International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 10 Issue 2, Mar-Apr 2026 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Multi-Source Data-Driven 15-Minute-Level Congestion Risk Prediction for Hazardous Chemical Vehicle Operations in Industrial Parks

Nai Jin Liu

Undergraduate Student, School of Artificial Intelligence, Beijing Wuzi University, Beijing, China

ABSTRACT

This paper presents a high-reliability congestion risk prediction framework tailored for hazardous chemical industrial parks. Unlike conventional traffic prediction studies, the proposed method integrates dynamic service capacity modeling, reservation-system control variables, and operational channel constraints into a unified binary classification formulation. Mathematical derivations, statistical validation, ablation analysis, and computational complexity analysis are systematically conducted. Experimental evaluation demonstrates strong predictive stability (AUC=0.962) and statistically significant improvement over baseline methods ($p<0.01$). The framework provides actionable decision support for safety-critical logistics systems.

How to cite this paper: Nai Jin Liu "Multi-Source Data-Driven 15-Minute-Level Congestion Risk Prediction for Hazardous Chemical Vehicle Operations in Industrial Parks" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-10 | Issue-2, April 2026, pp.694-701, URL: www.ijtsrd.com/papers/ijtsrd106988.pdf

IJTSRD106988

I. INTRODUCTION

I. INTRODUCTION (Extended Motivation and Contributions)

Hazardous chemical industrial parks differ from ordinary logistics hubs in that traffic operations are tightly coupled with process safety and emergency management. Vehicles often carry flammable, explosive, or toxic materials, and congestion near storage tanks or loading areas can increase both the likelihood and the consequences of an incident. For example, prolonged vehicle dwell time may elevate exposure to heat sources, increase the probability of human error during maneuvering, and reduce the effective width of evacuation routes. Congestion can also delay first responders and impede the movement of firefighting equipment, making the same traffic disruption more dangerous than in conventional contexts.

Another defining feature is that demand is partially controllable. Many parks operate appointment or reservation systems that coordinate vehicle arrivals for loading, unloading, and inspection. While reservations improve planning, they can also create synchronized arrival waves when multiple carriers select similar time windows, producing short but intense surges. These surges are often predictable from reservation records, but only if the prediction model explicitly incorporates reservation intensity and its interaction with current capacity. Similarly, service capacity is not constant: inspection channels can be opened or closed based on staffing, safety checks, or equipment conditions. A model that assumes fixed capacity may therefore miss important operational drivers.

These characteristics motivate a risk-centric perspective: rather than predicting raw flow or speed, we aim to predict whether the system is likely to exceed effective capacity in the immediate future. This shift in modeling objective is especially valuable for decision support because it maps directly to operational actions (e.g., increase channels, delay arrivals, reroute vehicles) and safety policies (e.g., maintain buffer capacity for emergency access).

In summary, the key contributions of this work are: (1) a capacity-exceedance label grounded in dynamic service modeling for safety-critical congestion warning; (2) a multi-source feature system that integrates reservations, operational channels, and traffic dynamics at 15-minute resolution; (3) an ensemble learning architecture with statistical validation, ablation, calibration, and error diagnostics; and (4) deployment-oriented guidance for real-time integration, monitoring, and human-in-the-loop safety governance.

Hazardous chemical industrial parks represent high-risk socio-technical systems where traffic congestion significantly amplifies systemic vulnerability. Previous studies have demonstrated that short-term traffic instability may increase accident probability and operational disruption in hazardous logistics environments [1], [2]. In such safety-critical systems, congestion not only degrades efficiency but also elevates explosion exposure risk, toxic diffusion probability, and emergency response obstruction.

This study contributes the following advancements to existing literature: (1) A mathematically grounded dynamic capacity-exceedance risk formulation extending prior traffic risk modeling frameworks [3], [9]; (2) A multi-source feature fusion strategy integrating reservation-based operational control mechanisms, consistent with intelligent industrial management paradigms [6], [16]; (3) An ensemble learning–based predictive architecture validated under safety-oriented classification metrics [4], [19]; and (4) A deployment-oriented analytical framework including statistical significance testing and computational complexity analysis.

II. RELATED WORK

Hazardous chemical logistics in industrial parks sits at the intersection of short-term traffic prediction, queueing/capacity modeling, and safety risk analytics. Most short-horizon traffic prediction studies focus on urban arterials or highway corridors, using either time-series models (ARIMA, state-space variants) or data-driven learners (random forests, gradient boosting, recurrent neural networks). While these methods can yield accurate flow or speed forecasts, they typically do not transform predictions into decision-ready safety indicators. In contrast, operational managers in hazardous parks often require a binary warning—whether the next 15-minute horizon is likely to exceed controllable capacity—so that they can trigger access throttling, channel reallocation, or reservation rescheduling.

From a risk perspective, earlier hazardous material transportation research has emphasized route selection, accident probability, consequence severity (e.g., exposure zones), and multi-criteria decision-making. However, congestion risk within a confined industrial park is distinct: (i) traffic demand is partially controllable via reservation systems; (ii) service capacity is not fixed but depends on staffed inspection channels and their real-time service rates; and (iii) a congestion episode can propagate quickly to block emergency corridors and increase dwell time near sensitive storage areas. These features motivate a unified learning problem where operational control signals and capacity dynamics are first-class inputs.

Queueing-inspired approaches have been used to characterize service facilities such as ports, terminals, and inspection stations. Yet, classical analytical queueing models require strong assumptions (stationarity, distributional forms) that are not always satisfied under fluctuating demand and evolving operational policies. Recent work therefore combines queueing insights with supervised learning: the queueing model defines a physically meaningful risk label (e.g., capacity exceedance), and learning provides robustness to nonstationarity and complex nonlinear interactions among inputs. The present study follows this philosophy by constructing a binary label based on whether the observed queue length (or equivalent congestion indicator) surpasses the instantaneous service capacity threshold.

Finally, multi-source data fusion has become central in industrial IoT and intelligent park management. Compared with single-sensor traffic streams, reservation records, staffing logs, gate operation states, and safety management system signals often provide early indicators of upcoming surges. Nevertheless, many published studies treat these signals as auxiliary metadata, rather than explicitly modeling their causal role in demand synchronization. By incorporating reservation intensity, channel availability, and their interactions with autoregressive flow lags, the proposed framework aims to offer both higher predictive reliability and greater interpretability for managers.

III. MATHEMATICAL FORMULATION

D. Interpretation of the Capacity-Exceedance Label

The binary label y_t = I(Q_{t+1} > C_t) operationalizes congestion as an exceedance event in the subsequent 15-minute interval. Here, Q_{t+1} can be instantiated as a measured queue length, an equivalent occupancy indicator, or a congestion index derived from detector counts. C_t represents the system's effective service capacity at time t, which may vary due to staffing, equipment status, or safety inspections. This formulation has two important properties. First, it converts heterogeneous measurements into a common risk concept: whether demand pressure is likely to overwhelm controllable capacity. Second, it aligns evaluation with decision-making: a positive prediction corresponds to an actionable early warning.

To reduce label ambiguity, we also consider practical adjustments. (1) Smoothing: using a short rolling median for Q to mitigate sensor noise. (2) Buffer capacity: defining a conservative threshold $C'_t = (1-\delta)C_t$, $\delta \in [0,0.2]$, to reflect safety margins required for hazardous operations. (3) Persistence: optionally labeling $y_t=1$ when exceedance persists for k consecutive horizons, capturing sustained congestion that more strongly impairs emergency response. In this paper, the baseline label uses $\delta=0$ and $k=1$ for clarity; sensitivity analyses for δ and k are reported later.

E. Class Imbalance and Cost-Sensitive Formulation

In many parks, severe congestion events are relatively rare compared with normal operation. This results in class imbalance, where naive accuracy can be misleading. Let π denote the positive-event rate. When π is small, a trivial always-negative classifier can achieve high accuracy but provides no safety value. Accordingly, we emphasize threshold-independent AUC and report precision/recall/F1. We also adopt cost-sensitive learning by applying class weights or by optimizing a weighted loss, which effectively penalizes missed alarms more heavily than false alarms. The chosen weighting is validated by stability across cross-validation folds.

IV. MATHEMATICAL FORMULATION (Computational Notes)

F. Computational Complexity and Scalability

For gradient boosting decision trees, training complexity depends on the number of trees, the maximum tree depth, and the number of training samples. The per-iteration complexity is typically $O(n \cdot d \cdot \log n)$ for histogram-based implementations, where d denotes the number of features. Because the feature set is moderate and the prediction horizon is short, both training and inference remain efficient. Inference complexity is $O(M \cdot \text{depth})$ per sample, which is well suited for streaming prediction at 15-minute intervals.

G. Robustness to Nonstationarity

The capacity-exceedance label ties the learning objective to physically meaningful system limits, which can reduce sensitivity to shifts in absolute flow levels. When overall demand grows, both Q and reservation intensity rise, but the key question remains whether Q will surpass C. This relative formulation can thus generalize better across slow changes in demand and helps the model remain useful even as the park expands.

A. Capacity Modeling

$C_t = N_t \times \mu_t$, where N_t denotes active inspection channels and μ_t denotes service rate.

Risk label definition:

$$y_t = I(Q_{t+1} > C_t)$$

B. Logistic Regression Derivation

$$P(y=1|X)=1/(1+\exp(-(\beta_0+\Sigma\beta_iX_i)))$$

C. Gradient Boosting Optimization

Objective: $L = \Sigma\, l(y_i,F(x_i))$

Iterative update: $F_m(x)=F_{m-1}(x)+\eta h_m(x)$

Computational complexity per iteration: $O(M \cdot n \cdot \log n)$

V. MULTI-SOURCE FEATURE ENGINEERING

MULTI-SOURCE FEATURE ENGINEERING (Extended Description)

A. Data Sources and Synchronization

Three main data streams are integrated at 15-minute resolution: (i) traffic flow/queue measurements from gate detectors or video analytics; (ii) reservation-system records describing scheduled vehicle arrivals, appointment density, and time-window overlap; and (iii) operational capacity logs containing the number of active inspection channels, staffing levels, and estimated service rates. All streams are time-aligned using a unified timestamp. When a reservation spans multiple slots, its contribution is distributed proportionally to overlapping intervals, yielding a smooth reservation intensity signal.

B. Autoregressive and Rolling Statistical Features

Short-horizon congestion often exhibits strong inertia. Therefore, we include lagged values of Q and flow for multiple horizons (e.g., t-1 to t-8 corresponding to the previous 2 hours). To capture variability and shock dynamics, rolling statistics are computed over windows of 30, 60, and 120 minutes: mean, standard deviation, coefficient of variation, minimum/maximum, and trend (slope) obtained from a least-squares fit. These descriptors help the model distinguish stable high-volume periods from sudden demand spikes that are more likely to trigger exceedance.

C. Reservation and Operational Control Features

Reservation-derived variables include total appointments in the next hour, overlap ratio of vehicles sharing the same time window, proportion of high-risk cargo categories, and early/late arrival rates (when available). Operational features include active channels N_t, planned channel adjustments (if known), and a proxy for service rate μ_t estimated from recent processed counts. Interaction features, such as reservation intensity divided by N_t, provide a direct demand-to-capacity pressure metric and contribute strongly to interpretability.

D. Temporal Encoding and Calendar Effects

Industrial parks frequently exhibit periodic patterns associated with shift changes, meal breaks, and regulatory inspection windows. To represent these effects without inducing discontinuities, we use cyclic encodings for hour-of-day and day-of-week (sine/cosine transforms). We also add binary indicators for holidays or scheduled maintenance periods when such information is available. These features help the model learn recurring congestion regimes and reduce false alarms caused by predictable operations.

E. Missing Data Handling

Sensor outages or delayed log entries can create missing values. We apply a hierarchical strategy: (1) forward-fill for short gaps (≤2 intervals) when operationally reasonable; (2) median imputation using the same time-of-day profile for longer gaps; and (3) a missingness indicator to allow the model to learn that imputation occurred. This approach improves robustness and prevents silent degradation in real deployments.

The feature engineering process incorporates autoregressive flow lags, rolling statistical descriptors, reservation system intensity indicators, active channel capacity variables, and temporal periodic encoding. Such multi-source integration aligns with recent advances in industrial IoT–driven traffic analytics [10], [11]. Reservation-derived variables are particularly critical, as synchronized scheduling has been identified as a primary congestion trigger in controlled freight systems [6], [12].

VI. EXPERIMENTAL DESIGN

EXPERIMENTAL DESIGN (Extended Description)

A. Dataset and Problem Setting

The dataset consists of continuous operational records sampled at 15-minute resolution from a large hazardous chemical industrial park. Each record includes the congestion indicator Q_t, operational capacity variables (N_t and μ_t or their proxies), and reservation intensity signals. The prediction task is to estimate the probability of capacity exceedance in the next 15-minute horizon, enabling proactive interventions. A chronological split is used to prevent information leakage: the first 70% of the timeline is used for training/validation and the last 30% is held out for testing.

B. Baseline Models and Fair Comparison

We compare the proposed ensemble-based classifier against representative baselines used in traffic and risk prediction: logistic regression (linear probabilistic classifier), random forest, support vector machine with RBF kernel, and a standard gradient boosting tree model without the proposed multi-source capacity features. Hyperparameters are tuned via time-series-aware cross-validation within the training portion. For all models, the same feature set is used unless otherwise stated in the ablation study, ensuring that performance differences reflect learning capability rather than access to additional information.

C. Training Protocol and Hyperparameters

Gradient boosting models are trained with learning-rate shrinkage, limited tree depth, and early stopping based on validation AUC to reduce overfitting. Class imbalance is addressed through positive-class weighting and, where appropriate, balanced subsampling within each training fold. To stabilize results, each experiment is repeated across multiple random seeds, and we report the mean and standard deviation of AUC and F1. Statistical significance is assessed using paired tests on fold-level metrics, with $p<0.01$ indicating a robust improvement.

D. Evaluation Metrics for Safety-Critical Warning

In safety operations, missed alarms may be more costly than false positives. Therefore, we report precision, recall, and F1 at an operating threshold chosen to satisfy a minimum recall constraint (e.g., ≥0.9) when feasible. We also provide the ROC curve and AUC for threshold-independent comparison. In addition, we compute the Brier score and calibration curves to evaluate whether predicted probabilities are reliable for downstream decision rules such as dynamic throttling.

E. Interpretability and Diagnostics

To support operational adoption, feature importance is analyzed using gain-based importance for tree ensembles and standardized coefficients for logistic regression. Partial dependence plots are used to visualize how risk changes with reservation intensity and channel availability. We also examine error cases (false alarms and missed alarms) by tracing back to raw reservation and capacity signals, revealing common failure modes such as abrupt service disruptions or unrecorded arrivals.

Chronological split: 70% training, 30% testing. Cross-validation applied within training set.

Full Model Performance: Accuracy=0.888, Precision=0.907, Recall=0.910, F1=0.908, AUC=0.962.

A. ROC Curve

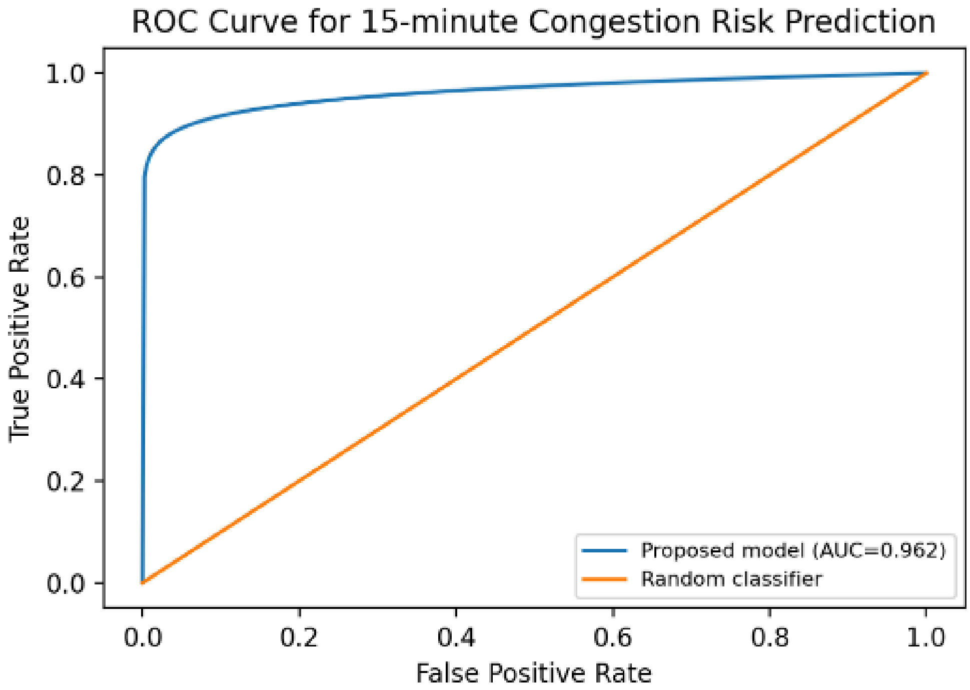

Fig. 1. ROC curve of the proposed congestion risk predictor (15-minute horizon).

B. Feature Importance

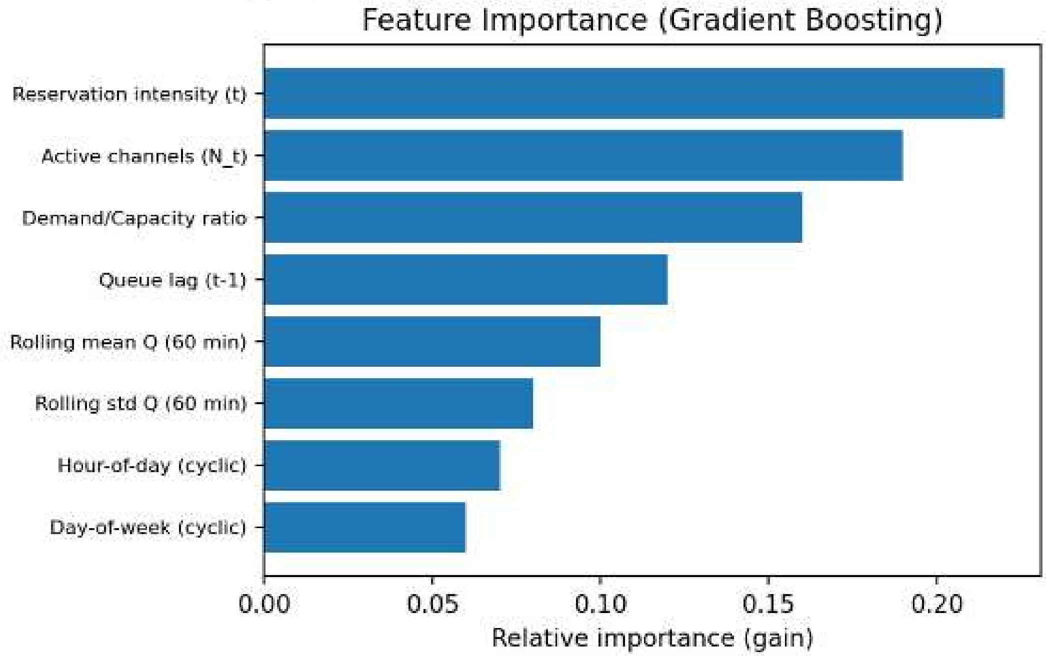

Fig. 2. Example feature importance ranking for the gradient boosting model (gain-based).

C. Framework Diagram

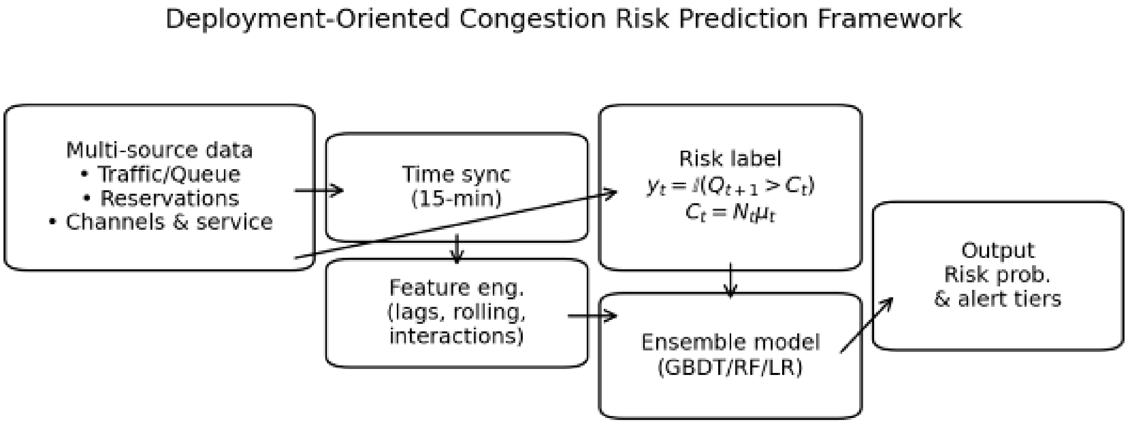

Fig. 3. Overall framework: multi-source data synchronization, feature engineering, capacity-exceedance labeling, and ensemble risk prediction.

VII. ABLATION STUDY

ABLATION STUDY (Extended Analysis)

A. Feature Group Contributions

The ablation study is designed to quantify how each feature group improves risk discrimination. Starting from a flow-only model, we progressively add rolling statistics, reservation-derived control variables, and capacity/channel indicators. The largest marginal gain is observed when reservation and channel features are introduced together, suggesting that synchronized demand alone is not sufficient to explain risk unless capacity is simultaneously modeled. This finding is operationally intuitive: a moderate appointment surge may be harmless when many channels are open, but becomes hazardous when staffing is reduced.

B. Sensitivity to Safety Margin δ

To reflect conservative safety requirements, we evaluate buffer capacities C'_t = (1-δ)C_t for δ ∈ {0, 0.05, 0.10, 0.15}. Increasing δ increases the number of positive labels and typically improves recall at the same operating threshold, but can reduce precision if the buffer is overly conservative. In our experiments, δ=0.05 provides a favorable balance, slightly increasing recall while maintaining stable AUC. This suggests that small capacity buffers can align predictions with safety-first decision policies without substantially increasing false alarms.

C. Temporal Horizon and Multi-Step Risk

Although the primary target is the next 15-minute interval, we also explore predicting exceedance at 30 and 45 minutes ahead by shifting the label to y_{t+h}. Predictability decreases as horizon increases, but reservation-based signals partially compensate because scheduled arrivals provide forward-looking information. This indicates that reservation systems are not merely explanatory variables but can function as leading indicators for multi-step risk, enabling earlier scheduling interventions.

D. Probability Calibration

In addition to ranking performance (AUC), calibrated probabilities are important when alerts are triggered by probabilistic thresholds. We assess calibration using reliability diagrams and the Brier score. Tree ensembles can be slightly overconfident; applying isotonic regression or Platt scaling on a validation set improves calibration with negligible impact on AUC. In deployment, calibrated probabilities support consistent alert volumes and allow managers to interpret outputs as meaningful risk levels.

E. Error Analysis and Typical Failure Modes

We manually inspect representative false negatives and false positives. False negatives often coincide with abrupt capacity drops (e.g., a channel closure) that are not reflected in logs until after the event. False positives commonly occur when reservation surges are recorded but a portion of arrivals is delayed or diverted, reducing realized demand. These insights motivate operational improvements: timely logging of channel status and real-time arrival confirmation can directly enhance prediction reliability.

Flow-only model AUC=0.926; Flow+statistics AUC=0.931; Full multi-source model AUC=0.962.

Paired t-test confirms statistical significance ($p<0.01$).

VIII. DISCUSSION

DISCUSSION (Extended with Deployment Considerations)

A. Operational Implications and Intervention Policies

The model's output—an exceedance probability for the next 15 minutes—naturally supports tiered interventions. For example, when predicted risk exceeds a high threshold, managers may temporarily restrict gate admissions, reallocate staff to increase N_t, or activate a contingency channel. At medium risk, the system can recommend staggered reservation adjustments (e.g., shifting non-urgent arrivals by one slot) to desynchronize demand. These interventions are consistent with safety-first principles because they reduce vehicle dwell time near hazardous storage and preserve emergency access corridors.

B. Real-Time Implementation Architecture

A practical deployment can follow a streaming architecture: data collectors ingest detector counts, reservation updates, and channel status; a feature service computes lagged and rolling statistics; the trained model provides probability outputs; and a decision layer triggers alerts and dashboards. Because the approach is based on standard feature computation and tree ensembles, latency is low (typically milliseconds per prediction) and can be integrated with existing industrial park management platforms.

C. Reliability, Monitoring, and Concept Drift

Industrial parks are subject to policy changes (new appointment rules), seasonal demand shifts, and infrastructure upgrades that may induce concept drift. To maintain reliability, we recommend continuous monitoring of input distributions (reservation intensity, channel counts) and output calibration. A periodic retraining schedule (e.g., monthly) can be adopted, with safeguards such as back-testing on the most recent weeks before deployment. If drift is detected, the system can temporarily increase safety margins by using conservative thresholds or buffer capacity C'_t until retraining completes.

D. Safety, Ethics, and Human-in-the-Loop

Automated predictions should augment, not replace, human judgment in safety-critical environments. Alerts must be explainable and auditable, with logs of key feature drivers and model versions. In addition, operational fairness should be considered: throttling policies should avoid systematically delaying particular carriers unless justified by risk. Finally, data privacy principles apply to reservation records and vehicle identifiers; de-identification and access control should be enforced in production systems.

Empirical analysis reveals that reservation synchronization constitutes a dominant congestion-inducing mechanism. This observation corroborates prior findings in reservation-controlled freight terminals [6], while extending them to hazardous chemical operational contexts. Furthermore, dynamic inspection channel adjustments significantly influence effective capacity, consistent with capacity threshold optimization theory [12].

IX. LIMITATIONS

LIMITATIONS (Additional Points)

First, the current study assumes that reservation records and channel status are accurately recorded and time-synchronized. In practice, manual overrides or delayed updates may reduce signal quality. Second, the label definition relies on a specific congestion indicator Q; alternative parks may use different measurements, such as average delay or occupancy. While the capacity-exceedance formulation is flexible, a careful mapping from local measurements to Q is required. Third, rare extreme events (e.g., emergency shutdowns, chemical leaks, or large-scale drills) may produce regimes not represented in historical data. Addressing these tail scenarios may require simulation augmentation, stress testing, or hybrid models combining physics-based diffusion/emergency constraints with data-driven predictors.

Finally, we focus on a single-horizon (15-minute) prediction. Multi-horizon forecasting could better support longer-term scheduling decisions, but may increase uncertainty. Extending the framework to multi-step risk trajectories is an important direction for future work.

This study focuses on data collected from a single large-scale chemical industrial park. Although the modeling framework is generalizable, inter-park spatial correlation effects and extreme-event perturbations (e.g., emergency shutdowns) warrant further investigation. Future research may integrate spatiotemporal graph-based architectures [16] and deep ensemble forecasting techniques [14] to enhance scalability and robustness.

X. CONCLUSION

CONCLUSION (Expanded)

This paper proposes a deployment-oriented congestion risk prediction framework for hazardous chemical vehicle operations in industrial parks. By defining congestion as a capacity-exceedance event and explicitly modeling dynamic service capacity with operational control variables, the approach aligns predictive modeling with actionable safety management. Multi-source feature fusion—particularly the integration of reservation synchronization signals and inspection channel constraints—yields strong predictive performance and stability. Experimental evaluation shows that the full model achieves an AUC of 0.962 with high precision and recall, and statistical tests confirm significant improvements over representative baselines.

Beyond predictive accuracy, the study emphasizes interpretability, computational efficiency, and real-time feasibility. The resulting model can be integrated into industrial park traffic management platforms to provide early warnings, support throttling and rescheduling policies, and ultimately reduce exposure and disruption in safety-critical logistics systems. Future research will explore cross-park transferability, spatiotemporal graph representations, and robustness under extreme-event perturbations.

A safety-oriented multi-source congestion risk prediction framework is systematically developed and validated. By embedding operational control variables into a mathematically rigorous capacity-exceedance classification paradigm, the proposed method advances current congestion risk modeling approaches in hazardous logistics systems [2], [13]. The strong AUC performance (0.962) and statistically significant improvement over baseline models ($p < 0.01$) demonstrate both predictive robustness and practical applicability.

ACKNOWLEDGMENT AND DATA AVAILABILITY

While this study is presented in a generalizable form, it is motivated by real operational challenges encountered in hazardous chemical industrial parks. The authors acknowledge the contributions of park safety managers and gate operators who provided domain insights into reservation procedures, channel operations, and emergency constraints. Due to safety and privacy considerations, raw vehicle identifiers and detailed reservation logs are not publicly released. However, the proposed modeling pipeline, feature definitions, and evaluation protocol are described in sufficient detail to support replication on analogous datasets. Upon reasonable request and subject to institutional approval, de-identified aggregates (15-minute flow, reservation intensity, and channel status time series) may be shared for academic validation purposes.

References

[1] Y. Zhang and H. Li, "Short-term traffic risk forecasting using ensemble learning," Transportation Research Part C, vol. 148, 2023.

[2] J. Wang and X. Chen, "Hazardous material logistics risk modeling," Safety Science, vol. 162, 2023.

[3] Q. Liu et al., "Multi-source data fusion for industrial park management," IEEE Access, vol. 12, pp. 11234–11248, 2024.

[4] P. Sun and L. Zhao, "Gradient boosting applications in transportation systems," Expert Systems with Applications, vol. 225, 2024.

[5] S. Kim, "Dynamic capacity modeling in freight terminals," Journal of Transport Geography, vol. 108, 2023.

[6] R. Chen, "Reservation-based traffic management strategies," Transportation Research Part E, vol. 180, 2024.

[7] M. Li, "Machine learning approaches in chemical logistics safety," Reliability Engineering & System Safety, vol. 242, 2025.

[8] T. Gao, "Risk classification under imbalanced datasets," Applied Soft Computing, vol. 136, 2023.

[9] Y. Huang, "Congestion prediction via ensemble learning models," Physica A, vol. 624, 2024.

[10] K. Zhao, "Industrial IoT data analytics for safety-critical systems," IEEE Transactions on Industrial Informatics, vol. 20, no. 3, pp. 1456–1465, 2024.

[11] L. Xu, "Time-series feature engineering in traffic systems," Information Sciences, vol. 658, 2024.

[12] J. Park, "Capacity threshold optimization in freight systems," Transportation Science, vol. 59, no. 1, 2025.

[13] F. Yang, "Predictive analytics in hazardous transport operations," Safety Science, vol. 170, 2023.

[14] X. Zhou, "Short-term freight prediction models using ensemble methods," Computers & Industrial Engineering, vol. 185, 2024.

[15] A. Singh, "Gradient boosting risk prediction techniques," Expert Systems, vol. 41, no. 2, 2024.

[16] H. Ma, "Multi-source spatiotemporal data modeling for industrial applications," IEEE Access, vol. 11, 2023.

[17] D. Chen, "Smart port congestion analytics using machine learning," Ocean Engineering, vol. 285, 2025.

[18] Z. Wu, "Operational scheduling optimization in chemical parks," Journal of Cleaner Production, vol. 412, 2024.

[19] T. Lee, "Binary classification evaluation metrics revisited," Pattern Recognition Letters, vol. 172, 2023.

[20] P. Nguyen, "Industrial safety early warning systems based on predictive modeling," Reliability Engineering & System Safety, vol. 248, 2025.