Chinese Journal of Chemical Engineering 84 (2025) 35–46

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Full Length Article

Hierarchical framework for predictive maintenance of coking risk in fluid catalytic cracking units: A data and knowledge-driven method

Nan Liu [1], Chunmeng Zhu [1, 2], Zeng Li [1], Yunpeng Zhao [1], Xiaogang Shi [1], Xingying Lan [1, 2, *]

[1] *State Key Laboratory of Heavy Oil Processing, China University of Petroleum (Beijing), Beijing 102249, China*
[2] *College of Artificial Intelligence, China University of Petroleum (Beijing), Beijing 102249, China*

ARTICLE INFO

Article history:
Received 22 December 2024
Received in revised form
24 April 2025
Accepted 28 April 2025
Available online 21 May 2025

Keywords:
Petroleum
Mixed-frequency data
Coking
Risk index
Neural networks
Predictive maintenance

ABSTRACT

The fractionating tower bottom in fluid catalytic cracking Unit (FCCU) is highly susceptible to coking due to the interplay of complex external operating conditions and internal physical properties. Consequently, quantitative risk assessment (QRA) and predictive maintenance (PdM) are essential to effectively manage coking risks influenced by multiple factors. However, the inherent uncertainties of the coking process, combined with the mixed-frequency nature of distributed control systems (DCS) and laboratory information management systems (LIMS) data, present significant challenges for the application of data-driven methods and their practical implementation in industrial environments. This study proposes a hierarchical framework that integrates deep learning and fuzzy logic inference, leveraging data and domain knowledge to monitor the coking condition and inform prescriptive maintenance planning. The framework proposes the multi-layer fuzzy inference system to construct the coking risk index, utilizes multi-label methods to select the optimal feature dataset across the reactor-regenerator and fractionation system using coking risk factors as label space, and designs the parallel encoder-integrated decoder architecture to address mixed-frequency data disparities and enhance adaptation capabilities through extracting the operation state and physical properties information. Additionally, triple attention mechanisms, whether in parallel or temporal modules, adaptively aggregate input information and enhance intrinsic interpretability to support the disposal decision-making. Applied in the 2.8 million tons FCCU under long-period complex operating conditions, enabling precise coking risk management at the fractionating tower bottom.

1. Introduction

Fluid catalytic cracking unit (FCCU) is the critical secondary process in petroleum refining, essential for converting heavy oil into lighter, more valuable products [1]. The deterioration of crude oil quality, complex process mechanisms, and harsh operating conditions in the FCCU result in over one-third of alarms and more than half of unplanned shutdowns, significantly increasing the risk of incidents [2]. Among them, coking failure deteriorates operating conditions, progressively decreasing decant oil circulation, reducing steam output, causing pipeline blockages, darkening diesel oil color, and ultimately limiting qualified product output and posing safety risks. The typical coking risk affecting the long period stable operation in the FCCU arises from the fractionating tower bottom and the decant oil circulation system. This coking failure increases energy consumption, reduces processing capacity, and disrupts the heat and gas-liquid balance, as excess heat from the decant oil cannot be removed. Such imbalances can result in tower flooding, reaction pressure fluctuations, and, in severe cases, unplanned shutdowns.

Maintenance strategies in oil and gas industry have shifted from traditional reactive approaches to predictive and prescriptive strategies, fundamentally transforming FCCU operations and maintenance [3,4]. This evolution aims to optimize decision-making in predictive maintenance (PdM) [5], supporting activity planning by leveraging insights from historical

This article is part of a special issue entitled: AI for Chemical Engineering published in Chinese Journal of Chemical Engineering.

* Corresponding author. State Key Laboratory of Heavy Oil Processing, China University of Petroleum (Beijing), Fuxue Road 18, Changping District, Beijing 102249, China.

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

https://doi.org/10.1016/j.cjche.2025.04.015
1004-9541/

degradation patterns and forecasting future operation states [6,7]. Effective use of data collected through digital transformation requires specialized knowledge to enhance decision-making, as maintenance must navigate numerous constraints from process mechanisms and operational adjustment, complicating the formulation and execution of maintenance plans [8,9]. To address this, multi-criteria decision analysis methods allow for independent assessments of the maintenance landscape [10]. Traditional analysis methods, however, often fall short due to data imprecision and uncertainty, motivating the adoption of fuzzy logic (FL) within PdM [11]. With applications in risk assessment [12], FL supports robust decision-making in complex industrial scenarios, transforming raw industrial data into actionable risk indicators by encoding expert knowledge in fuzzy sets and rules, managing semantic uncertainty effectively [13,14].

The complex external operation state and internal physical properties contributing to fractionating tower bottom coking failure necessitate the joint use of distributed control systems (DCS) and laboratory information management systems (LIMS) data to assess the coking condition accurately. This means that credible coking condition requires a comprehensive evaluation of operation states and physical properties. The encoder-decoder architecture effectively accommodates the mixed-frequency nature of DCS and LIMS data, with multiple encoder modules extracting distinct feature vectors to enhance information extraction and fusion [15–17]. This diversified feature learning captures a comprehensive operation state of the fractionating tower bottom and the decant oil circulation system. Specifically, high-quality feature embedding generated by the encoder supports reliable coking condition monitoring as the decoder separately processes each block before integrating their outputs [18]. To achieve this, the deep learning (DL) architecture extends feature vectors *via* concatenation, facilitating extracting features related to operation state and physical properties across parallel blocks [16,19]. Integrating multi-label feature selection (MFS) with architecture design, grounded in process mechanisms, enables optimal feature representation for comprehensive coking condition monitoring, capturing the full range of critical risk factors [20]. MFS methods-filter, wrapper, and embedded-differ in their interaction with learning algorithms. Filter methods, independent of specific algorithms, assign weights to feature dimensions using information theory or statistical techniques. This weighting aids interpretability, making it essential for models that incorporate expert knowledge and require transparency [21].

This study proposes a hierarchical framework combining data-driven DL and knowledge-driven FL methods, enabling a qualitative-to-quantitative transition in coking condition monitoring. and successfully applied it to optimize coking risk management in a large-scale refinery. The main contributions are highlighted below:

1) Developed a multi-layer fuzzy inference system to construct a coking risk index, effectively addressing the inherent uncertainty of coking processes.
2) Designed a parallel encoder-integrated decoder architecture with triple attention mechanisms (AMs) to extract and fuse features from reactor-regenerator and fractionation systems, resolving the challenge of mixed-frequency in DCS and LIMS data.
3) By integrating data-driven models with domain knowledge, the framework achieves a qualitative-to-quantitative shift in coking condition monitoring, enabling transparent and actionable decision-making for coking PdM.

The remainder of this paper is as follows: Section 2 describes the proposed framework, Section 3 details a case study in industrial scenarios, and Section 4 summarizes of the findings.

2. Methodology

2.1. Problem statement

Deterioration of crude oil is the primary cause of coking at the fractionating tower bottom. High operating temperatures and catalyst particles further contribute to thermal condensation reactions. Hydrogenation readily occurs under high temperatures, with the catalytic effects of oxygen and metals generating aromatic radicals and initiating chain reactions, forming polymers. These polymers accumulate, adhere to equipment surfaces, and deposit in low-flow areas like heat exchangers. The resulting coking particles accelerate mechanical wear and pose a more severe threat by blocking flow paths, disrupting the tower's thermal balance, and deteriorating product quality. Excess heat buildup further destabilizes heat and gas-liquid equilibrium, potentially leading to tower flushing, pressure fluctuations, unplanned shutdowns, increased energy consumption, and reduced processing capacity.

As illustrated in Fig. 1, coking frequently occurs at critical locations, including feed control valves, herringbone baffles, tower bottom outlets, return tower pipes, pumps, heat exchanger tubes, and steam generators. These deposits disrupt decant oil circulation, cause pump failures, block pipelines, and diminish steam generation. In extreme cases, uncontrolled coke accumulation necessitates unplanned shutdowns, resulting in significant economic losses, underscoring the urgent need for targeted monitoring and mitigation strategies to enhance system reliability and operational continuity. Consequently, quantitative risk assessment (QRA) and PdM are essential for effectively managing coking risks and ensuring long-term stable operation.

2.2. Overview architecture

The overview architecture, as shown in Fig. 2, integrates data and knowledge-driven methods for monitoring coking conditions. It consists of three sequentially connected components: the construction of the optimal feature dataset and coking risk index, which then imports the parallel encoder-integrated decoder architecture to address mixed-frequency data and enable PdM.

1) A robust feature dataset is constructed to monitor coking conditions through MFS. The label space is defined by risk factors derived from expert knowledge, while the feature space encompasses industrial datasets spanning the reactor-regenerator and fractionation systems. The MFS process systematically searches, evaluates, and ranks feature subsets to identify the most relevant input variables for prediction, ensuring accuracy and reliability.
2) A multi-layer fuzzy inference system is implemented to construct the coking risk index. This system integrates DCS and LIMS data with expert-derived knowledge. The determination of risk factors, criteria, and risk levels is guided by causal analysis and validated through operator consultations. Membership functions, fuzzy rule bases, and inference engines collaboratively quantify the coking risk, enabling precise and interpretable assessments of process safety.
3) To address the mixed-frequency challenges, the parallel encoder-integrated decoder architecture is proposed. This architecture effectively captures the dynamics of DCS data

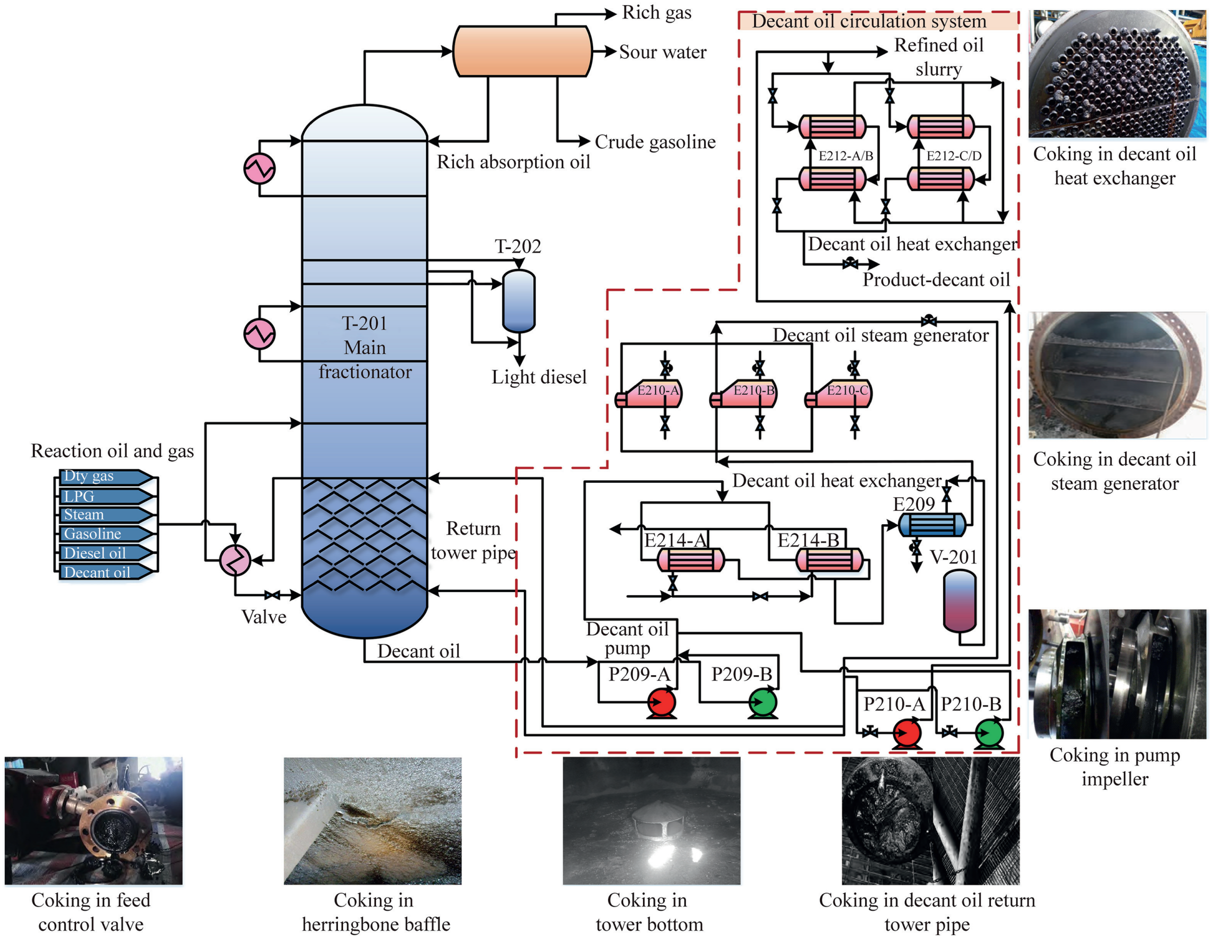

Fig. 1. Coking phenomenon at the fractionating tower bottom and the decant oil circulation system.

(representing operation states) and LIMS data (characterizing crude oil and decant oil properties). Triple AMs are embedded to enhance interpretability, handle non-stationarity, and provide insights into coking anomaly traceability. The resulting outputs support PdM by offering actionable countermeasures for mitigating coking risks.

2.2.1. Multi-layer fuzzy inference system

FL, proposed by Zadeh [22], utilizes fuzzy sets to analyze fuzzy reasoning, linguistic expressions, and underlying principles, serving as valuable tools for decision support. FL is used to represent uncertainty in opinions and address simulating nonlinear systems, theoretically enhancing the effectiveness of nonlinear conduction in the coking risk index. Converting qualitative assessments into quantitative evaluations through fuzzy set theory facilitates comprehensive evaluation of objects or processes influenced by multiple factors [12]. This study couples causal analysis and FL to construct the coking risk index. Fig. 3 depicts the multi-layer fuzzy inference system for the coking risk index.

The collection $F_i = \{f_1, f_2, \cdots, f_n\}$, consisting of n major risk factors. The multi-layer FL system takes into account physical properties and operation state collected from DCS and LIMS dataset as inputs. The procedural steps are as follows.

(1) Fuzzification: Based on expert knowledge and process mechanism, the criteria C_i $(i = 1, 2, \cdots, n)$ for each risk factors are aggregated, and the corresponding coking risk levels $L_i = \{l_1, l_2, \cdots, l_5\}$ are determined under abnormal coking conditions. Clear data F, C_i, and L are taken as input. The degree of the appropriate fuzzy set is determined using the Gaussian membership function, as shown in Eq. (1), which is used for nonlinear data, especially in industrial scenarios with high uncertainty and noise.

$$\mu(f) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \tag{1}$$

where c is the mean value and σ is the standard deviation.

(2) Inference engine: The inference engine implements the Mamdani inference algorithm [23]. A fuzzy rule base contains a set of fuzzy IF-THEN rules as shown in Eq. (2) which are generally derived from expert experience and process mechanism.

$$\text{Rule}^i : \text{if } f_1 \text{ is } S_1^i, \text{and } f_2 \text{ is } S_2^i, \text{and} \cdots f_n \textit{ is } S_n^i \text{ then } C_{\text{index}} \text{ is } I^i \tag{2}$$

where Rule^i represents the ith fuzzy rule, S_1^i and I^i are linguistic values defined by fuzzy sets, f_n and C_{index} are the risk factors and coking risk index linguistic variables.

(3) Defuzzification: The centroid method, as shown in Eq. (3), is applied for defuzzification because it exhibits the convenient property of continuity. It ensures that when the coking risk

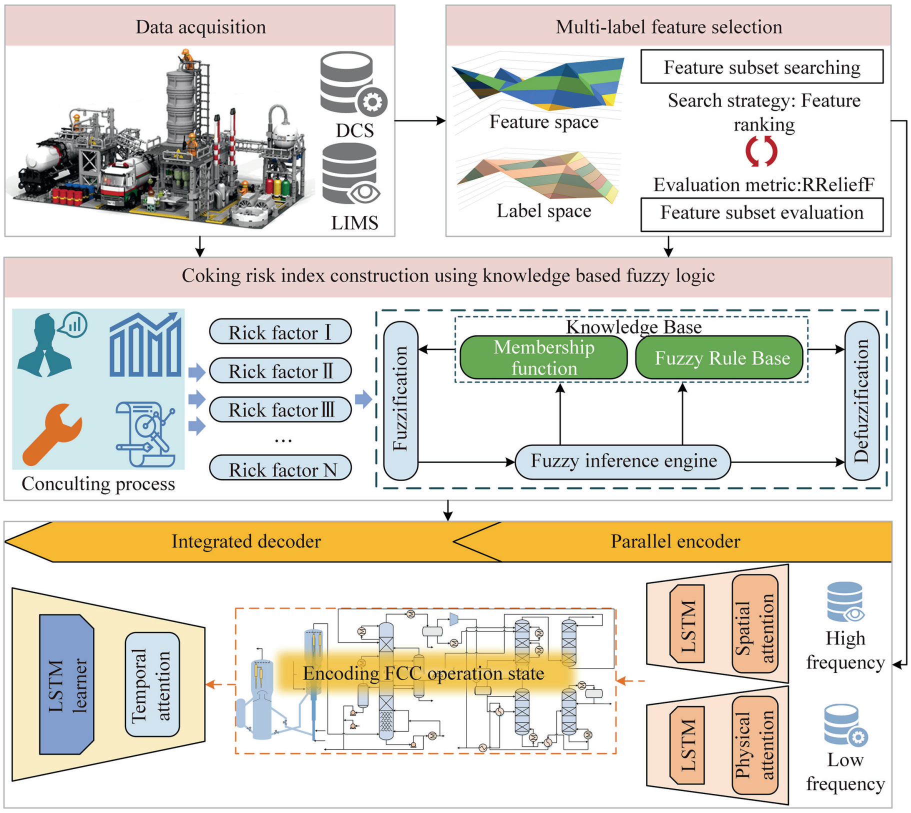

Fig. 2. The overall architecture for coking condition monitoring.

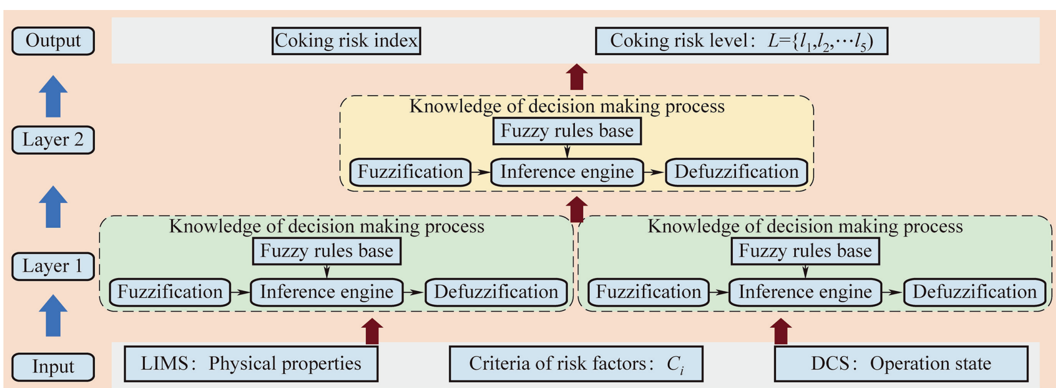

Fig. 3. Multi-layer fuzzy inference system for coking risk index.

factor is at a critical value, the fractionation system's risk index changes gradually, avoiding sudden and drastic adjustments.

$$Y = \frac{\int \mu(f) f \mathrm{d}f}{\int \mu(f) \mathrm{d}f} \tag{3}$$

where Y quantitative coking risk index; $\mu(f)$ is the membership value.

2.2.2. Multi-label feature selection

Multi-label learning allows instances to be associated with multiple labels simultaneously. To ensure comprehensive coverage of equipment units and process parameters, this study employs MFS methods to reduce dimensionality and enhance model performance for coking condition monitoring. Fig. 4 illustrates the form of a multi-label dataset. Suppose that $Z = \{(Y_i, X_i, F_i)|Y_i \in Y\}$ is the domain and range of multi-label data. $Y = \{Y_1, Y_2, Y_3, \cdots, Y_m\}$ represents the FCCU operation set of samples; $X = \{X_1, X_2, X_3, \cdots, X_m\}$ represents the FCCU operation set of features; $F = \{F_1, F_2, F_3, \cdots, F_m\}$ is the risk factors set of labels.

Relief is a classical filter-based feature selection method was initially designed for binary classification. Its extension, ReliefF, improves multi-label handling by searching for k nearest neighbors. RReliefF, a further refinement, is robust to noisy data and missing values, making it well-suited for regression tasks with continuous targets [24]. This adaptability renders RReliefF ideal for processing industrial big data.

RReliefF employs a dissimilarity function to identify nearest instances, using their proximity as partial weights to estimate

	X_1	X_2	X_3	$\cdots$	X_N	F_1	$\cdots$	F_L
Y_1	x_{11}	x_{12}	x_{13}	$\cdots$	x_{1N}	f_{11}	$\cdots$	f_{1L}
Y_2	x_{21}	x_{22}	x_{23}	$\cdots$	x_{2N}	f_{21}	$\cdots$	f_{2L}
Y_3	x_{31}	x_{32}	x_{33}	$\cdots$	x_{3N}	f_{31}	$\cdots$	f_{3L}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Y_M	x_M	x_{M2}	x_{M3}	$\cdots$	x_{MN}	f_{M1}	$\cdots$	f_{ML}

Fig. 4. The form of multi-label dataset.

feature importance. Unlike strictly univariate ranking measures, RReliefF accounts for attribute interactions, directly processing multi-label data without transformation. In this study, the normalized Hamming distance (HD) [25] is used as the dissimilarity measure, which counts differing labels. Formally, for two multi-label sets Y_a and Y_b, HD is defined in Eq. (4).

$$\mathrm{HD}(Y_a, Y_b) = \frac{|Y_a \cup Y_b| - |Y_a \cap Y_b|}{q} \tag{4}$$

where $|Y_a \bigcup Y_b| - |Y_a \bigcap Y_b|$ counts the number of labels which are different in Y_a and Y_b. q represents the total number of labels in the label space.

2.2.3. Parallel encoder-integrated decoder architecture

DCS and LIMS data are considered types of multi-source data characterized by a distinct sampling frequency. To extend feature vectors from the DL architecture design, simultaneous use of multiple encoder modules can bridge information islands and improve the model's processing capability. AMs, irrespective of whether they attend in a temporal or a parallel module, adaptively process and aggregate input information and make it accessible to the target, which circumvents the need for frequency alignment and provides intrinsic interpretability.

The target variables sequence is defined as $\boldsymbol{C} = (c_1, c_2, \cdots, c_n)^{\mathrm{T}} = (c_1, c_2, \cdots, c_T) \in R^{n\times T}$, where T represents the window size, and n denotes the number of features. The input optimal features is $\boldsymbol{X} = (x_1, x_2, \cdots, x_m)^{\mathrm{T}} = (x_1, x_2, \cdots, x_T) \in R^{m\times T}$ and $\boldsymbol{L} = (l_1, l_2, \cdots, l_n)^{\mathrm{T}} = (l_1, l_2, \cdots, l_T) \in R^{n\times T}$.

The parallel encoder architecture, depicted in Fig. 5, uses the gated recurrent unit (GRU)-base encoder to process the mixed-frequency input sequence. It generates the hidden layer state h'_t and h_t at current time step from the input x_t, l_t and previous hidden layer state h'_{t-1} and h_{t-1}. The encoder state is shown in Eq. (5).

$$\begin{aligned} h'_t &= \mathrm{GRU}_{\mathrm{enc}}(x_t, h'_{t-1}) \\ h_t &= \mathrm{GRU}_{\mathrm{enc}}(l_t, h_{t-1}) \end{aligned} \tag{5}$$

In encoder module I, the attention score e'^{m}_{t} for the mth feature of input sequence $\boldsymbol{x}^m = (x_1^m, x_2^m, \cdots, x_T^m)^{\mathrm{T}}$ at time t is computed using the previous time hidden state h'_{t-1} and the hyperbolic tangent function (tanh). This approach is similarly applied in encoder module II, as shown in Eq. (6).

$$\begin{aligned} \boldsymbol{e}'^{m}_{t} &= \boldsymbol{V}'^{\mathrm{T}}_{e} \tanh(W_e h'_{t-1} + U'_e x^m + b'_e) \\ \boldsymbol{e}^{n}_{t} &= \boldsymbol{V}^{\mathrm{T}}_{e} \tanh(\alpha_e h_{t-1} + U_e l^n + b_e) \end{aligned} \tag{6}$$

The attention scores, matching the length of input sequence, are computed at each step of the encoder. In encoder module I, $\boldsymbol{V}'_e$, W_e, U'_e, and b'_e are the weight matrixes and bias terms learned during training. The Softmax function normalizes these scores to derive the feature importance w_t^m. This process is analogous in encoder module II, as shown in Eq. (7).

$$\begin{aligned} w_t^q &= \mathrm{Softmax}(\boldsymbol{e}'^{m}_{t}) = \frac{\exp(e'^{m}_{t})}{\sum_{i=1}^{n} \exp(e'^{m}_{t})} \\ \alpha_t^n &= \mathrm{Softmax}(\boldsymbol{e}^{n}_{t}) = \frac{\exp(e^{n}_{t})}{\sum_{i=1}^{n} \exp(e^{n}_{t})} \end{aligned} \tag{7}$$

Using probabilities to weight the input sequence to obtain the update sequence $\tilde{x}_t$ and $\tilde{l}_t$, as shown in Eq. (8).

$$\begin{aligned} \tilde{\boldsymbol{x}}_t &= \left(w_t^1 x_t^1, w_t^2 x_t^2, \cdots, w_t^m x_t^m\right)^{\mathrm{T}} \\ \tilde{l}_t &= \left(\alpha_t^1 l_t^1, \alpha_t^2 l_t^2, \cdots, \alpha_t^n l_t^n\right)^{\mathrm{T}} \end{aligned} \tag{8}$$

Then, the hidden state of the encoder at time t can be updated, as shown in Eq. (9).

$$h'^{\mathrm{enc}}_{t} = \mathrm{GRU}_{\mathrm{enc}}\left(\tilde{x}_t, h'^{\mathrm{enc}}_{t-1}\right) h^{\mathrm{enc}}_{t} = \mathrm{GRU}_{\mathrm{enc}}\left(\tilde{l}_t, h^{\mathrm{enc}}_{t-1}\right) \tag{9}$$

The parallel encoder module performs feature extraction at different frequencies, and the input is reconstructed in the integrated decoder module, a channel fusion mechanism. Function concatenate$(\cdot)$ is mainly used to fuse the channel of different feature maps. The process is formulated as shown in Eq. (10):

$$H^{\mathrm{enc}}_{t} = \mathrm{concatenate}(h'^{\mathrm{enc}}_{t} + h^{\mathrm{enc}}_{t}) \tag{10}$$

where h'^{enc}_{t} represents the output feature map of high-frequency encoder module I, h^{enc}_{t} represents the output feature map of low-frequency module II.

The architecture of the integrated decoder module is illustrated in Fig. 6. The temporal AM assigns weights to different time steps, highlighting their influence on the target variables. This process generates time-weighted features, enhancing the model's sensitivity to the volatility patterns within the time series. Similar to the input AM, the attention score l_t^k at time t on the previous hidden state d_{t-1} of the kth hidden state, as defined in Eq. (11).

$$\boldsymbol{l}^{k}_{t} = \boldsymbol{V}^{\mathrm{T}}_{d} \tanh(W_d d_{t-1} + U_d H^{\mathrm{enc}}_{t} + b_d), 1 \le k \le T \tag{11}$$

where $\boldsymbol{V}_d$, W_d, U_d, and b_d are the weight matrixes and bias obtained by training process.

The attention scores are normalized by the Softmax function, indicating the importance of the kth hidden state at time t, as shown in Eq. (12).

$$\beta_t^k = \mathrm{Softmax}\left(\boldsymbol{l}^{k}_{t}\right) = \frac{\exp\left(l_t^k\right)}{\sum_{j=1}^{T} \exp\left(l_t^k\right)} \tag{12}$$

The context vector, defined in Eq. (13), is obtained through a weighted summation of attention scores and hidden states across all time steps in the decoder. It encapsulates both feature correlations and historical time-series information.

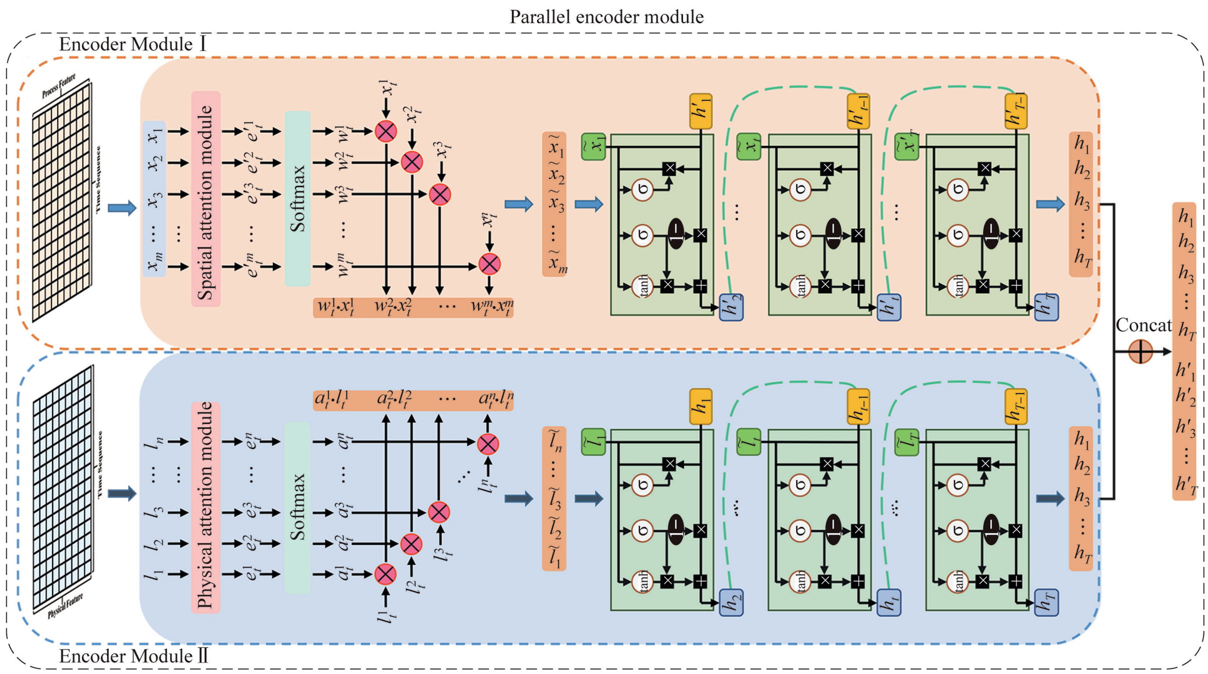

Fig. 5. The architecture of Parallel Encoder module.

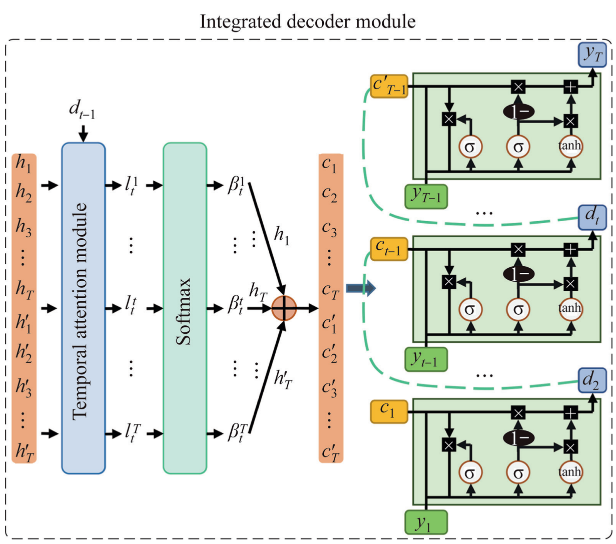

Fig. 6. The architecture of Integrated Decoder module.

$$c_t = \sum_{k=1}^{T} \beta_t^k h_k \tag{13}$$

Dynamic decoding in the integrated decoder module combines the previous output f_{t-1} to compute the hidden state d_{t-1}^{dec} after processing all hidden state outputs. The formula f is detailed in Eq. (14).

$$d_t^{\text{dec}} = \text{GRU}_{\text{dec}}\left(f_{t-1}, d_{t-1}^{\text{dec}}, c_t\right) \tag{14}$$

Finally, through the fully connected layer, previous hidden layer state d_{t-1}^{dec}, and context vector c_{t-1} are nonlinearly combined to obtain the output sequence y_t, as shown in Eq. (15). Moreover, $\tilde{\boldsymbol{W}}$ and $\tilde{b}$ are the weight matrix and bias of the memory cell in the model.

$$y_t = \tanh\left(\widetilde{\boldsymbol{W}} d_{t-1}^{\text{dec}} + \tilde{b}\right) \tag{15}$$

2.3. *Coking condition monitoring framework*

The framework consisting of the offline stage and online stage for monitoring coking condition that integrates DL, expert knowledge, and multi-source industrial data to achieve coking PdM is shown in Fig. 7. In the offline stage, causal analysis identify L_i, C_i, F_i, which are incorporated into a multi-layer fuzzy inference system to construct the coking risk index. Combined with MFS, the mixed-frequency dataset is constructed. To address the asynchronous nature of DCS and LIMS data, a parallel encoder-integrated decoder architecture is developed, enabling the fusion of mixed-frequency features and optimizing the model for accuracy and interpretability. In the online stage, real-time monitoring of the coking risk index triggers weight-based dynamic analysis of risk factors when the risk level is exceeded, enabling precise localization of anomaly variables and root causes identification combined causal analysis for targeted disposal.

3. Case Study

3.1. *Offline stage*

3.1.1. *Descriptions of FCCU*

This study investigates the coking condition at the fractionating tower bottom of a 2.8-million-ton FCCU. In Fig. 8, the dataset spans from 31 December 2021 to 15 December 2022, during which scheduled equipment maintenance occurs between 14 August and 10 September 2022. In collaboration with a refinery, comprehensive DCS and LIMS datasets are obtained that cover the entire FCC process, including the reactor-regenerator, fractionation, and refining units. The datasets include 603 DCS variables and 101 LIMS

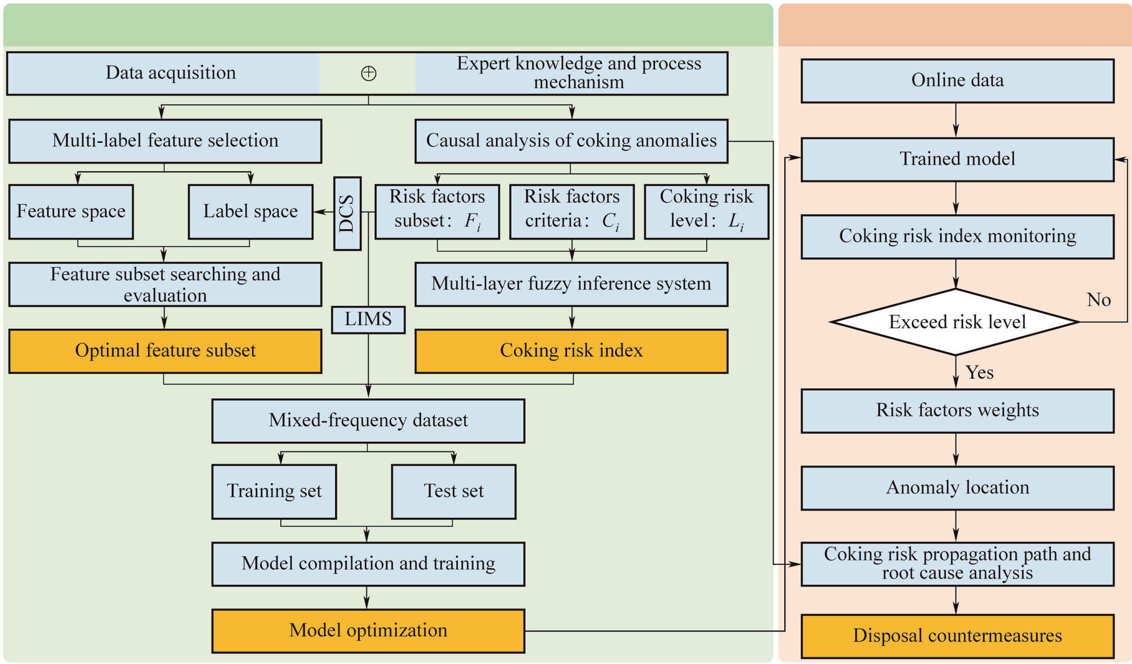

Fig. 7. The framework of coking condition monitoring.

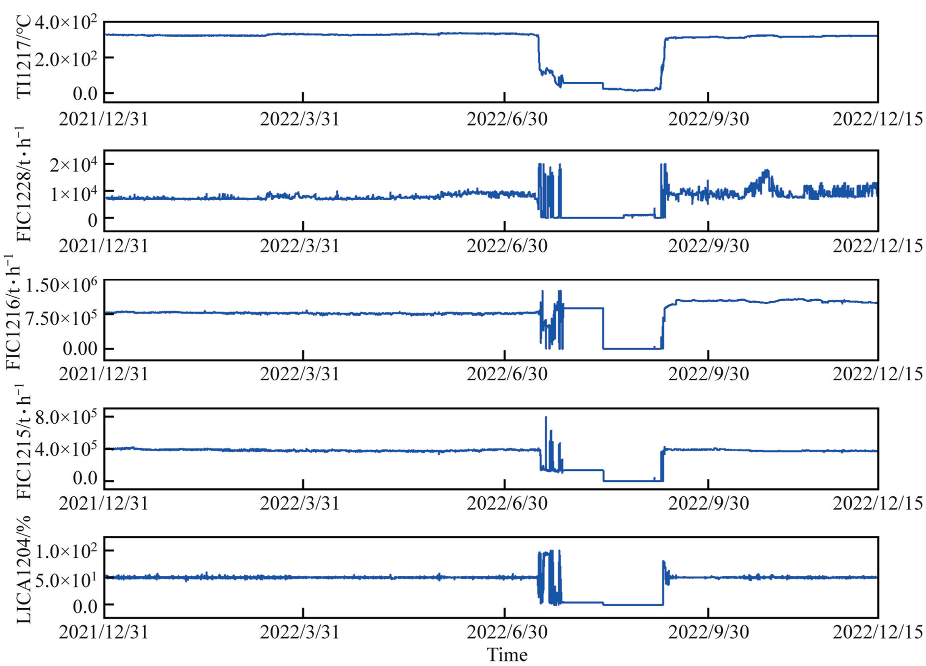

Fig. 8. Part of the risk factors.

indicators. Based on external operating conditions, 220 DCS variables are selected from upstream units of the decant oil circulation system to characterize the operational state of the plant, serving as the feature pool for monitoring the coking risk. In addition, five categories of crude oil and slurry physicochemical properties are identified that represent intrinsic factors contributing to coke formation. After preprocessing, the dataset contains 27158 records for each DCS variable and 1791 records for each LIMS indicator.

3.1.2. Coking risk index

As shown in Fig. 1, ensure stable operation of the fractionating tower bottom and decant oil circulation system, with adequate heat extraction to maintain operational flexibility amid changes in processing capacity. Based on fractionation system operation mechanisms and experience, the cause of coking at the fractionating tower bottom is attributed to five main categories: crude oil and decant oil properties, solid content in decant oil, decant oil circulation volume, bottom temperature, and liquid level. Thus, compromising the crude oil properties, product quality, and operation state, 11 risk factors F_i and their criteria C_i are defined, as detailed in Table 1.

Human language inherently involves semantic uncertainty; fuzzification addresses this by mapping input data to linguistic categories—fuzzy sets—with membership degrees ranging from 0 to 1. As detailed in the Supplementary Material (Fig. S1 for physical properties and Fig. S2 for operational states), these fuzzy sets stratify linguistic variables into three risk levels (low, medium, high) according to predefined criteria and thresholds, such that values below the lower bound denote low risk and those above the upper bound denote high risk. The outputs of this fuzzification process feed into the initial layer of the multi-layer fuzzy inference system, whose subsequent layer integrates physical and operational assessments to compute a unified coking risk index normalized on a 1–100 scale and partitioned into five categories (safety, minimal, low, moderate, high). Finally, the membership functions illustrated in Fig. 9 assign continuous weights across these five levels, enabling site operators to implement tiered countermeasures by progressive risk management principles—where intervention decisions are proportionally calibrated to the assessed severity of risk—thereby ensuring timely responses without overreaction or delay.

Fig. 10 illustrates the coking risk index constructed using the multi-layer fuzzy inference system. Prior to maintenance in August 2022, the index reaches the moderate to high risk regions, correlating with fluctuating operating conditions and increased coking tendency during low-load operations. During the equipment maintenance period (14 August-10 September 2022), the risk index drops sharply into the minimal risk and safety ranges. Following the restart, the index surges back into the moderate to high risk, mainly due to the purchased equilibrium catalyst containing more fines, which elevated catalyst loss and increased decant oil solid content. The coking risk index accurately reflects real-time operating conditions and coking tendencies, triggering a warning state when its value exceeds 60. By incorporating multiple risk factors and expert knowledge, the multi-layer fuzzy inference system effectively captures gradual and abrupt risk-level shifts. The adaptation of the coking risk index to adjust dynamically according to operational variables underscores its practical applicability for real-time monitoring and decision-making in industrial settings.

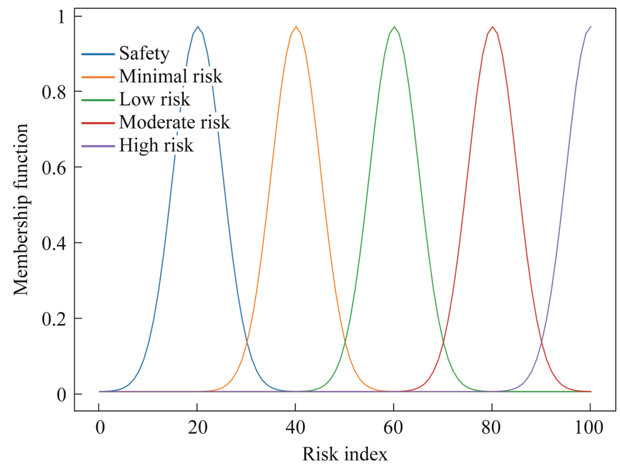

Fig. 9. Membership function for the coking risk index.

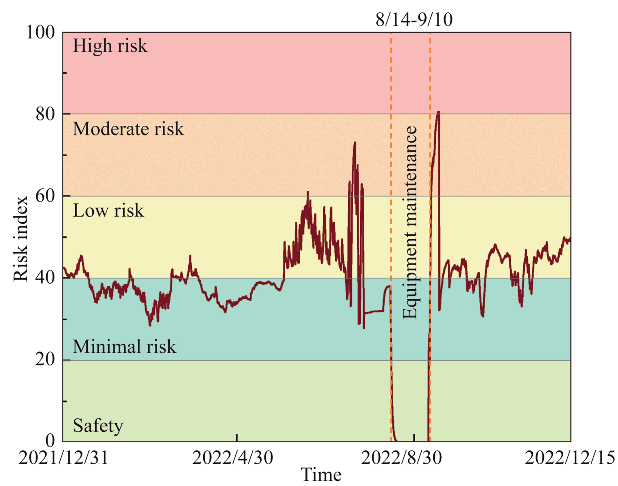

Fig. 10. The coking risk index under long-period operation conditions.

3.1.3. Optimal feature subset

The optimal feature dataset for coking condition monitoring is constructed by applying MFS, using risk factors as the label space and the industrial dataset across the reactor-regenerator and fractionation systems as the feature space. Fig. 11(a) presents the result of RReliefF-based MFS. The cumulative contribution of reconstructed feature weights is calculated to identify the crucial feature [26]. Specifically, the cumulative contributions of 95%, 97%, and 99.7% are achieved with 19, 23, and 69 variables, as shown in Fig. 11(b). The sharp rise in feature count between 24 and 69 variables, contributing only 2% to cumulative importance, demonstrates diminishing returns from including additional features beyond 23 variables. Selecting the 23 most important features, corresponding to the 95% cumulative contribution threshold, reduces the feature set size by 89.5% (from 220 to 23 variables), while

Table 1
Configurations of the coking risk factors and criteria.

	Name	Description	Criteria	Unit
Operation state	TI1217	Tower bottom temperature	330	°C
	FIC1227	Recycle oil flow rate	2500	$kg\cdot h^{-1}$
	FIC1228	Outlet device decant oil flow rate	7000	$kg\cdot h^{-1}$
	FIC1215	Decant oil up to the tower flow rate	400000	$kg\cdot h^{-1}$
	FIC1216	Decant oil down to the tower flow rate	800000	$kg\cdot h^{-1}$
	LICA1204	Tower bottom liquid level	30	%
Psychical property	YJ_GH	Decant oil solid content	6	$g\cdot L^{-1}$
	YJ_MD	Decant oil density	1100	$g\cdot m^{-3}$
	YJ_ND	Decant oil viscosity (100 °C)	60	$mm^2\cdot s^{-1}$
	HHYL_MD	Crude oil density	900	$g\cdot m^{-3}$
	HHYL_ND	Crude oil viscosity	12	$mm^2\cdot s^{-1}$

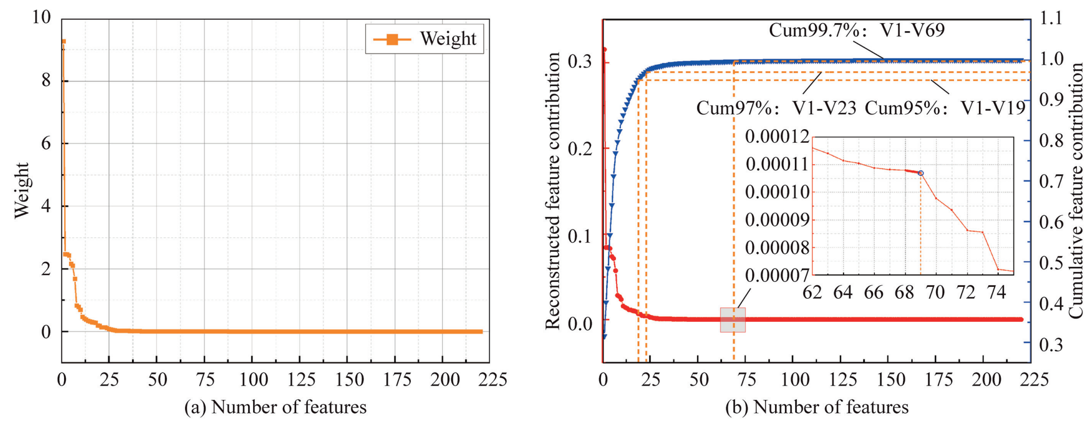

Fig. 11. The result of MFS.

maintaining model performance and eliminating redundant variables. The optimal feature subset related to the operation state of fractionation system is detailed in Table S1 of the Supplementary Material.

3.2. Online stage

3.2.1. Coking risk monitoring

The reactor-regenerator and fractionation systems are subject to dynamic fluctuations in the coking risk index due to variations in crude oil, load adjustments, and equipment conditions. The computational environment is running on an Intel(R) Core(TM) i9-10980XE CPU @ 3.00 GHz, 64.0 GB of RAM, and Windows 10. The model is trained for 80 epochs using the adaptive moment estimation (Adam) optimizer to ensure efficient convergence, with a batch size of 64 balancing computational efficiency and generalization, and a learning rate of 0.001 empirically determined from the dataset distribution. To ensure rigorous evaluation, a dual validation strategy combining quantitative assessment with qualitative interpretability is adopted to assess predictive performance and industrial applicability.

As depicted in Fig. 12(a), the blue, green, and orange curves illustrate the model's prediction performance during training, validation, and testing stages under varying operating conditions. Under the stable operating condition, the model effectively captures subtle fluctuations, enabling early detection of the coking risk and timely interventions to prevent accumulation. During operational adjustments, such as pre-maintenance stage, the model quickly adapts to variations in the coking risk index, demonstrating robust real-time tracking capabilities. Even under volatile conditions, including shutdowns and restarts, the model maintains high predictive accuracy by capturing both gradual trends and abrupt changes. These capabilities ensure robust adaptability, early hazard warnings, and data-driven decision-making, making the architecture highly effective for real-time monitoring and maintenance optimization in complex industrial environments.

Table 2
The result of ablation study.

Name	Encoder-decoder (with triple AMs)	Parallel encoder-integrated decoder (no triple AMs)	Parallel encoder-integrated decoder
MSE	1.2694	1.5358	1.1986
RMSE	1.4450	1.5191	1.2403
MAE	1.2205	1.2639	1.2176
MAPE	3.17%	3.89%	2.78%
R^2	0.9186	0.8854	0.9321

Quantitative validation includes residual distribution analysis and ablation studies. As depicted in Fig. 12(b), the residual normal distribution graph shows a mean value of 1.224, with an overall range between 0.4 and 2.0. The positive mean of the residuals indicates that the model's prediction is generally slightly higher than the actual values, aligning with the design goal of ensuring safety. A lower mean would result in systematic risk underestimation, potentially compromising safety. High residual values in the tail distribution are essential for capturing high-risk scenarios, as they help avoid underestimating extreme events. In coking risk assessment, a slight overestimation of risk provides decision-makers with more time and flexibility to implement precautionary measures, ultimately helping to prevent potential accidents and ensuring operational safety.

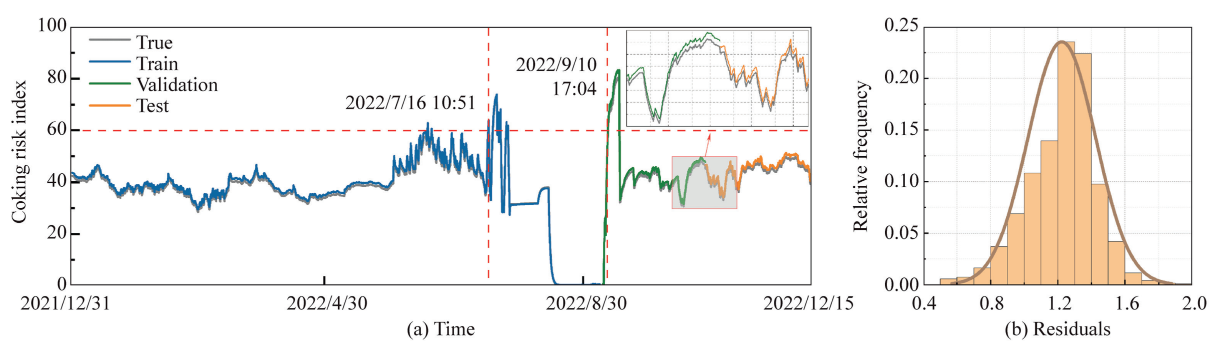

Fig. 12. The result of coking condition monitoring.

The ablation study demonstrates the performance advantages of the Parallel encoder-integrated decoder architecture and the triple-AMs, as shown in Table 2. By processing DCS and LIMS data separately, the parallel encoders preserve their unique characteristics while addressing frequency mismatches, significantly reducing MSE (1.1986) and RMSE (1.2403). The architecture achieves lower MAE (1.2176) and MAPE (2.78%), underscoring its accuracy and robustness in handling complex industrial processes. With an R^2 of 0.9321, the model demonstrates strong predictive effectiveness.

Qualitatively, the interpretability of the parallel encoder-integrated decoder is demonstrated through attention weight visualizations, revealing how the model dynamically adjusts focus across risk factors in response to changing operational states. The triple-AMs enhances intrinsic interpretability, supporting more informed decision-making. As shown in Fig. 13(a), critical risk factors such as LICA1204, FIC228, FIC1227, TI1217, FIC1216, and FIC1215 dominate under the steady-state condition. During the operational fluctuations of July 2022, the focus shifts from these key risk factors to variables representing the fractionation system's operation state, including changes in tower heat balance (FIC1207, FIC1208) and product flow rates (FIC1212, FIC1225, FIC1301). Conversely, as shown in Fig. S3 of the Supplementary Material, conventional data filling methods for aligning mixed-frequency inputs in the Encoder–Decoder (with triple AMs) architecture distort temporal dependencies. This results in attention weight shifts that deviate from domain knowledge, ultimately undermining both predictive accuracy and model interpretability in coking risk monitoring.

In Fig. 13(b), decant oil solid content serves as the primary risk indicator under the stable condition. Yet, during the fluctuations in July 2022, the coking risk shifts to crude oil properties, notably crude oil density. This shift occurred because the decant oil stored in the system is drained during equipment maintenance, resulting in its solid content being low. Upon restart in September 2022, higher catalyst loss and increased decant oil solid content cause the risk to alternate between decant oil and crude oil properties, reflecting the system's operation state. As shown in Fig. S4 of the Supplementary Material, the temporal AM dynamically reweights encoder hidden states to identify key time steps aligned with operational changes and maintenance events. By integrating long- and short-term dependencies, the model captures both abrupt shifts and gradual trends in coking tendencies. This enhances its ability to extract dynamic features, accurately predict risk trajectories, and rapidly adapt to evolving process conditions, ensuring robust performance in complex FCCU operations.

3.2.2. Disposal countermeasures

The outputs of the developed model—including the predicted coking risk index and attention-based risk factors weights—serve as the foundation for the entire decision-making process. Anomalies are identified through high attention weights, and further combined with causal analysis to trace fault propagation and guide

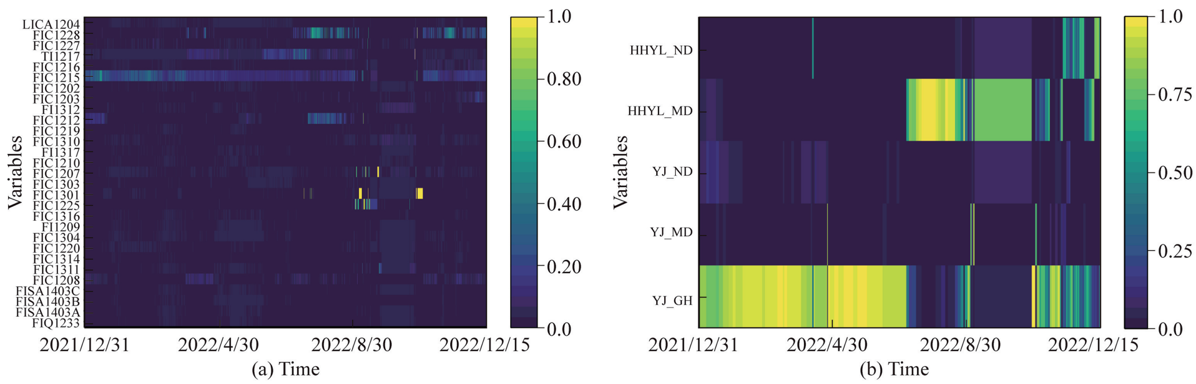

Fig. 13. The attention weight of ablation study architectures.

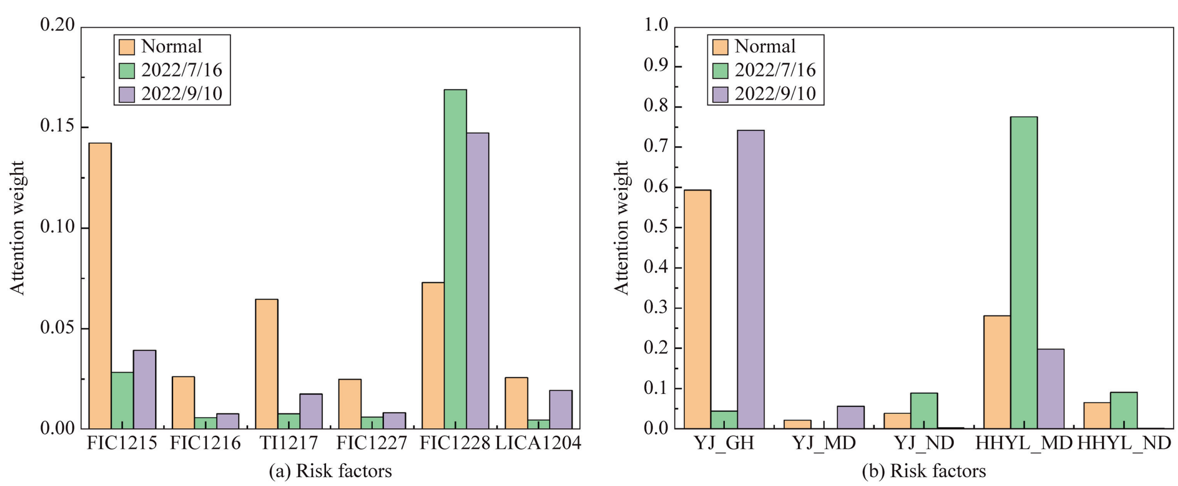

Fig. 14. The attention weight of risk factors during alarms.

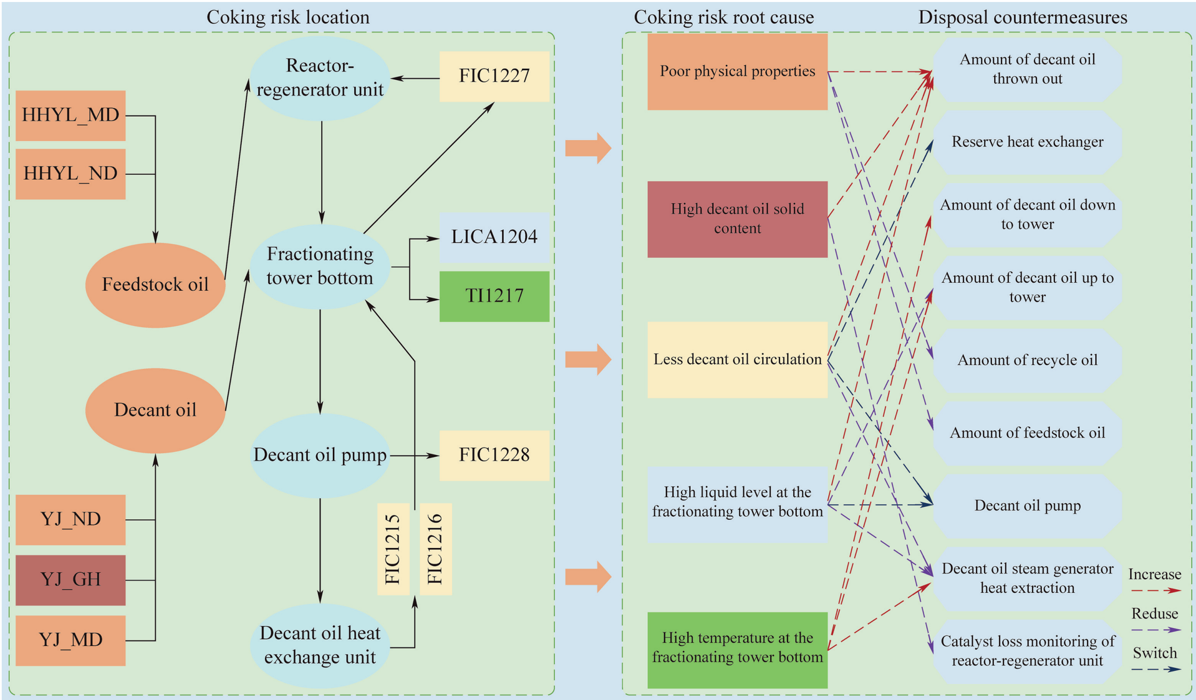

Fig. 15. The knowledge map of coking anomaly traceability and disposal.

prescriptive maintenance. Therefore, the decision-making strategy in this study is fully integrated with, and entirely dependent on, the proposed model architecture. When the risk index reaches the warning stage, anomaly variables are identified through high attention weights and combined with causal analysis to trace the propagation path, enabling prescriptive maintenance. Fig. 14 illustrates the risk factor weights for the two warning stages in Fig. 12 compared to the average values under normal conditions. Coking risk is primarily influenced by decant oil circulation flow (FIC1215, FIC1216, FIC1227, FIC1228), bottom temperature (TI1217), and bottom liquid level (LICA204). Fig. 14(a) shows during the warning stages before and after maintenance, FIC1228 dominates due to empty system during shutdown, where increased decant oil discharge is necessary. Fig. 14(b) reveals that decant oil solid content contributes most to coking risk during shutdown, while crude oil density dominates during startup. This shift occurs because decant oil is drained during shutdown, reducing its solid content. Upon restart, increased catalyst loss elevates the decant oil solid content, causing risk contributions to alternate between decant oil and crude oil properties, reflecting the system's operational state.

Fig. 15 presents the knowledge map for anomaly traceability and disposal of coking risk. The distribution of the risk factors is based on the FCCU process, which corresponds to the root cause of the fault according to the fill color. Once the anomaly variables are located, the coking risk propagation path is analyzed to trace the root cause and enable appropriate countermeasures. When a warning occurs, such as on 10 September 2022, Fig. 14 highlights the need to monitor decant oil circulation and solid content. If the decant oil solid content sharply rises above 6 $g\cdot L^{-1}$, immediate actions include reducing heat extraction from the decant oil steam generator to prevent coking caused by lower oil temperatures and increased viscosity after heat exchange. Simultaneously, increasing the decant oil discharge rate, checking for catalyst loss, and ensuring a low cyclone separator pressure difference are critical. If necessary, timely switching of the decant oil pump and heat exchanger must be performed to stabilize operations.

4. Conclusions

This study proposes a hierarchical framework integrating FL, MFS, and DL to monitor and manage coking conditions, presenting the explainable PdM model that captures critical signals reflecting production and condition evolution for informed decision-making. Its primary function is to remotely monitor coking risk and generate risk index, allowing managers to easily understand the system's processes and operations. This facilitates precise decision-making regarding the type of maintenance required in specific situations. Applied at the 2.8 million tons FCCU effectively facilitates coking risk management at the fractionator tower bottom. In the offline stage, causal analysis identified 11 coking risk factors F_i and corresponding criteria C_i, which are incorporated into a multi-layer fuzzy inference system to construct the coking risk index, providing the QRA under cross-cycle operating conditions. MFS is applied to a dataset of 220 operational variables, yielding 23 optimal features with a 95% cumulative contribution and achieving an 89.5% reduction in feature dimensionality. The parallel encoder-integrated decoder architecture, enhanced by triple AMs, processed mixed-frequency DCS and LIMS data, achieving superior accuracy and interpretability, with ablation studies confirming its performance across five error metrics. In the online stage, the framework enables real-time monitoring of the coking risk index, dynamic weighting of risk factors, and precise localization of root causes, ensuring effective PdM. Its robust adaptability to varying operational conditions demonstrates potential for extension to other industrial systems. Future advancements should incorporate reinforcement learning and process mechanisms to improve adaptability, operational optimization, and decision-making precision in dynamic environments.

CRediT Authorship Contribution Statement

Nan Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. Chunmeng Zhu:

Investigation, Conceptualization. Zeng Li: Investigation. Yunpeng Zhao: Supervision. Xiaogang Shi: Conceptualization. Xingying Lan: Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work.

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgements

The present work is financially supported by the Innovative Research Group Project of the National Natural Science Foundation of China (22021004), and Sinopec Major Science and Technology Projects (321123-1), and hereby their supports are sincerely acknowledged.

Supplementary Material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cjche.2025.04.015.

References

[1] C.W. Tang, J.L. Huang, M. Xu, X. Liu, F. Yang, W.T. Feng, Z.N. He, J.C. Lv, Attention-based early warning framework for abnormal operating conditions in fluid catalytic cracking units, *Appl. Soft Comput.* 153 (2024) 111275.
[2] S. Cai, L.B. Zhang, J.Q. Hu, Scale-reasoning based risk propagation analysis: an application to fluid catalytic cracking unit, *Proc. Saf. Environ. Prot.* 120 (2018) 155–165.
[3] M. Bulut, Ö. Evrencan, Planning of prescriptive maintenance types for generator with fuzzy logic-based genetic algorithm in a hydroelectric power plant, *Expert Syst. Appl.* 240 (2024) 122480.
[4] S. Gupta, A. Kumar, J. Maiti, A critical review on system architecture, techniques, trends and challenges in intelligent predictive maintenance, *Saf. Sci.* 177 (2024) 106590.
[5] M. Tanhaeean, S.F. Ghaderi, M. Sheikhalishahi, A decision-making framework for optimal maintenance management: an integrated simulation-mathematical programming-expert system approach, *Comput. Ind. Eng.* 185 (2023) 109671.
[6] R. Parkavi, P. Karthikeyan, A. Sheik Abdullah, Enhancing personalized learning with explainable AI: a chaotic particle swarm optimization based decision support system, *Appl. Soft Comput.* 156 (2024) 111451.
[7] J.R. Liang, Y. Tian, S.T. Yang, Y. Wang, R.Q. Yin, Y.F. Wang, Long-term operation optimization of circulating cooling water systems under fouling conditions, *Chin. J. Chem. Eng.* 65 (2024) 255–267.
[8] L. Basora, A. Viens, M.A. Chao, X. Olive, A benchmark on uncertainty quantification for deep learning prognostics, *Reliab. Eng. Syst. Saf.* 253 (2025) 110513.
[9] R.Q. Wang, Y.F. Wang, T. Gundersen, Y. Wu, X. Feng, M.X. Liu, A multi-objective optimization method for industrial park layout design: the trade-off between economy and safety, *Chem. Eng. Sci.* 235 (2021) 116471.
[10] F.N. Shahmohammad, Y. Pourrahimian, N. Akbari-Gharalari, Synthesizing complexity: trends, challenges, and future directions in fuzzy-based multi-criteria decision-making (FMCDM) methods, *Appl. Soft Comput.* 167 (2024) 112362.
[11] F. Mumali, J. Kałkowska, Intelligent support in manufacturing process selection based on artificial neural networks, fuzzy logic, and genetic algorithms: current state and future perspectives, *Comput. Ind. Eng.* 193 (2024) 110272.
[12] R. Xu, B.W. Kim, S.J.S. Moe, A.N. Khan, K. Kim, D.H. Kim, Predictive worker safety assessment through on-site correspondence using multi-layer fuzzy logic in outdoor construction environments, *Heliyon* 9 (9) (2023) e19408.
[13] E. Brazález, H. Macià, G. Díaz, M. Baeza_Romero, E. Valero, V. Valero, FUME: an air quality decision support system for cities based on CEP technology and fuzzy logic, *Appl. Soft Comput.* 129 (2022) 109536.
[14] Y.T. Qian, S. Vaddiraju, F. Khan, Inherent process risk index (IPRI)—A tool for analyzing inherently safer design using aspen plus simulation, *Proc. Saf. Environ. Prot.* 183 (2024) 399–416.
[15] J.H. Lin, G. Michailidis, A multi-task encoder-dual-decoder framework for mixed frequency data prediction, *Int. J. Forecast.* 40 (3) (2024) 942–957.
[16] R. Saeidimesineh, P. Adibi, H. Karshenas, A. Darvishy, Parallel encoder—decoder framework for image captioning, *Knowl. Base Syst.* 282 (2023) 111056.
[17] J.Y. Guo, Y.L. Yang, H. Li, L. Dai, B.K. Huang, A parallel deep neural network for intelligent fault diagnosis of drilling pumps, *Eng. Appl. Artif. Intell.* 133 (2024) 108071.
[18] Z.Y. Yao, Q.C. Jiang, X.S. Gu, Distributed process monitoring based on kantorovich distance-multiblock variational autoencoder and Bayesian inference, *Chin. J. Chem. Eng.* 73 (2024) 311–323.
[19] L. Jiang, T.A. Zhang, W. Lei, K.J. Zhuang, Y.B. Li, A new convolutional dual-channel transformer network with time window concatenation for remaining useful life prediction of rolling bearings, *Adv. Eng. Inform.* 56 (2023) 101966.
[20] J.H. Dai, W.Y. Huang, C.C. Zhang, J. Liu, Multi-label feature selection by strongly relevant label gain and label mutual aid, *Pattern Recogn.* 145 (2024) 109945.
[21] I. Niño-Adan, D. Manjarres, I. Landa-Torres, E. Portillo, Feature weighting methods: a review, *Expert Syst. Appl.* 184 (2021) 115424.
[22] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Set Syst.* 100 (1999) 9–34.
[23] E.H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, *Int. J. Man Mach. Stud.* 7 (1) (1975) 1–13.
[24] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, *Mach. Learn.* 53 (2003) 23–69.
[25] N. Spolaôr, E.A. Cherman, M.C. Monard, H.D. Lee, ReliefF for multi-label feature selection, in: 2013 Brazilian Conference on Intelligent Systems. 2013, Fortaleza, Brazil, IEEE, 2013.
[26] N. Liu, J. Wang, S.L. Sun, C.K. Li, W.D. Tian, Optimized principal component analysis and multi-state Bayesian network integrated method for chemical process monitoring and variable state prediction, *Chem. Eng. J.* 430 (2022) 132617.