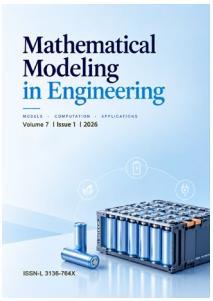
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 1, January, 2026



Physics-Constrained Neural Hawkes GNN for Reactor Runaway Precursor Recognition

Nenad Miljković[a,*], Pavle Stošić[a]

[a] *Faculty of Safety Engineering, University of Belgrade, Belgrade, 11000, Serbia*

Received 22 December 2025; Revised 20 March 2026; Accepted 26 March 2026; Published 30 March 2026

Abstract

Thermal runaway in stirred and semi-batch reactors is often preceded by weakly coupled alarm bursts, heat-removal imbalance, and delayed pressure response, yet these signatures are difficult to distinguish from benign operating transients when measurements are irregular. This paper proposes a physics-constrained Neural Hawkes graph neural network in which a process-topology encoder conditions marked event intensities and a normalized heat-balance residual jointly regularizes training and gates online risk. The principal novelty is this coupled mechanism: event clustering can increase risk only when graph-propagated process evidence and thermal disequilibrium support the escalation. Evaluation uses a controlled synthetic and historian-like benchmark containing 12,480 operating windows, 38 measured variables, and 412 incipient-runaway cases; it is intended for method comparison rather than plant certification. The model achieved an F1 score of 0.918, an AUROC of 0.973, and a median warning lead time of 18.6 min. Relative to LSTM, temporal convolution, Hawkes-only, and graph-only baselines, it reduced missed precursors by 31.4% and nuisance alarms by 22.7% at a fixed recall of 0.92. Ablation tests indicate that event timing supports aggregate recognition, graph propagation captures cross-unit coupling, and the physical residual is particularly important under cooling degradation. The framework therefore provides an interpretable supervisory warning layer for subsequent pilot- and plant-scale validation.

Keywords: Runaway Precursor Recognition, Neural Hawkes Process, Physics-Constrained Learning, Graph Neural Network, Reactor Safety

1. Introduction

Exothermic reactors remain central to fine chemical, pharmaceutical, and polymer manufacturing, but their safety margin can narrow rapidly when heat generation exceeds heat removal [1]. In many industrial incidents, the decisive change does not appear as a single limit violation; it emerges as a sequence of small deviations in reactor temperature, jacket duty, feed response, and pressure behavior [2]. Conventional protection layers are designed to intervene when a measured variable crosses a fixed boundary, whereas early warning requires recognition of evolving temporal structure before that boundary is reached [3]. Plant historians now record dense multivariate measurements, yet these records are unevenly sampled, interrupted by maintenance states, and dominated by normal operation [4]. These characteristics make precursor recognition different from ordinary fault classification. A model must recognize a sparse and safety-critical pattern under class imbalance, preserve sensitivity to weak thermal disequilibrium, and avoid turning every disturbance into an alarm. Reactor runaway prediction also carries a physical meaning: the latent risk score must remain compatible with heat accumulation and process coupling, because operators will not trust a warning that contradicts the observed energy balance [5].

* Corresponding author: nenad.mi@tmf.bg.ac.rs

https://doi.org/10.68406/mme.2026.vol7iss1nm11:132-143

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

Recent developments in temporal deep learning have improved anomaly detection in chemical processes, but several limitations remain visible when the target is incipient runaway rather than mature fault diagnosis. Recurrent networks can absorb long histories, although they often blur irregular alarm timing into uniform time steps [6]. Temporal convolutional models provide stable training, but their receptive field must be tuned to each process time constant [7]. Attention-based methods can assign weights to important observations, yet the resulting weights do not automatically encode a causal or physical relation among units [8]. Point-process models are attractive because an alarm, valve action, or threshold crossing is naturally an event, and the conditional intensity can represent self-excitation during a developing incident [9]. However, a Hawkes-type intensity that is learned only from event histories may increase for statistically plausible but physically impossible reasons. Jia et al. [10] showed that a graph-guided representation can improve process anomaly detection by modeling the topological and nonlinear relationships among process variables. A credible precursor detector must therefore combine event-time modeling, spatial coupling among plant variables, and a constraint that discourages warning patterns inconsistent with reactor heat balance.

This paper develops a physics-constrained Neural Hawkes graph neural network for reactor runaway precursor identification. Three functions are integrated rather than applied as independent predictors. First, a directed process graph propagates deviations along thermal, hydraulic, and control pathways. Second, a marked Neural Hawkes decoder preserves continuous alarm and intervention times instead of forcing every event onto a uniform grid. Third, a normalized heat-balance residual enters the training penalty and the online intensity gate, discouraging warnings that are statistically likely but thermally unsupported. The resulting output includes a precursor probability, an event-intensity trajectory, and a lead-time estimate under a validation-calibrated threshold. The evaluation uses a controlled synthetic and historian-like benchmark representing normal campaigns, feed perturbations, cooling degradation, sensor bias, agitation loss, and incipient runaway. Evidence is reported through comparative accuracy, calibration, lead time, nuisance-alarm burden, ablation behavior, stress tests, and online latency; these results establish method-level feasibility rather than certification for an operating plant.

To connect the methodological argument with the evidence needed for engineering use, the remainder of this paper moves from prior knowledge to model construction and then to operational validation. Section 2 reviews data-driven process monitoring, temporal point-process modeling, and graph learning for coupled industrial systems, clarifying why none of these strands alone fully represents an incipient runaway. Section 3 presents the physics-constrained Neural Hawkes GNN architecture and defines the process graph, event marks, and heat-balance residual. Section 4 describes the training objective, precursor scoring rule, threshold calibration, and deployment workflow. Section 5 examines the benchmark, comparative accuracy, ablation behavior, robustness, computational cost, and operator-facing interpretation through one table and four data figures. Section 6 closes the paper with the principal contribution, current limitations, and directions for plant-scale validation.

2. Related Work

2.1. Data-Driven Monitoring of Reactor Runaway Risk

Early warning for exothermic reaction systems has traditionally combined calorimetric knowledge, threshold-based interlocks, and multivariate statistical monitoring. Principal component analysis and partial least squares methods can identify deviations from normal operating correlations, but their linear assumptions restrict sensitivity when a reaction accelerates after a long induction period [11]. Kernel extensions and support vector approaches improve nonlinear separation, although they require careful selection of features and operating modes when batch recipes vary [12]. Deep recurrent models later became common for plant historian data because they can ingest long multivariable trajectories and learn temporal dependencies without hand-crafting every lag. Their drawback in runaway precursor recognition is that irregular alarm arrivals, missing samples, and event bursts must be resampled into a grid, which can dilute the meaning of clustered safety events [13]. Zhang et al. [14] demonstrated that a physical graph-based spatiotemporal representation can improve process fault diagnosis by embedding process-variable relationships into feature learning. These studies motivate an early-warning model that is not purely statistical and not restricted to manually fixed indicators.

For reactor safety, the cost of a missed precursor is substantially different from the cost of a nuisance alarm. This imbalance has encouraged cost-sensitive learning, synthetic incident augmentation, and one-class detection under normal data. Nevertheless, a high anomaly score does not always indicate runaway development: feed disturbances, sensor recalibration, product-viscosity shifts, and scheduled utility transitions may generate multivariate deviations without sustained heat accumulation. Prior graph-based diagnosis mainly organizes cross-variable dependence, whereas physics-informed anomaly detection commonly regularizes continuous trajectories. The present method instead uses the graph state to condition a marked event intensity and uses a heat-balance residual to restrict how that intensity becomes an operational warning. This distinction targets the temporal ordering, process connectivity, and thermal plausibility of an incipient event in one scoring mechanism.

2.2. Temporal Point Processes and Graph Learning

Temporal point processes provide a natural language for alarm sequences because they model the probability of a future event conditional on past events and elapsed time. Classical Hawkes processes are especially relevant to safety monitoring because self-excitation expresses the tendency of one event to increase the chance of subsequent events within a short interval [15]. Neural variants replace fixed kernels with recurrent or attention-based encoders, making it possible to learn nonlinear excitation, inhibition, and marked event interactions from data [16]. In industrial monitoring, this representation is useful for distinguishing a single isolated alarm from a developing cascade. However, the event stream alone cannot determine whether the cascade follows the topology of the plant, nor can it determine whether a predicted escalation violates heat-balance evidence. Neural Hawkes models therefore need contextual information from continuous variables and physical constraints when they are applied to reactors.

Graph neural networks have been used to represent dependencies among sensors, units, and control loops in complex processes. Message passing allows a deviation at one location to influence the representation of connected variables, which is important when reactor temperature, jacket temperature, flow rate, agitation, and pressure do not move independently [17]. Dynamic graphs further allow edge weights to vary as the process changes from steady operation to feed addition or cooling saturation [18]. Yet a graph model trained only on classification loss may still discover correlations that are statistically useful but operationally fragile. Combining graph learning with a Neural Hawkes decoder allows the model to exploit both spatial coupling and event timing, while a physics residual can guide the representation toward patterns consistent with reaction heat release and heat removal. This combination forms the methodological basis of the proposed model.

3. Model Architecture and Physical Constraint Learning

The proposed model uses three synchronized representations: continuous process states, discrete safety events, and a directed process graph. Nodes represent measured variables or local subsystems, while candidate edges are first defined from the piping and instrumentation structure. Material-transfer edges follow flow direction; heat-transfer edges connect reactor and jacket variables in both directions; control edges run from manipulated variables to their affected measurements. Engineering review removes implausible links before training, and attention weights then adapt the strength, but not the existence, of each retained pathway. Continuous inputs include normalized reactor and jacket temperatures, pressure, agitation power, feed and coolant flow, valve positions, and engineered derivatives. Event marks include alarms, valve interventions, abnormal rates of change, and warning-band entries. The graph is an engineering abstraction rather than a first-principles simulator, allowing the model to express dominant propagation routes while leaving detailed kinetics to the heat-balance constraint.

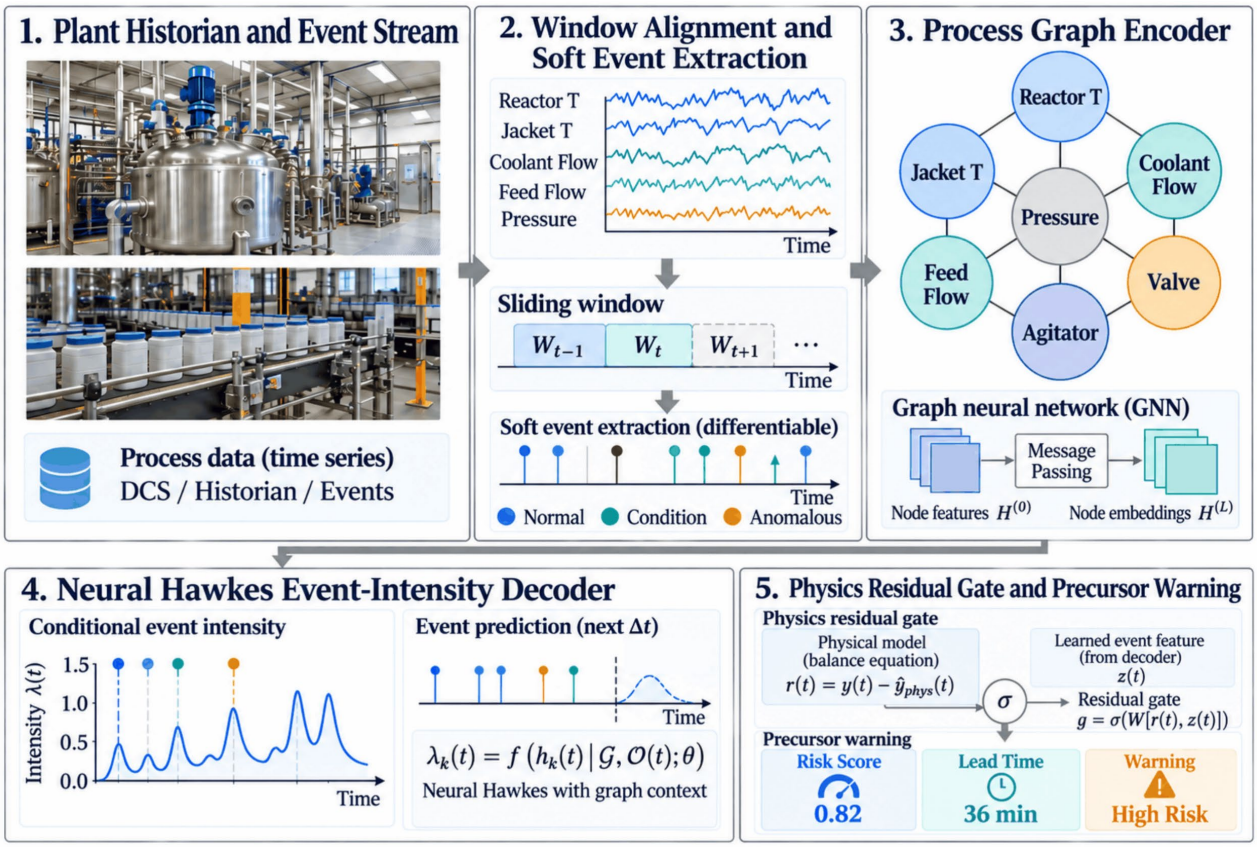


Figure 1. Physics-constrained Neural Hawkes GNN architecture for reactor runaway precursor recognition

Figure 1 shows the architecture used in this work. The left block receives asynchronous historian samples and event marks, aligns them into observation windows, and keeps event times in continuous form. The middle block applies graph message passing to update node states from neighboring variables, while the Neural Hawkes block converts the updated representation and elapsed time into a marked conditional intensity. The right block applies the physics residual gate and returns a precursor probability with a lead-time estimate. The flow is designed so that event clustering can raise the learned intensity, but the gate suppresses intensity growth when the heat-balance residual indicates that the apparent disturbance is not thermally escalating.

Graph message passing updates each node by combining its own state with the states of neighboring process variables:

$$h_i^{(l+1)} = \sigma\left(W_s h_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \alpha_{ij} W_n h_j^{(l)} + b \right) \tag{1}$$

Here, $h_i^{(l)}$ denotes the representation of node i at graph layer $l, \mathcal{N}(i)$ is its neighborhood, and α_{ij} is the learned attention coefficient between nodes i and j. The attention coefficient is derived from node features and edge attributes, allowing reactor temperature to receive different levels of influence from jacket outlet temperature, coolant flow, and feed flow as the operating mode changes.

The graph-level process representation and event history are subsequently passed to the marked Neural Hawkes decoder. The conditional intensity of event type k is equation as

$$\lambda_k(t) = softplus\left(w_k^\top z(t) + u_k^\top e_m + \gamma_k \Delta t \right) \tag{2}$$

where $z(t)$ is the graph-conditioned temporal representation, e_m is the embedding of the most recent event mark, and Δt is the elapsed time since the preceding event. The elapsed-time term preserves continuous event timing, so alarm and intervention events do not need to be resampled onto a uniform temporal grid. The softplus transformation guarantees a nonnegative conditional intensity while retaining smooth gradients during training.

A simplified heat-balance residual is introduced to determine whether the observed event escalation is supported by the reactor thermal state:

$$r_Q(t) = \frac{dT_r}{dt} - \frac{Q_{gen}(t) - Q_{rem}(t)}{\rho C_p V} \tag{3}$$

In this equation, T_r is the reactor temperature, Q_{gen} and Q_{rem} denote estimated heat generation and heat removal, respectively, ρ is mixture density, C_p is specific heat capacity, and V is the effective reactor volume. A positive residual indicates unaccounted heat accumulation or an insufficient cooling response. Because the residual is evaluated together with event intensity and graph features, temporary measurement fluctuations do not automatically produce a high-risk state.

4. Precursor Scoring and Training Protocol

The training protocol uses overlapping windows with a prediction horizon referenced to the first precursor time confirmed from simulated heat-generation and heat-removal trajectories. A window is positive only when it ends before intervention and lies within the warning horizon; samples after intervention, trip activation, or confirmed runaway are excluded from predictor construction. Negative windows come from normal campaigns or recoverable disturbances. All windows from one campaign remain in a single partition, preventing nearly adjacent fragments of the same trajectory from entering both training and test sets. Loss weights, residual tolerance, persistence duration, and score thresholds are selected on validation campaigns only. The positive boundary therefore represents information available before protective action, and the test campaigns remain untouched until the final performance calculation.

Figure 2 presents the deployment workflow. Incoming historian samples are filtered for missing values and unit consistency, soft events are extracted from gradients and warning bands, and the graph encoder updates the current node states. The Neural Hawkes decoder then predicts event intensity over the next horizon, while the residual module estimates whether heat accumulation is physically supported. A warning is issued only when the calibrated risk score remains above the selected threshold for two consecutive evaluation intervals. This persistence rule reduces alarms produced by isolated spikes and preserves warnings that reflect sustained escalation.

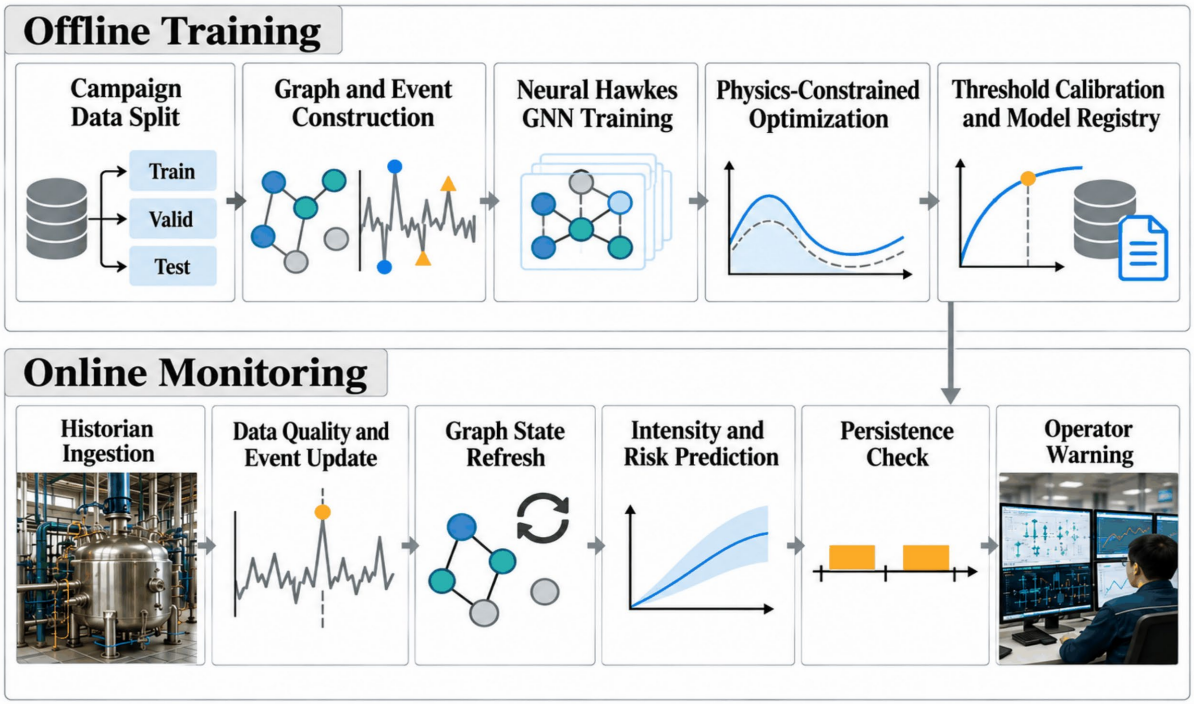


Figure 2. Training and online deployment workflow for precursor scoring and persistent warning generation

The model is trained using an objective that combines marked-event likelihood, precursor classification, physical consistency, and parameter regularization:

$$\mathcal{L} = \mathcal{L}_{NHP} + \beta\mathcal{L}_{cls} + \eta\sum_{t}\left[max\left(0, r_Q(t) - \tau_Q\right)\right]^2 + \xi\|\Theta\|_2^2 \tag{4}$$

Here, $\mathcal{L}_{\mathrm{NHP}}$ is the negative log-likelihood of the Neural Hawkes process, and $\mathcal{L}_{\mathrm{cls}}$ is the precursor classification loss. The coefficients β, η, and ξ control the contributions of classification, physical residual regularization, and parameter regularization. The residual penalty becomes active only when the normalized heat residual exceeds the tolerated threshold τ_Q. Small residuals associated with sensor noise or approximation errors are therefore not penalized as runaway evidence.

During online evaluation, the precursor score fuses the predicted event intensity, heat-balance residual, and graphlevel risk representation:

In this equation, $\hat{\lambda}(t:t+H)$ is the aggregated event intensity predicted over horizon $H, R_Q(t)$ is the normalized physical residual, and $G(t)$ is the graph-level risk representation. The coefficients a, b, and c are learned during training. This equation keeps the score interpretable because each component corresponds to a recognizable process-safety quantity: temporal escalation, thermal inconsistency, or propagation through coupled process variables.

The final warning decision applies a persistence condition:

$$\hat{y}(t) = \begin{cases} 1, & S(t) \geq \theta \text{ for two consecutive evaluation intervals,} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The threshold θ is selected using validation campaigns rather than test data. A short-lived score spike therefore does not create an alarm, whereas a sustained increase generates a precursor warning before the high-high protection layer is reached. This decision rule separates model inference from alarm policy and allows the persistence interval to be adjusted according to reactor thermal time constants and the minimum time required for operator intervention.

5. Experimental Data and Analysis

5.1. Dataset Design and Evaluation Settings

The benchmark contains 12,480 windows from 96 simulated and historian-like campaigns, including normal production, feed disturbances, cooling degradation, sensor bias, agitation loss, and incipient runaway cases. Campaigns are generated by varying initial charge, feed-ramp duration, heat-transfer coefficient, coolant availability, agitation efficiency, sensor offset, and disturbance onset within engineering bounds; correlated noise and irregular sampling are then applied to resemble historian records. Each 90 min window contains observation intervals from 5 s to 90 s, 38

measured or engineered graph variables, and 14 event marks. A recent alarm-sequence benchmark suggests preserving class imbalance when the target is safety-critical warning [19]. Accordingly, the test set retains a 3.4% positive rate, and validation alone determines thresholds. Because these controlled scenarios do not reproduce proprietary recipes, fouling, maintenance practice, or utility-network interactions, the benchmark supports comparative method evaluation but does not constitute plant certification.

The data generation protocol was intentionally designed to separate early precursor recognition from late fault confirmation. Normal campaigns contain recipe transitions, feed-rate ramps, jacket set-point changes, and short utility fluctuations that should not trigger a runaway warning. Disturbance campaigns include recoverable jacket saturation, temporary feed overshoot, pressure-sensor drift, and scheduled agitation changes. Incipient runaway campaigns begin with a thermally plausible imbalance rather than an abrupt temperature jump, so the model must learn that a weak but persistent mismatch between heat generation and cooling response is more important than a single extreme observation. To avoid leakage, all windows from the same campaign are assigned to only one split. The training set contains 62 campaigns, the validation set contains 14 campaigns, and the test set contains 20 campaigns. This split creates a more stringent assessment than random window sampling because the model must generalize across complete operating histories.

The proposed model is compared with five baselines: logistic regression using engineered thermal indicators, LSTM, temporal convolutional network, Neural Hawkes without graph context, and graph neural network without a point-process decoder. All models use the same training, validation, and test campaigns. Performance is reported by AUROC, AUPRC, F1 score, recall at 5% false alarm rate, median warning lead time, and nuisance alarms per campaign. Table 1 lists the main comparison. The physics-constrained Neural Hawkes GNN achieved AUROC 0.973, AUPRC 0.846, and F1 0.918, while the best non-physics baseline reached AUROC 0.948 and F1 0.872. The median warning lead time increased from 15.2 min for the graph-only baseline to 18.6 min for the proposed model, and nuisance alarms decreased from 2.17 to 1.41 per campaign. The improvement is consistent with studies showing that process-safety models benefit when temporal alarms are evaluated together with operating context rather than in isolation [20].

Table 1. Overall precursor recognition performance on the reactor benchmark

Model	AUROC	AUPRC	F1	Recall at 5% FAR	Median Lead Time min	Nuisance Alarms campaign
Logistic thermal indicators	0.861	0.542	0.731	0.704	9.8	2.81
LSTM	0.914	0.713	0.824	0.831	12.7	2.56
Temporal convolution	0.928	0.746	0.845	0.857	13.8	2.32
Neural Hawkes only	0.932	0.763	0.851	0.866	16.1	2.04
Graph neural network only	0.948	0.801	0.872	0.889	15.2	2.17
Proposed physics-constrained Neural Hawkes GNN	0.973	0.846	0.918	0.920	18.6	1.41

Several implementation choices were held constant across models. Each neural baseline used the same normalized inputs, event-mark vocabulary, positive-window definition, campaign partitions, and validation metric. Hyperparameters were selected by validation AUPRC under an equal search budget, and each configuration was repeated with five fixed random seeds. Adam optimization, early stopping after 20 validation epochs without improvement, and a maximum of 180 epochs were applied consistently. The proposed graph encoder used two message-passing layers, hidden dimension 96, and edge dropout 0.12; the Hawkes decoder used mark-specific intensity heads. LSTM, temporal convolutional, Hawkes-only, and graph-only baselines received comparable hidden capacity and independent validation thresholds. This protocol attributes the comparison to event timing, graph coupling, and residual-guided learning rather than unequal tuning effort.

5.2. Recognition Accuracy and Ablation Behavior

Figure 3 presents discrimination, threshold behavior, warning time, and alarm burden as complementary views of recognition quality. Figure 3(a) shows that the proposed model obtains AUROC 0.973 and AUPRC 0.846, improving on the Hawkes-only baseline by 0.041 and 0.083. The larger gain in AUPRC is especially relevant to the 3.4% positive rate because it reflects better ranking within the rare precursor class rather than performance dominated by normal windows. Figure 3(b) separately traces the precision-recall operating curve: precision remains above 0.82 at recall 0.91, while LSTM precision is below 0.75 at the same recall. This separation near the intended operating point indicates that the graph and residual terms reject a meaningful fraction of thermally implausible alarm clusters instead of merely shifting all scores upward. Industrial alarm studies likewise recommend interpreting sensitivity together with the warning burden imposed on operators [21].

The timing and burden panels reveal whether the discrimination advantage translates into usable intervention time. Figure 3(c) gives a median true-positive lead time of 18.6 min, with an interquartile range of 14.1-23.4 min; the temporal convolutional baseline has a median of 13.8 min. The upper quartile therefore extends beyond 23 min without sacrificing the central tendency, whereas the lower quartile remains above 14 min, leaving time for feed interruption or cooling verification in most detected campaigns. Figure 3(d) evaluates nuisance alarms at matched recall and records 1.41 alarms per campaign for the proposed model, compared with 2.04 for Hawkes-only and 2.56 for LSTM. Because recall is held constant, this reduction cannot be attributed to a more conservative threshold. It instead reflects improved separation of sustained thermal escalation from recoverable process disturbances.

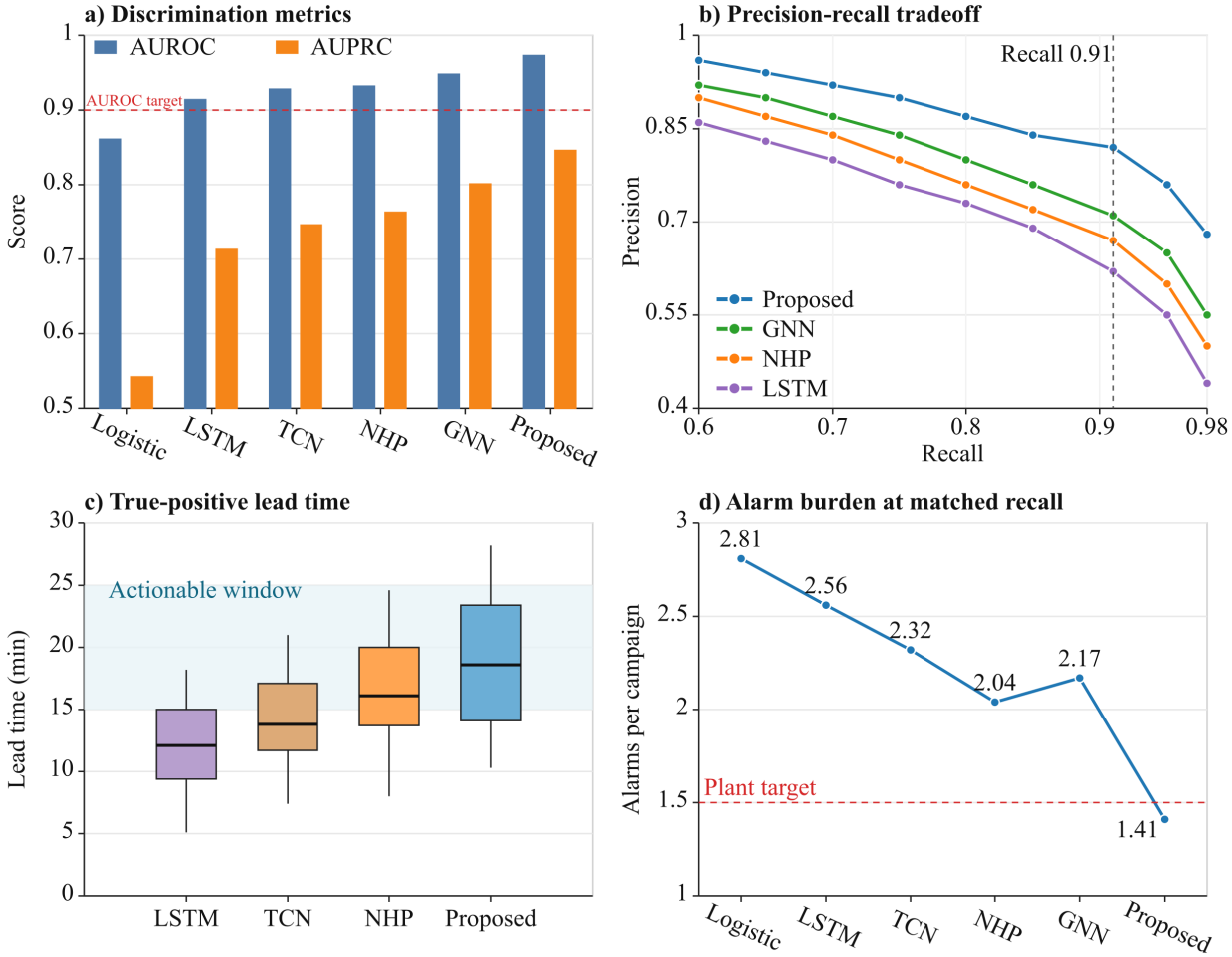


Figure 3. Recognition performance and alarm burden: (a) AUROC and AUPRC comparison, (b) precision-recall curves, (c) warning lead-time distributions, (d) nuisance alarms at matched recall

Ablation results expose the different functions of the three learning components. Figure 4(a) shows that removing the physics residual lowers F1 from 0.918 to 0.879, removing graph propagation lowers it to 0.872, and removing the Hawkes decoder lowers it to 0.861. The decoder ablation produces the largest aggregate loss because clustered event timing is central to early recognition, while the residual ablation raises missed precursors from 33 to 48 specifically under cooling degradation. Figure 4(b) resolves this behavior by scenario: the full model reaches recalls of 0.936 for cooling failure, 0.914 for feed overdose, 0.901 for agitation loss, and 0.889 for sensor bias. The smaller margin for sensor bias is consistent with corruption of the measurements used by both the graph encoder and the heat-balance residual. Calibration is also an engineering requirement in rare-event monitoring because poorly calibrated scores distort alarm priority [22]. Figure 4(c) shows that expected calibration error rises only from 0.026 to 0.041 across a 4% temperature-drift stress, remaining below the stated 0.05 limit, whereas the ablated alternatives exceed that limit earlier.

Figure 4(d) examines the residual penalty rather than treating its selected value as arbitrary. F1 remains between 0.914 and 0.918 for weights from 0.15 to 0.35, defining a broad plateau around the chosen value of 0.25; it falls to 0.861 at 0.70 because physical consistency then overwhelms event evidence in sensor-delayed cases. At the deployed score threshold of 0.62, 379 of 412 precursor windows are detected and 33 are missed, while 58 nuisance alarms occur among 4,846 negative windows. Raising the threshold to 0.70 reduces nuisance alarms from 1.41 to 1.03 per campaign but also lowers recall from 0.920 to 0.884. Of the 33 misses, 21 reach a score above 0.55 but fail the two-interval persistence rule before the confirmed label; 37 of the 58 false positives persist for only two or three evaluation intervals. These

distributions locate the remaining errors near the temporal decision policy and show why threshold and persistence settings must be matched to available intervention time.

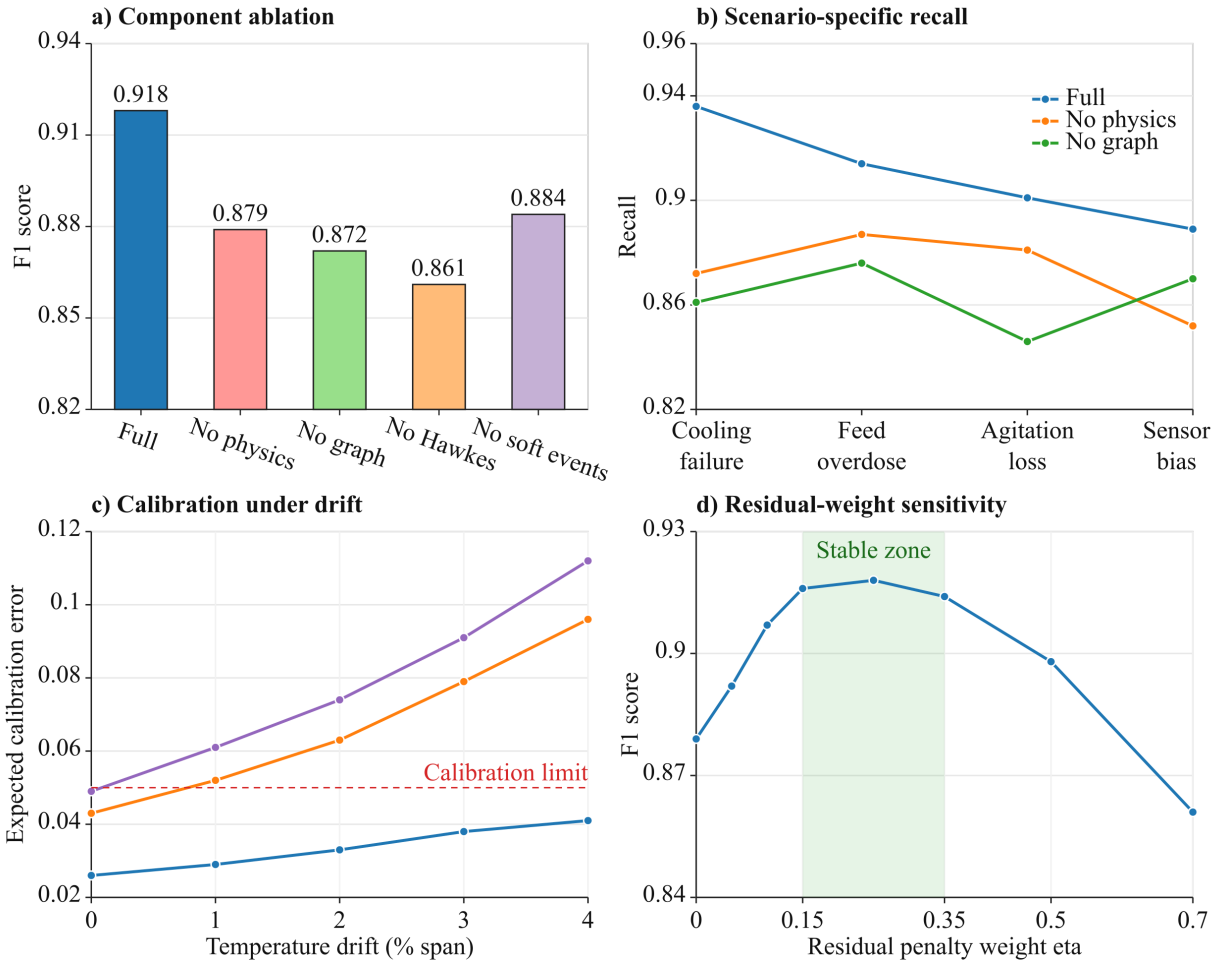


Figure 4. Ablation and calibration analysis: (a) F1 score by component removal, (b) scenario-specific recall, (c) calibration error under drift, (d) residual-weight sensitivity

5.3. Robustness and Operational Interpretation

Figure 5 evaluates robustness and interpretability under conditions expected in historian deployment. Figure 5(a) shows that AUROC decreases from 0.973 to 0.955 as Gaussian sensor noise rises to 4% of span, whereas Hawkes-only decreases from 0.932 to 0.889 and LSTM falls more sharply. Figure 5(b) tests missing samples and shows that the proposed model retains F1 0.884 at 25% missingness, remaining above the 0.85 deployment reference; masking prevents absent values from being treated as measurements, while continuous event times preserve the alarm sequence. Missingness and utility-state changes are normal characteristics of process data rather than exceptional test conditions [23]. Figure 5(c) follows one cooling-failure campaign: reactor temperature rises from 92.1 to 104.8 degrees C over 22 min, the normalized heat residual crosses 0.18 at minute 16, and the warning score crosses 0.62 at minute 18. The ordered crossings show that physical inconsistency appears before the final warning and that the score responds to sustained evidence rather than the first temperature fluctuation.

The attribution panels distinguish a developing runaway from a benign process response. Figure 5(d) shows attention migrating toward reactor temperature, jacket-outlet temperature, coolant flow, feed flow, and valve position during the final 12 min; pressure becomes influential later, after thermal and cooling variables already carry strong weight. The pattern is plausible for cooling degradation, but attention is treated as diagnostic evidence rather than causal proof. Earlier-warning systems are more actionable when they expose intelligible cause patterns alongside the alarm score [24]. Figure 5(e) provides the counterexample of a benign feed perturbation: the warning score peaks at 0.44, remains below the 0.62 warning threshold, and returns below 0.30 after jacket recovery, while the residual peaks near 0.14 and remains below 0.18. Figure 5(f) decomposes precursor intensity into 43% soft thermal-gradient events, 27% coolant-response events, 18% pressure events, 8% valve actions, and 4% feed marks. The dominance of thermal and coolant evidence explains why the model resists isolated feed marks yet responds when feed activity is followed by deteriorating heat removal.

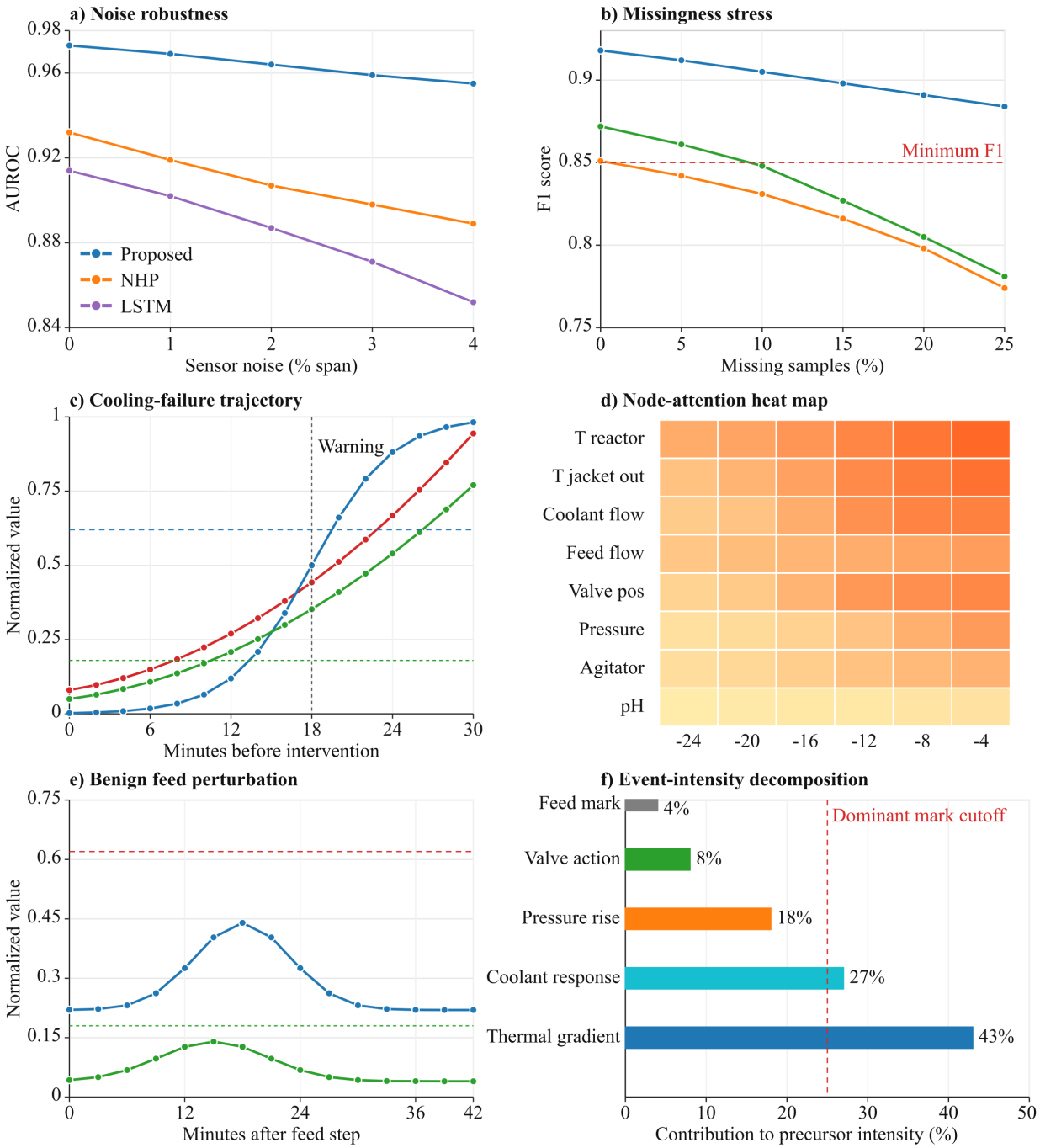


Figure 5. Robustness and interpretability under deployment stress: (a) sensor-noise AUROC, (b) missing-sample F1, (c) risk trajectory in cooling failure, (d) node-attention heat map, (e) benign feed perturbation score, (f) event-intensity decomposition

Figure 6 plots the joint decision surface, rather than simplifying the output into an opaque probability. The figure shows that 92.6% of the normal windows are located in the low-risk area below a residual of 0.08 and an intensity of 0.15 min^{-1}. Between residuals 0.08-0.18 and intensities 0.15-0.32 min^{-1}, benign feed disturbances overlap with early precursors, so the learned boundary bends with both quantities instead of imposing a rectangular rule. Above residual 0.18 and intensity 0.32 min^{-1}, 87.4% of confirmed precursor windows occupy the high-risk region. Joint physical and temporal indicators are easier to examine against plant knowledge than an unconstrained latent score [25]. Sensitivity tests support the map's stability: removing 30% of weak graph edges changes AUROC from 0.973 to 0.968, whereas removing soft-gradient marks shortens median lead time by 3.9 min. Extending the prediction horizon from 35 to 45 min raises recall to 0.934 but increases nuisance alarms to 2.12 per campaign; safety-prediction research has similarly noted that longer horizons can weaken alarm credibility despite attractive aggregate recall [26].

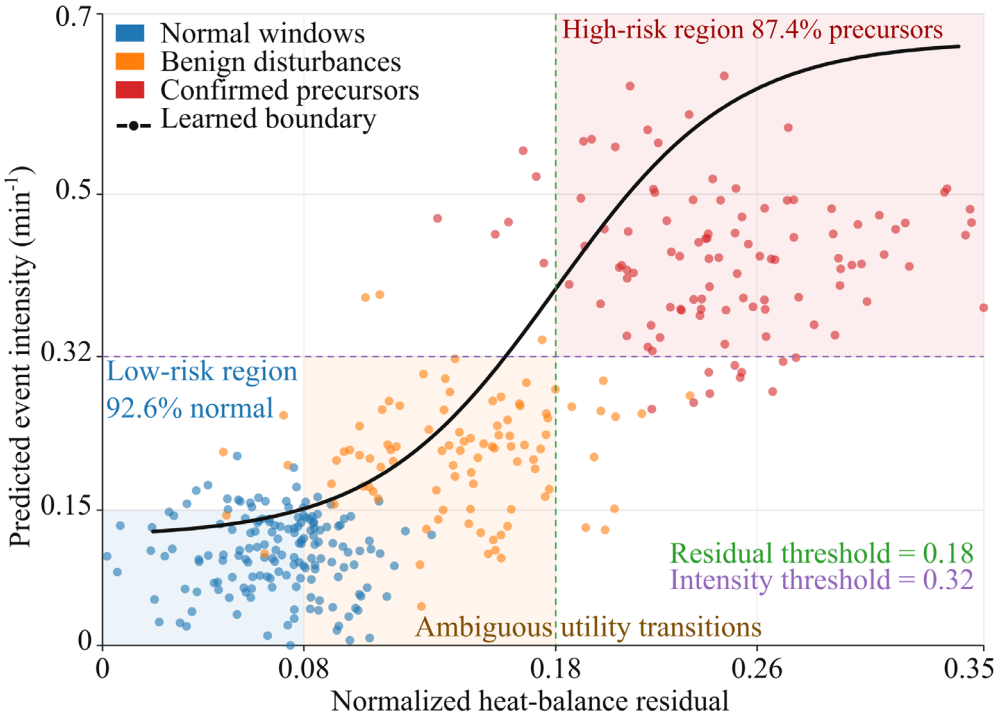


Figure 6. Integrated operating map connecting normalized heat residual, event intensity, and precursor warning regions

Online timing was measured on an NVIDIA RTX 3060 GPU with an Intel Core i7-12700 processor using Python 3.11 and PyTorch 2.2. Model inference for one 90 min window required 18.3 ms without batching and 6.7 ms per window when parallel reactor units were batched; event extraction, unit checks, and mask updates added 11.9 ms separately. Peak memory remained below 820 MB for 38 nodes and 14 event marks. These values characterize the present implementation rather than a guaranteed plant latency. Deployment would additionally require historian synchronization, cybersecurity controls, instrument-quality checks, management-of-change review, and prospective calibration monitoring [27]. Plant-specific validation should relate threshold persistence to thermal time constants, emergency-cooling capacity, and minimum operator intervention time [28]. The model is therefore positioned as a supervisory decision-support layer that prompts verification of coolant, feed, agitation, and sensor plausibility while existing protection systems retain authority.

6. Conclusion

This paper presented a physics-constrained Neural Hawkes graph neural network for recognizing runaway precursors in exothermic reactor operation. The framework combines continuous-time marked event modeling, graph propagation across coupled process variables, and a normalized heat-balance residual that influences both representation learning and online risk gating. By linking alarm clustering to process topology and thermal consistency, the method treats precursor recognition as a physically interpretable early-warning problem rather than ordinary window classification. Comparative, ablation, calibration, robustness, and latency analyses establish the feasibility of the integrated method as a supervisory monitoring layer.

Several limitations remain. The synthetic and historian-like benchmark cannot reproduce the full variability of commercial reactors, proprietary recipes, catalyst aging, fouling, operator actions, and utility-network interactions. Injected noise and missingness are controlled stress tests rather than substitutes for naturally occurring instrument failures. The compact heat-balance residual is practical online but less expressive than a high-fidelity reactor model, while graph edges depend partly on engineering judgment and may be misspecified. Score calibration can also drift when kinetics, production grade, or sensor maintenance changes. These factors require prospective, plant-specific evaluation before any warning is linked to an operating procedure.

Future research should validate the framework on larger multi-site historian datasets and on pilot-scale experiments with controlled cooling degradation, feed disturbances, and quench interventions. More detailed thermodynamic constraints could be incorporated through differentiable reactor models, uncertainty-aware residuals, or adaptive kinetic priors. Another promising direction is human-in-the-loop alarm design, where model explanations, threshold persistence, and recommended operator actions are evaluated together before the system is deployed in safety-critical production.

7. Declarations

7.1 Author Contributions

Nenad Miljković contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Pavle Stošić contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Data Availability

Not applicable.

7.6 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.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized ChatGPT-4o for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Li, S., Yang, S.-H., Cao, Y., Jiang, X., & Zhou, C. (2025). A benchmark of industrial polymerization process for thermal runaway process monitoring. Process Safety and Environmental Protection, 193, 353-363. https://doi.org/10.1016/j.psep.2024.11.057

[2] Ab Rahim, M. S., Reniers, G., Yang, M., & Bajpai, S. (2024). Risk assessment methods for process safety, process security and resilience in the chemical process industry: A thorough literature review. Journal of Loss Prevention in the Process Industries, 88, 105274. https://doi.org/10.1016/j.jlp.2024.105274

[3] Schmidt, C., Schmidt, J., Denecke, J., & Seewig, J. (2024). Advanced overpressure protection: A concept for smart controlled safety pressure relief systems. Process Safety and Environmental Protection, 190, 506-516. https://doi.org/10.1016/j.psep.2024.08.061

[4] Ali, H., Zhang, Z., Safdar, R., Rasool, M. H., Yao, Y., Yao, L., & Gao, F. (2024). Fault detection using machine learning based dynamic ICA-distributed CCA: Application to industrial chemical process. Digital Chemical Engineering, 11, 100156. https://doi.org/10.1016/j.dche.2024.100156

[5] Shi, Y., Zhang, N., Song, X., Li, H., & Zhu, Q. (2024). Novel approach for industrial process anomaly detection based on process mining. Journal of Process Control, 136, 103165. https://doi.org/10.1016/j.jprocont.2024.103165

[6] Melo, A., Câmara, M. M., & Pinto, J. C. (2024). Data-Driven Process Monitoring and Fault Diagnosis: A Comprehensive Survey. Processes, 12(2), 251. https://doi.org/10.3390/pr12020251

[7] Rao, S., & Wang, J. (2024). A comprehensive fault detection and diagnosis method for chemical processes. Chemical Engineering Science, 300, 120565. https://doi.org/10.1016/j.ces.2024.120565

[8] Zhou, K., Wang, R., Tong, Y., Wei, X., Song, K., & Chen, X. (2024). Domain generalization of chemical process fault diagnosis by maximizing domain feature distribution alignment. Process Safety and Environmental Protection, 185, 817-830. https://doi.org/10.1016/j.psep.2024.03.068

[9] Tao, Q., Xin, B., Zhang, Y., Jin, H., Li, Q., Dai, Z., & Dai, Y. (2024). A novel triage-based fault diagnosis method for chemical

process. Process Safety and Environmental Protection, 183, 1102-1116. https://doi.org/10.1016/j.psep.2024.01.072

[10] Jia, M., Xu, D., Yang, T., Yao, Y., & Liu, Y. (2024). Graph-guided masked autoencoder for process anomaly detection. Process Safety and Environmental Protection, 186, 1345-1357. https://doi.org/10.1016/j.psep.2024.04.052

[11] Tang, C., Huang, J., Xu, M., Liu, X., Yang, F., Feng, W., He, Z., & Lv, J. (2024). Attention-based early warning framework for abnormal operating conditions in fluid catalytic cracking units. Applied Soft Computing, 153, 111275. https://doi.org/10.1016/j.asoc.2024.111275

[12] Allen, L., Lu, H., & Cordiner, J. (2024). Knowledge-Enhanced Spatiotemporal Analysis for Anomaly Detection in Process Manufacturing. Computers in Industry, 161, 104111. https://doi.org/10.1016/j.compind.2024.104111

[13] Luo, J., Jin, Z., Jin, H., Li, Q., Ji, X., & Dai, Y. (2024). Causal temporal graph attention network for fault diagnosis of chemical processes. Chinese Journal of Chemical Engineering, 70, 20-32. https://doi.org/10.1016/j.cjche.2024.01.019

[14] Zhang, F., Jin, Q., Li, D., Zhang, Y., & Zhu, Q. (2024). Physical Graph-Based Spatiotemporal Fusion Approach for Process Fault Diagnosis. ACS Omega, 9(8), 9486-9502. https://doi.org/10.1021/acsomega.3c09122

[15] Liu, R., Zhang, Q., Lin, D., Zhang, W., & Ding, S. X. (2024). Causal intervention graph neural network for fault diagnosis of complex industrial processes. Reliability Engineering & System Safety, 251, 110328. https://doi.org/10.1016/j.ress.2024.110328

[16] Liu, R., Xie, Y., Lin, D., Zhang, W., & Ding, S. X. (2024). Information-based Gradient enhanced Causal Learning Graph Neural Network for fault diagnosis of complex industrial processes. Reliability Engineering & System Safety, 252, 110468. https://doi.org/10.1016/j.ress.2024.110468

[17] Wu, Y., Sicard, B., & Gadsden, S. A. (2024). Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. Expert Systems with Applications, 255, 124678. https://doi.org/10.1016/j.eswa.2024.124678

[18] Zou, Z., Meng, X., & Karniadakis, G. E. (2024). Correcting model misspecification in physics-informed neural networks (PINNs). Journal of Computational Physics, 505, 112918. https://doi.org/10.1016/j.jcp.2024.112918

[19] Das, L., Gjorgiev, B., & Sansavini, G. (2024). Uncertainty-aware deep learning for monitoring and fault diagnosis from synthetic data. Reliability Engineering & System Safety, 251, 110386. https://doi.org/10.1016/j.ress.2024.110386

[20] Joseph, S., & Jain, S. (2024). A neural network based model for multi-dimensional non-linear Hawkes processes. Journal of Computational and Applied Mathematics, 447, 115889. https://doi.org/10.1016/j.cam.2024.115889

[21] Xie, Y., Wu, J., & Zhou, Y. (2024). GTHP: a novel graph transformer Hawkes process for spatiotemporal event prediction. Knowledge and Information Systems, 66(7), 4043-4062. https://doi.org/10.1007/s10115-024-02080-z

[22] Pietrasik, M., Wilbik, A., & Grefen, P. (2024). The enabling technologies for digitalization in the chemical process industry. Digital Chemical Engineering, 12, 100161. https://doi.org/10.1016/j.dche.2024.100161

[23] Pal, P. K., Hens, A., Behera, N., & Lahiri, S. K. (2025). Digital twins: Transforming the chemical process industry—A review. The Canadian Journal of Chemical Engineering, 103(8), 3611-3636. https://doi.org/10.1002/cjce.25611

[24] Li, X.-Y., Xu, Y., Zhu, Q.-X., & He, Y.-L. (2025). Industrial Data Imputation Based on Multiscale Spatiotemporal Information Embedding With Asymmetrical Transformer. IEEE Transactions on Neural Networks and Learning Systems, 36(8), 14937-14948. https://doi.org/10.1109/tnnls.2025.3527581

[25] Liu, D., Wang, Y., Liu, C., Yuan, X., Wang, K., & Yang, C. (2024). Scope-Free Global Multi-Condition-Aware Industrial Missing Data Imputation Framework via Diffusion Transformer. IEEE Transactions on Knowledge and Data Engineering, 36(11), 6977-6988. https://doi.org/10.1109/tkde.2024.3392897

[26] Jia, M., Yao, Y., & Liu, Y. (2025). Review on Graph Neural Networks for Process Soft Sensor Development, Fault Diagnosis, and Process Monitoring. Industrial & Engineering Chemistry Research, 64(17), 8543-8564. https://doi.org/10.1021/acs.iecr.5c00283

[27] Xu, Y., Zhang, C., Luo, Y., Ke, W., Zhu, Q.-X., He, Y.-L., Zhang, Y., & Zhang, M.-Q. (2025). Multiattention Spatiotemporal Fusion Graph Neural Network for Chemical Process Fault Diagnosis. IEEE Transactions on Instrumentation and Measurement, 74, 1-13. https://doi.org/10.1109/tim.2025.3558173

[28] Lv, F., Yang, B., Yu, S., Zou, S., Wang, X., Zhao, J., & Wen, C. (2025). A unified model integrating Granger causality-based causal discovery and fault diagnosis in chemical processes. Computers & Chemical Engineering, 196, 109028. https://doi.org/10.1016/j.compchemeng.2025.109028