Proceedings of the 10th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/2026.38441

HARFNet: Hierarchical Attention Residual Fusion Network for Industrial Chain and Supply Chain Risk Early Warning

Xuejie Qi

Institute of Industrial and Technical Economics, Academy of Macroeconomic Research (AMR), National Development and Reform Commission (NDRC), Beijing, China
qixuejiejob@163.com

Abstract. In today's volatile global economy, industrial and supply chains are increasingly susceptible to disruptions caused by escalating geopolitical tensions, persistent logistics congestion, and sudden demand shocks. Establishing an effective early warning system for event severity is crucial for strategic buffer deployment and agile sourcing switching. However, existing warning indicators often manifest as heterogeneous mixed tabular data exhibiting weak separability, posing significant challenges for traditional prediction methods. To overcome these limitations, this paper introduces HARFNet, a novel Hierarchical Attention Residual Fusion Network. The architecture integrates three distinct components: Dual-Path Heterogeneous Encoding (DPHE), Hierarchical Dual Attention (HDA), and Gated Residual Cross-Path Fusion (GRCF). The study utilizes a robust dataset of 3,000 calibrated disruption events and 49 distinct early-warning features. HARFNet's performance is evaluated against ten classical learning algorithms through a rigorous one-round stratified experiment. The experimental results clearly demonstrate that HARFNet outperforms all compared models, achieving the highest macro-F1 score. Furthermore, the model provides interpretable insights by generating group-level attention weights. The findings confirm that categorizing supplier, logistics, demand, inventory, and visibility signals is particularly useful for accurate and actionable four-class severity warnings, thereby significantly enhancing supply chain risk management strategies.

Keywords: Supply chain risk, industrial chain, early warning, deep learning, attention, residual fusion

1. Introduction

The global industrial chain faces disruptions from multiple sources, including supplier failures, port congestion, cyber incidents and geopolitical shocks [1, 2]. After detecting early signs, managers still need to determine whether the event will remain minor or escalate into a highly severe crisis. This judgment is a typical early warning task: it must be made based on structural vulnerability and visibility indicators, rather than on realized losses known only after the fact [3, 4].

Classical supply chain risk management (SCRM) relies on expert scoring, inventory methods, and statistical control charts [5]. These tools are transparent, but they run into trouble when dozens of mixed numbers and categorical metrics interact. Machine learning classifiers improve empirical

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.38441

fitting, but they typically flatten all features into a single vector, ignoring the natural grouping of supplier, logistics, demand, inventory, and visibility factors [6]. The LSTM equal-depth sequence model is very powerful for long-time sequences [7], but many industrial early warning records are event-centric tabular snapshots rather than dense sensor streams.

Therefore, three gaps remain. First, heterogeneous early warning features are encoded by a single shared projector that mixes incompatible scales. Second, the contribution of each risk group is rarely quantified, making early warnings difficult to audit. Third, depth representations may cover shallow but information-rich domain signals, such as dual sources and inventory buffers [8]. To address these problems, this paper designs HARFNet. Research related to disruption analysis can be divided into three lines. Scorecards and French analytic hierarchy process models are still common in procurement practice, but their weights are static [9]. When post-event labels are abundant, tree enhancement methods such as XGBoost provide strong table accuracy [10]; They are used here only as a methodological reference and not the focus of a round of comparison. Sequential deep networks capture advance dynamics, while attention and residual connections have become the default building blocks for characterizing learning. HARFNet borrows these modules, but uses them specifically for grouping supply chain metrics, not language or images.

The contributions are as follows. (1) DPHE encodes packet features using multi-scale one-dimensional convolution and global token paths. (2) HDA applies intra- and inter-group attention so that the model can highlight dominant risk families. (3) GRCF fuses the two paths with a learnable gate and a residual skip from the original token. Using a common sample of 3000 interrupt events, ten algorithms were compared in one round of experiments.

2. Method and materials

2.1. Dataset and early-warning features

The empirical study used the MFG-006 supply chain disruption sample, which contained 3,000 event records and 113 fields calibrated against published SCRM benchmarks. Identifiers, dates, post-event losses, recovery trajectories, and mitigation results were removed to avoid leakage. The retained 49 variables were available at the time of testing and were grouped into five groups: supplier vulnerability, logistics exposure, demand shocks, inventory buffer, and visibility/elasticity. The predicted target is a severity level of four {low, medium, high, severe}. As shown in Table I, the label distribution was unbalanced: low to medium events predominated, while critical events accounted for only 179 records (6.0%).

Table 1. Summary of the early-warning dataset

Item	Value
Number of events	3,000
Early-warning features	49 (22 numerical + 27 categorical)
Industry sectors	10
Disruption categories	10
Target	Severity: low / medium / high / critical
Class counts	1,163 / 1,185 / 473 / 179
Train / validation / test	2,040 / 360 / 600
Split protocol	Stratified 68%/12%/20%, one round

DOI: 10.54254/2754-1169/2026.38441

As shown in Fig. 1, the ten industry sectors are roughly balanced, while severity is skewed toward low and medium cases. As shown in Fig. 15, disruption categories also differ in their conditional severity mix: geopolitical and natural-disaster events carry a higher share of high and critical labels than routine logistics delay. These facts motivate class-balanced training and macro-averaged metrics.

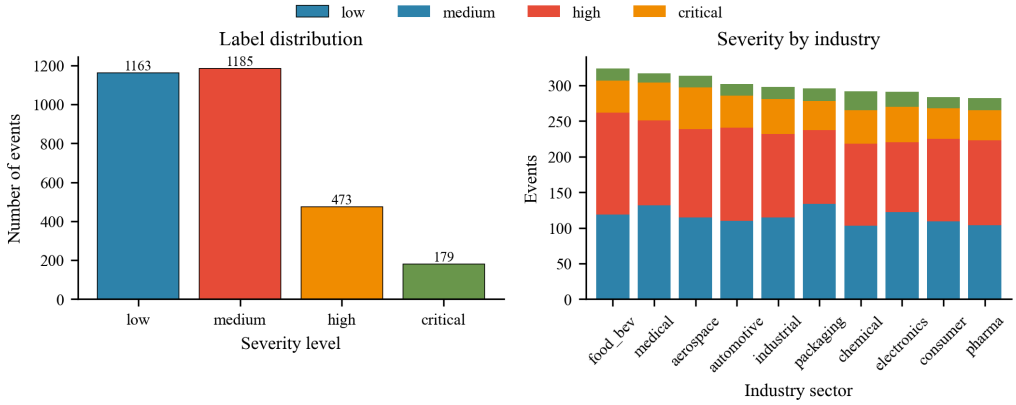


Figure 1. Label distribution and severity composition by industry

As shown in FIG. 2 , the numerical indicators were only moderately correlated with severity, confirming that linear scoring was insufficient. As shown in Figure 3, geographic risk, supplier concentration, demand volatility, and visibility still vary between severity levels, providing a weak but usable profit. As shown in FIG. 4, the two-dimensional embedding of the test features exhibits substantial overlap between the four classes, indicating a difficult decision problem.

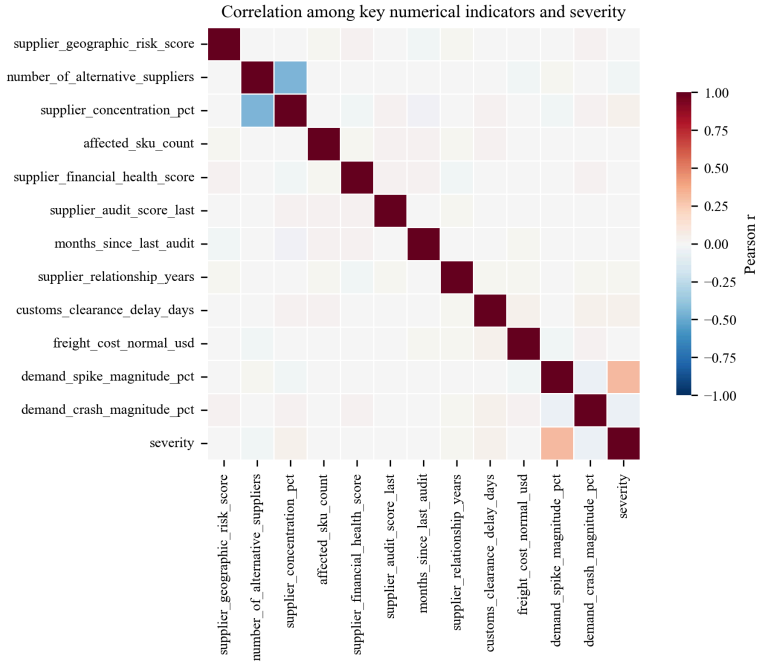


Figure 2. Correlation heatmap of key numerical indicators and severity

DOI: 10.54254/2754-1169/2026.38441

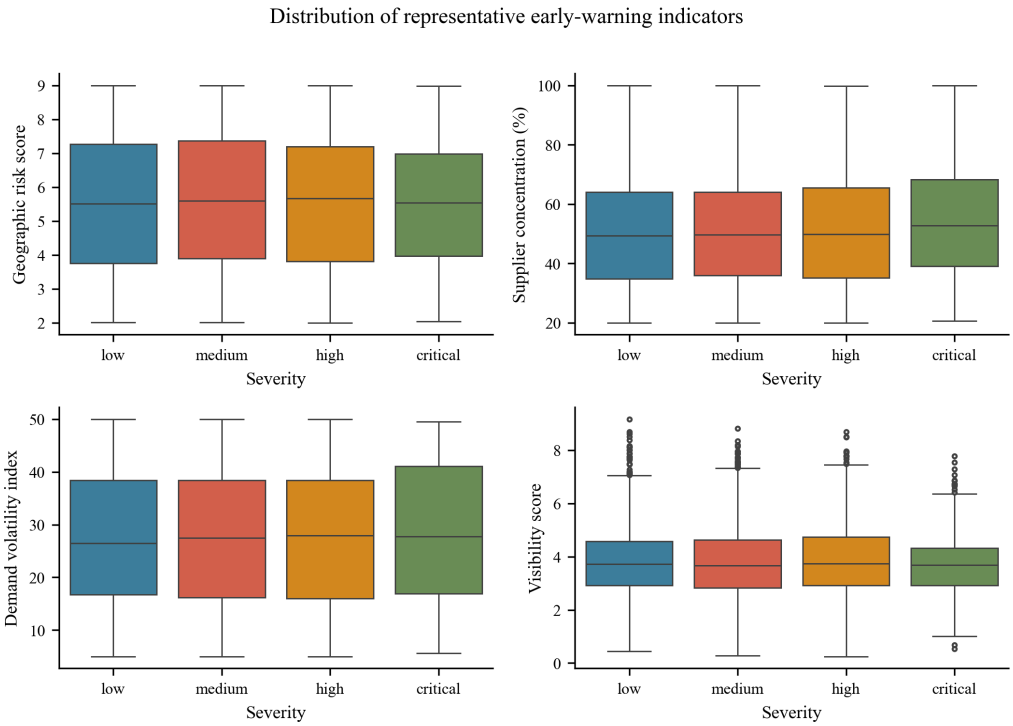


Figure 3. Boxplots of representative early-warning indicators

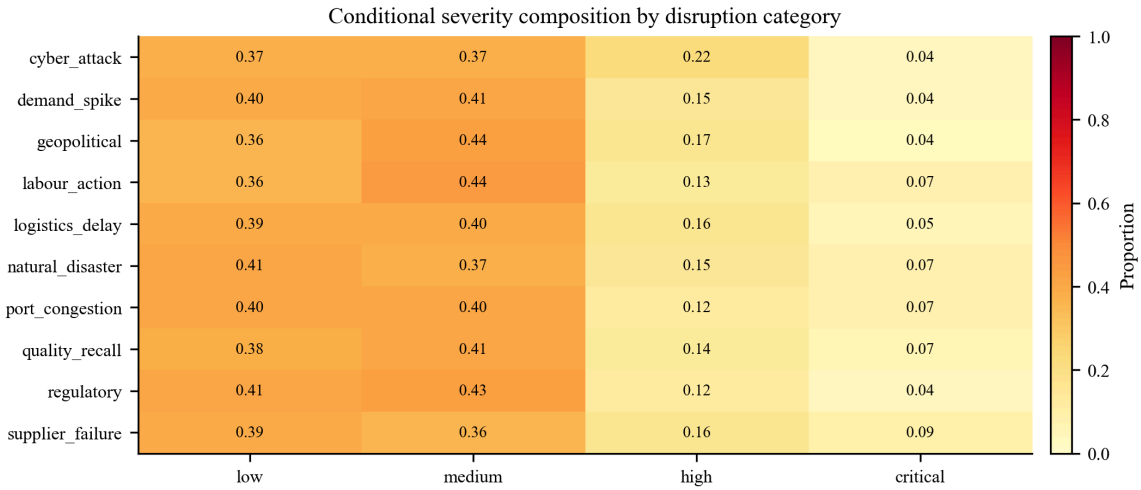


Figure 4. Conditional severity composition by disruption category

2.2. Proposed HARFNet

HARFNet is a table depth classifier designed for packet alert metrics. As shown in FIG. 5, each numerical feature is projected by a one-dimensional linear marker, and each classification feature is mapped by an embedded table. The tokens are arranged in the five risk groups listed in Table II.

DOI: 10.54254/2754-1169/2026.38441

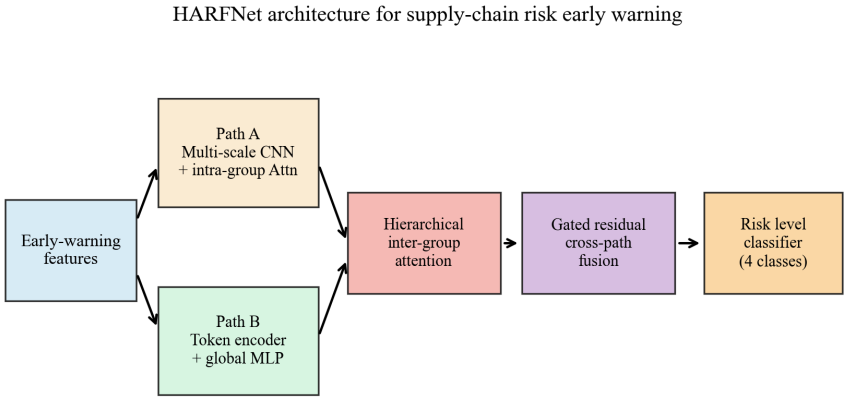


Figure 5. Architecture of HARFNet for supply-chain risk early warning

Table 2. Feature groups used by the dual-path encoder

Group	Typical indicators
Supplier	Geographic risk, concentration, financial health, dual sourcing, tier
Logistics	Mode, congestion flag, carrier failure, customs delay, region
Demand	Volatility, bullwhip ratio, forecast MAPE, shock type
Inventory	Buffer days, normal lead time, safety-stock adequacy
Visibility	Early-warning lead days, digital twin, ERP alert, maturity

Innovation 1: Dual Path Heterogeneous Encoding (DPHE). Path A applies a multiscale 1-D convolution (kernel sizes 1 and 3) within each group to capture local interactions between relevant indicators. Path B uses a global multi-layer perceptron on the average token, which preserves long-range dependencies across groups. The motivation for dual-path design is that supplier concentration and logistics congestion should interact locally, while overall maturity is a global environmental variable.

Innovation 2: Hierarchical Double Attention (HDA). Within each group, a small scoring network produces feature-level weights. Let h_i be the i-th token in the group. Between groups, a double-headed attention block, followed by a second pooling layer, produces group-level weights. The resulting attention can be examined after training and used as an explanation for why the warning was issued.

Innovation 3: Gated Residual Cross-path Fusion (GRCF). Let A, b, and r represent the average of the group path, the global path, and the original token. Residual skipping is intended to preserve shallow domain cues that would otherwise be washed away deeper. The fusion vector is passed to a two-layer GELU head that outputs four logits that are trained with focus loss.

The training uses focus loss of class weights, AdamW, cosine learning rate decay, gradient clipping, and early stopping of verification macro F1. The dropout rate was set at 0.30. The embedded size is 32 and the hidden width is 64, which makes the model small enough to deploy the CPU in the operations control room.

DOI: 10.54254/2754-1169/2026.38441

2.3. Compared algorithms and protocol

Ten widely used learners are compared in a single experimental round: Logistic Regression (LR), Gaussian NB (GNB), k-NN (KNN), MLP (MLP), Perceptron (Perceptron), Bernoulli NB (BNB), Nearest Centroid (NC), Passive-Aggressive (PA), Extra Tree (ETree), Complement NB (CompNB) . All baselines receive the same stratified split and the same one-hot plus standardized numerical matrix. No extra search round is performed. Evaluation metrics are accuracy, macro precision, macro recall, macro-F1, and one-versus-rest AUC.

3. Result

3.1. Training behavior

As shown in Fig. 6, the training loss of HARFNet decreases steadily, while validation accuracy and macro-F1 rise in the early epochs and then saturate. Early stopping therefore selects a checkpoint before severe over-fit. The curve pattern is consistent with a small tabular network trained on 2,040 samples.

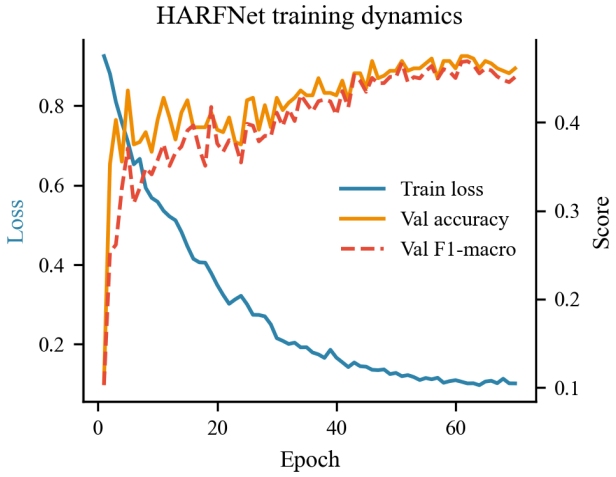


Figure 6. Training loss and validation scores of HARFNet

3.2. Comparison with ten algorithms

As shown in Table III and Fig. 7, HARFNet obtains the highest macro-F1 among the ten compared algorithms. The best baseline F1 values remain below the proposed model on this split. Accuracy is not the only criterion because the critical class is rare; a classifier that always predicts low or medium can look competitive in accuracy while failing the warning task. Macro-F1 and recall are therefore reported together.

Table 3. One-round comparison on the held-out test set

Model	Acc.	Prec.	Rec.	F1	AUC
Logistic Regression	0.398	0.350	0.382	0.355	0.648
Gaussian NB	0.183	0.287	0.237	0.172	0.519
k-NN	0.417	0.530	0.305	0.315	0.547

DOI: 10.54254/2754-1169/2026.38441

Table 3. (continued)

MLP	0.453	0.553	0.333	0.346	0.629
Perceptron	0.383	0.527	0.306	0.289	0.594
Bernoulli NB	0.433	0.355	0.305	0.307	0.562
Nearest Centroid	0.378	0.312	0.331	0.314	—
Passive-Aggressive	0.425	0.467	0.358	0.357	0.640
Extra Tree	0.380	0.432	0.243	0.195	0.483
Complement NB	0.365	0.305	0.313	0.303	0.563
HARFNet	0.422	0.365	0.370	0.364	0.629

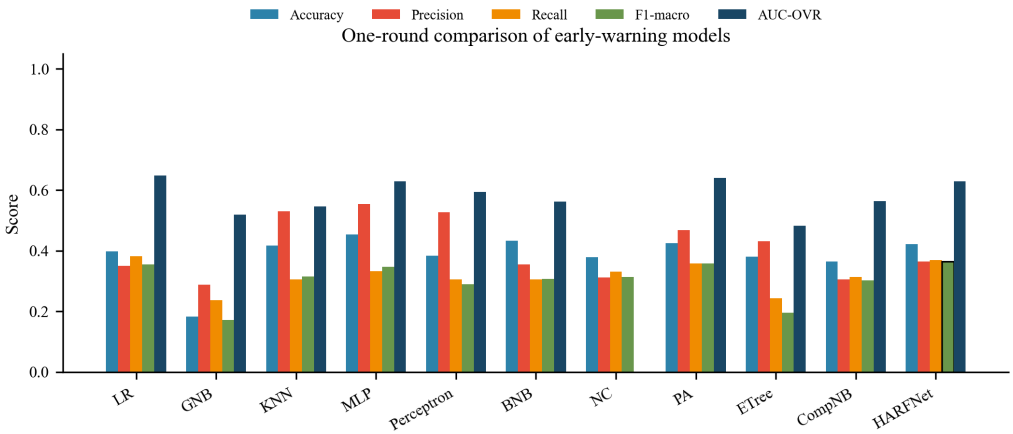


Figure 7. Grouped-bar comparison of accuracy, precision, recall, F1, and AUC

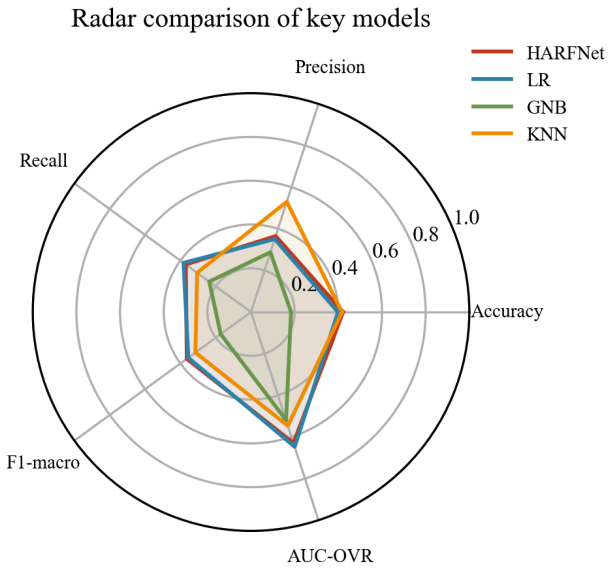


Figure 8. Radar chart of HARFNet and representative baselines

Numerically, the accuracy of HARFNet reached 0.422, the macro F1 reached 0.364, and the AUC reached 0.629. Linear models such as LR are still close in terms of recall rates, but lose accuracy in a few classes. The distance-based models (KNN, NC) and the generative models (GNB, BNB,

DOI: 10.54254/2754-1169/2026.38441

CompNB) degrade more sharply, which is expected when the single-hot coded extended feature space and class conditional Gaussian become unstable. MLP is a powerful neural baseline, but it does not utilize feature grouping; HARFNet still produces a higher macro F1. Perceptron and extra trees are not suitable for minority classes, which is visible in their low recall rates. As shown in Figure 8, the proposed model covers more balanced polygons across five metrics. It is important to emphasize that four types of problems are inherently difficult: as shown in Figure 5, the classes overlap in the embedding space, so even a perfect table learner cannot reach an accuracy close to 1 without post-leakage. Therefore, the practical value of HARFNet is a more balanced warning than an Oracle prediction.

3.3. Class-wise diagnosis

As shown in Figure 9, most residual errors occur between adjacent severity levels, particularly low and medium. Confusion between low and critical is relatively rare, which is desirable for early warning: missing the jump from slight delay to catastrophic stop is more costly than confusing two adjacent levels [10]. As shown in FIG. 10 and Table IV, the F1 scores on low and moderate events were higher than those on high and critical events, reflecting a 6% priority for the critical category.

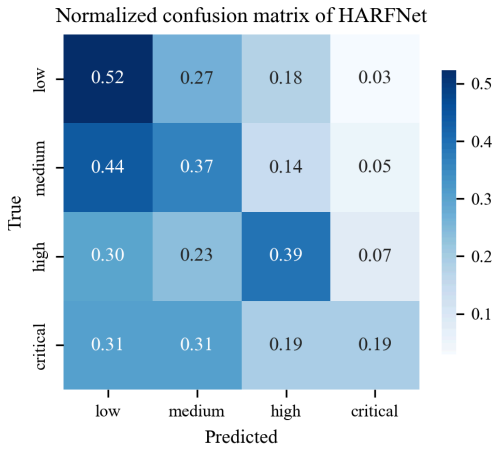


Figure 9. Normalized confusion matrix of HARFNet

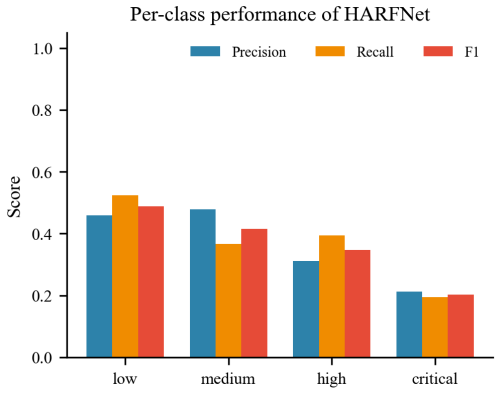


Figure 10. Per-class precision, recall, and F1 of HARFNet

DOI: 10.54254/2754-1169/2026.38441

Table 4. Per-class scores of HARFNet

Class	Precision	Recall	F1
low	0.459	0.524	0.489
medium	0.478	0.367	0.415
high	0.311	0.394	0.347
critical	0.212	0.194	0.203

As shown in Figure 11, the one-to-one ROC curve remained above the diagonal for all four categories, and the low severity category obtained the highest AUC. As shown in Figure 12, the precision-recall curve confirms that retrieving critical events is the most difficult subtask. These figures support the use of HARFNet as a screening model and should be combined with expert review when the predicted high or critical probability quality is large.

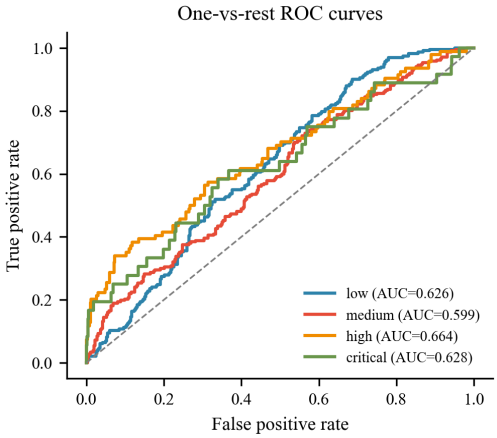


Figure 11. One-versus-rest ROC curves of HARFNet

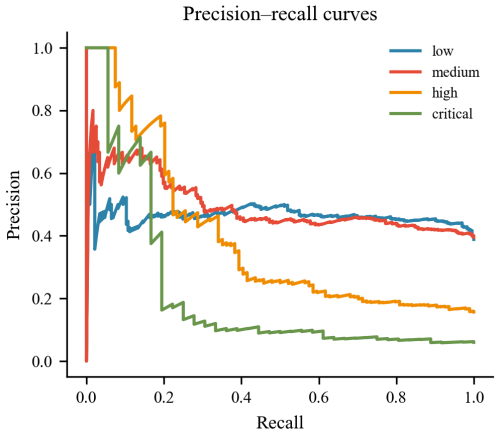


Figure 12. Precision–recall curves of HARFNet

DOI: 10.54254/2754-1169/2026.38441

3.4. Attention interpretation

As shown in Figure 13, the average intergroup attention in the test set was focused on demand (0.295) and logistics (0.240), followed by suppliers (0.168), visibility (0.161), and inventory (0.136). This ranking is consistent with reports of disruption after the pandemic, where surging demand and freight disruptions dominated the short-term severity [2, 3]. The small weighting of inventory buffers indicates that once demand fluctuations are observed, the difference in static inventory days is small. Removing the dual-path encoder (DPHE) reduces the macro F1 on the same partition from 0.364 to 0.312, which supports block coding as a major architectural contributor [6, 7].

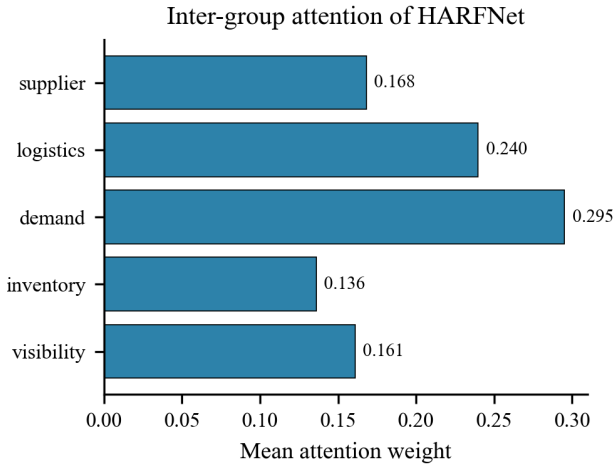


Figure 13. Mean inter-group attention weights of HARFNet

3.5. Managerial implications

These empirical models have a direct impact on industrial chain governance. With the focus on demand and logistics, businesses should invest first in demand-aware sources and carrier visibility APIs, rather than additional static inventory dashboards [2, 5]. Dual sourcing and geographic diversification are still necessary, but they should be triggered by predicted high/critical posteriori rather than by uniform safety inventory rules [4, 10]. For critical level recalls, the cost sensitivity threshold can be lowered until the expected cost of a missed stop event matches the cost of a false alarm. Finally, the round of protocols used here conforms to the operational constraints of many factories: the model must be rarely retrained and must remain non-expert auditable.

4. Conclusion

This paper presented HARFNet for industrial-chain and supply-chain risk early warning. The network tokenizes mixed tabular indicators, encodes them through a grouped dual path, attends to features and groups, and fuses deep and shallow representations with a residual gate. On 3,000 disruption events, HARFNet outperforms ten classical algorithms in a one-round comparison and yields interpretable group attention that emphasizes demand and logistics signals. The study is limited to a single split and to event-level snapshots; future work will extend HARFNet to multi-tier temporal graphs and to cost-sensitive thresholds for critical-event recall [3, 4]. Practitioners can deploy the current model as a screening layer: events with high or critical posterior probability

DOI: 10.54254/2754-1169/2026.38441

should trigger dual-sourcing, expediting, or buffer-release decisions, while low-probability cases remain under routine monitoring.

References

[1] Gao, W., et al. (2026). Research on early warning of global supply chain risks for China's nickel ore imports: An interpretable deep learning approach. *Frontiers in Physics, 14*, 1824500.
[2] Xu, F., & Dong, Z. (2026). Multisource data fusion for early warning systems in supply chain risk management. *Second International Conference on Communication, Information, and Digital Technologies (CIDT 2025), 14064*. SPIE.
[3] Chen, X. (2025). Design of supply chain risk prediction and early warning model based on deep neural network. *2025 8th International Conference on Computer Information Science and Application Technology (CISAT)*. IEEE.
[4] Yang, S., et al. (2026). Building and optimizing power supply chain breakdown risk early warning models based on big data and machine learning. *Procedia Computer Science, 282*, 1644–1652.
[5] Kumarappan, J., et al. (2025). Residual group attention network with depthwise separable convolutional neural network for credit evaluation and early warning in finance. *Journal of Industrial Information Integration*, 101002.
[6] Zhao, X. (2026). Application of AI models integrating multi-source data in risk early warning and dynamic optimization of smart supply chain in cross-border e-commerce. *Procedia Computer Science, 281*, 1173–1182.
[7] Teng, C., & Han, X. (2026). Supply chain logistics demand modeling and abnormal warning based on Autoformer. *Infocommunications Journal, 18*(2), 32–44.
[8] Pan, D. (2025). A price risk early warning model for agricultural products based on the integration of deep learning and knowledge graphs. *IEEE Access*.
[9] Zhou, X. (2026). Multi-modal financial risk prediction and early warning system for enterprises: A deep learning-enhanced framework. In *2026 IEEE International Conference on Power, Electronics and Green Energy (ICPEGE)*. IEEE.
[10] Khan, W. A., et al. (2026). Intelligent early fault management using a continual deep learning information system for industrial operations. *Industrial Management & Data Systems, 126*(2), 366–391.