

Article history:

Received 15 July 2026

Revised 26 September 2026

Accepted 06 October 2026

Initial Publish 29 October 2026

Published online 01 November 2027

Journal of Resource Management and Decision Engineering

Volume 6, Issue 6, pp 1-15



Modeling a Resilient Supply Chain Using Machine Learning and Evolutionary Algorithms

Mahta. Saffari¹, Ali. Rajabzadeh Ghatari^{2*}, Seyyed Hosein. Seyyed Esfahani³, Maryam. Khademi³

¹ Department of Industrial Management, NT.C., Islamic Azad University, Tehran, Iran.

² Department of Industrial Management, Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran.

³ Department of Systems Management, Iran Center for Management Studies, Tehran, Iran.

⁴ Department of Applied Mathematics, ST.C., Islamic Azad University, Tehran, Iran.

* Corresponding author email address: alirajabzadeh@modares.ac.ir

Article Info

Article type:

Original Research

How to cite this article:

Saffari, M., Rajabzadeh Ghatari, A., Seyyed Esfahani, S. H., & Khademi, M. (2027). Modeling a Resilient Supply Chain Using Machine Learning and Evolutionary Algorithms. *Journal of Resource Management and Decision Engineering*, 6(6), 1-15.

<https://doi.org/10.61838/kman.jrmde.446>



© 2027 the authors. Published by KMAN Publication Inc. (KMANPUB). This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

This study aimed to model a resilient supply chain using machine learning and evolutionary algorithms in the apparel, sporting goods, and electronic components industries. The research design employed a two-layer hybrid (predictive-prescriptive) framework with a quantitative, ex post facto approach. The population and sample comprised the standard DataCo database containing 180,519 logistics transactions, which was subsequently partitioned into three target subnetworks. Data analysis was based on the CRISP-DM methodology, K-means clustering, machine-learning models—particularly XGBoost (AUPRC = 0.8176)—and evolutionary multi-objective optimization algorithms (NSGA-II and MOPSO). NSGA-II demonstrated relative superiority in extracting a diverse Pareto front. The results showed that XGBoost provided acceptable performance in predicting delivery-delay risk, and the balanced strategy derived from NSGA-II yielded a proposed network design with an approximate cost of USD 327,650 and a mean risk of 0.519. Industry-level analysis further indicated that the sporting goods industry had the largest volume and cost, whereas the electronics industry exhibited the lowest cost and risk and therefore the most agile network. Future research is recommended to incorporate a disruption-scenario simulation layer to evaluate actual supply-chain resilience.

Keywords: resilient supply chain, machine learning, evolutionary algorithms, multi-objective optimization, apparel and sporting goods industries

1. Introduction

Global supply chains have become increasingly exposed to a complex combination of operational, technological, environmental, geopolitical, and market-related disruptions. The growing interdependence of suppliers, manufacturers, logistics providers, distribution centers, and customers has improved efficiency and market reach, but it has simultaneously increased the probability that localized disturbances will propagate throughout a network. Disruptions associated with transportation delays, demand volatility, supplier unreliability, climate-related events, geopolitical tensions, infrastructure failures, and unexpected market shocks can affect not only the continuity of material flows but also cost structures, customer service, inventory availability, and organizational competitiveness. Consequently, supply chain resilience has evolved from a supplementary risk-management concept into a central capability for maintaining continuity, adapting to disturbances, and restoring acceptable levels of performance after disruption. Recent studies emphasize that resilience should be understood as a multidimensional capability involving anticipation, preparedness, responsiveness, adaptation, and recovery rather than merely the ability to absorb isolated shocks (Golden et al., 2026; Wietschel et al., 2025).

The increasing frequency and interconnectedness of supply chain disruptions have also demonstrated the limitations of conventional risk-management approaches. Traditional models frequently rely on static assumptions, historical averages, deterministic parameters, or predefined disruption scenarios, whereas modern supply chains operate in environments characterized by dynamic uncertainty and nonlinear relationships among risk factors. Particularly in geographically dispersed logistics networks, disruption in one facility, transportation corridor, or supplier node may generate cascading consequences elsewhere in the network. Systematic evidence concerning cascading failures indicates that supply chain risks should therefore be evaluated not simply as independent events but as interconnected processes capable of propagating across multiple tiers and operational functions (Xiao et al., 2025). From this perspective, resilience-oriented network design requires analytical approaches capable of recognizing latent risk patterns, estimating disruption probabilities, and translating those predictions into actionable allocation and network-design decisions.

Digitalization and the emergence of Industry 4.0 have provided important opportunities for addressing these limitations. Artificial intelligence (AI), machine learning (ML), intelligent automation, advanced analytics, and data-driven decision-support systems allow organizations to analyze large-scale transactional and logistics data and identify patterns that may not be detectable through conventional statistical or managerial approaches. Within supply chain management, AI and ML can support demand forecasting, supplier evaluation, inventory control, transportation planning, risk detection, and strategic decision-making. Evidence suggests that these technologies can improve the quality, speed, and adaptability of supply chain decisions by converting large volumes of heterogeneous operational data into predictive information that can support managers under uncertainty (Nasirian, 2024; Paberja et al., 2026). Accordingly, the role of AI is increasingly shifting from isolated process automation toward the broader construction of intelligent and resilient supply chain systems.

The relationship between digital intelligence and supply chain resilience, however, extends beyond the adoption of individual technologies. Resilience emerges when technological capabilities are integrated with internal organizational competencies, external relationships, and appropriate decision structures. Research on intelligent manufacturing enterprises has shown that digital and intelligent development can strengthen resilience through mechanisms associated with internal capabilities and external linkages (Wang et al., 2026). Similarly, recent evidence indicates that technology or innovation alone may not be sufficient to produce resilient supply chains unless intelligent automation is embedded within broader organizational and configurational conditions (Mubarik & Maciukaite-Zviniene, 2026). These findings imply that the value of AI-based supply chain management depends not merely on predictive accuracy but also on the extent to which analytical outputs are incorporated into operational and strategic decisions.

Machine learning is particularly valuable in supply chain resilience because many operational risks have probabilistic, nonlinear, and high-dimensional characteristics. Models capable of learning from historical observations can estimate the likelihood of future outcomes, classify suppliers or transactions according to risk, and identify interactions among variables without requiring all relationships to be specified a priori. Supervised machine learning has, for example, been used to evaluate supplier performance and

develop risk profiles for resilience-oriented decision-making (Kang & Bhawna, 2025). Predictive analytics can likewise improve demand forecasting and allow organizations to respond more effectively to uncertainty in product requirements and market fluctuations (Prajapati, 2025). Such capabilities are especially important for logistics systems in which late delivery risk depends simultaneously on shipment characteristics, geographical distance, transportation methods, regional conditions, order attributes, and other operational factors.

Recent advances have further expanded supply chain analytics from conventional prediction toward causal and context-sensitive modeling. Causal machine learning and advanced data analytics have been proposed for resilience modeling under geopolitical, climate, and demand shocks, emphasizing the importance of distinguishing simple predictive associations from more informative representations of how disruptions emerge and propagate (Akinola, 2026). This development reflects a broader transition from descriptive analytics toward predictive and ultimately prescriptive intelligence. In a resilient supply chain, knowing that a delay is likely to occur is useful, but it is not sufficient. Managers also require guidance regarding how capacity should be allocated, which facilities should be activated, how flows should be redistributed, and what trade-offs should be accepted between cost and risk. Thus, the integration of prediction with optimization represents an important methodological direction in contemporary supply chain resilience research.

AI-supported resilience has also been increasingly connected with sustainability and the transition from Industry 4.0 to Industry 5.0. In manufacturing environments, AI and ML can contribute to both resilience and sustainability by improving resource utilization, enhancing visibility, anticipating disturbances, and supporting adaptive planning (Rane et al., 2025). The Industry 5.0 perspective further emphasizes the need for systems that combine technological intelligence with robustness, sustainability, and human-centered decision processes. This suggests that resilient supply chain design should not seek efficiency in isolation. Excessive concentration of inventory, transportation, or distribution activity may reduce costs under normal operating conditions but create substantial vulnerabilities when a disruption occurs. A resilient configuration therefore needs to balance efficiency with redundancy, diversification, flexibility, and adaptive capacity.

One of the most important analytical challenges is consequently the explicit management of trade-offs. Supply chain decision-makers often face conflicting objectives. A configuration that minimizes transportation or facility costs may increase dependence on a limited number of hubs, whereas a more distributed network may reduce disruption exposure but increase operating costs. Similarly, allocating demand to the geographically closest facility may reduce distance-related costs while potentially increasing concentration risk if too much volume is assigned to a single node. These conflicts make resilience-oriented supply chain network design inherently multi-objective. Rather than identifying one universally optimal solution, managers may need a set of non-dominated alternatives that represent different compromises between economic efficiency and risk reduction.

Simulation-optimization provides an important methodological foundation for addressing such problems. By combining models of uncertain supply chain behavior with optimization algorithms, researchers can evaluate alternative network configurations and identify robust decisions under multiple scenarios (Tordecilla et al., 2021). This approach is particularly appropriate where disruption probabilities and system responses cannot be adequately represented through simple deterministic formulations. Simulation-optimization also supports the evaluation of resilience across alternative conditions, enabling decision-makers to examine how network configurations perform when demand, capacity, transportation, or disruption patterns change. Nevertheless, the quality of optimization depends heavily on the parameters supplied to the model. If risk parameters are arbitrarily assumed or generated independently of observed operational data, the resulting network solution may have limited empirical relevance.

Evolutionary computation offers another powerful approach to supply chain network design because many resilience problems involve nonlinear relationships, multiple competing objectives, discrete and continuous decision variables, and complex constraints. Genetic algorithms have been applied to the development of reliable and resilient supply chain network models, demonstrating their suitability for searching large solution spaces where conventional analytical methods may be insufficient (Vishnu et al., 2020). Evolutionary algorithms are particularly useful for generating Pareto-optimal or near-Pareto-optimal solutions in multi-objective settings. Instead of collapsing cost and risk into a single arbitrary weighted objective, they can

preserve alternative trade-off solutions and thereby provide managers with a broader decision space.

Despite these advances, an important methodological gap remains between predictive risk modeling and prescriptive supply chain optimization. A substantial share of ML-oriented studies focuses on improving forecast or classification performance, while optimization studies often rely on externally assigned or simplified risk parameters. In practice, these two analytical layers should be connected. Risk probabilities estimated from actual transaction-level data can serve as empirically grounded parameters in optimization models, enabling the network-design process to respond to the same operational patterns detected by machine learning. Such integration is consistent with the AI-based framework proposed for manufacturing supply chain resilience, in which data-driven technologies are used to improve preparedness and adaptive capacity across interconnected supply chain activities (Belhadi et al., 2022). It is also aligned with recent work emphasizing enhanced logistics resilience through intelligent, integrated decision systems (Li & Ndiaye, 2026).

The relevance of integrated predictive-prescriptive approaches is further demonstrated across diverse sectoral contexts. Construction supply chains, for example, are vulnerable to disruptions arising from project interdependencies, material constraints, transportation problems, and uncertainty in supplier networks, highlighting the need for resilience-oriented planning (Golden et al., 2026). In healthcare, particularly intensive care environments, supply chain efficiency cannot be separated from resilience and patient safety because material shortages or delivery failures may directly affect service continuity and clinical operations (Hasmee et al., 2026). These sector-specific examples demonstrate that resilience is not limited to manufacturing or retail logistics; rather, it is a general network-design challenge in which resource availability, operational continuity, risk exposure, and cost must be jointly considered.

At the same time, sectoral heterogeneity means that a resilience strategy suitable for one industry may not be appropriate for another. Differences in order volume, product characteristics, value density, transportation requirements, geographic dispersion, and sensitivity to delay can produce substantially different cost-risk structures. Research examining resilience determinants in specific commodity systems has similarly shown that multiple structural and contextual factors contribute to supply chain robustness and adaptive capacity (Yao, 2025). Therefore,

analyzing industries as a single homogeneous network may conceal meaningful differences in logistical behavior. Industry-specific evaluation is particularly important when the same analytical framework is applied to categories with markedly different transaction volumes and distribution patterns, such as apparel, sporting goods, and electronic components.

Another conceptual issue concerns the measurement of resilience itself. Resilience is often treated as a broad and multidimensional construct that overlaps with reliability, flexibility, robustness, recovery, vulnerability, and criticality. Integrated conceptual approaches have therefore called for more systematic links between resilience assessment and supply criticality, especially where resource dependencies and disruptions can affect the continuity of downstream systems (Wietschel et al., 2025). In operational network design, however, resilience must ultimately be represented through measurable quantities that can be incorporated into mathematical models. Delivery-delay probability represents one such operationally interpretable indicator because late delivery reflects a disruption in planned logistics performance and can be estimated at the transaction level. When combined with concentration penalties or capacity constraints, predicted delay risk can provide a pragmatic basis for comparing alternative network configurations.

A hybrid analytical architecture that combines unsupervised learning, supervised machine learning, and multi-objective evolutionary optimization may therefore provide a more comprehensive representation of resilient supply chain decision-making. Unsupervised clustering can identify geographically meaningful distribution hubs; predictive models can estimate transaction- or hub-level delivery risk; and evolutionary algorithms can subsequently allocate demand while jointly minimizing economic cost and network risk. In this architecture, machine learning does not function as an isolated forecasting exercise. Instead, its outputs become endogenous inputs to prescriptive optimization. This creates a direct analytical pathway from historical logistics data to network configuration decisions and addresses the broader need for AI-supported decision processes in supply chain management (Kang & Bhawna, 2025; Paberja et al., 2026).

Such integration is particularly relevant when working with large logistics datasets containing substantial numbers of transactions, heterogeneous product categories, geographical information, transportation characteristics, and financial variables. These datasets provide an empirical

foundation for identifying distribution structures and estimating late-delivery risk, but they also create methodological challenges, including class imbalance, outliers, leakage from post-outcome variables, and nonlinear relationships. Carefully designed machine learning pipelines can address these issues and yield calibrated risk estimates for subsequent optimization. Furthermore, comparing alternative predictive algorithms and alternative evolutionary optimization procedures is important because the effectiveness of the entire hybrid framework depends on both the quality of the estimated risks and the diversity and feasibility of the resulting Pareto solutions.

The existing literature therefore points toward a clear convergence of three research streams: supply chain resilience, data-driven prediction, and multi-objective optimization. Yet these streams are still frequently examined separately. Predictive models may identify risk without prescribing network-level responses, whereas optimization models may prescribe configurations without deriving their risk coefficients from observed transaction-level behavior. Bridging this separation can strengthen the empirical realism and managerial usefulness of resilience modeling. It can also provide decision-makers with transparent strategic alternatives, such as minimum-cost, minimum-risk, and balanced configurations, rather than imposing a single solution irrespective of organizational priorities. In addition, evaluating these solutions separately across industries can reveal how demand intensity and logistics characteristics alter the cost-risk frontier and thus contribute to more context-sensitive supply chain planning.

Accordingly, the aim of this study is to develop a hybrid predictive-prescriptive model for resilient supply chain design by integrating machine learning-based delivery-risk prediction with multi-objective evolutionary optimization and to evaluate the resulting cost-risk configurations across the apparel, sporting goods, and electronic components industries.

2. Methods and Materials

The methodology of this study was developed by extending the standard CRISP-DM framework and was designed as a two-layer hybrid approach (predictive and

prescriptive) using a quantitative, ex post facto research design.

In the first stage, the standard DataCo database containing 180,519 logistics transactions was used. After domain-specific partitioning into three target subnetworks (apparel, sporting goods, and electronic components), data-cleaning rules were applied. These included removing post-outcome variables to completely prevent data leakage and winsorizing financial variables at the 1st and 99th percentiles. Next, unsupervised K-means clustering was used to identify five regional distribution hubs based on geographic coordinates, and Haversine great-circle distances between customers and hubs were calculated. In the predictive modeling layer, machine-learning algorithms including logistic regression, random forest, and Extreme Gradient Boosting (XGBoost) were trained using cost-sensitive techniques to address class imbalance. XGBoost ultimately achieved the highest performance, with an area under the precision–recall curve (AUPRC) of 0.8176, and was therefore selected as the core model for estimating the probability of delivery-delay risk.

In the second, prescriptive modeling layer, the probability outputs generated by the machine-learning model were introduced as risk-penalty parameters, together with logistics distances and the demand scale of each industry, into a bi-objective nonlinear optimization model. The two objectives were to minimize total network cost and to minimize aggregate risk while applying a load-concentration penalty. The problem was solved over 200 generations using the Non-dominated Sorting Genetic Algorithm II (NSGA-II) and Multi-Objective Particle Swarm Optimization (MOPSO). Because NSGA-II generated a more diverse Pareto front, three load-allocation strategies—minimum cost, minimum risk, and balanced—were extracted from its solutions.

Mathematical Model of the Study

The supply-chain network design problem was formulated as a hybrid multi-objective nonlinear programming model. Its fundamental distinction from conventional models is that risk parameters were not assigned subjectively or generated randomly; instead, they were derived from the probabilistic outputs of the XGBoost machine-learning model and entered the optimization layer as a risk-penalty component.

Table 1

Sets and indices of the model

Index symbol	Set	Values / members in this study
i	Candidate regional distribution hubs (identified by clustering)	$i = \{\text{Hub 1, Hub 2, Hub 3, Hub 4, Hub 5}\}$
s	Target industries in the supply chain	$s = \{\text{Apparel, Sporting goods, Electronic components}\}$
k	Customer orders recorded within each industry	Order 1 through the total number of orders in that industry

The model parameters are presented in Table 2.

Table 2

Model parameters

Parameter	Description	Unit	Derivation in this study
FC_i	Annual fixed cost of opening and activating hub i	USD	Regional warehousing and inventory-holding cost data
d_i	Mean Haversine geographic distance to hub i	km	Great-circle distance between customers and the hub
c_{tr}	Variable freight transportation cost rate	USD per km	Baseline logistics rate (equivalent to USD 500)
V_s	Demand-volume scaling coefficient for industry s	Weighting coefficient	Ratio of the number of orders in each industry to the baseline value (10,000)
$\hat{R}_i$	Probability of delivery-delay risk at hub i	0–1	Mean delay probability predicted by XGBoost
Cap_{max}	Maximum permitted load allocation to a single hub (concentration threshold)	Percent	60% of total network capacity (0.60)
θ	Minimum load required to economically justify opening a hub	Percent	5% of total network capacity (0.05)
λ	Concentration/overload penalty coefficient (resilience loss)	Penalty coefficient	0.1
ε	Very small value used to avoid division-by-zero errors	Mathematical constant	0.000001

The decision variables are presented in Table 3.

Table 3

Decision variables

Variable	Type	Description and application
x_i	Continuous (0 to 1)	Proportion or share of demand load allocated to distribution hub i.
y_i	Binary (0 or 1)	Activation status of hub i; equal to 1 if load is allocated to hub i and 0 otherwise.

Objective Functions of the Model

The model contains two simultaneous objective functions. NSGA-II is used to solve the multi-objective problem and extract the Pareto front.

Objective Function 1: Minimization of Total Logistics Network Cost (f_1)

This objective minimizes the sum of the fixed warehousing costs of active hubs and the variable transportation costs, which are functions of distance and industry order volume.

Minimize total cost (f_1): $f_1 = \sum_i (FC_i \times y_i) + \sum_i (x_i \times d_i \times c_{tr} \times V_s)$

The first term calculates the fixed cost of hubs that are actually opened ($y_i = 1$).

The second term represents freight transportation cost, obtained by multiplying the load share assigned to hub i (x_i) by the distance to customers (d_i), the freight rate (c_{tr}), and the industry volume scale (V_s).

Objective Function 2: Minimization of Delivery-Delay Risk and Enhancement of Network Resilience (f_2)

This objective minimizes the mean probability of delivery delay while simultaneously penalizing excessive load concentration on any single hub.

$$f_2 = \sum_i (x_i \hat{R}_i) + \varepsilon + \lambda \sum_i \max(0, x_i - Cap_{max})$$

In this formulation, $Cap_{max} = 0.60$ and $\lambda = 0.10$.

Mathematical Constraints

Constraint 1: Full demand coverage

$$\sum_{i=1}^5 x_i = 1$$

Constraint 2: Hub activation/opening condition

The activation variable y_i equals 1 when hub i receives an economically meaningful allocation; the fixed cost is included in the first objective function only for active hubs.

Constraint 3: Feasible ranges of the decision variables

$$0 \leq x_i \leq 1; y_i \in \{0, 1\}$$

Linking Machine Learning to Mathematical Optimization (Hybrid Mechanism)

The parameter $\hat{R}_i$ is obtained directly from the XGBoost model as follows:

$\hat{R}_i$ = mean predicted delay probability generated by XGBoost for all orders assigned to hub i .

Mathematically: $\hat{R}_i = (1/N_i) \times \sum_k P(\text{delay of order } k \mid \text{its logistics features})$

Predictive-model inputs: Haversine distance, shipping mode, region, committed delivery days, and winsorized financial variables.

Output: a probability between 0 and 1 for each hub, entered directly into the risk objective function.

3. Findings and Results

After examining the initial distributions, the next stage of data preparation involved filtering irrelevant records and applying data-cleaning rules to prevent modeling errors, including data leakage and the influence of extreme values. Given the broad range of products in the DataCo dataset, the distribution network was divided into three specialized subnetworks.

The Category Name variable was used as the filtering basis, and unrelated categories (e.g., books, gardening, music, and pet products) were excluded from the problem space. After this filtering process, 175,327 of the original 180,521 records (97.1% of the data) remained within the study framework, indicating that the three selected industries represented a very large share of the dataset:

Sporting goods industry: 122,095 orders (largest subnetwork).

Apparel industry: 47,772 orders.

Electronics industry: 5,459 orders.

This partitioning enabled the evolutionary algorithms in the subsequent phase to optimize hub structure and logistics distances in accordance with the demand characteristics of each industry—high-volume demand in sporting goods and low-volume demand in electronics. To support the external validity of the models, two mathematical and structural operations were applied to the filtered data.

a) Prevention of data leakage: In accordance with the methodological rules, post-event variables derived from the

outcome, including actual shipping days and delivery status, were completely removed from the feature matrix (X). Retaining these variables during training could have produced an artificially inflated accuracy approaching 100%, because they encode information about the actual delivery outcome.

b) Winsorization of financial variables: To prevent distortion of gradient-sensitive algorithms such as logistic regression, sales and profit per order were winsorized at the 1st and 99th percentiles. Rather than deleting observations and losing potentially valuable information, this technique limits extreme observations to the nearest acceptable boundary while preserving the heavy-tailed distributional structure common in supply-chain data.

Customer Clustering Using K-Means

To identify distribution centers capable of covering customer demand at minimum distance, K-means clustering was applied to customer geographic coordinates (latitude and longitude). Assuming five macro-regional hubs, the algorithm was executed and the cluster centroids were treated as distribution hubs. The coordinates obtained for the five hubs were:

Hub 0: latitude 34.03, longitude -97.23.

Hub 1: latitude 18.30, longitude -66.24.

Hub 2: latitude 35.32, longitude -119.44.

Hub 3: latitude 38.79, longitude -79.61.

Hub 4: latitude 29.10, longitude 86.11.

Calculation of Exact Logistics Distance

After identifying hub coordinates, the direct distance between each customer and the hub assigned to that customer was calculated in kilometers. Because the Earth is spherical, conventional Euclidean distance can introduce substantial error over long distances; therefore, the Haversine formula was used. The resulting Haversine-distance variable served not only as one of the most important inputs to the machine-learning model for predicting delay probability, but also as the principal distance component in the cost function used during evolutionary optimization. Descriptive analysis showed that the mean logistics distance in the distribution network was 364.8 km, with a minimum of 8.3 km and a maximum shipping route of 9,973.1 km.

Evaluation of Predictive Models (Machine Learning)

This section directly addresses the second sub-objective and the first study hypothesis: predicting disruption and delivery-delay risk using machine learning. The input data consisted of a logistics feature matrix including Haversine distance, shipping mode, order region, scheduling variables,

and winsorized financial variables, together with the target label for delivery-delay risk. Three models—logistic regression, random forest, and XGBoost—were trained and compared using cost-sensitive weighting to address class imbalance and were evaluated primarily using AUPRC. The best-performing model was selected, and its probabilistic outputs were passed directly to the second objective function of the evolutionary optimization model as risk-penalty parameters. The dataset was split into training (80%) and test (20%) subsets, and cost-sensitive weighting was

implemented within the algorithms to control class imbalance.

Comparison of Machine-Learning Algorithm Performance

Table 4 presents the performance of the predictive models on the test data. Because the outcome variable was imbalanced, emphasis was placed on AUPRC (area under the precision–recall curve) and F1 score, as overall accuracy can be misleading under class imbalance.

Table 4

Performance of machine-learning predictive models on the test data

Algorithm	Overall accuracy	F1 score (risk)	Recall (risk)	AUPRC (risk)
Logistic regression	0.6962	0.6612	0.5410	0.8014
Random forest	0.6964	0.6617	0.5418	0.8089
XGBoost	0.6979	0.6641	0.5450	0.8176

As shown in Table 4, XGBoost achieved an AUPRC of 0.8176, outperforming logistic regression (0.8014) and random forest (0.8089). Although the differences were relatively small, XGBoost was selected because of its stronger capacity to model nonlinear relationships and

handle class imbalance. Its probabilistic outputs were therefore used as risk parameters in the optimization layer.

Precision–Recall Curve Analysis

For imbalanced data, the precision–recall curve provides an informative visualization of model performance.

Figure 1

Comparison of precision–recall curves across algorithms

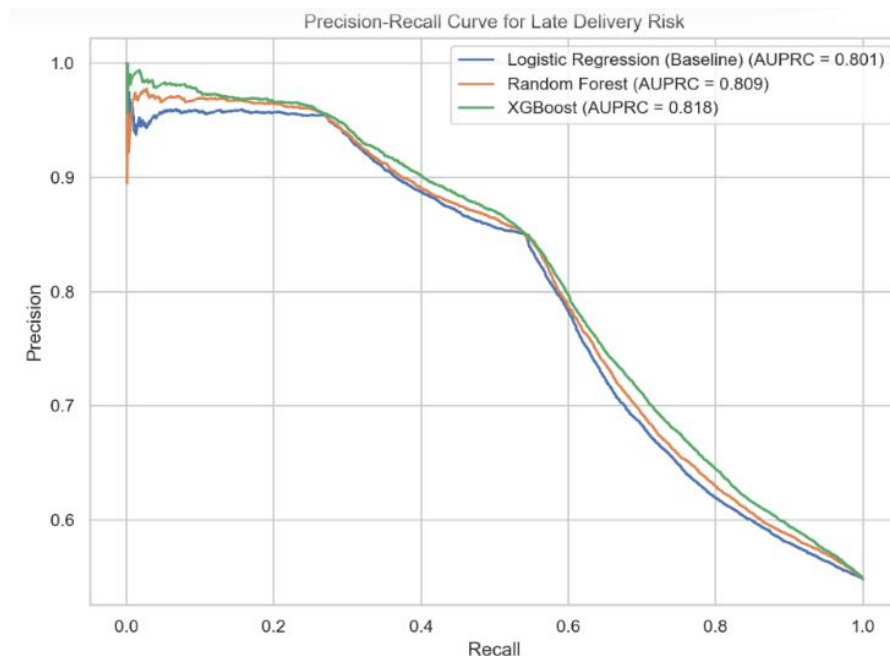


Figure 1 shows that the XGBoost curve (AUPRC $\approx$ 0.817) lies above the other model curves over most of the range. This indicates that the model experiences the smallest reduction in precision as recall increases. Accordingly, the

probabilistic outputs of XGBoost were treated as the reference estimates of delivery-delay risk for the next phase, namely network modeling using evolutionary algorithms.

Prescriptive Network Modeling and Pareto-Front Extraction Using an Evolutionary Algorithm (NSGA-II)

This section addresses the third sub-objective and the principal research question: integrating machine-learning predictions with evolutionary algorithms to design an optimized network. Inputs included the mean Haversine distances of the hubs, fixed hub-opening costs, and delay-risk probabilities produced by XGBoost in the previous phase. The procedure comprised formulation of a bi-objective problem (cost minimization and risk minimization with an overload penalty), comparison of NSGA-II and

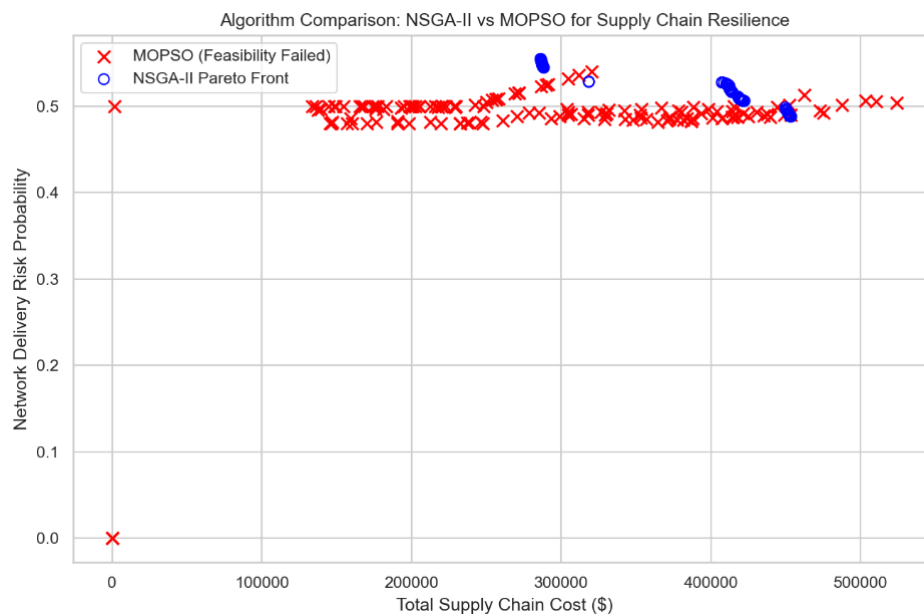
MOPSO, and extraction of Pareto solutions over 200 evolutionary generations. The resulting macro-level load-allocation strategies were subsequently intended for validation under disruption conditions in a simulation phase.

Comparison of Evolutionary Algorithms (NSGA-II vs. MOPSO)

The logistics-network optimization problem was modeled and solved using both NSGA-II and MOPSO. The algorithms were run under identical conditions—150 chromosomes/particles over 200 generations—to ensure a fair comparison.

Figure 2

Comparison of NSGA-II and MOPSO Pareto fronts in supply-chain network design



Balanced performance of NSGA-II: The NSGA-II solutions produced a Pareto front with broad diversity because the algorithm uses a crowding-distance operator. Its 150 solutions extended across the decision spectrum, from the lowest-cost/highest-risk configuration to the highest-cost/lowest-risk configuration.

One-dimensional trapping in MOPSO: By contrast, MOPSO explored a wide cost range—from approximately USD 150,000 to USD 500,000—but its population formed a dense horizontal band. This pattern suggests premature convergence in the second objective (risk minimization) and entrapment in a local optimum; consequently, MOPSO failed to identify some of the low-risk solutions found by NSGA-II.

Based on this empirical output, NSGA-II was selected as the basis for extracting the final network-design strategies. Whereas machine-learning algorithms provide probabilistic

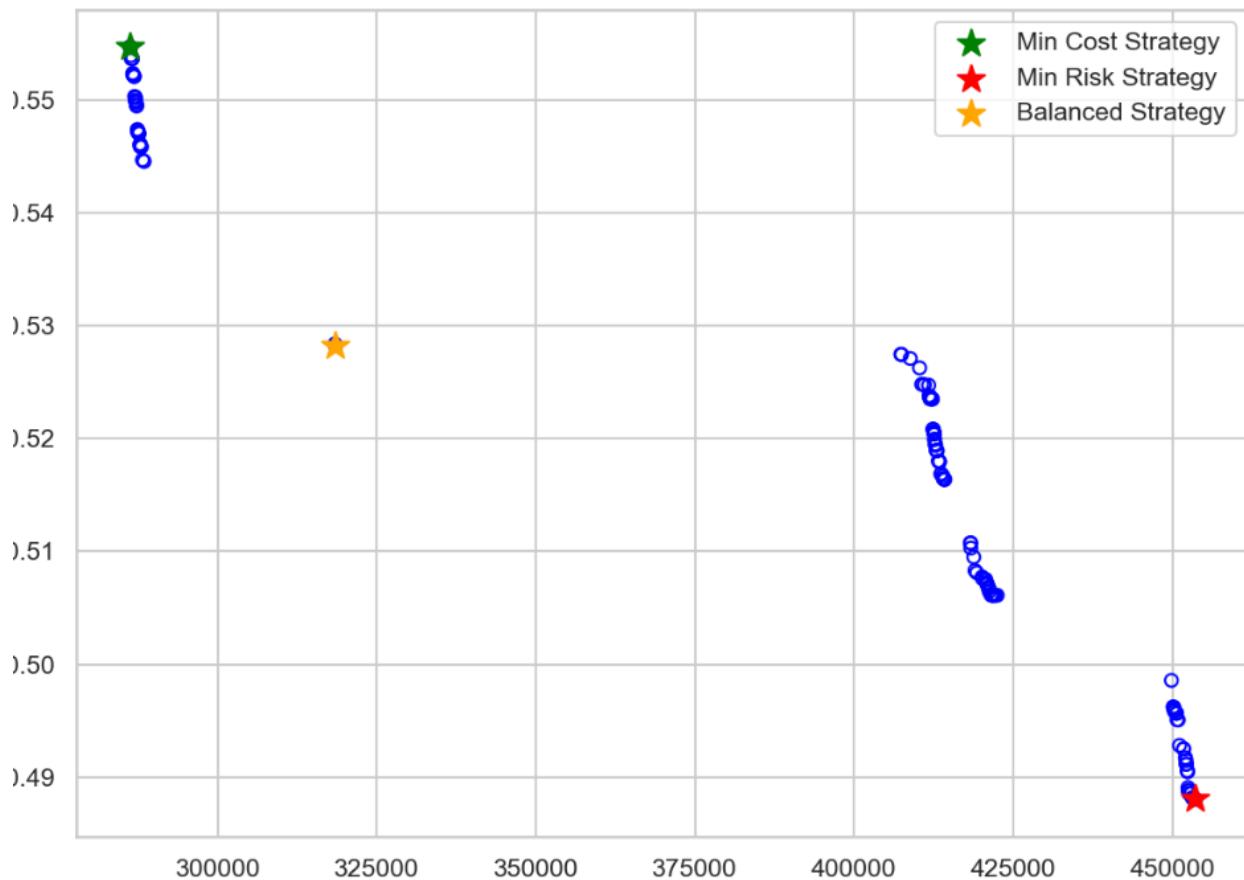
predictions, designing a resilient supply-chain network requires prescriptive decision-making under uncertainty. In this phase, the objective was to identify an optimal logistics-network structure by allocating demand across five regional hubs. The multi-objective optimization problem simultaneously minimized: (1) total network cost, including fixed hub-maintenance costs and variable transportation costs that are linear functions of Haversine distance; and (2) mean delivery-delay probability derived from XGBoost, with an additional penalty for excessive loading of a single hub. NSGA-II was run with an initial population of 150 chromosomes for 200 generations.

Pareto-Front Analysis

NSGA-II successfully extracted a set of non-dominated solutions representing the trade-off between lower operating costs and greater resilience (i.e., lower delivery-delay risk).

Figure 3

Pareto front extracted by NSGA-II (network cost versus network risk)



As shown in Figure 3, supply-chain managers face a classic trade-off: no single solution simultaneously provides both the minimum cost and the minimum risk. To

operationalize the results, three macro-level strategies were extracted from the Pareto front.

Table 5

Strategic points extracted from the Pareto front

Proposed strategy	Total network cost (USD)	Mean delay risk (probability)	Capacity-allocation weights across five hubs
Strategy A: Minimum cost	286,222	0.5547	[0.02, 0.00, 0.93, 0.02, 0.03]
Strategy B: Minimum risk	453,600	0.4881	[0.00, 0.40, 0.00, 0.60, 0.00]
Strategy C: Balanced	318,488	0.5283	[0.03, 0.84, 0.05, 0.05, 0.03]

Under the minimum-cost strategy, the maximum share assigned to a hub is constrained by the 60% cap, and hubs receiving less than 5% of the load are considered inactive so that the activation constraint is satisfied. The minimum-risk strategy distributes the load more evenly and reduces risk below 0.50, but at a higher cost. The balanced strategy represents an intermediate Pareto compromise as the preferred resilient design. These allocations are intended to ensure that no hub is loaded beyond its permitted capacity

and that fixed costs are incurred only for hubs that are actually active.

Industry-Specific Analysis of the Target Sectors (Sporting Goods, Electronics, and Apparel)

According to the methodological framework, supply chains in different industries may respond differently to shocks and delays because of differences in demand volume, product weight, and product characteristics. Accordingly, the performance of the XGBoost model and the balanced

strategy extracted from NSGA-II were evaluated separately for each of the three target industries.

Table 6

Comparison of predictive and optimized network-design indicators by industry

Industry	Order volume	Predictive accuracy (AUPRC)	F1	Network cost (USD)	Network risk
Apparel	47,772	0.8001	0.6646	638,404	0.5489
Sporting goods	122,095	0.8095	0.6614	1,050,541	0.5604
Electronics	5,459	0.7870	0.6581	158,462	0.5216

A comparative analysis across industries produced the following findings:

Sporting goods: Because this industry had the largest order volume (more than 122,000 orders), it achieved the highest predictive performance (AUPRC = 0.8095). However, the large demand volume resulted in the highest optimized logistics cost (more than USD 1 million) and the highest delay risk (0.5604) among the three industries.

Electronics: This industry was characterized by low volume but high value, with only 5,459 orders. NSGA-II designed its logistics network at the lowest cost (USD 158,462) and with the lowest risk (0.5216), suggesting a more agile distribution network than those of apparel and sporting goods.

Apparel: In terms of volume, cost, and predictive performance, the apparel industry occupied an intermediate and comparatively balanced position between the other two industries.

The study ultimately specified the network-design model for each industry's subject to the corresponding demand-allocation, activation, and capacity constraints. The final constraint shown in the source manuscript is:

$$\sum_i x_{i,s} \geq 0.60$$

Additional variable-domain and hub-activation conditions are those defined above: $0 \leq x_{i,s} \leq 1$ and $y_{i,s} \in \{0,1\}$, with the activation threshold applied when $x_{i,s} \geq 0.05$.

4. Discussion and Conclusion

The present study developed and evaluated a hybrid predictive–prescriptive framework for resilient supply chain design by integrating machine learning with multi-objective evolutionary optimization. The findings demonstrate that the proposed architecture can connect transaction-level prediction of delivery-delay risk with strategic network-design decisions concerning logistics cost, risk exposure, hub activation, and allocation of demand. At the predictive level, three machine-learning algorithms were compared—

logistic regression, random forest, and Extreme Gradient Boosting (XGBoost)—using metrics suitable for an imbalanced classification problem. XGBoost provided the strongest overall performance, achieving an accuracy of 0.6979, an F1-score of 0.6641, recall of 0.545, and an AUPRC of 0.8176. By comparison, logistic regression produced an AUPRC of 0.8014 and random forest an AUPRC of 0.8089. Although the differences were moderate, the consistently higher performance of XGBoost justified its selection as the predictive core of the optimization framework. These results indicate that nonlinear ensemble learning can provide a sufficiently discriminative estimate of late-delivery probability for subsequent use as an empirically derived risk parameter in resilient network optimization.

The superior predictive performance of XGBoost is consistent with the growing literature emphasizing the value of machine-learning methods for supply chain risk prediction, supplier evaluation, and resilience-oriented decision support. Kang and Bhawna showed that supervised machine-learning approaches can enhance resilience by improving supplier performance analysis and multidimensional risk profiling (Kang & Bhawna, 2025). Similarly, Prajapati demonstrated that machine learning-based predictive analytics can improve demand forecasting and strengthen the capacity of supply chains to respond proactively to uncertainty (Prajapati, 2025). The present results extend this evidence from prediction alone to an integrated decision architecture in which predicted probabilities are not treated merely as descriptive indicators but are transferred directly into the optimization layer. This is important because predictive accuracy becomes operationally meaningful only when it changes how resources, facilities, and flows are configured.

The findings are also compatible with evidence that AI and machine learning can improve the speed and quality of supply chain decision processes. Paberja et al. highlighted the influence of AI and machine learning on supply chain

decision-making, particularly through improved analysis of complex data and support for more responsive managerial decisions (Paberja et al., 2026). Nasirian likewise emphasized the potential of machine learning and artificial intelligence for supply chain optimization (Nasirian, 2024). From this perspective, the contribution of the present study lies in operationalizing this relationship: machine learning identifies the probability of disruption in the form of delivery delay, while evolutionary optimization transforms that predictive information into a structured set of network alternatives. Rather than treating AI as an independent forecasting mechanism, the study positions it as part of a broader prescriptive analytics system.

The predictive findings can additionally be interpreted through the broader Industry 4.0 and Industry 5.0 literature on intelligent and resilient supply chains. Belhadi et al. proposed an AI-based framework for strengthening manufacturing supply chain resilience, emphasizing the role of digital technologies in anticipating, responding to, and recovering from disruption (Belhadi et al., 2022). Rane et al. similarly argued that AI and machine learning can simultaneously support resilience and sustainability by enhancing visibility, prediction, flexibility, and adaptive planning within increasingly digitalized manufacturing environments (Rane et al., 2025). In the present study, the predictive layer contributes primarily to anticipation by estimating delivery-delay probabilities, whereas the optimization layer contributes to response and adaptation by identifying alternative logistics configurations. The integration of these capabilities therefore reflects a more complete interpretation of resilience than a purely predictive application of machine learning.

A second major finding concerns the performance of the evolutionary optimization algorithms. The comparative analysis showed that NSGA-II generated a more diversified and informative Pareto frontier than MOPSO. NSGA-II produced a broad range of non-dominated solutions spanning configurations from relatively low-cost, higher-risk designs to higher-cost, lower-risk designs. In contrast, the MOPSO solutions were concentrated within a narrower horizontal region, suggesting reduced diversity in the risk dimension and premature convergence. The study therefore selected NSGA-II as the basis for deriving strategic network configurations. This finding reinforces the value of evolutionary computation for resilient supply chain network design, especially when the decision problem involves conflicting objectives that cannot be adequately represented through a single optimization criterion.

This result aligns with previous research demonstrating the applicability of genetic and evolutionary algorithms to complex network-design problems. Vishnu et al. developed a reliable and resilient supply chain network design model using a genetic algorithm and showed that evolutionary search is well suited to problems involving multiple operational constraints and resilience considerations (Vishnu et al., 2020). Tordecilla et al. likewise emphasized the usefulness of simulation-optimization approaches for designing and evaluating resilient supply chain networks under uncertainty (Tordecilla et al., 2021). The current findings add to this literature by demonstrating the value of a Pareto-based evolutionary algorithm when network risk itself is derived from machine-learning outputs. This combination is particularly useful because the risk component is not assigned through arbitrary scoring but is informed by observed logistics data.

The Pareto analysis further demonstrated that no single configuration simultaneously minimized both logistics cost and delivery risk, confirming the existence of a fundamental economic-resilience trade-off. Three strategic solutions were extracted from the Pareto frontier. The minimum-cost strategy produced a total network cost of approximately \$286,222 with an average risk of 0.5547. The minimum-risk strategy reduced average risk to 0.4881 but increased total cost to approximately \$453,600. The balanced strategy resulted in an intermediate cost of approximately \$318,488 and an average risk of 0.5283. These results illustrate a central principle of resilient supply chain design: reducing vulnerability generally requires additional resources, diversification, redundancy, or redistribution of capacity, all of which may increase operating expenditure.

The observed cost-risk trade-off is conceptually consistent with the broader literature on resilience and criticality. Wietschel et al. emphasized that resilience assessment should be integrated with criticality assessment because supply networks must balance efficiency against exposure to structural vulnerabilities (Wietschel et al., 2025). Similarly, Xiao et al., in their review of cascading failures, demonstrated that supply chain disruption can propagate through interconnected systems and that excessive dependence on specific nodes or flows can magnify systemic vulnerability (Xiao et al., 2025). The present findings provide an operational illustration of this principle. A lower-cost network may be economically attractive under normal conditions, but concentration of flows can increase exposure to disruption. By contrast, a

more diversified allocation structure can reduce risk, although at a higher financial cost.

The balanced solution is therefore particularly important from a managerial perspective because it provides a compromise between purely economic and purely risk-minimizing objectives. This reflects the argument advanced by Mubarik and Maciukaite-Zviniene that technology or innovation alone is insufficient to generate resilience; technological capabilities must be combined with appropriate organizational configurations and decision structures (Mubarik & Maciukaite-Zviniene, 2026). The balanced strategy can be interpreted as such a configuration because it does not pursue maximal efficiency or maximal risk avoidance in isolation. Instead, it integrates both considerations within a feasible network architecture. The same interpretation is compatible with Wang et al., who found that intelligent development contributes to supply chain resilience through both internal capabilities and external linkages (Wang et al., 2026). In practice, resilience emerges not from the prediction model alone, but from how predictive information is converted into allocation choices.

The results of the industry-specific analysis provide another important contribution. Sporting goods represented the largest subnetwork, with 122,095 orders, and achieved the highest predictive AUPRC of 0.8095. However, it also generated the highest optimized network cost, exceeding \$1.05 million, and the highest network risk of 0.5604. Apparel occupied an intermediate position, with 47,772 orders, an AUPRC of 0.8001, a network cost of \$638,404, and a risk level of 0.5489. Electronic components formed the smallest subnetwork, with 5,459 orders, and had an AUPRC of 0.7870, but achieved the lowest network cost of \$158,462 and the lowest network risk of 0.5216. These results indicate that resilience outcomes are strongly conditioned by sector-specific demand volume and logistics structure.

The higher cost and risk observed in sporting goods can be interpreted as a consequence of scale and network load. Larger transaction volumes impose greater pressure on transportation and distribution capacity, potentially increasing congestion, concentration, and exposure to late delivery. This observation is compatible with Yao's finding that supply chain resilience is affected by multiple interacting structural factors rather than by a single isolated determinant (Yao, 2025). It also supports Akinola's argument that advanced analytics for resilience should account for interacting sources of uncertainty, including demand shocks and external disruptions (Akinola, 2026). Accordingly, a high-volume industry may achieve strong

predictive performance because more observations are available for model learning, while simultaneously experiencing greater operational vulnerability because of the scale and complexity of its logistics network.

The relatively low cost and risk found for electronic components should also be interpreted cautiously. Although the network appeared more agile in the optimization results, lower transaction volume may partially explain this pattern. Resilience should not therefore be equated simply with lower estimated cost or risk. Sector-specific characteristics such as value density, product criticality, dependence on specialized suppliers, and sensitivity to disruption may modify the practical consequences of failure. The importance of sectoral context is illustrated by Golden et al., who showed that resilience challenges in construction supply chains are shaped by sector-specific dependencies and disruption patterns (Golden et al., 2026), and by Hasmees et al., who demonstrated that healthcare and ICU supply chains require resilience strategies that simultaneously protect efficiency, continuity, and patient safety (Hasmees et al., 2026). These studies reinforce the need to interpret network metrics in relation to the operational environment in which the supply chain functions.

The present study also supports recent work emphasizing intelligent logistics as an integrated capability rather than a collection of isolated digital tools. Li and Ndiaye described the enhancement of logistics resilience as dependent on coordinated improvements in information processing, decision support, and adaptive logistical capabilities (Li & Ndiaye, 2026). In the current model, clustering establishes a geographically meaningful hub structure, machine learning estimates operational risk, and NSGA-II determines cost-risk-efficient demand allocations. This layered architecture illustrates how digital techniques can be combined sequentially to support resilience. The value of the model therefore lies not only in the performance of XGBoost or NSGA-II independently, but also in the linkage between data preparation, predictive modeling, and prescriptive optimization.

Overall, the findings indicate that resilient supply chain design benefits from integrating predictive intelligence with multi-objective decision-making. The stronger performance of XGBoost supports the use of nonlinear machine-learning models for late-delivery risk estimation, while the superior Pareto diversity obtained from NSGA-II confirms the suitability of evolutionary algorithms for revealing alternative cost-risk configurations. The industry-level findings further demonstrate that the optimal resilience

strategy is not uniform across sectors but depends on transaction volume, logistics characteristics, and network structure. Taken together, these findings extend the existing literature by showing how machine-learning-derived probabilities can become operational risk coefficients within a multi-objective optimization framework, thereby creating a direct connection between observed logistics data and resilient network design.

Several limitations should be considered when interpreting the findings. First, the analysis was based on a secondary transactional dataset, and the available variables were therefore constrained by the original data structure rather than being designed specifically for resilience research. Second, delivery delay was used as the principal operational representation of risk, whereas supply chain resilience is multidimensional and may also involve supplier failure, inventory shortage, transportation interruption, cyber disruption, infrastructure breakdown, recovery time, and financial exposure. Third, the network was represented through five geographically derived hubs, which simplified the underlying logistics structure and may not fully reflect actual facility locations, contractual restrictions, or capacity heterogeneity. Fourth, the analysis was retrospective and did not include real-time adaptation following a disruption. Finally, although three industries were analyzed separately, the dataset does not necessarily capture all sector-specific characteristics, such as product criticality, perishability, substitution possibilities, supplier concentration, or strategic inventory policies.

Future studies should extend the proposed framework by incorporating dynamic simulation and explicit disruption scenarios to examine how optimized networks perform during and after actual shocks. Researchers could introduce stochastic variations in demand, transportation capacity, lead times, facility availability, and supplier reliability and then evaluate recovery trajectories under alternative network configurations. Additional predictive approaches, including deep learning, causal machine learning, survival models, and calibrated ensemble methods, could also be compared with XGBoost. Future research should further investigate multi-period optimization in which capacity and allocation decisions can change over time rather than remaining static. Incorporating additional resilience indicators such as recovery time, service-level loss, unmet demand, redundancy, supplier diversification, and disruption propagation would provide a more comprehensive representation of network resilience. External validation with datasets from different countries, industries, and

logistics systems would also help determine the generalizability of the proposed framework.

Supply chain managers should integrate predictive risk analytics directly into network planning rather than using machine-learning outputs only for reporting or monitoring. Organizations can use predicted delivery-risk probabilities to prioritize vulnerable flows, reconsider hub allocations, and evaluate whether current cost-minimizing structures create excessive concentration. Decision-makers should also avoid treating the lowest-cost solution as automatically preferable; instead, they should compare alternative Pareto-efficient configurations and select a cost-risk balance consistent with service requirements and strategic priorities. High-volume industries may require greater diversification, contingency capacity, and closer monitoring of overloaded distribution nodes. Firms should periodically retrain predictive models as new transactions become available and rerun the optimization process when demand patterns, transportation costs, or network conditions change. In this way, resilience planning can become a continuously updated decision process rather than a one-time network-design exercise.

Authors' Contributions

Authors contributed equally to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethics Considerations

In this research, ethical standards including obtaining informed consent, ensuring privacy and confidentiality were considered.

References

- Akinola, O. (2026). Causal Machine Learning and Advanced Data Analytics for Supply Chain Resilience Modeling under Geopolitical, Climate, and Demand Shocks. *International Journal of Science and Engineering Applications*, 15(6), 1012-1025. <https://doi.org/10.7753/IJSEA1406.1012>
- Belhadi, A., Kamble, S., Jabbour, C. J. C., Mani, V., Khan, S. A. R., & Touriki, F. E. (2022). Manufacturing Supply Chain Resilience in the Industry 4.0 Era: An Artificial Intelligence-Based Framework. *International Journal of Production Research*, 60(14), 4385-4405. <https://doi.org/10.1080/00207543.2021.1950935>
- Golden, J., Mahdavishtarif, M., Castelblanco, G., Mangano, G., & Ferrari, A. (2026). Disruptions and Resilience in Construction Supply Chains. *Automation in Construction*, 181, 106577. <https://doi.org/10.1016/j.autcon.2025.106577>
- Hasmee, N., Joshi, H., Gurung, M., Singh, B., & Naseema, S. (2026). Optimising ICU Supply Chains: Evidence-Based Strategies for Efficiency, Resilience and Patient Safety—a Structured Narrative Review. *BMJ Open Quality*, 15(2), e003422. <https://doi.org/10.1136/bmjopen-2025-003422>
- Kang, P., & Bhawna, B. (2025). Enhancing Supply Chain Resilience through Supervised Machine Learning: Supplier Performance Analysis and Risk Profiling for a Multi-Class Classification Problem. *Business Process Management Journal*, 31(2), 345-368. <https://doi.org/10.1108/BPMJ-03-2024-0174>
- Li, Y., & Ndiaye, A. (2026). Enhancing Resilience in Logistics. *International Journal of Information Systems and Supply Chain Management*, 19(1), 1-21. <https://doi.org/10.4018/ijisscm.401244>
- Mubarik, M., & Maciukaite-Zviniene, S. (2026). Technology or innovation alone is not enough: A SEM-fsQCA based study on intelligent automation driven supply chain resilience. *Journal of Innovation & Knowledge*, 13, 100930.
- Nasirian, M. (2024). Supply Chain Optimization Using Machine Learning and Artificial Intelligence. Smart Governance Conference (IGC), Tehran.
- Paberja, M., Goudarzi, M., Nouri, M., & Monavarian, A. (2026). The Impact of Artificial Intelligence and Machine Learning on Supply Chain Decision-Making Processes. 28th National Conference on Economics, Management and Accounting, Shirvan.
- Prajapati, V. (2025). Enhancing Supply Chain Resilience through Machine Learning-Based Predictive Analytics for Demand Forecasting. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 11(2), 285-294. <https://doi.org/10.32628/CSEIT25112857>
- Rane, J., Chaudhari, R., & Rane, N. (2025). *Enhancing Sustainable Supply Chain Resilience through Artificial Intelligence and Machine Learning: Industry 4.0 and Industry 5.0 in Manufacturing*. Academic Source Publishing. <https://doi.org/10.70593/978-93-7185-621-8>
- Tordecilla, R. D., Juan, A. A., Montoya-Torres, J. R., Quintero-Araujo, C. L., & Panadero, J. (2021). Simulation-Optimization Methods for Designing and Evaluating Resilient Supply Chain Networks under Uncertainty Scenarios. *Simulation Modelling Practice and Theory*, 106, 102166. <https://doi.org/10.1016/j.simpat.2020.102166>
- Vishnu, C. R., Sridharan, R., & Kumar, P. R. (2020). Supply Chain Risk Management: Development of a Reliable and Resilient Network Design Model Using Genetic Algorithm. *International Journal of Logistics Systems and Management*, 37(1), 112-138.
- Wang, S., Chen, X., & Han, N. (2026). Intelligent Development of Manufacturing Enterprises and Supply Chain Resilience: A Perspective Based on Internal Capabilities and External Linkages. *Expert Systems with Applications*, 310, 131338. <https://doi.org/10.1016/j.eswa.2026.131338>
- Wietschel, L., Helbig, C., Hillenbrand, M., & Thorenz, A. (2025). Integrated conceptual framework for resilience and criticality assessments for raw material supply chains. *Resources, Conservation and Recycling*, 218, 108249-108249. <https://doi.org/10.1016/j.resconrec.2025.108249>
- Xiao, X., Zhu, J., Regan, D., & Wen, Z.-m. (2025). Exploring Cascading Failures in Supply Chain Risk Management: A Systematic Review, 2013-2024. *Journal of Safety Science and Resilience*, 100234. <https://doi.org/10.1016/j.jnlssr.2025.100234>
- Yao, J. (2025). Analysis of the Factors Influencing Grain Supply Chain Resilience in China Using Bayesian Structural Equation Modeling. *Sustainability*, 17(7), 3250. <https://doi.org/10.3390/su17073250>