Supply Chain and Operations Decision Making

www.scodm.reapress.com

Sup. Chain Oper. Decis. Mak. Vol. 3, No. 1 (2026) 26–45.

Paper Type: Original Article

Sustainable Design of a Closed-Loop Automotive Spare Parts Supply Chain Using Machine Learning–Based Product Return Rate Forecasting

Abbas Foroozanfar[1,*], Amirhossein Amou Jafari[2], Mahsa Mehrabi[3]

[1] Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran; forouzanfar.abbas@ie.sharif.edu; jafaryamir558@gmail.com; Mehrabi.mahsa@ie.sharif.edu.

Citation:

Received: 29 July 2025 *Revised: 01 September 2025* *Accepted: 16 November 2025*	Foroozanfar, A., Amou Jafari, A., & Mehrabi, M. (2026). Sustainable design of a closed-loop automotive spare parts supply chain using machine learning–based product return rate forecasting: *supply chain and operations decision making, 3*(1), 26-45.

Abstract

Faced with mounting ecological issues alongside tighter financial constraints, businesses are turning toward circular supply models that handle goods moving both ways - outward to customers, backward from returns. One of the key challenges in designing such supply chains is accurately forecasting product return rates, as it plays a critical role in location, operational, and capacity planning decisions. In this study, a two-stage framework is proposed based on real-world data obtained from a company operating in the automotive spare parts industry, with a specific focus on brake pads as the target product. In the first stage, product return rates are predicted using three machine learning algorithms—Ridge Regression, Random Forest, and Support Vector Regression—based on six key customer- and order-related features. The model evaluation results indicate that the Random Forest algorithm outperforms the other two models, achieving lower Mean Absolute Error (MAE) and a higher coefficient of determination (R^2), thereby providing more accurate predictions. In the second stage, the output of the forecasting model is incorporated as input into a weighted multi-objective mathematical programming model that simultaneously considers economic, environmental, and social objectives for the design of a Closed-Loop Supply Chain (CLSC) network, encompassing eight facility types. Solving the model with real-world data demonstrates that more accurate return-rate forecasts lead to a 16.5% reduction in operational costs and a 33.7% decrease in environmental impacts compared to models with lower predictive accuracy. These results highlight that integrating machine learning techniques with mathematical optimization can significantly enhance the effectiveness of strategic decision-making in CLSC design.

Keywords: Closed-loop supply chain, Random forest, Ridge regression, Support vector regression, Product return rate.

1 | Introduction

In recent decades, growing environmental concerns and the need for more efficient resource use have motivated firms to design and implement Closed-Loop Supply Chains (CLSCs). Unlike traditional supply

Corresponding Author: forouzanfar.abbas@ie.sharif.edu

https://doi.org/10.48313/scodm.v3i1.50

Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

chains that focus solely on forward flows—such as procurement, production, and customer distribution—CLSCs explicitly incorporate reverse flows. These reverse processes include the collection, return, repair, remanufacturing, or recycling of used, defective, or end-of-life products [1].

The effective design of a CLSC requires careful consideration of several critical factors, among them the product return rate, which plays a pivotal role. The return rate directly influences capacity planning, resource allocation, and logistics-related decisions. Inaccurate estimation of this parameter can lead to misleading modeling outcomes and suboptimal strategic decisions [2]. Consequently, reliable prediction of product returns is a fundamental prerequisite for robust CLSC network design.

Product returns occur for various reasons, including product malfunction, mismatch with customer expectations, warranty claims, or customer dissatisfaction. These returns significantly increase operational costs and reduce supply chain efficiency. Therefore, designing and implementing a CLSC capable of efficiently managing recovery, remanufacturing, and redistribution processes has become an unavoidable necessity for modern industries [3]. However, due to the inherent complexities and uncertainties associated with consumer behavior and product life cycles, traditional statistical forecasting approaches often fail to provide sufficiently accurate predictions of return rates.

Today, managers' viewpoints on supply chain management have changed, and the use of information technology in this field is becoming more widespread [4]. In recent years, the emergence of Machine Learning (ML) techniques, along with their strong capability to analyze large-scale, complex datasets and uncover hidden patterns, has created new opportunities to address this challenge. By leveraging historical data on sales, returns, product attributes, and customer characteristics, ML models can predict product return rates with greater accuracy and robustness than conventional methods [5].

In this study, a CLSC model is developed for the automotive spare parts industry, a particularly suitable context for such an approach given the nature of its products and the relatively high return rates of defective or remanufacturable components. Within the proposed framework, the spare parts return rate is predicted using historical data and ML algorithms and subsequently incorporated as a key input parameter into the supply chain optimization model.

This research proposes an integrated two-stage approach that combines ML-based predictive capabilities with strategic decision-making in CLSC management. The primary objective of this study is to evaluate the effectiveness of various ML models for forecasting customer return rates and to provide actionable insights to improve supply chain performance from economic, environmental, and social sustainability perspectives. The overall structure of the proposed approach is illustrated in *Fig. 1*.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the problem definition and model formulation. Section 4 reports the computational results and validation of the proposed approach. Finally, Section 5 concludes the paper and outlines directions for future research.

Based on the objectives and scope of this study, the following Research Questions (RQ) are formulated:

RQ1: How accurately can different ML algorithms predict product return rates in a CLSC for the automotive spare parts industry?

RQ2: What is the impact of incorporating ML-based return rate predictions on strategic decisions and the overall performance of CLSC network design?

RQ3: How does the proposed integrated approach contribute to improving economic efficiency, environmental sustainability, and social performance within CLSCs?

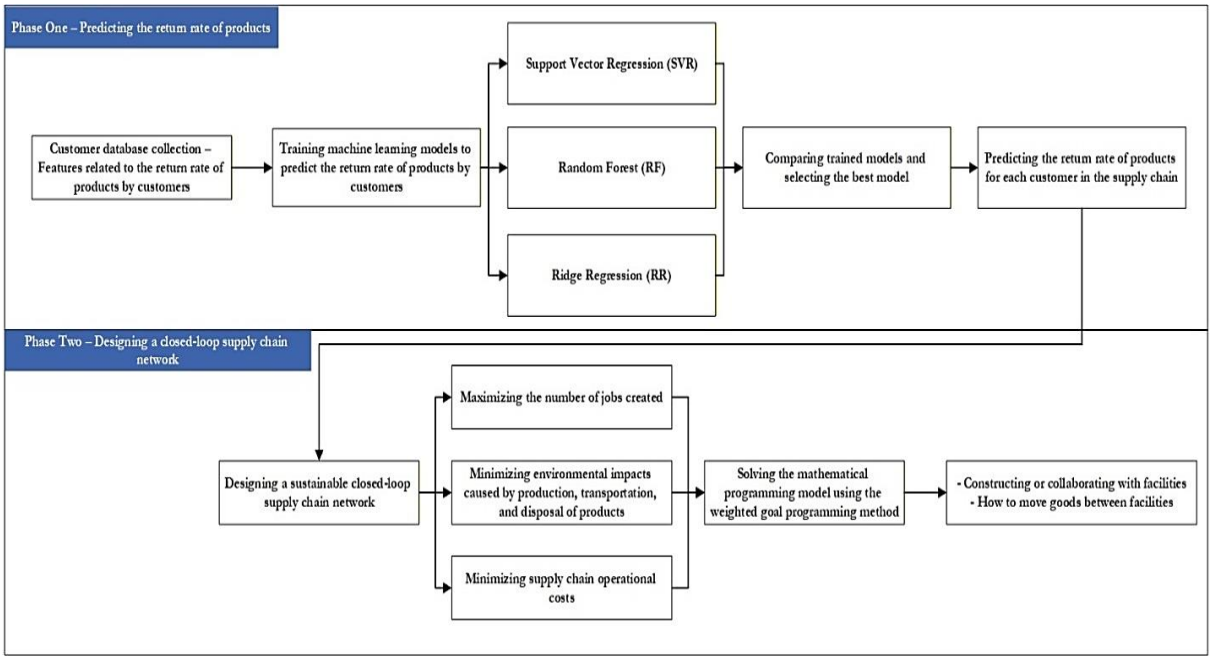

Fig. 1. Graphical abstract of the methodology.

2 | Literature Review

2.1 | Learning-Based Return Forecasting in Closed-Loop Supply Chains

In CLSCs, accurately forecasting product return rates is fundamental to designing and operating an efficient reverse flow, because return volumes, timing, and condition directly shape collection capacity, remanufacturing plans, inventory policies, and transportation decisions. Early work in this area primarily relied on statistical forecasting to characterize return behavior and reduce uncertainty. In this stream, Carrasco-Gallego et al. [6], working in the Liquefied Petroleum Gas (LPG) context, introduced dynamic regression models to forecast the return of reusable cylinders in Spain by explicitly linking sales patterns to subsequent return flows, thereby translating market demand signals into reverse logistics expectations. Building on the need to represent uncertainty and heterogeneity in return processes better, Krapp et al. [7] adopted Bayesian estimation for forecasting product returns in CLSCs, showing that probabilistic inference can offer higher flexibility than conventional approaches and improve decisions—particularly in inventory management—where planners benefit from uncertainty-aware predictions; their evidence also suggested practical value for recycling and return-intensive industries in reducing operational variability.

As researchers began to recognize that return forecasting often requires going beyond a single dimension, attention shifted toward integrated predictions that jointly capture multiple return attributes. From this perspective, Liang et al. [8] proposed a method that simultaneously forecasts the quantity, quality, and timing of returns, focusing on electric vehicle battery returns and using information on sales, usage patterns, customer behavior, and product lifetime to improve predictive performance; importantly, they validated their approach via Monte Carlo simulation, reinforcing its applicability under realistic conditions. In parallel, other studies examined the planning implications of returns, particularly in remanufacturing systems where production decisions depend heavily on the availability of returned cores. Matsumoto and Komatsu [9] addressed this challenge by forecasting demand for returned products in remanufacturing, deliberately moving away from strong assumptions about new-product sales timing and fixed lifetimes; instead, they employed time-series techniques—Holt–Winters and Autoregressive Integrated Moving Average (ARIMA)—to learn trends and seasonality directly from historical signals, demonstrating improved forecasts using a long-horizon dataset of 12 years of sales across 160 types of remanufactured alternators and starters.

To better capture the process structure underlying returns and recovery, another line of research modeled remanufacturing operations as systems whose stochastic behavior can be characterized analytically. In this direction, Zhou et al. [10] developed a model that jointly predicts return amount, timing, and probability, while also estimating recoverable parts/materials and waste generation; they converted the operational remanufacturing process into a stochastic network and used the Graphical Evaluation and Review Technique (GERT) together with Mason's rule to derive an equivalent transfer function, aiming to increase forecast fidelity by embedding process logic into the prediction mechanism. Closely related, yet focused on the dependence between sales/demand signals and delayed returns, Geda and Kwong [11] tackled forecasting used-product returns to meet remanufacturing demand by employing Distributed Lag Models (DLMs) with a normal–negative binomial delay function, and they estimated delay parameters using Bayesian Monte Carlo simulation; their evaluation using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) indicated strong predictive capability and apparent usefulness for closed-loop planning.

While many studies addressed returns at the product level, others emphasized sector-specific return streams and the macro drivers behind them. For instance, Pereira et al. [12] examined the closed-loop tire supply chain by modeling the relationship between tires introduced into the market and tires returned for recycling, using a transfer function approach and studying its application in Brazil, where aftermarket tire volumes and new vehicle counts served as key inputs, and collected tires were treated as the output. Alongside forecasting models, inventory-focused research explored how return predictions should be operationalized into policies. Chou et al. [13] investigated periodic-review inventory models over a finite horizon with two remanufacturable part types—buyback returns and ordinary parts—treating buyback returns as a function of past demand and sales; they derived an optimal inventory policy when returns are forecast from past demand and then introduced a practical, structured policy for the case where returns are forecast from past sales, evaluating performance to ensure operational feasibility.

As methodological sophistication increased, researchers also revisited how return-forecasting parameters should be estimated—particularly in delay-based models where complexity can challenge standard estimators. Geda and Kwong [14] extended their earlier work by improving forecasts of the quantity and timing of used-product returns within the DLM framework, proposing a Bayesian inference approach coupled with Monte Carlo simulation to estimate delay-function parameters more efficiently; crucially, they showed that this method can be more robust than Maximum Likelihood Estimation (MLE) when the delay structure is complex, and their real case study reported improvements in MAPE and error variance. More recently, return forecasting has been investigated in domains where intermittency, legal constraints, and high uncertainty complicate planning. Turki et al. [15], focusing on healthcare CLSCs, addressed forecasting the supply capacity of recovered spare parts extracted from returned systems and proposed a dynamic, monthly updated forecasting procedure that combines statistical models—TSB-Croston, a 12-month moving average, ARIMA, and Seasonal ARIMA (SARIMA)—with a business-knowledge-based model; using over 1,400 intermittent time series, they compared performance against standalone methods using MAPE, MAE, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), highlighting the importance of hybrid, continuously updated forecasting in practice.

Alongside these statistical and process-based approaches, ML has increasingly been adopted to capture nonlinearities and complex interactions in return behavior, especially when rich feature data are available. Within this line, ADIGÜZEL TÜYLÜ and Eroğlu [16] studied the prediction of product return rates in reverse logistics for the textile industry, demonstrating that ML models can accurately infer return likelihood from product attributes and customer preferences, with practical implications for reducing unnecessary production and transportation costs while improving planning. In the automotive accessories context, Cui et al. [5] forecasted product return volumes using real industrial data and compared multiple ML approaches—Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, Smoothly Clipped Absolute Deviation (SCAD), Random Forest (RF), and Gradient Boosting—finding that LASSO performed best due to its ability to select essential interactions and its implementation simplicity. At the same time, overall results supported the operational and financial value of accurate return forecasts. For settings where return behavior

is driven by multiple interacting factors and fast-changing customer preferences, Tüylü and Eruglo [17] employed ensemble learning in the apparel industry's reverse supply chain, using Stacking and Vote (Voting) to improve prediction accuracy beyond traditional time-series methods; tested on data from an international apparel company, their approach supported better control of production, distribution, warehousing, and cost in reverse logistics. Finally, with the rise of deep learning, Gao [18] explored optimization strategies for CLSC management by developing a forecasting model based on a Multi-head Attention Gated Recurrent Unit (GRU) architecture for time-series supply chain data, and then using Genetic Algorithms (GA) and Simulated Annealing (SA) to optimize decisions; the proposed hybrid pipeline illustrates the direction toward combining deep predictive models with metaheuristic optimization to improve sustainability and efficiency in industrial contexts.

Taken together, this body of work shows a steady progression from classical statistical forecasting toward richer machine-learning and hybrid paradigms—ranging from Bayesian estimation and time-series methods such as Holt–Winters and ARIMA, to stochastic-network modeling via GERT and Mason's rule, and DLM-based delay modeling supported by Bayesian Monte Carlo inference. However, despite these advances, much of the literature still focuses on limited-return aspects (often quantity or timing) or treats forecasting as a standalone task. At the same time, the explicit, operationally linked integration of forecasting outputs into closed-loop network design and optimization decisions—such as inventory control and remanufacturing capacity planning—remains comparatively underdeveloped.

2.2|Network Design and Decision Models for Closed-Loop Supply Chains

Alongside forecasting studies, a significant limitation in the CLSC literature has been that predictive results were often not translated into prescriptive network and policy decisions, even though the real value of forecasting emerges only when it informs choices such as inventory control, disposition planning, and reverse-flow configuration. Early decision-oriented work began to operationalize return forecasts within policy design: Chou et al. [13] investigated periodic-review inventory models for remanufacturable parts and embedded return forecasts—modeled as functions of past demand and sales—into the structure of optimal and implementable inventory policies. More recently, research has increasingly emphasized the need to integrate data-driven insights into broader sustainability and planning decisions. Ashtab and Tosarkani [19] linked predictive/behavioral drivers to CLSC decision-making by analyzing how brand-related information affects consumer recycling awareness and return behavior and then incorporating these insights into a closed-loop optimization setting, thereby illustrating how predictive signals can be converted into design-relevant inputs.

A key challenge in return-intensive retail systems is that product returns are often sales-dependent and delayed across multiple periods, so inventory decisions based only on "net demand" may ignore when and from which past sales returns will arrive. Addressing this gap, Gökbayrak et al. [20] developed a finite-horizon inventory model in which stochastic returns arrive from multiple previous sales periods with time-varying return probabilities, and the policy explicitly tracks a return-potential state alongside inventory. They formulated the ordering problem as a dynamic program with standard operational costs and constraints. They proposed an approximate dynamic programming solution to handle the large state space induced by the multi-period return window.

At the process-integration level, Cortes et al. [21] addressed the fragmentation of AI applications in reverse logistics by framing the reverse logistics process around three macro-stages—planning, execution, and control—and mapping AI tools to the key decision points across these stages, which clarified how predictive and decision-support methods should be positioned within activities such as collection and transport, inspection and sorting, disposition, and performance monitoring. Complementing this process lens with an explicit predictive–prescriptive pipeline, Ibrahim and Abdul-kader [22] developed an integrated framework for sustainable cellphone return management in which return-rate prediction was connected to downstream disposition planning through condition-uncertainty modeling and decision-oriented categorization of returned items. Overall, the progression of this stream shows a clear shift from using forecasts in isolated

settings toward structuring and integrating return-related predictions into coherent CLSC decision models, particularly in the most recent work.

2.3 | Gap Analysis

Existing research on CLSCs has generally advanced along two streams that have remained only loosely connected in terms of data, modeling assumptions, and decision interfaces. In the first stream, return-forecasting studies have focused on improving predictive accuracy for return quantity, timing, and, sometimes, quality/condition by leveraging statistical regression, time-series approaches, Bayesian inference, and, more recently, machine-learning and deep-learning methods. However, these studies have often been formulated as self-contained forecasting exercises: they typically report error metrics and predictive improvements, but stop short of specifying how the forecasted return rates should be embedded into the subsequent decision layers of a closed-loop system. In practice, this means that forecasts are frequently produced at an aggregate level (e.g., product category, period, or market region) or treated as descriptive outputs, without an explicit mapping from predicted returns to the concrete decision variables that govern reverse logistics, such as facility utilization, collection and inspection capacity, disposition volumes, and reverse-flow routing.

In the second stream, network design and prescriptive optimization models have been widely used to determine facility locations, allocation rules, and forward/reverse flows in pursuit of economic and sustainability objectives. Yet these models have often represented product returns through simplified constructs—such as fixed return ratios, scenario assumptions, stationary distributions, or deterministic averages—chosen mainly for tractability rather than fidelity to customer behavior. Even when uncertainty is considered, it is commonly introduced through generic uncertainty sets or high-level scenarios that do not originate from observed, customer-level return mechanisms. As a result, the return process is frequently exogenous to the model: the reverse flows are "assumed" rather than generated from behavioral or transactional evidence, and they do not respond endogenously to the same signals that drive forward sales. This disconnect is non-trivial because returns in many industries are delayed, sales-dependent, and heterogeneous across customers, products, and sales channels; ignoring these properties can distort strategic decisions by misestimating how much material will arrive, when, and in what condition.

To address this structural disconnect, the present work introduces an integrated forecasting–optimization framework in which a learning-based return model is positioned as a decision-relevant input rather than a descriptive add-on. Specifically, a supervised machine-learning module is developed to estimate customer/order-level return rates from historical transactional data (by benchmarking models such as Ridge Regression (RR), Support Vector Regression (SVR), and RF). The resulting predicted return rate is then embedded as a parameter within a multi-period, multi-echelon mathematical programming model for an automotive spare-parts CLSC. In this formulation, the predicted returns directly govern reverse-flow generation and, consequently, influence facility activation decisions and forward/reverse shipment plans across suppliers, plants, distribution centers, customer zones, and recovery facilities (collection/inspection, repair, recycling, and disposal). The optimization layer is structured as a sustainability-oriented multi-objective model—cost, environmental impact, and social employment—and is operationalized via Weighted Goal Programming (WGP) to obtain a single implementable plan consistent with decision-maker priorities, thereby ensuring that the learning output is used in the precise place where strategic network decisions are made.

3 | Problem Definition

In the automotive industry, the timely, efficient, and sustainable supply of spare parts has always been a critical strategic challenge, as poor performance in this area can lead to customer dissatisfaction, reduced vehicle lifespans, increased technical waste, and ultimately damage to manufacturers' brand reputation. Given the complexity of automotive spare parts supply chains, the high-risk nature of demand, the variability of customer return rates, and the necessity of simultaneously addressing economic, environmental, and social

dimensions, designing a CLSC network that meets sustainability requirements has become unavoidable. In this context, the present study focuses on developing a comprehensive and practical model for a sustainable automotive spare parts supply chain, in which customer product return rates are forecast using ML techniques such as Ridge Linear Regression, RF, and SVR, and subsequently incorporated into the mathematical model as input data. The network under investigation comprises a variety of facilities, including manufacturing, distribution, collection, recycling, repair, and final disposal centers, within which product flows from production to consumption and from consumption to recovery or disposal are modeled in an integrated manner.

The proposed model is formulated as a multi-objective mathematical programming framework to minimize operational costs (including production, transportation, and facility establishment costs), minimize environmental impacts (such as carbon emissions), and maximize social benefits (notably job creation across different regions). To transform the multi-objective model into a solvable formulation, a WGP approach is employed to derive an optimal solution that maximizes alignment with the decision maker's priorities. Subject to constraints such as facility capacity, flow balance, optimal resource allocation, and spatial and temporal limitations, the model seeks to design a sustainable, cost-effective, and environmentally compatible network for managing the life cycle of automotive spare parts over a specified planning horizon. The overall structure of the CLSC examined in this study is illustrated in *Fig. 2*.

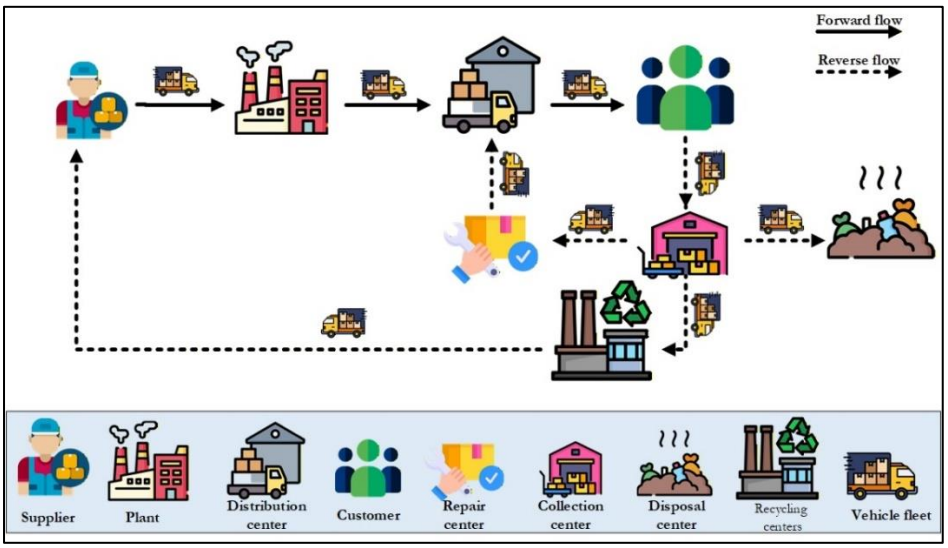

Fig. 2. Supply chain network structure.

3.1 | Machine Learning Models

This section introduces the ML models used to predict product return rates. Three models are examined, namely SVR, RF, and RR. These models are selected due to their strong capabilities in analyzing complex data patterns and providing accurate predictions, making them well-suited for modeling product return behavior. In the following subsections, the structure, key characteristics, and underlying mathematical foundations of each model are briefly described.

3.1.1 | Support vector regression

SVR is an ML–based method that follows principles similar to those of the Support Vector Machine for classification, but is explicitly designed for regression tasks. The primary objective of SVR is to identify a function whose prediction errors do not exceed a predefined threshold, denoted by ε. The general form of the SVR model is expressed in *Eq. (1)*, where $w \in \mathbb{R}^n$ represents the weight vector, $b \in \mathbb{R}^n$ denotes the bias term, and $x \in \mathbb{R}^n$ is the input feature vector. Under this formulation, it is assumed that the data in the feature space can be separated or approximated by a hyperplane.

$$f(x) = b + \langle w, x \rangle. \quad (1)$$

The optimization objective function in SVR is defined as shown in *Eq. (2)*, subject to the constraints presented in *Eqs. (3)-(5)*. In these formulations, y_i denotes the actual target value of the 3th sample. The variables ξ_i^* and ξ_i represent positive slack variables that penalize deviations exceeding the ε-insensitive margin, capturing both positive and negative deviations from the observed data, respectively. The parameter C is a tunable regularization constant that controls the trade-off between model complexity and the allowable level of prediction error.

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*). \quad (2)$$

Subjected to

$$y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i, \quad \text{for all } i \in I, \quad (3)$$
$$\langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*, \quad \text{for all } i \in I, \quad (4)$$
$$\xi_i, \xi_i^* \geq 0, \quad \text{for all } i \in I. \quad (5)$$

When the data are not linearly separable or cannot be adequately modeled by a linear function, a nonlinear mapping to a higher-dimensional feature space can be used. This argument is achieved using kernel functions. Under this formulation, the objective is to identify a function of the form shown in *Eq. (6)*, where $\phi(X)$ is a mapping function that transforms the data from the original input space into a higher-dimensional (or potentially infinite-dimensional) feature space. As a result, a nonlinear problem in the original space is converted into a linear problem in the transformed feature space.

$$f(x) = b + \langle w, \phi(X) \rangle. \quad (6)$$

The optimization objective function of SVR is defined in *Eq. (7)*, subject to the constraints presented in *Eqs. (8)-(10)*.

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*). \quad (7)$$

Subjected to

$$y_i - \langle w, \phi(x_i) \rangle - b \leq \varepsilon + \xi_i, \quad \text{for all } i \in I, \quad (8)$$
$$\langle w, \phi(x_i) \rangle + b - y_i \leq \varepsilon + \xi_i^*, \quad \text{for all } i \in I, \quad (9)$$
$$\xi_i, \xi_i^* \geq 0, \quad \text{for all } i \in I. \quad (10)$$

This optimization problem is typically solved in its dual form, yielding a prediction function expressed as in *Eq. (11)*. In this formulation, α_i^* and α_i denote the Lagrange multipliers, and $K(x_i, x)$, which can equivalently be written as $\langle \phi(x_i), \phi(x_j) \rangle$, represents the kernel function.

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b. \quad (11)$$

3.1.2 | Random forest

The RF model is an ensemble learning method composed of a collection of independent decision trees. Let the training dataset consist of n samples. $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in \mathbb{R}^n$ denotes the feature vector of the i-th sample and $y_i \in \mathbb{R}^n$ represents the corresponding target value. Each decision tree f_t is constructed using bootstrap sampling (i.e., random sampling with replacement) from the training data, along with the random selection of a subset of features at each split, to preserve diversity and independence among the trees.

Each tree partitions the feature space into several disjoint regions R_{tm}, for $m = 1, \ldots, M_t$, where M_t denotes the number of terminal nodes (leaves) of the tree t. The prediction of each tree for a new input $x$is given by

the average target value of the training samples falling into the leaf corresponding to x. The final prediction of the RF is obtained by averaging the predictions of all decision trees and is expressed as shown in *Eq. (12)*, where $f_t(x)$ denotes the prediction of the t-th decision tree for the sample x, and the total number of trees in the forest is denoted by T.

$$\hat{f}(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x). \quad \textbf{(12)}$$

3.1.3 | Ridge regression

RR is a regularized linear regression method primarily used to mitigate overfitting in regression models. This approach reduces model complexity and shrinks the regression coefficients toward zero by incorporating a penalty term into the cost function of ordinary linear regression.

Suppose that the linear regression model is expressed as shown in *Eq. (13)*, where y denotes the vector of observed response values, X represents the feature matrix (an $n \times p$ matrix in which n is the number of observations and p is the number of features), β is the vector of model coefficients, and ε denotes the random error term.

$$y = X\beta + \varepsilon. \quad \textbf{(13)}$$

In linear regression, the cost function (loss function) is given by *Eq. (14)*, which aims to minimize the sum of squared errors.

$$\min_{\beta} ||y - X\beta||^2. \quad \textbf{(14)}$$

But in RR, the cost function is given by *Eq. (15)*.

$$\min_{\beta}(||y - X\beta||^2 + \lambda||\beta||^2). \quad \textbf{(15)}$$

where $||\beta||^2$ is computed as shown in *Eq. (16)*. The parameter λ is a regularization coefficient that controls the strength of the penalty term.

$$||\beta||^2 = \sum_{j=1}^{P} \beta_j^2. \quad \textbf{(16)}$$

By differentiating *Eq. (15)*, *Eq. (17)* is obtained. In this expression, X^T denotes the transpose of the feature matrix, I is the identity matrix, and λ is a non-negative scalar. When $\lambda = 0$, the formulation reduces to ordinary linear regression. As the value of λ increases, the regression coefficients are increasingly shrunk toward zero, although they do not become exactly zero, resulting in a simpler model. This method is particularly effective in situations where multicollinearity exists among the features or when the number of features exceeds the number of observations.

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y. \quad \textbf{(17)}$$

3.2 | Mathematical Modeling

The symbols used for mathematical modeling of the problem are given in *Table 1*.

Table 1. Notation.

Symbol	Description
Sets	
$s \in \{1, \dots, S\}$	Set of suppliers.

Table 1. Continued.

Symbol	Description
Sets	
$p \in \{1, \dots, P\}$	Set of manufacturing plants.
$d \in \{1, \dots, D\}$	Set of distribution centers.
$c \in \{1, \dots, C\}$	Set of customers.
$i \in \{1, \dots, I\}$	Set of collection and inspection centers.
$k \in \{1, \dots, K\}$	Set of repair centers.
$l \in \{1, \dots, L\}$	Set of recycling centers.
$m \in \{1, \dots, M\}$	Set of disposal centers.
$t \in \{1, \dots, T\}$	Set of time periods.
Parameters	
$C\gamma_{\gamma}^{t}$	Capacity of facility $\gamma \epsilon \{s, p, d, i, m, l, k\}$ in period t.
FJ_{γ}	Number of fixed jobs created by establishing facility $\gamma \epsilon \{s, p, d, i, m, l, k\}$.
VJ_{γ}	Number of variable jobs created by establishing facility $\gamma \epsilon \{s, p, d, i, m, l, k\}$.
FC_{γ}	Fixed cost of establishing facility $\gamma \epsilon \{s, p, d, i, m, l, k\}$.
TC_{sp}	Transportation cost per unit from suppliers to the plant p.
TC_{pd}	Transportation cost per unit from plant p to distribution center d.
TC_{dc}	Transportation cost per unit from distribution center d to customer c.
TC_{ci}	Transportation cost per unit from customer c to the collection and inspection center i.
TC_{il}	Transportation cost per unit from the collection and inspection center i to the recycling center l.
TC_{im}	Transportation cost per unit from the collection and inspection center i to the disposal center m.
TC_{ik}	Transportation cost per unit from the collection and inspection center i to the repair center k.
TC_{ls}	Transportation cost per unit from the recycling center l to the supplier s.
TC_{kd}	Transportation cost per unit from repair center k to distribution center d.
EIS_{sp}	Environmental impact per unit transported from suppliers to the plant p.
EIS_{pd}	Environmental impact per unit transported from plant p to distribution center d.
EIS_{dc}	Environmental impact per unit transported from the distribution center d to customer c.
EIS_{ci}	Environmental impact per unit transported from customer c to the collection and inspection center i.
EIS_{il}	Environmental impact per unit transported from the collection and inspection center i to the recycling center l.
EIS_{im}	Environmental impact per unit transported from the collection and inspection center i to the disposal center m.
EIS_{ik}	Environmental impact per unit transported from the collection and inspection center i to the repair center k.
EIS_{ls}	Environmental impact per unit transported from the recycling center l to suppliers.
EIS_{kd}	Environmental impact per unit transported from repair center k to distribution center d.
EIP_{p}	Environmental impact of product manufacturing at plant p.
EID_{m}	Environmental impact of product disposal at the disposal center m.
MC_{p}^{t}	Production cost at plant p in period t.
DC_{m}^{t}	Disposal cost at the disposal center, m, in period t.
RC_{l}^{t}	Recycling cost at the recycling center l in period t.
BC_{k}^{t}	Repair cost at repair center k in period t.
DE_{c}^{t}	Customer demand c in period t.

Table 1. Continued.

Symbol	Description
Parameters	
RY_t	Recycling rate in period t.
RV_t	Repair rate in period t.
RD_t	Disposal rate in period t.
ρ_c	Product return rate of customer c, predicted by the ML model.
Variables	
$O\gamma_\gamma^t$	Binary variable equal to 1 if facility $\gamma\epsilon\{s, p, d, i, m, l, k\}$ is used in period t, and zero otherwise.
QS_{sp}^t	Quantity shipped from suppliers to plant p in period t.
QS_{pd}^t	Quantity shipped from plant p to distribution center d in period t.
QS_{dc}^t	Quantity shipped from distribution center d to customer c in period t.
QS_{ci}^t	Quantity shipped from customer c to collection and inspection center i in period t.
QS_{il}^t	Quantity shipped from collection and inspection center i to recycling center l in period t.
QS_{im}^t	Quantity shipped from collection and inspection center i to disposal center m in period t.
QS_{ik}^t	Quantity shipped from collection and inspection center i to repair center k in period t.
QS_{ls}^t	Quantity shipped from recycling center l to supplier s in period t.
QS_{kd}^t	Quantity shipped from repair center k to distribution center d in period t.

The following is a mathematical programming model for the supply chain network.

$$\begin{aligned}\text{Min OC} = {} & \sum_p \sum_t FC_p OP_p^t + \sum_d \sum_t FC_d OD_d^t + \sum_i \sum_t FC_i OI_i^t + \sum_m \sum_t FC_m OM_m^t \\ & + \sum_l \sum_t FC_l OL_l^t + \sum_k \sum_t^t FC_k OK_k^t + \sum_p \sum_d \sum_t MC_p^t QS_{pd}^t + \sum_i \sum_l \sum_t RC_l^t QS_{il}^t \\ & + \sum_k \sum_i \sum_t BC_k^t QS_{ik}^t + \sum_m \sum_i \sum_t DC_m^t QS_{im}^t + \sum_d \sum_p \sum_t TC_{pd} QS_{pd}^t \\ & + \sum_d \sum_c \sum_t TC_{dc} QS_{dc}^t + \sum_c \sum_i \sum_t TC_{ci} QS_{ci}^t + \sum_i \sum_m \sum_t TC_{im} QS_{im}^t \\ & + \sum_i \sum_l \sum_t TC_{il} QS_{il}^t + \sum_i \sum_k \sum_t TC_{ik} QS_{ik}^t + \sum_s \sum_d \sum_t TC_{kd} QS_{kd}^t \\ & + \sum_s \sum_p \sum_t TC_{sp} QS_{sp}^t,\end{aligned} \tag{18}$$

$$\begin{aligned}\text{Min EI} = {} & \sum_p \sum_d \sum_t EIP_p QS_{pd}^t + \sum_i \sum_m \sum_t EID_m QS_{im}^t + \sum_s \sum_p \sum_t EIS_{sp} QS_{sp}^t \\ & + \sum_p \sum_d \sum_t EIS_{pd} QS_{pd}^t + + \sum_d \sum_c \sum_t EIS_{dc} QS_{dc}^t + \sum_c \sum_i \sum_t EIS_{ci} QS_{ci}^t \\ & + \sum_i \sum_m \sum_t EIS_{im} QS_{im}^t + \sum_i \sum_l \sum_t EIS_{il} QS_{il}^t + \sum_i \sum_k \sum_t EIS_{ik} QS_{ik}^t \\ & + \sum_l \sum_s \sum_t EIS_{ls} QS_{ls}^t + \sum_k \sum_d \sum_t EIS_{kd} QS_{kd}^t,\end{aligned} \tag{19}$$

$$\begin{aligned}\text{MAX SI} = {} & \sum_p \sum_t FJ_p OP_p^t + \sum_d \sum_t FJ_d OD_d^t + \sum_i \sum_t FJ_i OI_i^t + \sum_m \sum_t FJ_m OM_m^t \\ & + \sum_l \sum_t FJ_l OL_l^t + \sum_k \sum_t^t FJ_k OK_k^t + \sum_p \sum_d \sum_t VJ_p QS_{pd}^t / CP_p^t + \sum_d \sum_c \sum_t VJ_d QS_{dc}^t / CD_d^t \\ & + \sum_c \sum_i \sum_t VJ_i QS_{ci}^t / Ci_i^t + \sum_i \sum_m \sum_t VJ_m QS_{im}^t / CM_m^t + \sum_i \sum_l \sum_t VJ_l QS_{il}^t / CL_l^t \\ & + \sum_i \sum_k \sum_t \frac{VJ_k QS_{ik}^t}{CK_k^t}.\end{aligned} \tag{20}$$

s.t.

$$\sum_p QS_{sp}^t \le OS_s^t CS_s^t, \quad \text{for all } s, t, \tag{21}$$

$$\sum_d QS_{pd}^t \le OP_p^t CP_p^t, \quad \text{for all } p, t, \tag{22}$$

$$\sum_{p} QS_{pd}^{t} + QS_{kd}^{t} \leq OD_{d}^{t} CD_{d}^{t}, \quad \text{for all } d, t, \tag{23}$$

$$\sum_{c} QS_{ci}^{t} \leq OI_{i}^{t} CI_{i}^{t}, \quad \text{for all } i, t, \tag{24}$$

$$\sum_{i} QS_{im}^{t} \leq OM_{m}^{t} CM_{m}^{t}, \quad \text{for all } m, t, \tag{25}$$

$$\sum_{s} QS_{ls}^{t} \leq OL_{l}^{t} CL_{l}^{t}, \quad \text{for all } l, t, \tag{26}$$

$$\sum_{s} QS_{sp}^{t} = \sum_{d} QS_{pd}^{t}, \quad \text{for all } p, t, \tag{27}$$

$$\sum_{k} QS_{kd}^{t} + \sum_{p} QS_{pd}^{t} = \sum_{c} QS_{dc}^{t}, \quad \text{for all } d, t, \tag{28}$$

$$\sum_{c} QS_{ci}^{t} = \sum_{m} QS_{im}^{t} + \sum_{l} QS_{il}^{t} + \sum_{k} QS_{ik}^{t}, \quad \text{for all } i, t, \tag{29}$$

$$\sum_{l} QS_{il}^{t} = \sum_{c} QS_{ci}^{t} RY_{t}, \quad \text{for all } i, t, \tag{30}$$

$$\sum_{i} QS_{il}^{t} = \sum_{s} QS_{ls}^{t}, \quad \text{for all } l, t, \tag{31}$$

$$\sum_{k} QS_{ik}^{t} = \sum_{c} QS_{ci}^{t} RV_{t}, \quad \text{for all } i, t, \tag{32}$$

$$\sum_{i} QS_{ik}^{t} = \sum_{d} QS_{kd}^{t}, \quad \text{for all } k, t, \tag{33}$$

$$\sum_{m} QS_{im}^{t} = \sum_{c} QS_{ci}^{t} RD_{t}, \quad \text{for all } i, t, \tag{34}$$

$$\sum_{d} QS_{dc}^{t} = DE_{c}^{t}, \quad \text{for all } c, t, \tag{35}$$

$$\sum_{i} QS_{ci}^{t} = DE_{c}^{t} * \rho_{c}, \quad \text{for all } c, t, \tag{36}$$

$$\sum_{s} OS_{s}^{t} \leq S, \quad \text{for all } t, \tag{37}$$

$$\sum_{p} OP_{p}^{t} \leq P, \quad \text{for all } t, \tag{38}$$

$$\sum_{i} OI_{i}^{t} \leq I, \quad \text{for all } t, \tag{39}$$

$$\sum_{d} OD_{d}^{t} \leq D, \quad \text{for all } t, \tag{40}$$

$$\sum_{l} OL_{l}^{t} \leq L, \quad \text{for all } t, \tag{41}$$

$$\sum_{m} OM_{m}^{t} \leq M, \quad \text{for all } t, \tag{42}$$

$$\sum_{m} OK_{m}^{t} \leq K, \quad \text{for all } t, \tag{43}$$

$$QS_{sp}^{t}, QS_{pd}^{t}, QS_{dc}^{t}, QS_{ci}^{t}, QS_{im}^{t}, QS_{il}^{t}, QS_{ik}^{t}, QS_{ls}^{t}, QS_{kd}^{t} \geq 0, \quad \text{for all } i, s, k, l, m, p, t, c, d, \tag{44}$$

$$OS_{s}^{t}, OP_{p}^{t}, OD_{d}^{t}, OI_{i}^{t}, OM_{m}^{t}, OL_{l}^{t}, OK_{k}^{t} \in \{0,1\}, \quad \text{for all } i, s, k, l, m, p, t, c, d. \tag{45}$$

The first objective function computes the total cost of the CLSC network. It includes three major components: 1) fixed establishment/activation costs for the different facility types considered in the network, 2) operating costs associated with production, recycling, repair, and disposal processes, and 3) transportation costs incurred across all forward and reverse links between the network echelons. The second objective function captures the environmental impacts generated by closed-loop operations. In this function, the first two terms reflect the environmental burdens of manufacturing at plants and disposing of returned items at disposal centers, respectively. In contrast, the remaining terms quantify the transportation-related environmental impacts arising from flows between facilities. The third objective function represents the social dimension of sustainability. Social impact is measured through employment creation by considering 1) fixed jobs generated by establishing facilities and 2) variable jobs induced by operational throughput. In particular, the seventh to twelfth terms correspond to variable employment created by operational activities at the facilities; therefore, this objective seeks to maximize social benefits.

The mathematical program is given in *Eqs. (18)-(45). Eq. (18)* minimizes total cost, *Eq. (19)* minimizes total environmental impact, and *Eq. (20)* maximizes social impact (employment). *Eqs. (21)-(26)* enforce capacity limits for the corresponding facilities in each period, including output capacities for suppliers, plants, and recycling centers, and input capacities for distribution, collection/inspection, and disposal centers. *Eqs. (27)-(34)* impose flow-balance relationships, ensuring conservation of flow at each node: manufacturing plants, distribution centers, and collection/inspection centers *(27)-(29)*; recycling centers *Eqs. (30)-(31)*; repair centers *Eqs. (32)-(33)*; and disposal centers *Eq. (34). Eq. (35)* guarantees full demand satisfaction for each customer and period. *Eq. (36)* defines the collected return quantity as a function of demand and the predicted return rate. *Eqs. (37)-(43)* limit the allowable number of facilities that can be opened/operated for each facility type. Finally, *Eqs. (44)-(45)* specify the domains of decision variables: shipment quantities are non-negative continuous variables, and facility activation variables are binary, ensuring that the model remains consistent with practical operating conditions.

3.3 | Weighted Goal Programming

Since the proposed CLSC network design model is multi-objective, an explicit preference-articulation mechanism is required to obtain implementable solutions and to manage the inherent trade-offs among economic, environmental, and social performance. In reverse logistics and CLSC settings, these trade-offs are particularly pronounced because decisions such as facility activation, capacity allocation, and forward and reverse flow routing simultaneously influence total cost, emissions generated by processing and transportation, and employment creation across network echelons. Accordingly, a solution approach is needed that can incorporate decision-maker priorities transparently and operationally while producing a single plan suitable for implementation. To address this requirement, we adopt the WGP approach. WGP enables decision makers to specify aspiration (target) levels for each objective function in advance and then derive a balanced solution by minimizing weighted deviations from these targets. For minimization-type objectives (e.g., total cost and environmental impact), the goal constraints are formulated so that the deviation variables capture exceedances above the aspiration level. In contrast, for maximization-type objectives (for example, social impact measured through employment), the deviation variables represent shortfalls from the aspiration level (and, if necessary, exceedances). This formulation preserves the original direction of each objective while providing a coherent mechanism to reconcile conflicting sustainability goals within a single, implementable CLSC network design.

Step 1 (solving individual objective functions). Initially, the model is solved separately for each objective function as formulated in *Eq. (46):*

$$\min f_i(x), \qquad s.t., i = 1,2,3, \dots, n. \tag{46}$$

This step provides the optimal values of each objective function when considered in isolation. These individually optimized solutions constitute the components of the ideal objective vector and are commonly used as reference points for defining aspiration levels (target values) in goal programming models [23, 24].

Accordingly, the obtained optimal values were adopted as the target values. g_i in the present study, allowing decision-makers to evaluate trade-offs among conflicting objectives based on these reference benchmarks.

Step 2 (defining the WGP model). The aspiration levels are set based on the reference values obtained in Step 1 and can be adjusted to reflect decision-maker preferences and policy priorities. Once target values for each objective function are established, the goal programming model is formulated as follows:

$$\min\left[\sum_{i=1}^{n}\left(\frac{w_i^- d_i^- + w_i^+ d_i^+}{k_i}\right)^p\right]^{\frac{1}{p}}. \quad (47)$$

Subject to:

$$Ax \le b, \quad (48)$$

$$f_i(x) + d_i^- - d_i^+ = g_i, \qquad \text{for all } i \in \{1,2,3,..,n\}, \quad (49)$$

$$d_i^-, d_i^+, x \ge 0, \quad (50)$$

where $f_i(x)$ represents the i-th objective function. g_i does the decision-maker set the target value for the i-th objective? d_i^- and d_i^+ are negative and positive deviations, representing underachievement and overachievement of the goal, respectively. w_i^- and w_i^+ are assigned to the negative and positive deviations, reflecting the relative importance of minimizing these deviations. $k_i = f_i^* - f_i^-$ is the normalization constant, with f_i^* and f_i^- being the positive and negative ideal points. p determines the aggregation type. In line with prior goal programming formulations that adopt the $P = 1$ achievement function, we set $p = 1$ to obtain a linear weighted deviation measure [25, 26].

4 | Numerical Results

4.1 | Training of Machine Learning Models

In this study, six numerical features extracted from customer historical records and order information are used to predict the product return rate at the time of order placement. These features include: the number of the customer's previous orders, historical return rate, average value of past orders, current order value, percentage discount applied to the order, and the duration of the customer relationship (measured in months). The dataset comprises 845 order records, of which 80% are used for model training and the remaining 20% for model evaluation.

To identify the most effective ML approach, three models—SVR, RF, and RR—are examined, and their predictive performances are compared using four evaluation metrics: MSE, MAE, RMSE, and the coefficient of determination (R^2). The comparative results are presented in *Fig. 3*. Model training and evaluation are conducted in Python using the scikit-learn library.

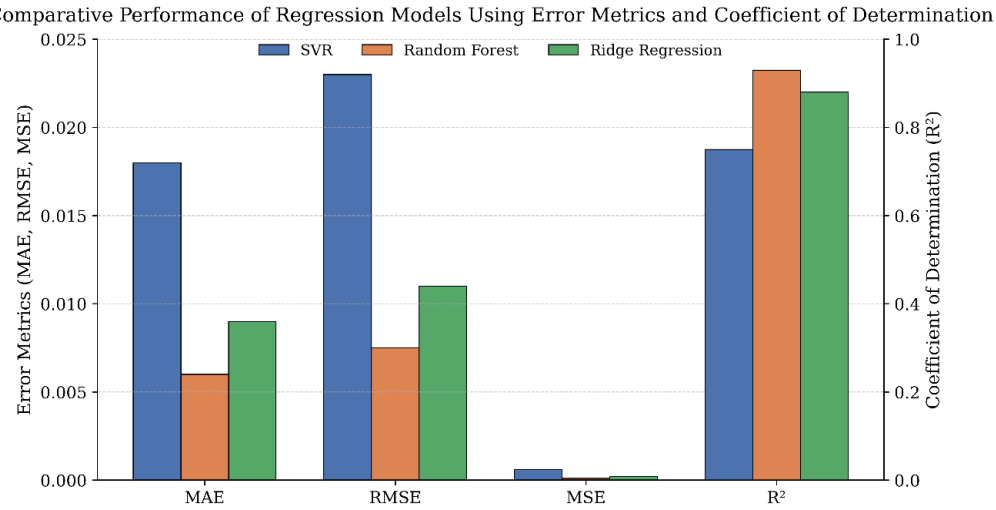

Fig. 3. Comparing the performance of machine learning algorithms in predicting return rates.

According to the figure, the RF model outperforms both SVR and RR, with lower MAE and RMSE and a higher R^2. These performance indicators demonstrate that RF not only provides greater predictive accuracy but also explains a larger proportion of the data's variance. Therefore, it can be concluded that RF is a more suitable modeling approach for the problem under investigation.

4.2 | Solution of the Multi-Objective Mathematical Programming Model

In this study, a CLSC redesign model is implemented and solved for a company operating in Iran's automotive spare parts industry, with a particular focus on brake pads. The company previously operated a logistics network comprising suppliers, manufacturing plants, distribution centers, customers, and recycling, repair, and disposal facilities; however, due to inefficiencies in the existing structure, a network redesign was deemed necessary. The proposed model is solved over six monthly planning periods.

The solution to the mathematical model is obtained by incorporating predicted product return rates from the RF algorithm. The multi-objective mathematical programming model simultaneously addresses three primary objectives: minimizing total operational costs, minimizing environmental impacts, and maximizing social benefits through job creation. To solve the multi-objective formulation, a WGP approach is employed. The model is implemented in the Pyomo optimization framework using Python and solved with the CPLEX solver.

All computational experiments are conducted on a computer system equipped with a 12th-generation Intel Core i7 processor, 32 GB of RAM, and Windows 11. Given the moderate network size and the reasonable number of facilities, the model reaches the optimal solution within an acceptable computational time.

The results of solving the model indicate that the product return rate directly affects the location of collection centers, with regions with higher return rates requiring more facilities. *Fig. 4* illustrates the geographical distribution of the selected facilities in Iran for the first planning period, clearly highlighting the critical role of customer return data in the optimal design of the CLSC network.

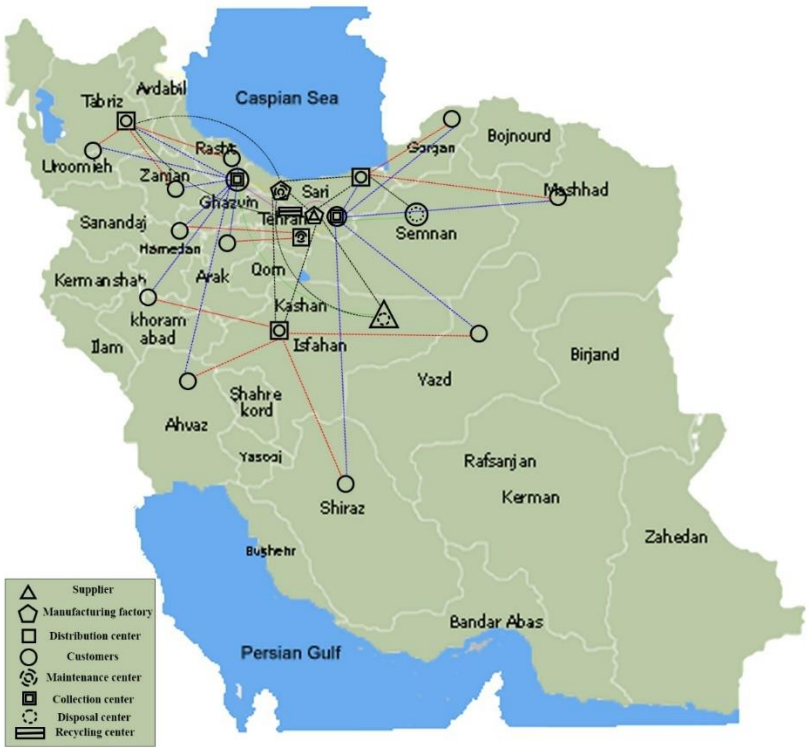

Fig. 4. Distribution map of supply chain facilities in the first period after the mathematical model limit.

4.3 | Validation

In this analysis, to validate the proposed approach for CLSC design, product return rates are predicted using three ML methods—SVR, RF, and RR—and subsequently incorporated as input into the mathematical model. A comparison of the results across the four objective functions indicates that the RF model achieves the best overall performance, yielding an operational cost of 1,456,785, an environmental impact value of 450,123, a social index of 1,452, and a weighted goal objective function value of 842,351.

In contrast, the SVR model exhibits inferior performance, with an operational cost of 1,758,635 and an environmental impact of 678,963. These quantitative differences demonstrate that selecting an appropriate forecasting model can lead to substantial improvements in network-level decision-making. Therefore, this evaluation not only confirms the superior accuracy and effectiveness of the RF model in predicting product return rates but also implicitly validates the robustness, coherence, and generalizability of the proposed integrated framework. The comparative results of the forecasting methods and their impact on solving the mathematical model are presented in *Fig. 5*.

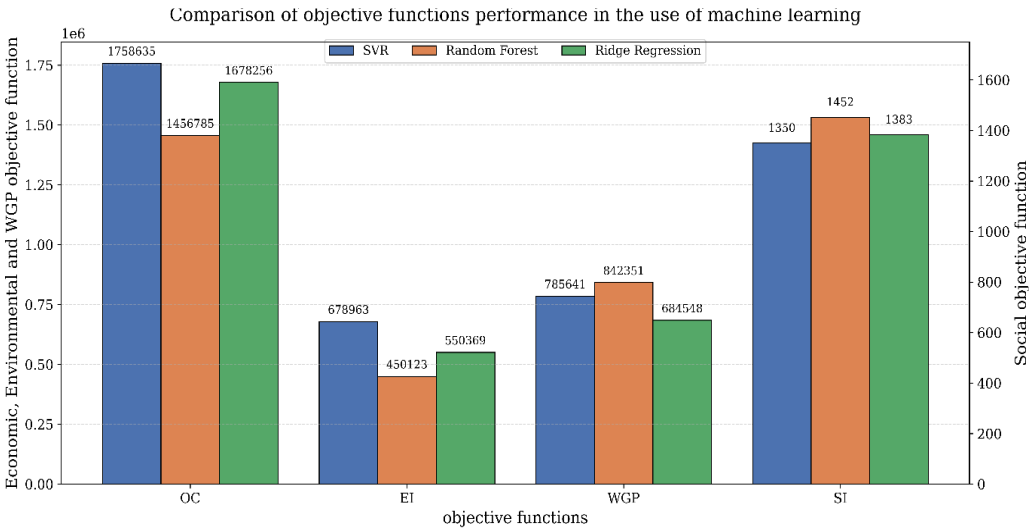

Fig. 5. Validating the use of machine learning models in mathematical modeling.

4.4 | Sensitivity Analysis

To examine the model's responsiveness to variations in key parameters, a sensitivity analysis is conducted with respect to the product return rate and the transportation cost from customers to collection centers. The left-hand diagram in *Fig. 6* shows that as the product return rate increases, both the economic and environmental objective functions exhibit an increasing trend due to the higher volume of reverse flows. In contrast, the social objective function increases significantly due to greater employment generated by collection, repair, and recycling centers. Nevertheless, the weighted goal objective function indicates that when the weight assigned to the social objective is lower than that of the economic and environmental objectives, the substantial growth in social benefits is insufficient to offset the associated economic and environmental losses. Consequently, the overall value of the weighted goal function increases.

In the right-hand diagram of *Fig. 6*, an increase in the transportation cost from customers to collection centers leads to a rise in the economic objective value. However, its impact on the environmental objective is negative, as the network tends to reduce the number of transportation routes, resulting in a relative decrease in environmental impacts. Additionally, reduced utilization of certain facilities leads to a decline in the social objective function.

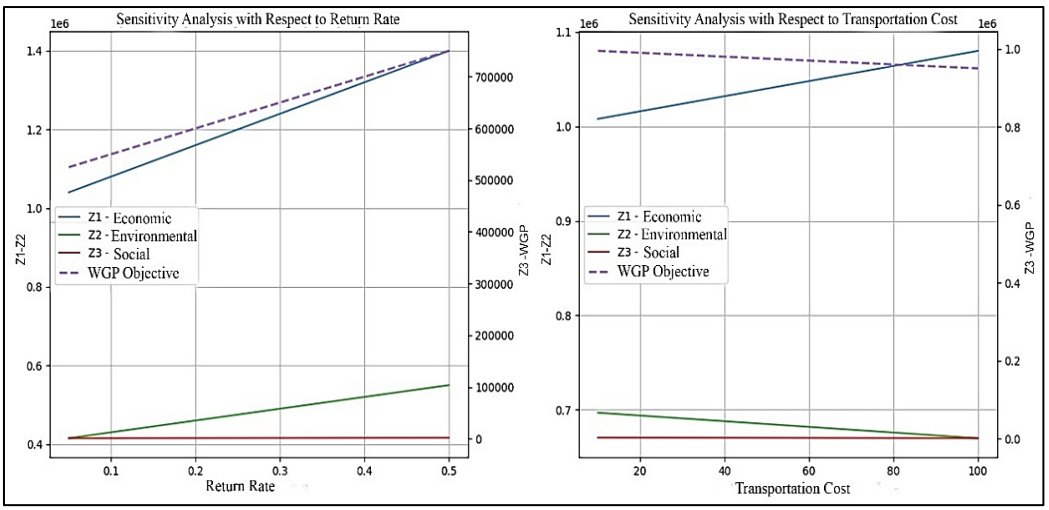

Fig. 6. Sensitivity analysis of the return rate and the cost of transportation from the customer to the collection center.

5 | Discussion and Conclusion

This study aims to develop an integrated, data-driven framework for designing a CLSC in the automotive spare parts industry. The proposed approach combines ML techniques with multi-objective mathematical optimization and is implemented using real-world data from a company operating in the brake pad sector. In the first phase, product return rates are predicted using historical data comprising 845 customer order records through three ML algorithms: RR, SVR, and RF. The results indicate that the RF model significantly outperforms the other methods, achieving a coefficient of determination (R^2) of 0.81 and a Mean Absolute Error (MAE) of 0.093.

In the subsequent phase, the forecasting model's output is incorporated into a multi-objective mathematical programming model that simultaneously considers economic, environmental, and social objectives. Numerical results from solving the model demonstrate that using accurate return-rate predictions from the RF model, rather than SVR, leads to a 16.5% reduction in operational costs (from 1,758,635 to 1,456,785 units) and a 33.7% reduction in environmental impacts (from 678,963 to 450,123 units). Moreover, the social benefit index increases to 1,452 under this scenario. These findings clearly illustrate that integrating accurate ML models with mathematical optimization frameworks can substantially enhance decision-making quality in logistics network design.

Beyond providing optimal location and operational decisions, the proposed framework shows strong potential for generalization to other industries with reverse supply chains, such as household appliances, electronics, and pharmaceuticals. For future research, it is recommended to incorporate hybrid deep learning models, such as GRU or LSTM, and robust optimization techniques to better account for environmental uncertainties. Such extensions could further strengthen the resilience and adaptability of CLSC networks under real-world conditions.

Author Contribution

"Conceptualization, Abbas. Foroozanfar; Methodology, Amirhossein. Amou Jafari; Software, Mahsa. Mehrabi; Validation, Abbas. Foroozanfar, Amirhossein. Amou Jafari, writing and creating the initial design, Mahsa. Mehrabi, writing, reviewing, and editing, Abbas. Foroozanfar, visualization, Mahsa. Mehrabi. All authors have read and agreed to the published version of the manuscript.

Funding

This work was carried out without any external financial support from governmental, commercial, or non-profit organizations.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest related to the publication of this research.

References

[1] Yadegari, M., Tavakkoli-Moghaddam, R., & Ahmadi, G. (2018). Closed-loop supply chain inventory-location problem with spare parts in a multi-modal repair condition. *International Journal of Engineering (IJE), Transactions B: Applications, 31*(2), 346–356. https://doi.org/10.5829/ije.2018.31.02b.20

[2] Masoudipour, E., Amirian, H., & Sahraeian, R. (2017). A novel closed-loop supply chain based on the quality of returned products. *Journal of cleaner production*, *151*, 344–355. https://doi.org/10.1016/j.jclepro.2017.03.067

[3] Ritola, I., Krikke, H., & Caniëls, M. C. J. (2020). Learning from returned products in a closed loop supply chain: A systematic literature review. *Logistics*, *4*(2), 7. https://doi.org/10.3390/logistics4020007

[4] Saadi, M. K., & Kazemi, A. (2025). Artificial intelligence and supply chain management of small and medium-sized enterprises. *Supply chain and operations decision making*, *2*(1), 12–20. https://doi.org/10.48313/scodm.v2i1.25

[5] Cui, H., Rajagopalan, S., & Ward, A. R. (2020). Predicting product return volume using machine learning methods. *European journal of operational research*, *281*(3), 612–627. https://doi.org/10.1016/j.ejor.2019.05.046

[6] Carrasco-Gallego, R., & Ponce-Cueto, E. (2009). Forecasting the returns in reusable containers' closed-loop supply chains. A case in the LPG industry. *3rd international conference on industrial engineering and industrial management*, 311–320. http://adingor.es/congresos/web/uploads/cio/cio2009/311-320.pdf

[7] Krapp, M., Nebel, J., & Sahamie, R. (2013). Forecasting product returns in closed-loop supply chains. *International journal of physical distribution & logistics management*, *43*(8), 614–637. https://doi.org/10.1108/IJPDLM-03-2012-0078

[8] Liang, X., Jin, X., & Ni, J. (2014). Forecasting product returns for remanufacturing systems. *Journal of remanufacturing*, *4*(1), 8. https://doi.org/10.1186/s13243-014-0008-x

[9] Matsumoto, M., & Komatsu, S. (2015). Demand forecasting for production planning in remanufacturing. *The international journal of advanced manufacturing technology*, *79*, 161–175. https://doi.org/10.1007/s00170-015-6787-x

[10] Zhou, L., Xie, J., Gu, X., Lin, Y., Ieromonachou, P., & Zhang, X. (2016). Forecasting return of used products for remanufacturing using Graphical Evaluation and Review Technique (GERT). *International journal of production economics*, *181*, 315–324. https://doi.org/10.1016/j.ijpe.2016.04.016

[11] Geda, M. W., & Kwong, C. K. (2018). Forecasting of used product returns for remanufacturing. *2018 ieee international conference on industrial engineering and engineering management (IEEM)* (pp. 889–893). IEEE. https://doi.org/10.1109/IEEM.2018.8607362

[12] Pereira, M. M., Machado, R. L., Ignacio Pires, S. R., Pereira Dantas, M. J., Zaluski, P. R., & Frazzon, E. M. (2018). Forecasting scrap tires returns in closed-loop supply chains in Brazil. *Journal of cleaner production*, *188*, 741–750. https://doi.org/10.1016/j.jclepro.2018.04.026

[13] Chou, M. C., Sim, C. K., & Yuan, X. M. (2020). Policies for inventory models with product returns forecast from past demands and past sales. *Annals of operations research*, *288*(1), 137–180. https://doi.org/10.1007/s10479-020-03545-4

[14] Geda, M., & Kwong, C. K. (2021). An MCMC based Bayesian inference approach to parameter estimation of distributed lag models for forecasting used product returns for remanufacturing. *Journal of remanufacturing*, *11*(3), 175–194. https://doi.org/10.1007/s13243-020-00099-3

[15] Turki, E., Jouini, O., Jemai, Z., Traiy, Y., Lazrak, A., Valot, P., & Heidseick, R. (2024). Forecasting spare part extractions from returned systems in a closed-loop supply chain. *International journal of production research*, *62*(21), 7860–7876. https://doi.org/10.1080/00207543.2024.2333052

[16] Adigüzel Tüylü, A. N., & Eroğlu, E. (2019). Using Machine Learning Algorithms For Forecasting Rate of Return Product In Reverse Logistics Process. *Alphanumeric journal*, *7*(1), 143–156. https://doi.org/10.17093/alphanumeric.541307

[17] Tuylu, A. N. A., & Eroglu, E. (2022). The prediction of product return rates with ensemble machine learning algorithms. *Journal of engineering research*, 1–14. https://doi.org/10.36909/jer.13725

[18] Gao, C. (2024). Research on Optimization Strategies for Closed-Loop Supply Chain Management Based on Deep Learning Technology. *International journal of information systems and supply chain management*, *17*(1), 1–22. https://doi.org/10.4018/IJISSCM.341802

[19] Ashtab, S., & Tosarkani, B. M. (2023). Scenario-based multi-objective optimisation model based on supervised machine learning to configure a plastic closed-loop supply chain network. *International journal of business performance and supply chain modelling*, *14*(1), 106. https://doi.org/10.1504/ijbpscm.2023.10055639

[20] Gökbayrak, E., Kayış, E., & Güllü, R. (2025). Unlocking the value in product return data: Inventory management with sales dependent stochastic product return flows from multiple periods. *International journal of production economics*, *285*, 109618. https://doi.org/10.1016/j.ijpe.2025.109618

[21] Cortes, P., Alarcon, F., & Pérez-perales, D. (2025). Integration of artificial intelligence in reverse logistics process : enhancing decision-making across planning , execution , and control stages, *13*, 214–235. https://doi.org/10.4995/ijpme.2025.23262

[22] Ibrahim, A. A., & Abdul-kader, W. (2025). A predictive and prescriptive analytics approach for sustainable cellphone return management. *Decision analytics journal*, *17*(July), 100656. https://doi.org/10.1016/j.dajour.2025.100656

[23] Olfati, M., Snášel, V., Krömer, P., Fanati, S., & Seyedali, R. (2025). A goal programming-based algorithm for solving multi objective optimization problems. *Annals of operations research*. https://doi.org/10.1007/s10479-025-06646-0

[24] Oliveira, W. A., Fiorotto, D. J., Song, X., & Jones, D. F. (2021). An extended goal programming model for the multiobjective integrated lot-sizing and cutting stock problem. *European journal of operational research*, *295*(3), 996–1007. https://doi.org/10.1016/j.ejor.2021.03.049

[25] Carrizosa, E., & Fliege, J. (2002). Generalized goal programming: Polynomial methods and applications. *Mathematical programming*, *93*(2), 281–303. https://doi.org/10.1007/s10107-002-0303-4

[26] Merchant, J., & Jones, D. F. (2023). Goal programming duality revisited: Formulation of a set of variant duals. *Journal of the operational research society*, *74*(11), 2350–2361. https://doi.org/10.1080/01605682.2022.2147031