TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Sun, H., Huang, J., Qi, M., Yang, M., & Wang, F. (2026). A hybrid risk prediction method for chemical process systems based on enhanced feature engineering and XGBoost. *Journal of Loss Prevention in the Process Industries*, *102*, Article 106036. https://doi.org/10.1016/j.jlp.2026.106036

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

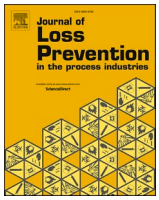
Journal of Loss Prevention in the Process Industries 102 (2026) 106036



Contents lists available at ScienceDirect

Journal of Loss Prevention in the Process Industries

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



A hybrid risk prediction method for chemical process systems based on enhanced feature engineering and XGBoost

Hao Sun [a,*], Jin Huang [a], Meng Qi [b], Ming Yang [c], Fuyu Wang [a]

[a] School of Management Science and Engineering, Anhui University of Technology, Ma'anshan, Anhui, 243002, China
[b] College of Chemistry and Chemical Engineering, China University of Petroleum (East China), Qingdao, 266580, China
[c] Safety and Security Science Section, Department of Values, Technology, and Innovation, Faculty of Technology, Policy, and Management, Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Chemical process safety
Risk prediction
Dynamic process simulation
Machine learning

ABSTRACT

Deviations of process parameters from their normal ranges are the primary causes of accidents in chemical process systems. Traditional risk assessment methods largely rely on static probability analysis based on historical data, which struggles to capture the dynamic influence of real-time parameter variations on risk and lacks the capability to predict risk evolution trends. This paper proposes a dynamic risk prediction method for chemical process systems based on enhanced feature engineering and XGBoost. First, key process parameters (KPPs), including temperature, liquid level and flow rate are identified through process analysis. A comprehensive risk indicator is then constructed using Dempster-Shafer (D-S) evidence theory to achieve dynamic quantification of system risk. Second, a dynamic simulation model is established using Aspen Plus, simulating operations under normal, disturbed, and extreme conditions to generate time-series data of KPPs. On this basis, multi-scale sliding window techniques are employed to extract enhanced features, including temporal, statistical, trend, and disturbance features. Finally, an XGBoost-based risk prediction model is developed. A continuous stirred-tank reactor (CSTR) is employed to demonstrate the proposed methodology. The results indicate that the proposed methodology achieves an RMSE of 0.159, MAE of 0.122 and an R^2 of 0.7237, outperforming traditional methods by significant margins. The results validate the effectiveness of combining enhanced feature engineering with XGBoost for risk prediction in chemical processes.

1. Introduction

As a core component of modern industrial systems, chemical process systems involve complex physicochemical reactions, multiphase flow, and heat transfer processes, characterized by high temperature, high pressure, flammability, explosiveness, and continuous operation (Shi et al., 2026). Major safety incidents in the chemical industry are directly associated with the loss of control over process parameters (Li et al., 2026; Khan et al., 2021). Once leakage, fire, or explosion accidents occur, they often result in severe casualties, property losses, and environmental damage (Sun et al., 2025; Yang et al., 2023). Therefore, real-time monitoring and risk prediction of the operational status of chemical process systems constitute a critical link in ensuring chemical safety production and achieving proactive risk prevention and control.

Traditional risk assessment methods for chemical processes can be broadly classified into two categories: probabilistic statistical methods and mechanistic model-based methods (Reniers and Amyotte, 2012; Aven and Zio, 2021; Zio, 2022). Probabilistic statistical methods (such as fault tree analysis, event tree analysis, and Bayesian networks) estimate the probability of risk events based on historical accident data or expert experience, thereby evaluating system risk levels (Nehal et al., 2023; Lai et al., 2026). While these methods can quantify risk levels, they exhibit certain limitations: (1) those methods rely heavily on the completeness of historical data and the accuracy of expert judgment, leading to considerable ambiguity and subjectivity in assessing novel processes or scenarios lacking accident records (Hu et al., 2021); (2) they focus on static probability assessment of risk, making it difficult to capture the dynamic influence of real-time variations in process parameters on risk evolution.

Mechanistic model-based methods describe the dynamic behavior of

This article is part of a special issue entitled: Process Safety in the Era of AI published in Journal of Loss Prevention in the Process Industries.

* Corresponding author.
E-mail address: sunhao@ahut.edu.cn (H. Sun).

https://doi.org/10.1016/j.jlp.2026.106036
Received 13 April 2026; Received in revised form 7 May 2026; Accepted 7 May 2026
Available online 9 May 2026
0950-4230/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

chemical processes by establishing differential equations for material balance, energy balance, and reaction kinetics, thereby assessing the risk evolution process (Liao and Shu, 2026). Although these methods can characterize the variation patterns of process parameters from a physical perspective, their application faces several challenges (Yuan et al., 2026): (1) accurate modeling depends on complete knowledge of reaction mechanisms and precise acquisition of physical property parameters. This becomes increasingly difficult for complex reaction systems or scenarios with frequent operating condition changes. Consequently, such approaches suffer from high modeling difficulty and computational complexity; (2) mechanistic models struggle to integrate sensor data in real time and lack adaptive capabilities to address model errors and parameter uncertainties.

In recent years, with the rapid advancement of the industrial internet of things and intelligent sensing technologies, chemical process systems have accumulated massive amounts of process operation data, providing a foundation for data-driven risk prediction. Algorithms such as support vector machines (SVM) (Jiao et al., 2023), artificial neural networks (ANN) (Raeihagh et al., 2020), and random forests (RF) (Bellè et al., 2023) are employed for state identification and fault diagnosis of process parameters. Gao et al. (2025) presented a method based on machine learning to analyze the causes of production accidents in CPS systems. Li et al. (2024) proposed a risk assessment method based on SVM. This method was then applied to the risk assessment and management of hazardous chemical warehouses. Abdalla et al. (2025) introduced a ANN-based approach for real-time assessment of CO2 leak and dispersion from injection wells and surface facilities. The dataset is generated from PHAST simulations. Hassen et al. (2025) presented a RF-based approach for evaluating accidental release. The dataset is driven from two thousand simulations by ALOHA. The research results show that in terms of predicting gas leaks, the performance of RF is better than that of SVM. Li e t al. (2021) introduced a SVM-based risk assessment for chemical process systems. The research results show that the proposed method can identify system risks and reduce the accident probability. Although the above studies can effectively assess the risk of chemical process systems, existing research still suffers from the following limitations:

(1) Insufficient attention to dynamic risk assessment driven by process parameters. Most current data-driven approaches continue to follow the conventional risk assessment paradigm, treating risk as a static probabilistic event based on historical data. The essence of risk in chemical processes lies in the gradual deviation of process parameters (such as temperature, pressure, flow rate, and liquid level) from safe ranges, rather than in sudden discrete events. How to continuously quantify the degree of risk based on real-time process parameter sequences and identify its evolutionary trajectory remains a challenging problem that needs to be solved.
(2) Research on predicting the trend of risk evolution is limited. Existing studies predominantly focus on system risk assessment at the current moment. However, chemical accidents usually undergo a gradual evolution process: from a normal state to deviation, then to early warning, then to critical condition, and finally to accident. If the system risk could be predicted before it reaches the alarm threshold, operators would be afforded valuable time for proactive intervention. Currently, the research on the dynamic prediction of process parameters in chemical systems is still at an early stage, and the methods that combine the prediction results with risk assessment are even rarer. Developing risk prediction models that can provide early warnings is of crucial importance for enhancing chemical process safety.

To address the aforementioned research gaps, this paper proposes a dynamic risk prediction method for chemical process systems based on enhanced feature engineering and eXtreme Gradient Boosting (XGBoost) integrated with process parameter fusion. The main contributions are as follows:

(1) Construction of a risk quantification indicator that considers the multi-dimensionality and time-varying characteristics of process parameters. This indicator retains the physical meaning of the original parameters, and at the same time provides a continuous target variable for subsequent risk prediction, overcoming the limitation of traditional methods that can only output discrete risk states.
(2) By conducting simulation to obtain real-time data on the system's process parameters, the uncertainty in data acquisition has been reduced to a certain extent.
(3) Development of a time-series prediction method for predicting the risks of a chemical process system. The proposed approach enables early perception of risk evolution trajectories, providing technical support for proactive risk prevention and control.

The remainder of this paper is organized as follows. The proposed methodology is described in Section 2. Section 3 shows the application of the proposed method in a case study. Section 4 concludes the study.

2. The proposed methodology

The impact of disturbance events causing key process parameters in chemical process systems to exceed safety thresholds is a primary cause of accidents such as leakages, fires, and explosions. An accurate risk assessment and prediction method can assist operators in taking effective preventive measures before process parameters exceed the thresholds. The framework of the proposed methodology is depicted in Fig. 1. First, a comprehensive risk indicator for the chemical process system is established through a two-step procedure: (i) identification of KPPs that reflect the system's operational status, along with their corresponding safety thresholds, based on process hazard analysis and engineering considerations; and (ii) aggregation of these KPPs into a unified risk indicator according to their respective importance. Subsequently, a dynamic simulation model of the chemical process system is developed to generate time-series data of the KPPs under various disturbance scenarios. Finally, a data-driven risk assessment and prediction model is constructed using a hybrid XGBoost framework. Fig. 1 illustrates the main steps of the proposed methodology, which are detailed in the following sections.

2.1. Construction of comprehensive risk indicators

2.1.1. Identification of key process parameters

The risk evolution of chemical process systems is a complex process involving multiple factors that are coupled together. The contribution of different process parameters to system safety varies significantly (Qi et al., 2025). Reasonably determining the KPPs is the basis for constructing comprehensive risk indicators and is directly related to the accuracy of risk assessment. The selection of key process parameters should follow the following basic principles (Moslem et al., 2025):

(1) The selected parameters should be directly related to the safety of the system, and their abnormal changes should be able to reflect potential risks. For example, temperature directly determines the reaction rate and the risk of thermal runaway.
(2) The selected parameters should meet the requirements for real-time online monitoring and be capable of being continuously collected through sensors.
(3) The selected parameters should be sensitive to changes in the operation conditions and be able to detect abnormal signals at the early stage of risk evolution.
(4) Relatively independent parameters should be selected to avoid information redundancy.

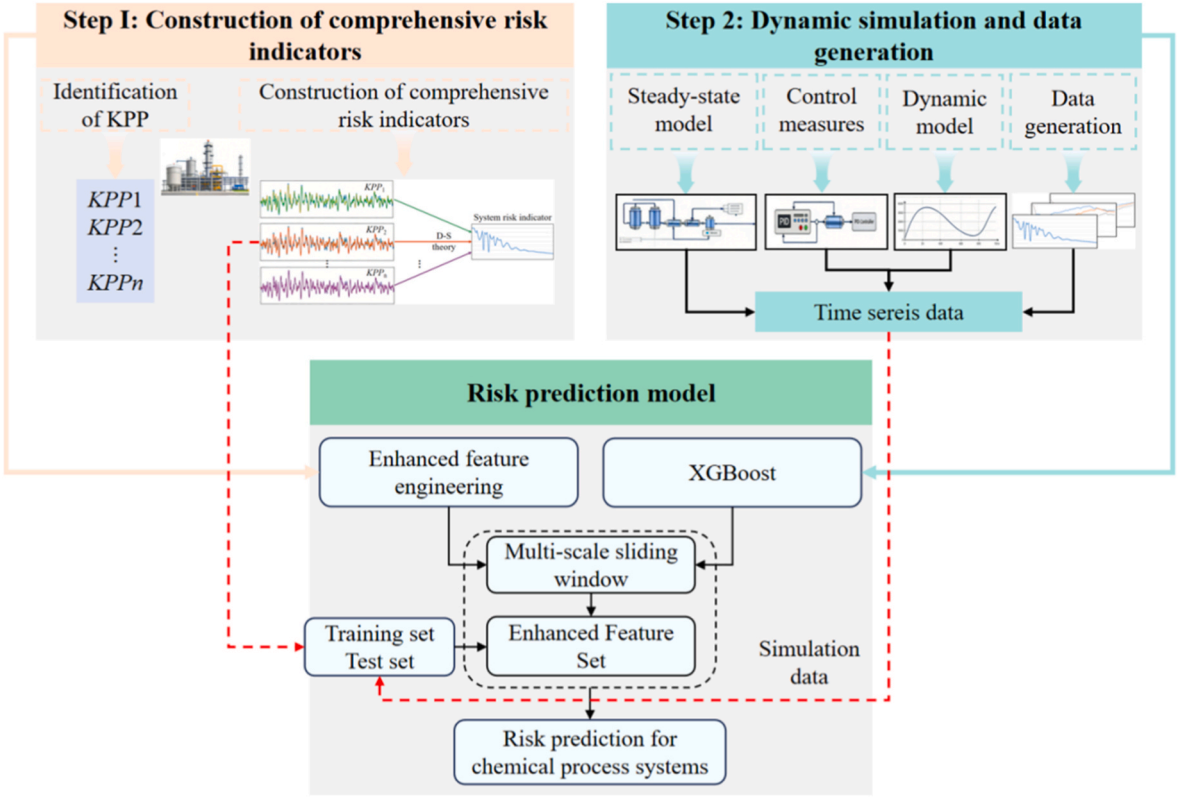


Fig. 1. The framework of the proposed methodology.

After determining the key process parameters, the safety threshold needs to be determined based on the design specifications, industry standards, and historical operation data (Yadiz et al., 2024). Accordingly, for a given KPP_i, the corresponding potential disruptive event D is defined as:

$$KPP_i(t) \geq KPP_i^{\max} \text{ or } KPP_i(t) \leq KPP_i^{\min}, \forall t \in T \tag{1}$$

where $KPP_i^{\max}$ and $KPP_i^{\min}$ represent the upper and lower operating limits for the KPP.

2.1.2. Representation of system risk based on D-S theory

D-S theory is employed to integrate various process parameters into a single indicator that can comprehensively represent the system risk. D-S theory can effectively address issues related to incomplete information, imprecision, and conflicts among information sources (Zeng et al., 2022; Kalathil et al., 2020). Due to the ability of Dempster's combination rule to synthesize knowledge and data from different experts and information sources, D-S theory has been widely applied across various fields (Sezer et al., 2023). In risk analysis, expert knowledge is integrated by defining a belief structure and employing the reasoning algorithm of evidence theory, thereby determining the degree of belief in events.

D-S theory encompasses four fundamental concepts: the frame of discernment, basic probability assignment, belief function, and plausibility function.

(1) Frame of discernment

Let $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ be a non-empty set consisting of n complete and mutually exclusive elements. Then Θ is called the frame of discernment. In D-S theory, 2^Θ is referred to as the power set of Θ, which is composed of all subsets of the frame of discernment, satisfying $2^\Theta = \{\varnothing, (\theta_1), (\theta_2), ..., (\theta_n), (\theta_1, \theta_2), (\theta_1, \theta_3), ..., (\theta_1, \theta_n), ...\Theta\}$.

(2) Basic probability assignment

Let A be a subset of Θ. If the mapping m: $2^\Theta \rightarrow [0, 1]$ satisfies Eq. (2), then m is called the basic probability assignment function on Θ. Any A with $m(A) > 0$ is termed a focal element, representing the degree of support that the evidence provides for A.

$$\begin{cases} m(\varnothing) = 0 \\ \sum_{A \subseteq \Theta} m(A) = 1 \end{cases} \tag{2}$$

(3) Belief function and plausibility function

Let A and B be subsets of Θ. Then the belief function Bel is defined as shown in Eq. (3), and the plausibility function Pl is defined as shown in Eq. (4):

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{3}$$

$$Pl(A) = \sum_{B \cap A \neq \varnothing} m(B) \tag{4}$$

where the interval [0, $Bel(A)$] represents the degree of belief in proposition A, the interval [$Bel(A)$, $Pl(A)$] denotes the uncertainty of the evidence, and the interval [$Pl(A)$, 1] indicates the degree of disbelief in proposition A.

(4) Dempster's Combination Rule

The combination rule is the core of D-S theory and is used to synthesize information from different sources. After the belief function and plausibility function are determined, Dempster's combination rule is employed to fuse multiple pieces of evidence. If two basic probability assignment functions m_1 and m_2 are derived from two distinct pieces of

evidence, they can be combined according to Dempster's rule, as shown in Eq. (5):

$$m(A) = m_1 \oplus m_2(A) = \begin{cases} \dfrac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1 - K}, A \neq \varnothing, A \subseteq \Theta \\ 0, A = \varnothing \end{cases} \tag{5}$$

where K is the conflict coefficient, which represents the degree of discrepancy between different pieces of evidence and satisfies Eq. (6):

$$K = \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C) \tag{6}$$

The process of establishing the system risk indicators based on D-S theory is shown in Fig. 2.

2.2. Dynamic simulation and data generation

Dynamic simulation is based on rigorous thermodynamic models and comprehensive property databases. It generates high-fidelity time-series data of process parameters (Sun et al., 2024). This enables accurate characterization of nonlinear coupling relationships and dynamic evolution patterns among process variables. Unlike historical data, which are limited by actual operating records, dynamic simulation allows systematic modeling of various operating conditions under controlled settings (Qi et al., 2025; Cui et al., 2023). These include normal operations, disturbances, and extreme scenarios such as feed fluctuations, cooling failures, and valve malfunctions. This approach achieves comprehensive coverage of the operating condition space. Aspen PLUS is used to develop the dynamic simulation model for chemical process systems (Salehifar and Farsi, 2026). The model development process is shown in Fig. 3.

2.3. Risk prediction based on enhanced feature engineering and XGBoost

The overall framework of the dynamic risk prediction method for chemical process systems consists of two main modules: enhanced feature engineering and XGBoost, this approach enables dynamic prediction of the comprehensive risk indicator for chemical processes.

Enhanced feature engineering is a method that extracts rich features from raw process parameters using multi-scale sliding window techniques. Specifically, the method first employs sliding windows of different sizes (e.g., window sizes of 3, 5, and 8) to extract raw temporal features, capturing the variation patterns of parameters at different time scales. On this basis, statistical features within the windows (mean, standard deviation, maximum, minimum, and range) are calculated to characterize the overall distribution of the parameters. First-order and second-order difference features are constructed to capture the variation trends of the parameters. Disturbance detection and decay features are introduced to quantify the persistent impact of disturbance events on the system. Interaction features are also constructed to reflect the coupling relationships among parameters. Through the above multi-dimensional feature extraction and construction, enhanced feature engineering can fully exploit the implicit information contained in the raw data, providing rich information input for subsequent prediction models.

XGBoost is an ensemble learning algorithm based on the gradient boosting framework. It can be understood as "combining multiple simple decision tree models to form a powerful predictive model". Its core principle involves iteratively constructing decision trees, where each new tree is trained to fit the residuals of the previous model, thereby progressively optimizing the prediction results, as shown in Fig. 4. Compared with traditional gradient boosting methods, XGBoost approximates the loss function using first-order and second-order Taylor expansions, significantly accelerating convergence. At the same time, regularization terms (including the number of leaf nodes and the L2 norm of leaf node weights) are incorporated into the objective function to effectively control model complexity and prevent overfitting. Furthermore, XGBoost supports optimization strategies such as column subsampling, row subsampling, and parallel computing, enabling high computational efficiency when handling high-dimensional features. Combining enhanced feature engineering with XGBoost, on the one hand, improves the model's ability to capture complex nonlinear relationships through rich feature inputs; on the other hand, leverages XGBoost's powerful fitting and regularization mechanisms to achieve high-precision and high-robustness risk prediction.

3. Case study

To illustrate the application of the proposed risk prediction method, a continuous stirred-tank reactor (CSTR) case study originally presented by Al-Malah (2016) is adopted. CSTR exhibit typical strong nonlinearity, multivariable coupling, and thermodynamic characteristics. Their state variables, such as reactant concentration and reaction temperature, are highly coupled. This can lead to complex dynamic behaviors. In addition, the reaction process inside a CSTR involves significant exothermic or endothermic effects. These effects make temperature control difficult. As a result, system operation carries a high level of risk. The primary reaction taking place in this CSTR is an exothermic process, in which ethylene (E) and benzene (B) react to form the target product, ethylbenzene (EB). In addition, two undesired side reactions occur concurrently: first, E further reacts with the already-formed EB to produce an unwanted by-product diethylbenzene (DEB); second, B reacts with DEB to regenerate EB. The overall reaction scheme is as follows:

$$C_2H_4 + C_6H_6 \rightarrow C_8H_{10} \tag{7}$$

$$C_2H_4 + C_8H_{10} \rightarrow C_{10}H_{14} \tag{8}$$

$$C_6H_6 + C_{10}H_{14} \rightarrow 2C_8H_{10} \tag{9}$$

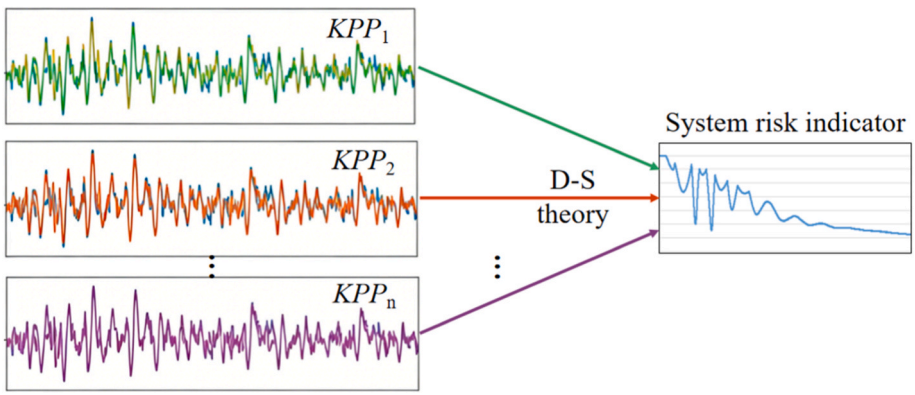


Fig. 2. Diagram of system risk indicator establishment.

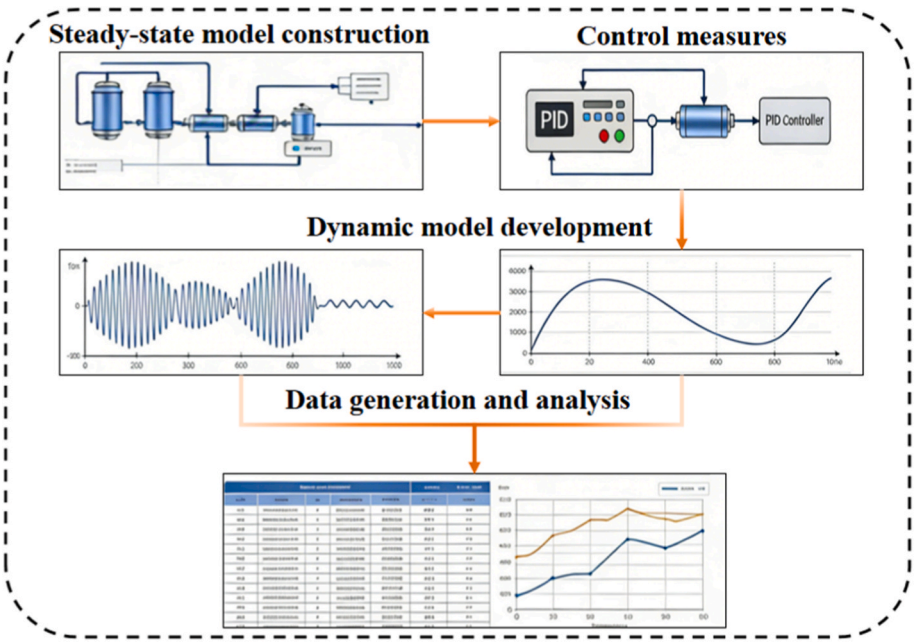


Fig. 3. The performance curves of CPPs and the overall system performance.

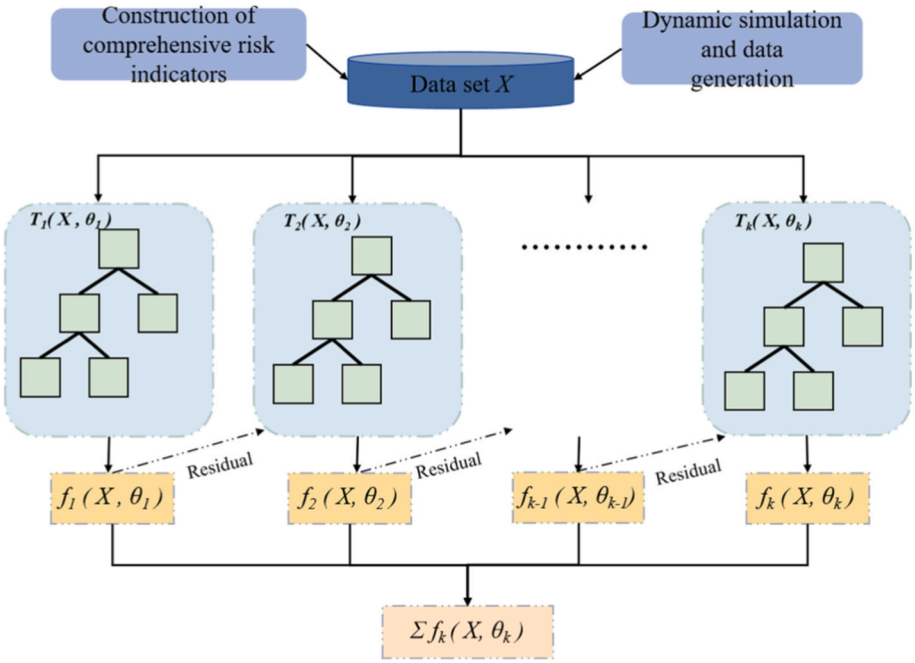


Fig. 4. XGBoost flowchart.

Fig. 5 depicts the process flow and control configuration of the CSTR system. Table 1 lists the specifications and parameter values used for modeling the CSTR.

3.1. Construction of comprehensive risk indicators

To effectively evaluate and predict the dynamic risk of the CSTR system, three KPPs are identified as critical indicators, including the CSTR liquid level, the CSTR temperature and the product flow rate. The liquid level and temperature are closely related to process safety, as they reflect the mass balance and thermal stability of the CSTR. In contrast, the product flow rate serves as an indicator of production efficiency, directly representing the system's output performance. The combination of these parameters provides a reliable foundation for reflecting CSTR behavior, covering process safety, operational stability, and system output. Table 2 lists the established safety thresholds for liquid level, temperature and product flow rate, which are essential for accurately assessing and predicting systemic risks.

The operational state of CSTR is inherently subject to various disturbances. Accurate identification of these disturbances serves as a fundamental prerequisite for assessing and predicting the CSTR system risk. Disturbances can be broadly categorized into human errors and equipment failures. The impact of such disturbance events on the system is primarily manifested through deviations in the KPPs of CSTR. For example, variations in feed flow rate or cooling water temperature can significantly alter the value of KPPs for CSTR, thereby posing potential

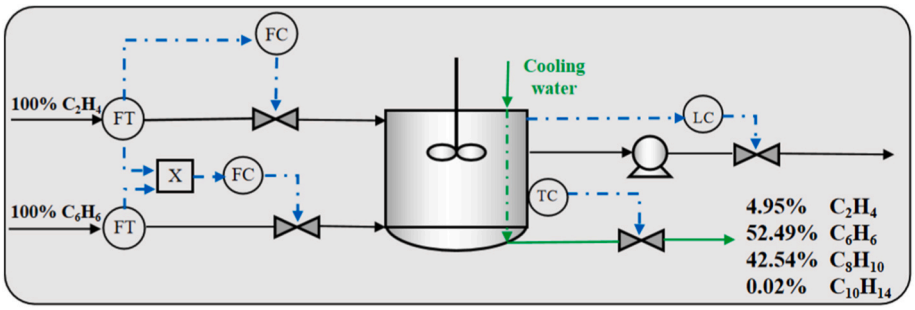


Fig. 5. CSTR with basic control structure.

Table 1
Parameter values of the CSTR.

Parameter	value	Parameter	value
Flow rate of E	360 kmol/h	Reactor volume	60 m^3
Flow rate of B	720 kmol/h	Heat transfer option	Dynamic
B/E feed temperature	21.85 °C	Medium temperature	24.85 °C
B/E feed pressure	2 MPa	Temperature approach	77.33 °C
Reactor temperature	21.85 °C	Heat capacity	4200 J/kg/K
Reactor pressure	1.5 MPa	Medium holdup	1000 kg
Reactor level	8.64m	Equipment mass	6800 kg

Table 2
Values of CPPs for the CSTR.

Parameter	Expected value	Safety threshold
CSTR level	8.64 m	[8.0, 10]
CSTR Temperature	126.85 °C	[120, 140]
Product flow rate	757.509 kmol/h	[600, 900]

risks to operational safety. To simulate the effects of unexpected disruptions on the CSTR operation, the values of selected input variables can be systematically varied.

Feed temperature variations directly affect the reaction rate and thermal balance of a CSTR. An increase in feed temperature introduces additional sensible heat into the reaction system, leading to a corresponding rise in the initial reaction temperature within the CSTR. Since both the primary reaction (E + B→EB) and the side reaction (E + EB→DEB) are exothermic, a higher feed temperature further accelerates the reaction rate and releases more reaction heat. Under severe conditions, this may lead to CSTR runaway and pose thermal safety hazards. Conversely, an excessively low feed temperature lowers the initial temperature of the reaction mixture, suppresses the reaction rate, and reduces the conversion of the primary reaction. Moreover, the relative extent of the side reaction may change, thereby affecting product distribution and yield. Therefore, feed temperature is a key operating variable that significantly influences the thermal stability, reaction selectivity, and process safety of the CSTR.

3.2. Dynamic simulation

3.2.1. Identification of potential disruption

Taking the temperature variation of Feed B as the disruptive event, its temporal evolution exhibits distinct jump characteristics, as shown in Fig. 6. The disruptive event occurs in two stages. In the first stage, at the 3rd hour, the temperature jumps and then remains stable. In the second stage, at the 6th hour, the temperature jumps again and stays stable. Within the range of 0 to 3 h, the temperature fluctuates steadily within the range of 24.06-25.88 °C, indicating a relatively steady state. After 3 h, the temperature suddenly rises to 30.43 °C. It then fluctuates violently between 29.27 °C and 31.26 °C for about 3 h. This marks the first strong disruptive event. Subsequently, After 6 h, the temperature undergoes

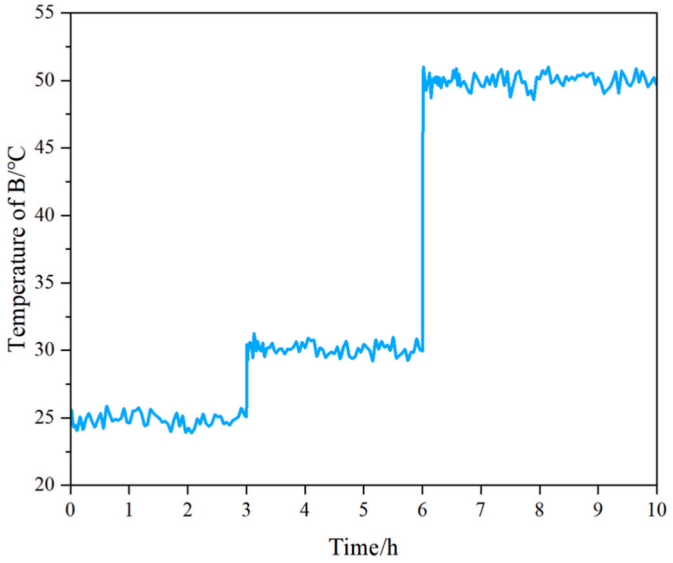


Fig. 6. The temperature variation of feed B.

another abrupt change, surging sharply from 29.96 °C to 49.90 °C, reaching a peak of 50.99 °C at 6.02 h. Thereafter, it continues to oscillate with high frequency within the range of 48.73-50.99 °C. In summary, the temperature of Feed B experiences two significant step-like disturbances, with the second disturbance being larger in magnitude and longer in duration.

During the simulation period of 0-10 h, both the temperature of Feed E and the cooling water temperature remained within their normal operating ranges and exhibited a dynamically stable pattern with small fluctuations. Specifically, the temperature of Feed E varied between 23.58 °C and 26.36 °C, while the cooling water temperature fluctuated within the range of 23.71 °C to 25.97 °C. The variation amplitude of both temperatures was controlled within approximately 3 °C, with no abrupt step changes or pulsed disturbances observed. These results indicate that the system possesses good thermal stability and controllability under normal operating conditions. The temperature of Feed E and the cooling water temperature change over time are shown in Figs. 7 and 8.

3.2.2. Simulation data generation

During the 0-10 h simulation period, the CSTR liquid level remained between 7.6 m and 9.5 m, as shown in Fig. 9. During the period of 0-6 h, the liquid level fluctuated steadily within the range of 7.6 m to 9.3 m, indicating a relatively steady state. After the first disturbance event at 3 h, the CSTR liquid level did not show any significant change. The fluctuation remained stable. Following the second disturbance event at 6 h, the liquid level dropped to a minimum of approximately 7.9 m and rose to a maximum of approximately 9.4 m. The fluctuation did not increase

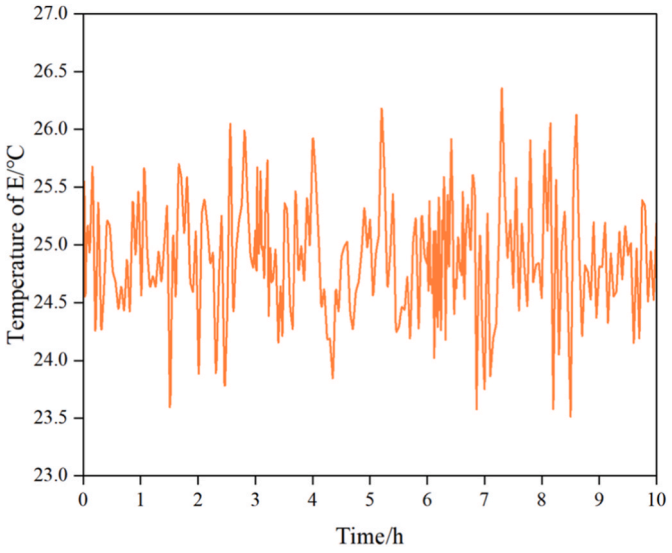


Fig. 7. The temperature variation of feed E.

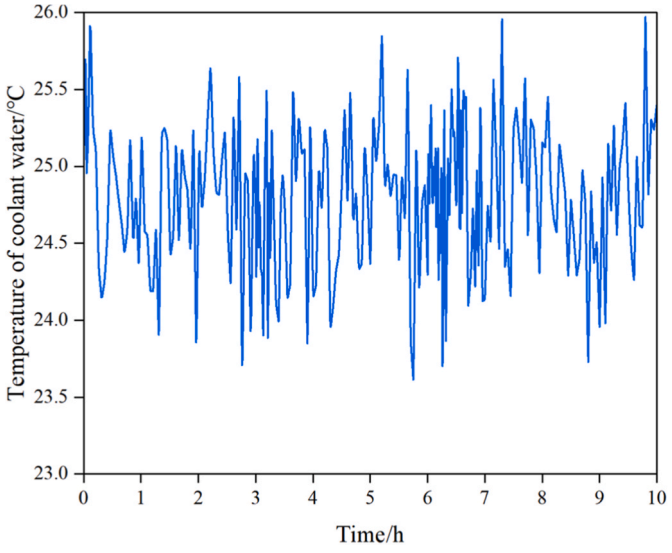


Fig. 8. The temperature variation of cooling water.

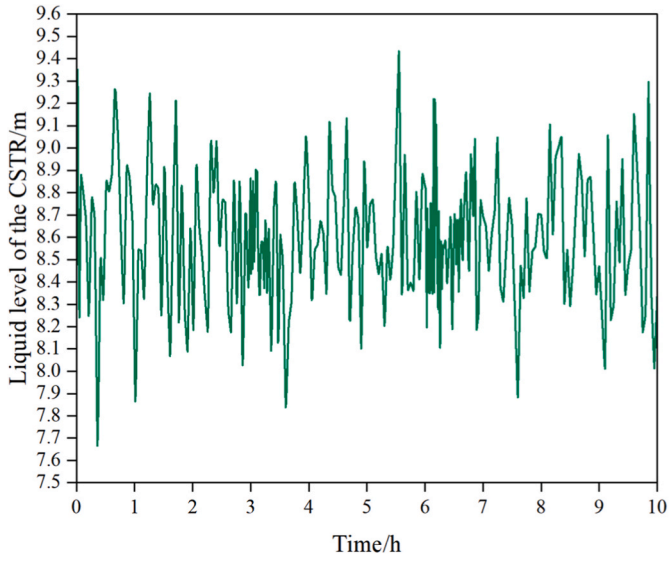


Fig. 9. The liquid level variation of the CSTR under the disruption.

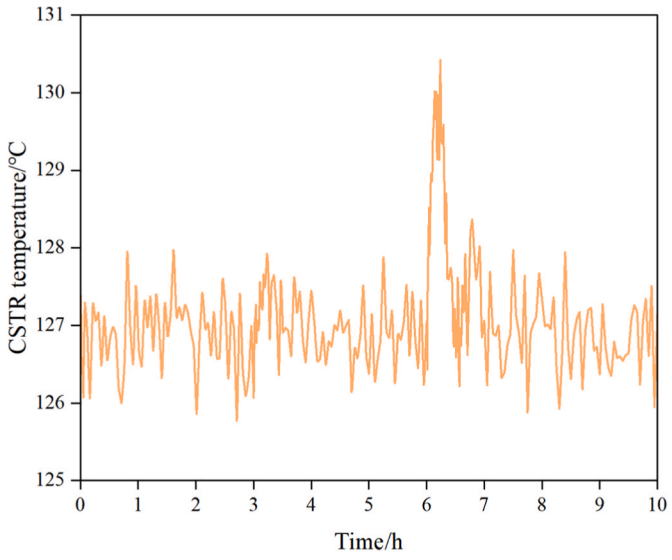


Fig. 10. The temperature variation of the CSTR under the disruption.

significantly. Overall, both disturbance events caused some degree of fluctuation in the liquid level. However, the liquid level remained within the safe threshold throughout the period. This indicates that the temperature variation of Feed B did not have a significant impact on the CSTR liquid level.

The CSTR temperature was relatively sensitive to the disruption. During the first 3 h, the temperature fluctuated slightly between 126.8 °C and 127.4 °C, indicating stable operation. After 3 h, the CSTR temperature increased slightly. It then oscillated between 125.8 °C and 128.0 °C. The fluctuation did not increase significantly compared to that before the disruption. However, after 6 h, the CSTR temperature exhibited significant fluctuations. The temperature increased sharply from approximately 126 °C to 130.4 °C, as shown in Fig. 10. This indicates that temperature changes of feed B will significantly intensify the exothermic reaction, preventing the reaction heat from being removed in time, thereby causing the CSTR temperature to increase sharply.

The response of the CSTR product flow rate to the two disturbance events was not significant, as shown in Fig. 11. During the period of 0-3 h, the flow rate fluctuated steadily within the range of 743 to 775 kmol/h. After the first disturbance event at 3 h, the flow rate did not show any obvious fluctuation. Following the second disturbance event at 6 h, the first peak and lowest point appeared in the flow rate fluctuation. The flow rate dropped to a minimum of approximately 738 kmol/h and rose to a maximum of approximately 779 kmol/h, with a fluctuation amplitude exceeding 40 kmol/h. This indicates that the disruption events had a certain effect on the outlet flow rate by influencing the reaction rate and material balance. Although the fluctuation amplitude was relatively large, the flow rate remained within the acceptable operating range, as the safe range of the product flow rate is 700-900 kmol/h.

3.3. Risk assessment and prediction

As shown in Fig. 12, during the 0-10 h simulation period, the normalized liquid level, temperature, and product flow rate of the CSTR all exhibited different degrees of fluctuation. The normalized liquid level mainly varied between 0.9 and 1.0, with relatively stable fluctuations. Only a few moments showed values below 0.9, indicating that the temperature fluctuation of Feed B had a limited impact on the CSTR liquid level. Notably, after the disturbance events, the CSTR temperature

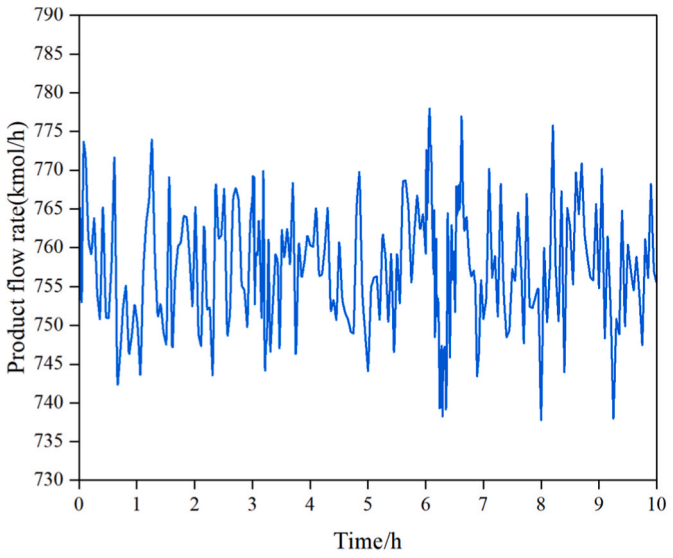


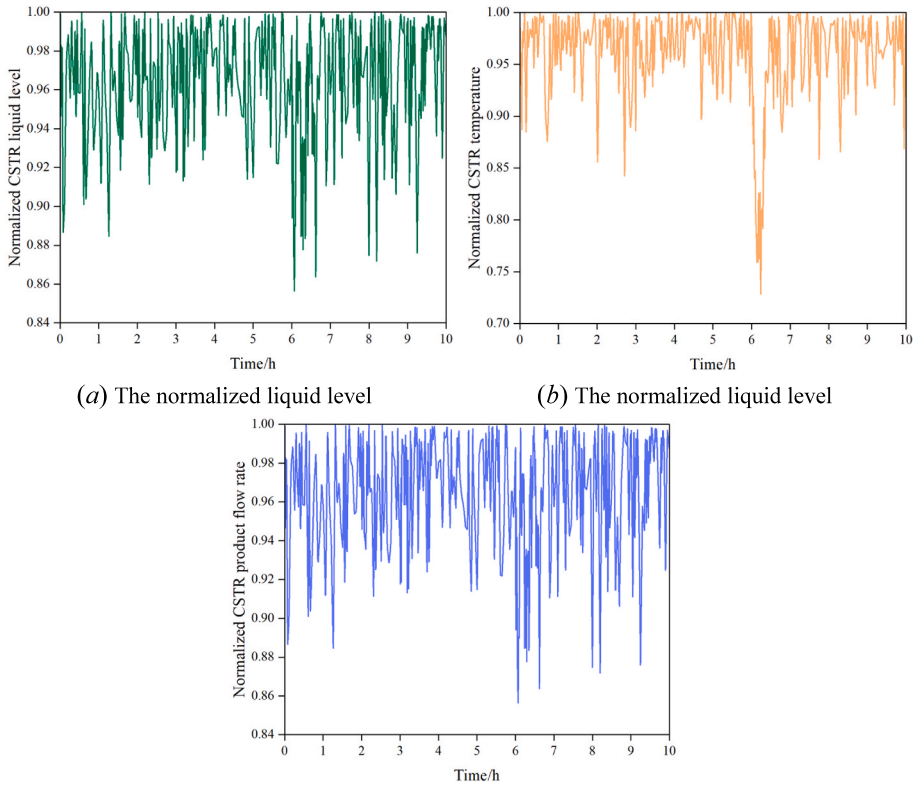
Fig. 11. The product flow rate variation of the CSTR under the disruption.

showed relatively large fluctuations, especially around 3 h and 6 h, reflecting its strong sensitivity to disturbances. The product flow rate also showed some correlation with the disturbance events. Its fluctuations were moderate, and the flow rate remained between 0.8 and 1.0 for most of the time, indicating a generally high level of product output. However, the flow rate reached its minimum at 6 h, indicating that the temperature change of Feed B influenced the reaction in the CSTR and thus reduced product generation to some extent. Overall, all three variables responded to the disruption. The temperature showed the most significant response, while the fluctuations in liquid level and flow rate were relatively mild.

Based on the D-S evidence theory, this study takes the normalized CSTR liquid level, temperature and product flow rate as three independent evidence sources and constructs their basic probability assignment functions to represent the initial support of each parameter for the system risk. The basic probability assignment of each piece of evidence is directly determined by its normalized value, with the remaining portion assigned to the frame of discernment to represent uncertainty. On this basis, Dempster's combination rule is applied to fuse the three pieces of evidence, comprehensively considering the contribution of liquid level, temperature, and flow rate to the system risk, thereby obtaining a comprehensive system risk indicator. The system risk under the disruption is shown in Fig. 13.

On the basis of the assessed system risk, the proposed risk prediction model is employed to forecast the CSTR system risk under the influence of disruptive events. A comparison between the predicted results and the actual values is presented in Fig. 14. From an overall trend perspective, the predicted values are generally consistent with the actual values, both fluctuating within the range of 0.9 to 1, indicating that the proposed model can effectively capture the overall evolution pattern of system risk. For the vast majority of sample points, the predicted values are close to the actual values, with relatively small errors. However, deviations between the predicted and actual values are observed around sample points 60-80, with a maximum error of approximately 0.06. Overall, the model maintains high prediction accuracy on the entire test set, validating the effectiveness of combining enhanced feature engineering with XGBoost for system risk prediction in chemical process systems.

According to the comparison results presented in Table 3, the proposed method outperforms the traditional algorithms across all evaluation metrics. Specifically, the proposed method achieves an RMSE of 0.159, which is 23.2% lower than that of linear regression (0.207) and 54.7% lower than that of SVM (0.351). The MAE is 0.122, representing reductions of 26.1% and 58.9% compared to linear regression (0.165)



(*a*) The normalized liquid level (*b*) The normalized liquid level

(*c*) The normalized CSTR product flow rate

Fig. 12. The normalized liquid level, temperature and flow rate of CSTR under the disruption.

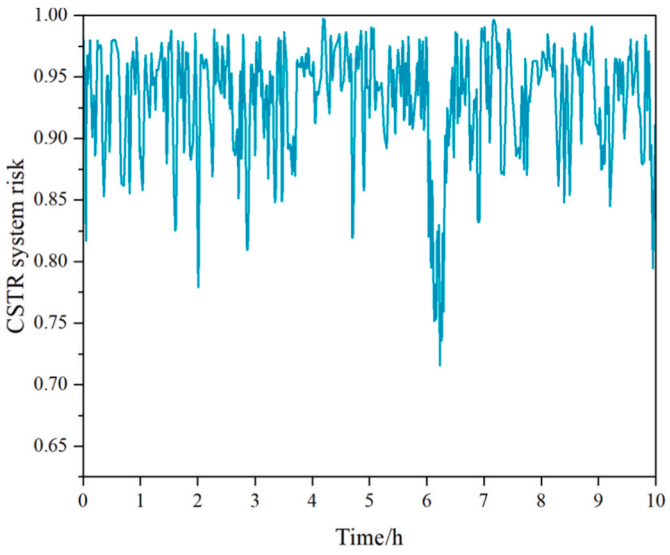


Fig. 13. The CSTR risk under the disruption.

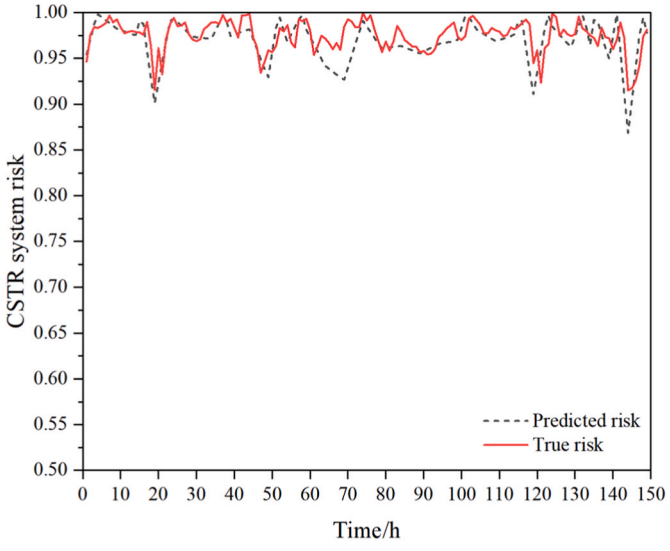


Fig. 14. Comparison of actual and predicted CSTR risk under disruptive events.

Table 3
Comparison of results among different algorithms.

Algorithm	RSME	MAE	R2	MAPE
The proposed methodology	**0.159**	**0.122**	**0.7237**	**0.100**
SVM	0.351	0.297	0.348	0.230
Linear regression	0.207	0.165	0.534	0.130

and SVM (0.297), respectively. The proposed method attains an R^2 of 0.7237, significantly higher than those of linear regression (0.534) and SVM (0.348). Furthermore, the MAPE is 0.100, which is 23.1% and 56.5% lower than those of linear regression (0.130) and SVM (0.230), respectively. These results demonstrate the clear superiority of the proposed approach. By integrating enhanced feature engineering with XGBoost, this method effectively captures complex nonlinear relationships and improves prediction accuracy for system risk in chemical process systems.

4. Conclusion

This paper presents a systematic study on the dynamic risk prediction method for chemical process systems, with the main research findings and conclusions summarized as follows.

(1) A comprehensive risk indicator construction method based on D-S evidence theory is proposed. By taking the normalized CSTR liquid level, temperature, and product flow rate as three independent evidence sources, the basic probability assignment functions are constructed and fused using Dempster's combination rule. This method effectively integrates the contributions of multiple key parameters to system risk, overcoming the limitation of using a single parameter for risk characterization. The resulting comprehensive risk indicator provides a reliable quantitative basis for subsequent risk prediction.
(2) An enhanced feature engineering method is developed to fully exploit the temporal dependencies and nonlinear coupling relationships among process parameters. Multi-scale sliding window techniques are employed to extract raw temporal features, statistical features, trend features, and disturbance features. The construction of disturbance detection and decay features effectively quantifies the impact of abrupt changes in temperature B. Interaction features are also constructed to capture the coupling relationships among parameters. These enhanced features provide rich information input for the prediction model.
(3) A risk prediction model combining enhanced feature engineering with XGBoost is established. The proposed method achieves superior performance in predicting CSTR system risk under disturbance events. Experimental results demonstrate that the proposed method significantly outperforms traditional algorithms such as linear regression and SVM. These results validate the effectiveness of the proposed method in capturing complex nonlinear relationships and improving prediction accuracy.

In summary, the proposed dynamic risk prediction method for chemical process systems, which integrates D-S evidence theory, enhanced feature engineering, and XGBoost, effectively addresses the limitations of traditional static probability assessment methods. It achieves a transition from static probability assessment to dynamic risk prediction, providing decision support for proactive risk prevention and control in chemical processes. Future work will focus on further improving the model's responsiveness to rapid risk fluctuations and extending the method to more complex chemical process systems.

CRediT authorship contribution statement

Hao Sun: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Jin Huang:** Writing – original draft. **Meng Qi:** Validation, Software. **Ming Yang:** Writing – review & editing, Supervision, Methodology. **Fuyu Wang:** Investigation, Formal analysis.

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.

Acknowledgments

The authors gratefully acknowledge the financial support provided by National Natural Science Foundation of China (52504194), Natural Science Foundation of Anhui, China (2508085QG281), Key Project of the Scientific Research Program of Universities in Anhui Province (2024AH050162) and the Anhui Province Philosophy and Social

Sciences Planning Project (AHSKYQ2024D014).

Data availability

No data was used for the research described in the article.

References

Abdalla, M., Alresheed, M., Wang, Q.S., 2025. Modeling CO2 leak and dispersion from injection wells and surface facilities using machine learning. J. Loss Prev. Process. Ind. 97, 105709. https://doi.org/10.1016/j.jlp.2025.105709.

Al-Malah, K.I., 2016. Aspen Plus: Chemical Engineering Applications. John Wiley & Sons, Hoboken, NJ, USA.

Aven, T., Zio, E., 2021. Globalization and global risk: how risk analysis needs to be enhanced to be effective in confronting current threats. Reliab. Eng. Syst. Saf. 205, 107270. https://doi.org/10.1016/j.ress.2020.107270.

Bellè, A., Abdin, A.F., Fang, Y.P., Zeng, Z.G., Barros, A., 2023. A data-driven distributionally robust approach for the optimal coupling of interdependent critical infrastructures under random failures. Eur. J. Oper. Res. 309, 872–889. https://doi.org/10.1016/j.ejor.2023.01.060.

Cui, C.T., Qi, M., Shu, C.M., Liu, Y., 2023. Rigorous dynamic simulation methodology for scenario-based safety analysis of pressure-swing distillation considering independent protections. Process Saf. Environ. Prot. 172, 282–304. https://doi.org/10.1016/j.psep.2023.02.010.

Gao, W.W., Yang, J.X., Fang, Y.D., Guo, W., Wang, Z.C., Chen, X.C., 2025. Insights into ensemble learning based data-driven model for prediction of chemical safety production accidents. Process Saf. Environ. Prot. 204, 108033. https://doi.org/10.1016/j.psep.2025.108033.

Hassen, O., Khan, Z.T., Irfan, M., Rashid, M.I., 2025. Beyond ALOHA- quickly predict accidental release of toxic chemicals using machine learning. J. Loss Prev. Process. Ind. 94, 105542. https://doi.org/10.1016/j.jlp.2024.105542.

Hu, J.Q., Dong, S.H., Zhang, L.B., Chen, Y.Y., Xu, K.K., 2021. Cyber-physical-social hazard analysis for LNG port terminal system based on interdependent network theory. Saf. Sci. 137, 105180. https://doi.org/10.1016/j.ssci.2021.105180.

Jiao, Z.R., Zhang, Z.R., Jung, S., Wang, Q.S., 2023. Machine learning based quantitative consequence prediction models for toxic dispersion casualty. J. Loss Prev. Process. Ind. 81, 104952. https://doi.org/10.1016/j.jlp.2022.104952.

Kalathil, M.J., Renjith, V.R., Augustine, N.S., 2020. Failure mode effect and criticality analysis using dempster shafer theory and its comparison with fuzzy failure mode effect and criticality analysis: a case study applied to LNG storage facility. Process Saf. Environ. Prot. 138, 337–348. https://doi.org/10.1016/j.psep.2020.03.042.

Khan, F., Yarveisy, R., Abbassi, R., 2021. Risk-based pipeline integrity management: a road map for the resilient pipelines. J. Pipeline Sci. Eng. 1, 74–87. https://doi.org/10.1016/j.jpse.2021.02.001.

Lai, C.C., Liu, C.C., Xiao, H.Z., Shu, C.M., 2026. The second residual hydrodesulphurisation of the refinery at the sixth naphtha cracker industrial park accident using fault tree analysis: a case study from a chemical complex in Taiwan. J. Loss Prev. Process. Ind. 99, 105812. https://doi.org/10.1016/j.jlp.2025.105812.

Li, S.M., Chen, G.H., Men, J.K., Li, X.F., Zhao, Y.M., Xu, Q.M., et al., 2024. A novel dynamic risk assessment method for hazardous chemical warehouses based on improved SVM and mathematical methodologies. J. Loss Prev. Process. Ind. 89, 105302. https://doi.org/10.1016/j.jlp.2024.105302.

Li, X.H., Zhang, Y.Q., Zhang, Y.H., Zhang, R.R., 2026. Leak diagnosis of string in underground gas storage using hybrid machine learning algorithms. J. Loss Prev. Process. Ind. 99, 105816. https://doi.org/10.1016/j.jlp.2025.105816.

Liao, S.W., Shu, C.M., 2026. Atmospheric effects on ABS powder thermokinetics and explosibility: mechanistic insights from integrated thermoanalysis and adiabatic-sphere tests. J. Loss Prev. Process. Ind. 99, 105825. https://doi.org/10.1016/j.jlp.2025.105825.

Moslem, S., Gholamizadeh, K., Zarei, E., Pasman, H., Martinez-Pastor, B., Pilla, F., 2025. A comparative assessment of domino accident analysis methods in process industries using LMAW and DNMA techniques. Reliab. Eng. Syst. Saf. 260, 110981. https://doi.org/10.1016/j.ress.2025.110981.

Nehal, N., Lounis, Z., Bouhadiba, B., Lounis, Z., 2023. Modelling of heating furnace fire scenarios using fault tree analysis, a bayesian network, and a thermal transfer method for system reliability analysis. J. Loss Prev. Process. Ind. 83, 104995. https://doi.org/10.1016/j.jlp.2023.104995.

Qi, M., Kim, Y., He, T.B., Lee, I., Park, J., 2025. Sustainable LNG supply chain enabled by clean and cost-effective energy self-integration via cold storage and the Allam cycle. Energy 320, 135126. https://doi.org/10.1016/j.energy.2025.135126.

Raeihagh, H., Behbahaninia, A., Aleagha, M.M., 2020. Risk assessment of sour gas inter-phase onshore pipeline using ANN and fuzzy inference system – case study: the south pars gas field. J. Loss Prev. Process. Ind. 68, 104238. https://doi.org/10.1016/j.jlp.2020.104238.

Reniers, G., Amyotte, P., 2012. Prevention in the chemical and process industries: future directions. J. Loss Prev. Process. Ind. 25, 227–231. https://doi.org/10.1016/j.jlp.2011.06.016.

Salehifar, F., Farsi, M., 2026. Development of a hydrogen production plant from H2S: process synthesis and simulation. Int. J. Hydrogen Energy 203, 152955. https://doi.org/10.1016/j.ijhydene.2025.152955.

Sezer, S.I., Elidolu, G., Akyuz, E., Arslan, O., 2023. An integrated risk assessment modelling for cargo manifold process on tanker ships under FMECA extended Dempster–Shafer theory and rule-based Bayesian network approach. Process Saf. Environ. Prot. 174, 340–352. https://doi.org/10.1016/j.psep.2023.04.024.

Shi, J.H., Li, J.J., Li, Y.X., Xie, Z.H., Li, Y.T., Zhang, L.B., 2026. Proactive decision-making agent for industrial leakage and explosion emergencies powered by Physics_GNN and LLM. J. Loss Prev. Process. Ind. 102, 105976. https://doi.org/10.1016/j.jlp.2026.105976.

Sun, H., Yang, M., Zio, E., Li, X.H., Lin, X.F., Huang, X.J., et al., 2024. A simulation-based approach for resilience assessment of process system: a case of LNG terminal system. Reliab. Eng. Syst. Saf. 249, 110207. https://doi.org/10.1016/j.ress.2024.110207.

Sun, H., Qi, M., Yang, M., Wang, F.Y., Wang, H.P., 2025. Multiparametric resilience assessment of chemical process systems incorporating process dynamics and independent protection layers. Process Saf. Environ. Prot. 197, 107018. https://doi.org/10.1016/j.psep.2025.107018.

Yadiz, M., Zarei, E., Pirbalouti, R.H., Li, H., 2024. A comprehensive resilience assessment framework for hydrogen energy infrastructure development. Int. J. Hydrogen Energy 51, 928–947. https://doi.org/10.1016/j.ijhydene.2023.06.271.

Yang, M., Sun, H., Geng, S.Y., 2023. On the quantitative resilience assessment of complex engineered systems. Process Saf. Environ. Prot. 174, 941–950. https://doi.org/10.1016/j.psep.2023.05.019.

Yuan, S.Q., Li, J.P., Wang, C.J., Zhang, X.G., 2026. Conformal machine learning for reliable anomaly detection in industrial cyber-physical systems. Reliab. Eng. Syst. Saf. 274, 112417. https://doi.org/10.1016/j.ress.2026.112417.

Zeng, T., Chen, G.H., Reniers, G., Men, J.K., 2022. Developing a barrier management framework for dealing with Natech domino effects and increasing chemical cluster resilience. Process Saf. Environ. Prot. 168, 778–791. https://doi.org/10.1016/j.psep.2022.10.024.

Zio, E., 2022. Prognostics and health management (PHM): where are we and where do we (need to) go in theory and practice. Reliab. Eng. Syst. Saf. 218, 108119. https://doi.org/10.1016/j.ress.2021.108119.