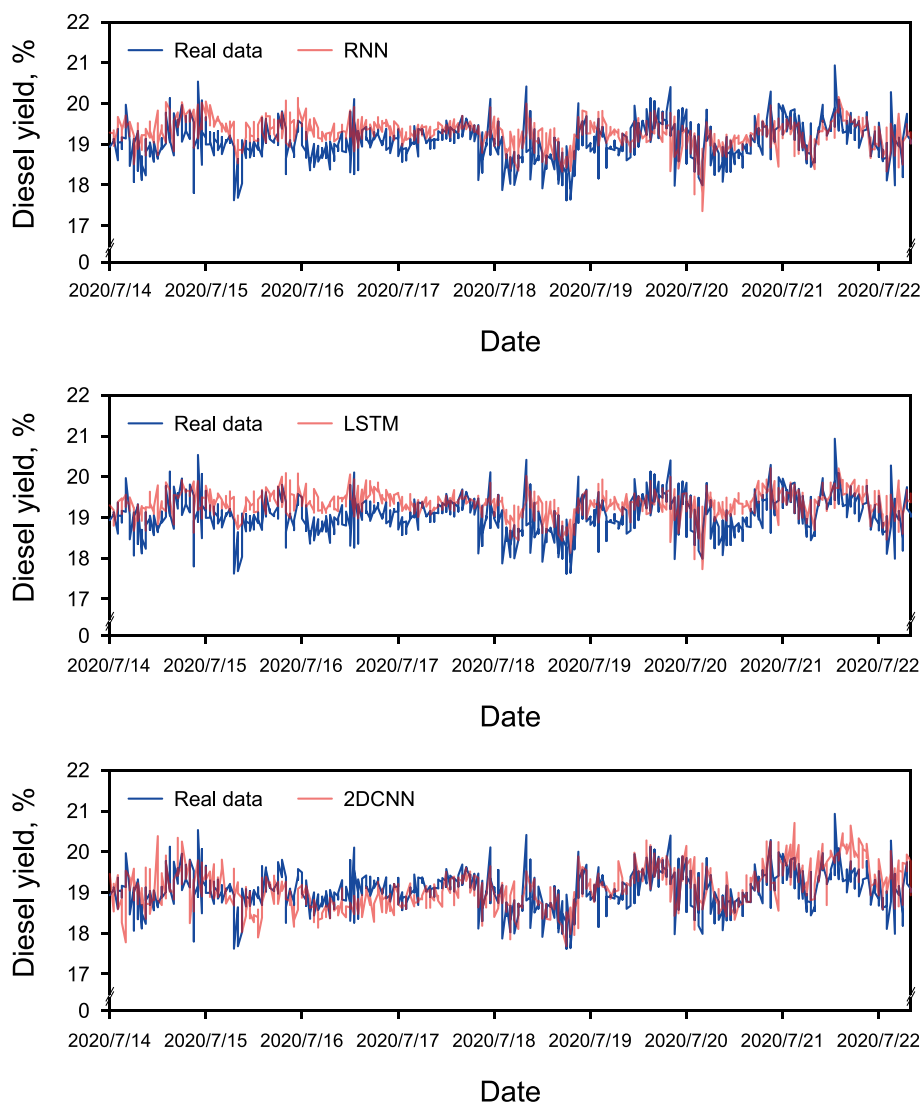


Fig. 20. Detailed prediction results of diesel yield by RNN, LSTM, and 2DCNN.

4.2.3. Results and discussion

As indicated in Table 5, we evaluated of four distinct neural network models. Among these, MsrtNet demonstrates the most noteworthy performance on the training set, exhibiting the lowest RMSE and MAE, underscoring its robust fitting capability to the training data. Notably, MsrtNet's performance on the test set remains highly commendable, boasting relatively low RMSE and MAE values, indicative of its exceptional generalization proficiency on new data. In contrast, the other three models exhibit relatively limited performance on the test set.

MsrtNet is run on Intel(R) Core(TM) i9-12900K (24CPUs) @ 3.2 GHz 128G RAM, Windows 10. As indicated in Table 6, the training time for MsrtNet is 2678 s using the same computational resources. When comparing Tables 5 and 6 results, MsrtNet demonstrates a lower prediction error. However, its training time is less advantageous than the other three models. There is a trade-off between model size and performance in practical industrial applications. MsrtNet is targeted in industrial applications for scenarios that require high model performance, such as advanced control, intelligent monitoring, and fault diagnosis. Considering the industrial application context of MsrtNet and comparing it with the 2DCNN models reported in the literature at this stage, the training

time remains acceptable (Ning et al., 2023; Singh et al., 2020).

Fig. 20 further highlights the performance of each model by exhibiting the results of diesel yield prediction and actual values of RNN, LSTM, and 2DCNN to observe the distribution intuitively. RNN has the disadvantage of capturing long-term dependencies in sequence data, making the performance of the training set of RNN poorer than LSTM. The generalization ability of RNN is superior to LSTM, as demonstrated by the higher test set accuracy of RNN than LSTM. Although the RNN has only short-term memory compared to the LSTM, it performs better at local details while ensuring trend accuracy. The LSTM prediction is relatively smooth, lacking details of local fluctuation. 2DCNN demonstrates comparable overall prediction accuracy to RNN and LSTM, but it notably excels in accurately capturing the finer details of high-frequency fluctuations. This is particularly evident between 19th and 21st July 2020, where the predicted values of 2DCNN closely align with the fluctuation trend of the actual values, showcasing its proficiency in capturing nuanced variations.

The absence of overfitting in MsrtNet is further demonstrated by expanding the diversity and size of the dataset in confluence with the operational characteristics of the FCC unit shown in Fig. 18(a). The prediction results of MsrtNet are illustrated in Fig. 21(a)–(c).

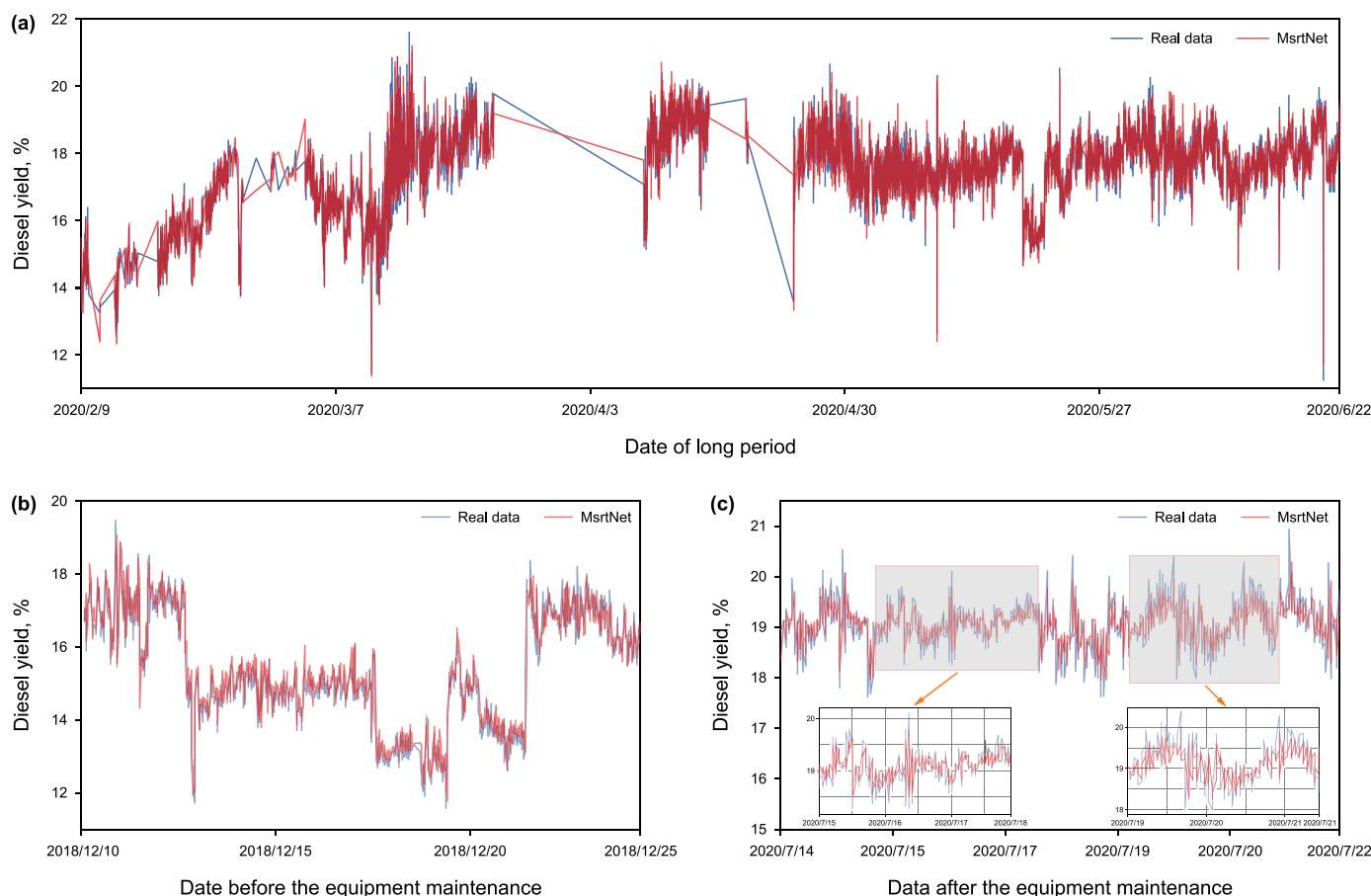


Fig. 21. Detailed prediction results of diesel yield by MsrtNet.

Fig. 21(b) and (c) depict the test results under different operational conditions before and after the equipment maintenance, while Fig. 21(a) shows the test results over a long period. MsrtNet exhibits outstanding prediction accuracy and the capability to capture high-frequency details. As illustrated from the partially enlarged detail in Fig. 21(c), the prediction for the long-period low amplitude high-frequency details from 15th to 18th July exhibits excellent accuracy. Meanwhile, the long-period high amplitude high-frequency details prediction from 19th to 21st July performs well. It demonstrates that MsrtNet is adept at handling strongly nonlinear FCC units and exhibits robustness across various operating conditions.

The generalizability of MsrtNet is a crucial indicator of its adaptability to the industrial environment. We include a case study on predicting gasoline yield to assess this further. Since crude petrol and diesel are both products of fractionation systems, the datasets used remain consistent. The fundamental architecture and parameters of MsrtNet are not adapted, corroborating the transferability of the model. Given that issues related to productivity prediction often involve high-frequency data fluctuations, a lower learning rate is deemed necessary. This selection is evident in Fig. 18(b), where the final learning rate is appropriately tiny due to the reduced fluctuations in gasoline yield compared to diesel. The prediction results obtained by adjusting the epoch and learning rate to 500 and 0.0001 are depicted in Fig. 22(a)–(c). Consistent with the performance validation process for the diesel case, MsrtNet maintains not only stable performance in long-period prediction, as shown in Fig. 22(a), but also fits high-frequency details very well, as demonstrated in the partially enlarged detail of Fig. 22(a) and (b).

MsrtNet decomposes a single time-series data into a series of IMFs, discarding the components containing high-frequency random noise. Extracting feature across temporal and spatial scales with different convolutional kernel structures can better capture the deeper patterns of the industrial data and decouple the complex non-linear relationships between variables. In general, MsrtNet, the adaptive multiscale convolutional neural network with multiple time and energy scales, can better predict long-period, multi-condition, non-linear industrial time series data. MsrtNet's prediction overlaps better with the actual values because it perceives more splicing between variables by setting up convolutional kernels with different sizes of receptive fields accompanied by different moving strides. The local information is then spatiotemporally integrated into higher dimensions, which better perceives the coupled information of the industrial data.

The fundamental ideology of MsrtNet involves creating the composite 2D convolutional structure for spatiotemporal multi-scale feature extraction through various padding and stride configurations. The Multiscale-CNN module in MsrtNet utilizes three convolutional structures, providing flexibility that can be tailored to meet the requirements of industrial processes. The formulas for the associated convolutional and pooling layers are presented in Eqs. (8) and (9), with the relevant parameters detailed in Tables 3 and 4. The modest parameter count of MsrtNet, totaling 635,092, underscores the model's capabilities in capturing intricate relationships, offering flexibility, and ensuring generalizability.

The crucial aspect of MsrtNet in processing complex industrial data is tuning the convolutional and pooling layers, which is accomplished during offline modeling. The performance of MsrtNet

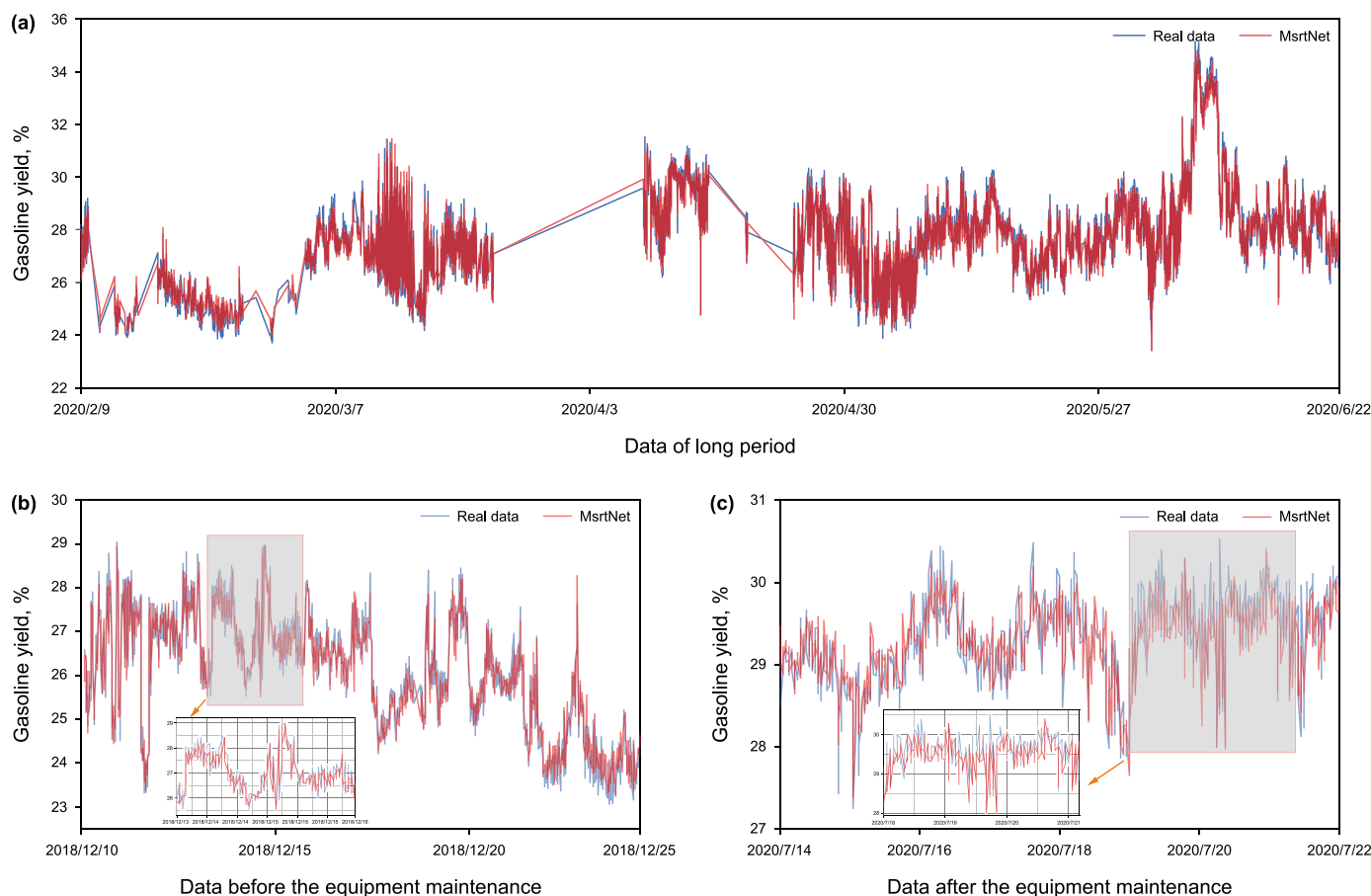


Fig. 22. Detailed prediction results of gasoline yield by MsrtNet.

in FCC complex processes suggests that the three convolutional structures will meet the requirements of most industrial monitoring scenarios, with a high likelihood of structural simplification for other situations. The Multiscale-CNN and Adaptive-CNN modules of the MsrtNet are trained offline to adapt to the process without altering their fundamental architecture and parameters. Once deployed in an industrial system, the epochs and learning rate only need to be periodically adjusted offline to maintain MsrtNet performance. The optimization period needs to be determined in conjunction with the industrial site. Furthermore, reference values for epochs and learning rates convenient for industrial site tuning can be determined based on the process and data characteristics during the offline training. For instance, the epoch for diesel and gasoline yield prediction is 500, with learning rates of 0.001 and 0.0001, respectively. In conclusion, the lightweight architecture and tuning pattern of MsrtNet ensure responsiveness, robustness, and transferability in practical industrial scenarios.

5. Conclusion

Accurate prediction for product yield is of great significance to the long-period stable and safe operation of the FCC unit. Multiscale structure of MsrtNet measures the accuracy of the high-frequency detail capture and long-period fluctuation trend prediction in monitoring critical chemical industry sites. According to the research conducted in this paper, the following conclusions are obtained.

- MsrtNet introduces CEEMDAN to denoise the feature dataset and decompose the signal at the spatiotemporal scale. On this basis, a Multiscale-CNN module is constructed to capture feature on spatiotemporal states of variable length. Then, the Adaptive-CNN module receives the feature map obtained by the Multiscale-CNN module and integrates the target feature from all receptive fields through non-linear weighting in higher dimensions to complete the feature fusion from the entire feature subset.
- The significance of capturing multiscale feature is illustrated in validating the TE case. It is demonstrated that the sequence order of the model influences the performance of time series models, underlining the importance of multiscale feature extraction for developing effective data-driven models. The study demonstrates that MsrtNet enhances feature identifiability and mitigates the impact of feature sequence order. It excels at capturing spatiotemporal multiscale feature, as evidenced by its proficient prediction performance in accurately reproducing high-frequency details.
- The performance of MsrtNet is validated using two years of operation data from a 2.80×10^6 t/a refinery. In particular, we consider the missing data due to equipment maintenance and the condition shifts for the long-period operation, and the performance of MsrtNet is unaffected. Compared to LSTM, the optimal prediction results are reduced by 39.88% and 32.35% in RMSE and MAE, respectively. In summary, MsrtNet shows excellent robustness and transferability on industrial data and can be

used as a fundamental monitoring model in different industrial scenarios.

CRedit authorship contribution statement

Nan Liu: Writing – review & editing, Writing – original draft.
Chun-Meng Zhu: Methodology. **Meng-Xuan Zhang:** Supervision.
Xing-Ying Lan: Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.petsci.2024.01.014>.

References

- Andooz, A., Egbalpour, M., Kowsari, E., Ramakrishna, S., Ansari Cheshmeh, Z., 2023. A comprehensive review on pyrolysis from the circular economy point of view and its environmental and social effects. *J. Clean. Prod.* 388, 136021. <https://doi.org/10.1016/j.jclepro.2023.136021>.
- Ascher, S., Wang, X., Watson, I., Sloan, W., You, S., 2022. Interpretable machine learning to model biomass and waste gasification. *Bioresour. Technol.* 364, 128062. <https://doi.org/10.1016/j.biortech.2022.128062>.
- Byun, M., Lee, H., Choe, C., Cheon, S., Lim, H., 2021. Machine learning based predictive model for methanol steam reforming with technical, environmental, and economic perspectives. *Chem. Eng. J.* 426, 131639. <https://doi.org/10.1016/j.cej.2021.131639>.
- Chai, T., Ding, J., 2018. Smart and optimal manufacturing for process industry. *Chinese Journal of Engineering Science* 20, 51. <https://doi.org/10.15302/j-SSCAE-2018.04.009>.
- Chen, J., Lin, C., Yao, B., Yang, L., Ge, H., 2023. Intelligent Fault diagnosis of rolling bearings with low-quality data: a feature significance and diversity learning method. *Reliab. Eng. Syst. Saf.* 109343. <https://doi.org/10.1016/j.ress.2023.109343>.
- Cui, Y., Kara, S., Chan, K.C., 2020. Manufacturing big data ecosystem: a systematic literature review. *Robot. Comput. Integrated Manuf.* 62, 101861. <https://doi.org/10.1016/j.rcim.2019.101861>.
- Gulay, E., Duru, O., 2020. Hybrid modeling in the predictive analytics of energy systems and prices. *Appl. Energy* 268, 114985. <https://doi.org/10.1016/j.apenergy.2020.114985>.
- He, G., Dang, Y., Zhou, L., Dai, Y., Que, Y., Ji, X., 2020. Architecture model proposal of innovative intelligent manufacturing in the chemical industry based on multi-scale integration and key technologies. *Comput. Chem. Eng.* 141, 106967. <https://doi.org/10.1016/j.compchemeng.2020.106967>.
- Hong, B.Y., Liu, S.N., Li, X.P., Fan, D., Ji, S.P., Chen, S.H., Li, C.C., Gong, J., 2022. A liquid loading prediction method of gas pipeline based on machine learning. *Pet. Sci.* 19 (6), 3004–3015. <https://doi.org/10.1016/j.petsci.2022.05.002>.
- Kang, J., Li, N.Y., Zhao, L.Q., Xiong, G., Wang, D.C., Xiong, Y., Luo, Z.F., 2022. Construction of complex digital rock physics based on full convolution network. *Pet. Sci.* 19 (2), 651–662. <https://doi.org/10.1016/j.petsci.2021.11.018>.
- Kawai, E., Sato, H., Furuichi, K., Toru, T., Yoshioka, T., 2022. Maximizing margins and optimizing operational conditions for residue fluid catalytic cracking with an artificial intelligence hybrid reaction model. *J. Adv. Manuf. & Process* 4, e10118. <https://doi.org/10.1002/amp2.10118>.
- Khalidi, M.K., Al-Dhaifallah, M., Taha, O., 2023. Artificial intelligence perspectives: a systematic literature review on modeling, control, and optimization of fluid catalytic cracking. *Alex. Eng. J.* 80, 294–314. <https://doi.org/10.1016/j.aej.2023.08.066>.
- Lee, J., Cameron, I., Hassall, M., 2019. Improving process safety: what roles for Digitalization and Industry 4.0? *Process Saf. Environ. Protect.* 132, 325–339. <https://doi.org/10.1016/j.psep.2019.10.021>.
- Li, C.Q., Chen, Y., Shang, Y., 2022. A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal* 29, 101021. <https://doi.org/10.1016/j.jestch.2021.06.001>.
- Li, Chang, Huang, X., Song, R., Qian, R., Liu, X., Chen, X., 2022. EEG-based seizure prediction via Transformer guided CNN. *Measurement* 203, 111948. <https://doi.org/10.1016/j.measurement.2022.111948>.
- Li, D., 2016. Perspective for smart factory in petrochemical industry. *Comput. Chem. Eng.* 91, 136–148. <https://doi.org/10.1016/j.compchemeng.2016.03.006>.
- Li, D., Jiang, B., Suo, H., Guo, Y., 2015. Overview of smart factory studies in petrochemical industry. In: *Computer Aided Chemical Engineering*. Elsevier, pp. 71–76. <https://doi.org/10.1016/B978-0-444-63578-5.50009-8>.
- Li, Q., Zhang, M., Shi, X., Lan, X., Guo, X., Guan, Y., 2021. An intelligent hybrid feature subset selection and production pattern recognition method for modeling ethylene plant. *J. Anal. Appl. Pyrol.* 160, 105352. <https://doi.org/10.1016/j.jaap.2021.105352>.
- Liu, W., Chen, Z., Hu, Y., Xu, L., 2023. A systematic machine learning method for reservoir identification and production prediction. *Petrol. Sci.* 20 (1), 295–308. <https://doi.org/10.1016/j.petsci.2022.09.002>.
- Liu, W.Y., Gao, Q.W., Ye, G., Ma, R., Lu, X.N., Han, J.G., 2015. A novel wind turbine bearing fault diagnosis method based on Integral Extension LMD. *Measurement* 74, 70–77. <https://doi.org/10.1016/j.measurement.2015.06.005>.
- Liu, Z., Cai, W., Xu, Z.Q., 2020. Multi-scale deep neural network (MscaledDNN) for solving Poisson-Boltzmann equation in complex domains. *CICP* 28, 1970–2001. <https://doi.org/10.4208/cicp.OA-2020-0179>.
- Liu, J., Huang, X., Li, Q., Chen, Z., Liu, G., Tai, Y., 2023. Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD. *Energy Convers. Manag.* 280. <https://doi.org/10.1016/j.enconman.2023.116804>.
- Lung, S.-Y., 2007. Wavelet feature domain adaptive noise reduction using learning algorithm for text-independent speaker recognition. *Pattern Recogn.* 40, 2603–2606. <https://doi.org/10.1016/j.patcog.2007.01.028>.
- Min, Q., Lu, Y., Liu, Z., Su, C., Wang, B., 2019. Machine learning based digital twin framework for production optimization in petrochemical industry. *Int. J. Inf. Manag.* 49, 502–519. <https://doi.org/10.1016/j.iinfomgt.2019.05.020>.
- Mohammadpoor, M., Torabi, F., 2020. Big Data analytics in oil and gas industry: an emerging trend. *Petroleum* 6, 321–328. <https://doi.org/10.1016/j.petlm.2018.11.001>.
- Ni, P., Liu, B., He, G., 2021. An online optimization strategy for a fluid catalytic cracking process using a case-based reasoning method based on big data technology. *RSC Adv.* 11, 28557–28564. <https://doi.org/10.1039/D1RA03228C>.
- Ning, C., Xie, Y., Sun, L., 2023. LSTM, WaveNet, and 2D CNN for nonlinear time history prediction of seismic responses. *Eng. Struct.* 286, 116083. <https://doi.org/10.1016/j.engstruct.2023.116083>.
- Peng, S., Chen, R., Yu, B., Xiang, M., Lin, X., Liu, E., 2021. Daily natural gas load forecasting based on the combination of long short term memory, local mean decomposition, and wavelet threshold denoising algorithm. *J. Nat. Gas Sci. Eng.* 95, 104175. <https://doi.org/10.1016/j.jngse.2021.104175>.
- Qu, H.Y., Zhang, J.L., Zhou, F.J., Peng, Y., Pan, Z.J., Wu, X.Y., 2023. Evaluation of hydraulic fracturing of horizontal wells in tight reservoirs based on the deep neural network with physical constraints. *Pet. Sci.* 20 (2), 1129–1141. <https://doi.org/10.1016/j.petsci.2023.03.015>.
- Rayhan Ahmed, Md, Islam, S., Muzahidul Islam, A.K.M., Shatabda, S., 2023. An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition. *Expert Syst. Appl.* 218, 119633. <https://doi.org/10.1016/j.eswa.2023.119633>.
- Singer, G., Cohen, Y., 2021. A framework for smart control using machine-learning modeling for processes with closed-loop control in Industry 4.0. *Eng. Appl. Artif. Intell.* 102, 104236. <https://doi.org/10.1016/j.engappai.2021.104236>.
- Singh, N., Vyjayanthi, C., Modi, C., 2020. Multi-step short-term electric load forecasting using 2D convolutional neural networks. In: *2020 IEEE-HYDICON*. Presented at the 2020 IEEE-HYDICON. IEEE, Hyderabad, India, pp. 1–5. <https://doi.org/10.1109/HYDICON48903.2020.9242917>.
- Soni, D., Kumar, N., 2022. Machine learning techniques in emerging cloud computing integrated paradigms: a survey and taxonomy. *J. Netw. Comput. Appl.* 205, 103419. <https://doi.org/10.1016/j.jnca.2022.103419>.
- Stratiev, D., Ivanov, M., Chavdarov, I., Argirov, G., Strovegli, G., 2023. Revamping fluid catalytic cracking unit, and optimizing catalyst to process heavier feeds. *Appl. Sci.* 13 (2017). <https://doi.org/10.3390/app13032017>.
- Sun, B., Liu, X., Wang, J., Wei, X., Yuan, H., Dai, H., 2023. Short-term performance degradation prediction of a commercial vehicle fuel cell system based on CNN and LSTM hybrid neural network. *Int. J. Hydrogen Energy* 48, 8613–8628. <https://doi.org/10.1016/j.ijhydene.2022.12.005>.
- Sun, C., Luo, Y., Huang, Y., Ouyang, X., 2017. A comparative study on the production efficiencies of China's oil companies: a true fixed effect model considering the unobserved heterogeneity. *J. Clean. Prod.* 154, 341–352. <https://doi.org/10.1016/j.jclepro.2017.03.222>.
- Taira, G.R., Park, S.W., Zanin, A.C., Porfirio, C.R., 2022. Fault detection in a fluid catalytic cracking process using Bayesian Recurrent Neural Network. *IFAC-PapersOnLine* 55, 715–720. <https://doi.org/10.1016/j.ifacol.2022.07.528>.
- Teh, H.Y., Kempa-Liehr, A.W., Wang, K.I.K., 2020. Sensor data quality: a systematic review. *J. Big Data* 7 (11). <https://doi.org/10.1186/s40537-020-0285-1>.
- Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Presented at the ICASSP 2011–2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Prague, Czech Republic, pp. 4144–4147. <https://doi.org/10.1109/ICASSP.2011.5947265>.
- Usman, A.G., Tanimu, A., Abba, S.I., Isik, S., Aitani, A., Alasiri, H., 2023. Feasibility of the optimal design of AI-Based Models integrated with ensemble machine learning paradigms for modeling the yields of light olefins in crude-to-chemical conversions. *ACS Omega* 8, 40517–40531. <https://doi.org/10.1021/acsomega.3c05227>.
- Wu, C.L., Chau, K.W., 2010. Data-driven models for monthly streamflow time series prediction. *Eng. Appl. Artif. Intell.* 23, 1350–1367. <https://doi.org/10.1016/j.engappai.2010.04.003>.

- Xie, J., Yang, R., Gooi, H.B., Nguyen, H.D., 2023. PID-based CNN-LSTM for accuracy-boosted virtual sensor in battery thermal management system. *Appl. Energy* 331, 120424. <https://doi.org/10.1016/j.apenergy.2022.120424>.
- Xie, Y., Sun, W., Ren, M., Chen, S., Huang, Z., Pan, X., 2023. Stacking ensemble learning models for daily runoff prediction using 1D and 2D CNNs. *Expert Syst. Appl.* 217, 119469. <https://doi.org/10.1016/j.eswa.2022.119469>.
- Xu, Z.-Q.J., 2020. Frequency principle: fourier analysis sheds light on deep neural networks. *CiCP* 28, 1746–1767. <https://doi.org/10.4208/cicp.OA-2020-0085>.
- Yang, F., Dai, C., Tang, J., Xuan, J., Cao, J., 2020. A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance. *Chem. Eng. Res. Des.* 155, 202–210. <https://doi.org/10.1016/j.cherd.2020.01.013>.
- Yang, F., Xu, M., Lei, W., Lv, J., 2023. Artificial intelligence methods applied to catalytic cracking processes. *Big Data Min. Anal.* 6, 361–380. <https://doi.org/10.26599/BDMA.2023.9020002>.
- Yang, T., Yi, X., Lu, S., Johansson, K.H., Chai, T., 2021. Intelligent manufacturing for the process industry driven by industrial artificial intelligence. *Engineering* 7, 1224–1230. <https://doi.org/10.1016/j.eng.2021.04.023>.
- Yao, Z., Zhao, C., 2022. FedTMI: knowledge aided federated transfer learning for industrial missing data imputation. *J. Process Control* 117, 206–215. <https://doi.org/10.1016/j.jprocont.2022.08.004>.
- Yuan, Z., Qin, W., Zhao, J., 2017. Smart manufacturing for the oil refining and petrochemical industry. *Engineering* 3, 179–182. <https://doi.org/10.1016/J.ENG.2017.02.012>.
- Zhang, M., Cao, D., Lan, X., Shi, X., Gao, J., 2022. An ensemble-learning approach to predict the coke yield of commercial FCC unit. *Ind. Eng. Chem. Res.* 61, 8422–8431. <https://doi.org/10.1021/acs.iecr.1c04735>.
- Zhang, M., Yang, Z., Zhao, Y., Lv, M., Lan, X., Shi, X., Gao, J., Li, C., Yuan, Z., Lin, Y., 2023. A hybrid safety monitoring framework for industrial FCC disengager coking rate based on FPM, CFD, and ML. *Process Saf. Environ. Protect.* 175, 17–33. <https://doi.org/10.1016/j.psep.2023.05.004>.
- Zhang, Z., Zhu, J., Zhang, S., Gao, F., 2023. Process monitoring using recurrent Kalman variational auto-encoder for general complex dynamic processes. *Eng. Appl. Artif. Intell.* 123, 106424. <https://doi.org/10.1016/j.engappai.2023.106424>.
- Zhao, Z., Yun, S., Jia, L., Guo, J., Meng, Y., He, N., Li, X., Shi, J., Yang, L., 2023. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng. Appl. Artif. Intell.* 121, 105982. <https://doi.org/10.1016/j.engappai.2023.105982>.
- Zheng, Z., Ali, M., Jamei, M., Xiang, Y., Karbasi, M., Yaseen, Z.M., Farooque, A.A., 2023. Design data decomposition-based reference evapotranspiration forecasting model: a soft feature filter based deep learning driven approach. *Eng. Appl. Artif. Intell.* 121, 105984. <https://doi.org/10.1016/j.engappai.2023.105984>.