

Article

An Explainable AI-Driven Framework for Sustainable Supplier Selection in Healthcare Systems: A Methodological Framework and Proof of Concept

Lara J M Naser, Alper Göksu * and Berrin Denizhan 

Department of Industrial Engineering, Faculty of Engineering, Sakarya University, Sakarya 54050, Turkey; lara.naser@ogr.sakarya.edu.tr (L.J.M.N.); denizhan@sakarya.edu.tr (B.D.)

* Correspondence: agoksu@sakarya.edu.tr

Abstract

Supplier selection in healthcare is a complex multi-criteria decision-making (MCDM) challenge requiring a balance of sustainability, resilience, and operational efficiency. Traditional methods struggle with scalability and subjectivity when applied to large administrative datasets. This study introduces a transparent hybrid Machine Learning–MCDM (ML–MCDM) framework, validated using a U.S. Medicare dataset of 661 suppliers. The framework integrates eXtreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP) for criterion prioritization, the Full Consistency Method (FUCOM) for mathematically consistent weighting, and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) for final ranking. As the dataset lacks direct sustainability metrics, seven indicators were synthetically generated; thus, the results serve as proof-of-concept demonstration of the framework’s architecture. Specifically, XGBoost–SHAP is trained to predict a synthetically constructed Overall Performance Score (OPS), meaning that the resulting feature importance output constitutes an algorithmic consistency check—confirming that the pipeline correctly recovers importance signals deliberately embedded in the training target. For interpretability, suppliers were segmented into five performance profiles via K-Means: Strategic Partners (17.7%), Green Leaders (18.6%), Reliable Emergency Suppliers (18.2%), Balanced Performers (20.4%), and Developing Suppliers (25.1%). Carbon Footprint Score (0.408) and Emergency Response Capability (0.316) achieved the highest feature importance. FUCOM-derived weights prioritized On-Time Delivery Rate (0.272), Carbon Footprint Score (0.222), and Emergency Response Capability (0.220). The top supplier attained a TOPSIS closeness coefficient of 0.800, showing strong discrimination. Sensitivity analysis across four scenarios confirmed ranking robustness, maintaining Spearman correlations $\rho \geq 0.977$. This ML–FUCOM–TOPSIS approach provides an auditable, scalable, and policy-relevant decision-support tool, enabling procurement managers to navigate high-dimensional data while ensuring operational continuity and environmental responsibility in healthcare supply chains.



Academic Editors: Antonio Pesqueira and Andreia De Bem Machado

Received: 8 May 2026

Revised: 15 June 2026

Accepted: 16 June 2026

Published: 20 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: sustainable healthcare; supplier selection; machine learning; multi-criteria decision-making; XGBoost; SHAP; FUCOM; TOPSIS; Supply Chain Resilience

1. Introduction

Supplier selection has long been recognized as a strategic decision-making challenge in supply chain management, requiring organizations to evaluate multiple, often conflicting criteria such as cost, quality, risk, reliability, and sustainability [1,2]. Traditional

supplier selection approaches have relied heavily on expert judgment and static evaluation techniques, which often struggle to handle high-dimensional datasets, complex variable interactions, and uncertainty in supplier performance [3]. In recent years, the emergence of advanced data-driven methodologies, particularly Machine Learning (ML) and Multi-Criteria Decision-Making (MCDM), has significantly transformed the way supplier performance is assessed [4,5]. ML techniques contribute to predictive accuracy, automated pattern discovery, and objective criterion weighting, while MCDM provides structured, explainable, and transparent decision frameworks that can synthesize diverse performance indicators into coherent evaluation and ranking schemes.

Integrating ML with MCDM yields a hybrid decision-support paradigm that can overcome the limitations of either approach alone. ML algorithms such as Random Forest, XGBoost, Support Vector Machines, and CatBoost can analyze large-scale, heterogeneous datasets to identify the most influential features affecting supplier performance. These data-driven insights are then incorporated into MCDM methods such as Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Measurement of Alternatives and Ranking according to the Compromise Solution (MARCOS), Complex Proportional Assessment (COPRAS), or Fuzzy-TOPSIS to generate interpretable and theoretically grounded supplier rankings. Such hybrid approaches have been shown to enhance accuracy, reduce human subjectivity, and provide more robust, evidence-based evaluation models across various industries—including automotive, oil and gas, aerospace, manufacturing, and pharmaceuticals [6–8]. Recent studies demonstrate that ML–MCDM integration is especially advantageous in contexts where supplier performance is multifaceted, data-rich, and influenced by operational, financial, environmental, and contextual factors.

Despite the growing body of literature on hybrid ML–MCDM supplier selection models, their application to large-scale healthcare administrative datasets remains largely unexplored. Existing studies predominantly focus on manufacturing, automotive, or energy supply chains and rely on expert-driven or small-sample datasets [9–11]. In contrast, healthcare supplier ecosystems—such as those represented in the U.S. Medicare system—are characterized by thousands of suppliers, high-dimensional claim-based indicators, and strong policy implications [12]. This gap is particularly consequential for several reasons. First, healthcare supply chains are subject to stringent regulatory oversight and public accountability requirements, making transparent, auditable decision-making tools a necessity rather than a convenience. Second, the scale and complexity of public health administrative data exceed what conventional expert-driven MCDM methods can effectively handle, creating an urgent need for objective, data-driven weighting mechanisms. Third, disruptions in healthcare supply chains carry direct patient safety consequences, underscoring the urgency of resilience-aware procurement frameworks. Addressing this gap, therefore, holds both methodological significance and practical societal value. Building This work builds on this gap by proposing a transparent, scalable, and ML–MCDM framework that integrates XGBoost–SHAP for objective criterion prioritization, FUCOM for fully consistent weight determination, and TOPSIS for interpretable supplier ranking.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on traditional supplier selection methods, machine learning applications in supplier evaluation, and hybrid ML–MCDM frameworks and highlights a research gap in large-scale healthcare administrative datasets. Section 3 presents the proposed methodology, including data preparation, feature engineering, sustainability and delivery performance modeling, clustering analysis, ML-based criterion prioritization, FUCOM weight derivation, and TOPSIS ranking. Section 4 reports and discusses the results. Finally, Section 5 concludes the study by summarizing key findings, outlining managerial and policy implications, discussing limitations, and suggesting directions for future research.

This study addresses the identified gap by proposing a novel hybrid framework—XGBoost–SHAP–FUCOM–TOPSIS—that systematically integrates machine learning, explainable AI, and MCDM for sustainable healthcare supplier selection under large-scale administrative data. By shifting from purely subjective expert elicitation to a hybrid subjective–objective paradigm, this framework not only enhances the accuracy and objectivity of supplier selection but also ensures full explainability and auditability, offering a powerful and practical decision-support tool for sustainable procurement in data-rich environments like public health systems.

This study makes several significant and interconnected contributions to the fields of sustainable supplier selection, healthcare supply chain management, and decision science by proposing and validating a novel hybrid framework—XGBoost–SHAP–FUCOM–TOPSIS. The primary contribution lies in the innovative and systematic fusion of machine learning, Explainable Artificial Intelligence (XAI), and multi-criteria decision-making (MCDM) to address the specific challenges of supplier evaluation using large-scale administrative data, a context largely unexplored in prior literature.

Unlike traditional models that rely heavily on subjective expert elicitation or treat ML as a black-box predictor, this framework establishes a fully integrated, end-to-end analytical pipeline. This pipeline moves from raw, high-dimensional data to transparent and auditable supplier rankings by: (1) employing XGBoost to model complex, non-linear relationships within the data, objectively learning the predictive importance of supplier attributes; (2) utilizing SHAP to demystify the ML model, providing a granular, interpretable quantification of each criterion’s marginal contribution to supplier performance; (3) using these objectively derived SHAP values to inform the FUCOM pairwise comparisons, creating a robust foundation for weighting that minimizes expert bias; and (4) leveraging TOPSIS to perform a transparent, mathematically grounded ranking of suppliers based on these empirically derived weights. Beyond its methodological innovation, the study makes a crucial contribution by applying this framework to a large-scale U.S. Medicare supplier dataset. The analysis yields practically relevant insights that advance the understanding of sustainable healthcare procurement.

2. Literature Review

2.1. Evolution from Traditional Supplier Selection

Traditional supplier selection research has historically focused on operational and economic efficiency, emphasizing measurable performance attributes rather than broader sustainability concerns. Early work by Dickson identified 23 evaluation criteria, with quality, delivery reliability, and supplier performance history emerging as the most influential determinants of sourcing decisions [1]. Subsequent analyses confirmed that quality, cost, and delivery remained the dominant criteria across manufacturing and purchasing contexts [2]. Methodologically, the problem has long been modeled as a multi-criteria decision-making task, leading to the adoption of structured analytical tools such as AHP and TOPSIS to balance conflicting quantitative and qualitative factors [3].

Overall, supplier selection has evolved from a Quality, Cost, and Delivery-dominated task to a multi-criteria, sustainability-aware decision problem that increasingly includes environmental, social, risk, and circularity concerns. However, the core decision structure and many practical priorities still resemble traditional supplier selection, with economic performance and familiar MCDM tools remaining dominant and sustainability criteria often layered on rather than fundamentally transforming the process. Critically, while these tools are well-validated in small expert-panel contexts, their reliance on subjective weight elicitation creates reproducibility concerns in the data-rich, high-volume procurement environments that characterize modern public health systems.

2.2. Sustainable Supplier Selection

Research on Sustainable Supplier Selection (SSS) has evolved from early green and environmental emphases to broader triple-bottom-line and resilience perspectives. Foundational reviews synthesize three decades of work, showing Sustainable Supplier Selection (SSS) as a multi-criteria decision-making problem in which economic, environmental, and social criteria must be evaluated simultaneously, primarily using compensatory MCDM methods such as the Analytic Hierarchy Process (AHP), the Analytic Network Process (ANP), TOPSIS, and their fuzzy extensions [4,5]. More recent reviews focusing on 2018–2022 confirm that TOPSIS, AHP, the VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) method, the Best-Worst Method (BWM), Data Envelopment Analysis (DEA), the Decision-Making Trial and Evaluation Laboratory (DEMATEL), and the Multi-Objective Optimization on the basis of Ratio Analysis plus full multiplicative form (MULTIMOORA)—often in hybrid, fuzzy, or grey forms—are the dominant tools [13]. A broader triple-theme review of sustainability, resilience, and artificial intelligence in supplier selection highlights how economic, environmental, social, and resilience criteria are increasingly combined, and points to the growing feasibility of mixed approaches (e.g., AI + MCDM) to handle uncertainty and complexity [14].”

Beyond classical MCDM, intelligent approaches leverage multi-agent systems to decentralize sustainable supplier evaluation in Industry 4.0 supply chains [6]. Dynamic and learning-oriented systems further integrate fuzzy BWM, FIS, and machine learning to update supplier scores over time in circular petrochemical chains [7]. Several studies embed sustainability into Industry 4.0 and circular economy settings using Pythagorean fuzzy entropy–Stepwise Weight Assessment Ratio Analysis (SWARA)–COPRAS frameworks [9]. Despite this methodological progress, the overwhelming majority of SSS studies remain anchored in expert-elicited weights applied to small, curated datasets—a fundamental limitation when the target domain involves thousands of heterogeneous suppliers and no direct environmental data, as is the case in public health procurement.

2.3. Methods: From Simple Tools to Hybrid, Uncertainty-Aware Models

A critical reading of the methodological evolution reveals a clear but incomplete progression. Classical MCDM methods, while transparent, are limited by sensitivity to subjective weights. Fuzzy extensions partially address linguistic uncertainty yet introduce parameter complexity. Advanced fuzzy variants capture hesitation among decision-makers but at the cost of interpretability. Subjective–objective hybrid weighting models integrate expert judgment with data-driven signals, yet still require reliable expert input. Crucially, recent hybrid ML–MCDM frameworks [7,14] improve accuracy but often sacrifice transparency, functioning as black boxes—a critical barrier in public healthcare, where auditability is non-negotiable. This study’s XGBoost–SHAP–FUCOM–TOPSIS framework is specifically designed to close this transparency gap. Table 1 reveals the methodological evolution of SSS approaches classified in 5 main segments.

Table 1. Methodological evolution of supplier selection approaches.

Method	Representative Approaches	Main Purpose	Main Limitations	Citation
Classical MCDM	AHP, SAW, TOPSIS, VIKOR, MABAC compared on the same dataset	Rank suppliers on cost, quality, and delivery	Sensitive to subjective weights; limited support for ambiguous linguistic judgments	[8]

Table 1. Cont.

Method	Representative Approaches	Main Purpose	Main Limitations	Citation
Single fuzzy extensions	Fuzzy AHP vs. Fuzzy TOPSIS [15]; fuzzy DEA vs. fuzzy TOPSIS [16]; fuzzy TOPSIS–VIKOR–GRA for green suppliers [17]; fuzzy AHP–TOPSIS in the auto sector [18]	Extend established MCDM to handle linguistic/uncertain ratings	Added parameter complexity; choice of membership functions and distance metrics can affect results	[15–18]
Advanced fuzzy variants	Hesitant fuzzy VIKOR variants [19]; intuitionistic fuzzy entropy–BWM–WASPAS [10]; PFAHP–TOPSIS–LAM hybrid [11]	Capture hesitation and disagreement among decision makers	Higher data and computational requirements; more difficult to explain to practitioners	[10,11,19]
Subjective–objective hybrid weighting	BWM + fuzzy Shannon entropy + fuzzy MULTIMOORA [20]; FBWM + CoCoSo‘B [21]; intuitionistic fuzzy BWM–entropy–WASPAS [10]	Integrate expert judgment and data-driven information for sustainability criteria	Requires both expert input and reliable data; additional steps for weight combination	[10,20,21]
Recent hybrid BWM/AHP/TOPSIS/others	AHP–BWM integration vs. TOPSIS/VIKOR/ELECTRE/SWARA [22]; fuzzy DEMATEL–BWM + CO-COSO/MOORA/TOPSIS [23]	Address resilience, sustainability, and complexity	Methodological complexity; need for software support and expert training	[22,23]

Despite the methodological richness in sustainable supplier selection (SSS) literature, a critical gap persists in integrating data-driven insights from large-scale administrative data with structured multi-criteria decision-making (MCDM) frameworks. Existing research is predominantly anchored in expert-driven subjective weighting and fuzzy-based uncertainty modeling, as evidenced by the prevalence of methods like AHP, BWM, and their fuzzy extensions [15,18,21]. While these approaches effectively handle linguistic judgments, they are inherently limited by expert bias and fail to leverage the vast, untapped potential of objective, real-world procurement data. Furthermore, recent hybrid models that combine artificial intelligence with MCDM [7,14] often lack transparency, functioning as “black boxes” that obscure how individual supplier attributes drive final selection outcomes. This lack of explainability is a significant barrier to adoption in sectors like public healthcare, where decisions must be auditable and justified. The cumulative evidence in Table 1 reveals a persistent blind spot: no existing hybrid framework has been designed to extract criterion weights objectively from large-scale administrative data while maintaining full explainability and auditability throughout the pipeline. This study addresses that gap.

3. Methodology

The dataset preparation involved a structured, systematic data selection process to ensure that only analytically relevant, non-redundant, and high-quality variables entered the modeling pipeline. The initial raw dataset contained a wide range of supplier-level attributes, including identification fields, demographic information, financial metrics, service indicators, geographic descriptors, and administrative metadata.

3.1. Data Preparation and Cleaning

The analysis begins with importing the original supplier performance dataset, which contains financial, operational, and service-related indicators. Since the dataset uses European numeric formatting (a comma as the decimal separator and a dot as the thousands separator), all relevant numeric variables were systematically cleaned. The cleaning procedure involved removing the thousand separators, replacing decimal commas with decimal points, and finally converting the column into a proper numeric format. Missing values in core numeric fields were imputed using the median to preserve the overall distribution while ensuring model fitness.

3.2. Feature Selection

To enhance the interpretability and analytical usefulness of the CMS supplier dataset, several new features were derived from the raw operational and financial variables. The derivation process focused on transforming high-volume raw totals into normalized performance indicators that enable fair comparisons across thousands of suppliers of varying sizes.

Table 2 shows the first set of derived features, Services per Beneficiary (SPB) and Services-to-Claims Ratio (SCR), which convert operational totals into utilization ratios. The Cost Efficiency Index (CEI) measures supplier financial efficiency by converting the average payment per service into a normalized inverse score, with higher values indicating better cost performance. The Geographically Normalized Payment (GNP) adjusts the payment-per-claim for regional price variation using RUCA-based factors, producing a cost indicator that is independent of geographic economic conditions.

Table 2. Dataset-derived features.

Feature	Definition	Purpose	Unit	Category
Services per Beneficiary (SPB)	Ratio of the total number of services provided to the total number of beneficiaries.	Indicates service intensity or utilization rate.	Services per patient	Capability
Services-to-Claims Ratio (SCR)	Ratio of the total services provided to the total claims submitted.	Measures of how many services are provided over each claim on average.	Services per claim	Capability
Cost Efficiency Index (CEI)	Normalized inverse of average payment per service.	Provides a unit-free efficiency score: higher values represent better cost performance.	Index (0–1)	Financial Efficiency
Geographic Normalized Payment (GNP)	Supplier's payment per claim is divided by the state or regional RUCA.	Adjust cost metrics for regional pricing variations.	Ratio	Geographic /Context

3.3. Feature Engineering for Sustainability and Delivery Performance

To enrich the dataset with meaningful metrics relevant to supplier sustainability and supply chain reliability, seven new indicators were developed.

Because public administrative datasets, such as U.S. Medicare claims, primarily track financial and operational volumes rather than granular environmental metrics, these specific sustainability features were generated using statistical methods. This approach bridges the gap between available public health data and the realities of modern sustainable procurement. It is important to acknowledge that all subsequent analyses (clustering, XGBoost,

SHAP, FUCOM, TOPSIS) were conducted on these synthetically generated indicators. Consequently, this study should be understood as a proof-of-concept validation of the framework's architecture and analytical pipeline, rather than a definitive assessment of actual Medicare suppliers' environmental performance. The framework is fully replicable with authentic sustainability data—such as real environmental disclosures or verified ESG metrics—once such data becomes available. The new features include:

- Green Packaging Index (GPI);
- Emergency Response Capability (ERC);
- Order Accuracy Rate (OAR);
- On-Time Delivery Rate (OTD);
- Carbon Footprint Score (CFS);
- Waste Management Score (WMS);
- Supplier Certification Level (SCL).

Normalization factors derived from existing features were used to create realistic performance distributions. Controlled Gaussian noise was added to maintain natural variability and prevent model overfitting, ensuring the simulated variables mirror the stochastic nature of real-world supplier behavior.

3.4. Data Normalization

Since clustering algorithms are highly sensitive to variable magnitude, all sustainability and delivery features were standardized using Z-score normalization. StandardScaler was applied via Equation (1) to transform each indicator to zero mean and unit variance.

$$x'_i = (x_i - \mu) / \sigma \quad (1)$$

3.5. K-Means Clustering ($k = 5$)

To segment suppliers into meaningful sustainability–delivery profiles, the K-Means algorithm was applied using five clusters. While the elbow method was plotted for exploratory purposes, the number of clusters was explicitly fixed at five to ensure a richer segmentation pattern suitable for multi-criteria supplier evaluation studies. Clustering is performed on the subset of the new seven features, focusing on packaging, emergency capability, accuracy, timeliness, and emission performance.

The selection of K-Means with $k = 5$ was based on methodological suitability for this study's specific context, rather than on empirical comparison with alternative algorithms. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) was not adopted for two structural reasons: its requirement to set density hyperparameters (ϵ and min_samples) is problematic when features exhibit moderately uniform density distributions—as is the case with Z-score-normalized, partially correlated sustainability indicators—and more fundamentally, DBSCAN does not guarantee exhaustive cluster assignment, meaning a subset of suppliers would receive no profile at all, which is operationally unacceptable in a procurement evaluation framework where every supplier must be classified.

Hierarchical clustering was similarly not pursued, as it is a non-partitioning method that does not iteratively optimize cluster compactness, tends to produce ambiguous dendrogram cut-points requiring subjective interpretation, and is generally better suited to small exploratory datasets than to large operational classification tasks such as the 661-supplier dataset analyzed here [24,25].

K-Means, by contrast, is specifically designed for exhaustive, mutually exclusive, and centroid-interpretable segmentation at scale, making it the most appropriate choice for producing discrete, actionable supplier tiers. The selection of $k = 5$ is further supported by

the Elbow Method curve showing within-cluster sum of squares (WCSS) flattening beyond $k = 4$ – 5 , and by its natural alignment with the five-tier supplier classification structure established in the sustainable supply chain literature [24,26].

To determine the optimal number of clusters for segmenting suppliers, we evaluated the Silhouette Score for $k = 2$ to $k = 10$. The Silhouette Score measures the average intra-cluster cohesion relative to inter-cluster separation, with higher values indicating better-defined clusters. The Silhouette Score peaked at $k = 5$ (0.441) and remained very close at $k = 4$ (0.423). Also, the elbow method (within-cluster sum of squares) indicated a flattening trend beyond $k = 5$. While $k = 5$ yields a higher statistical score, the five-cluster solution provides a rich, actionable segmentation for healthcare procurement managers. Therefore, we selected $k = 5$ as the optimal trade-off between statistical validity and managerial interpretability, following the principle that cluster analysis in applied research should balance quantitative criteria with domain-specific utility [24,25].

3.5.1. Choosing the Number of Clusters (Elbow Method)

Although the Elbow Method was used to inspect how the WCSS changes across different values of k , the analysis ultimately adopted a fixed number of clusters to ensure meaningful segmentation for the supplier context. In this study, the WCSS curve displayed a gradual decline without a sharply defined elbow, which is common in high-dimensional or moderately homogeneous datasets. Therefore, instead of relying solely on the visual elbow, the number of clusters was set explicitly to $k = 5$ to produce interpretable and practically relevant supplier profiles [24].

Based on Figure 1, the Elbow Method curve shows a steep decline in WCSS from $k = 1$ to $k = 3$, indicating substantial improvements in clustering compactness as additional clusters are introduced. Beyond $k = 4$ or $k = 5$, the reduction in WCSS becomes progressively smaller, and the curve begins to flatten, reflecting diminishing returns in model performance. This pattern suggests that while increasing the number of clusters improves fit, the most meaningful segmentation occurs around four to five clusters, where the balance between complexity and explanatory power is optimized. Therefore, selecting $k = 5$ provides a reasonable and interpretable clustering structure for the supplier sustainability and delivery performance data.

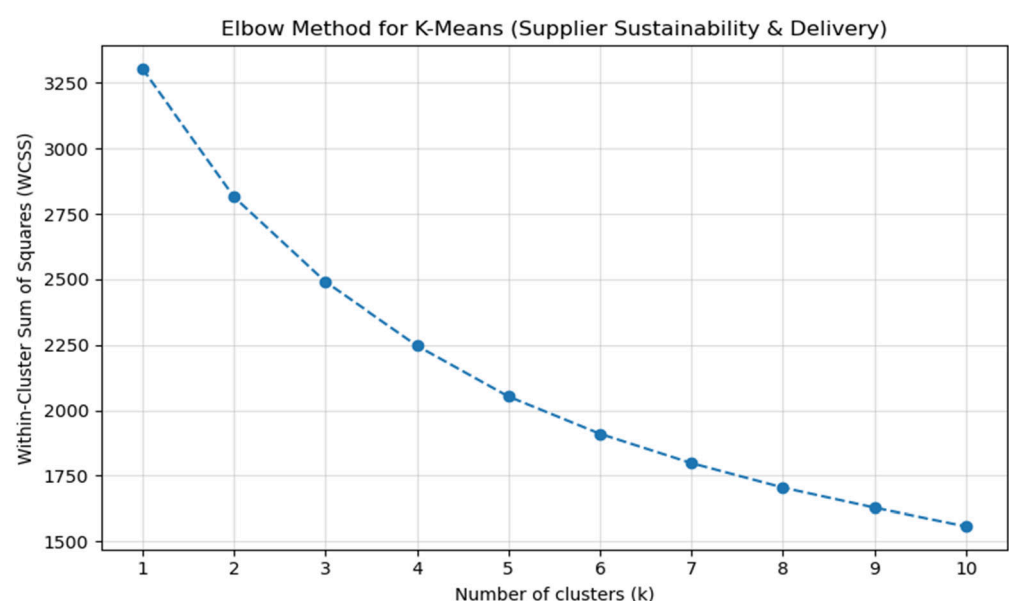


Figure 1. Elbow method WCSS curve for K-Means clustering.

3.5.2. Final Clustering and Profiles

Once the optimal number of clusters was determined, each supplier was assigned to a cluster based on the K-Means algorithm's label output. After cluster assignment, descriptive profiles were generated for each group by computing the mean performance values, the number of suppliers, and their proportional representation within the dataset [25].

3.6. Cluster Interpretation

Following the numerical clustering process, the study applied a rule-based interpretation function to translate the K-Means cluster outputs into business-meaningful categories. This step evaluates the average sustainability and delivery metrics within each cluster and assigns descriptive labels that reflect their performance characteristics [26]. The five selected features together form a five-dimensional performance vector for each supplier. Table 3 lists the features used for supplier segmentation, along with their descriptions and ranges. Equation (2) defines the five-dimensional performance vector for each supplier.

$$x_i = [GPI_i, ERC_i, OAR_i, OTD_i, CFS_i] \quad (2)$$

Table 3. Clustering features and descriptions.

Feature	Description	Range
Green Packaging Index	Percentage of packaging that is recyclable/biodegradable	0–1
Emergency Response Capability	Ability to handle emergency/rush orders	1–5
Order Accuracy Rate	Percentage of orders shipped without errors	95–100%
On-Time Delivery Rate	Percentage of orders delivered by the promised date	85–99%
Carbon Footprint Score	Environmental performance score based on emissions	0–100
Waste Management Score	Effectiveness in waste management and recycling	0–100
Supplier Certification Level	Level of sustainability certifications (0 = None, 1 = Some, 2 = Comprehensive)	0–2
Overall Performance Score	Composite score combining delivery and sustainability performance	Varies

Suppliers are classified into strategic performance tiers based on a multi-criteria evaluation of their operational and sustainability metrics. Strategic Partners are assigned to suppliers that exceed an *OTD* of 94%, an *OAR* of above 98%, and a *CFS* of greater than 75. Reliable Emergency Suppliers are designated for those with a high *ERC* (above 3.5) and a strong *OTD* (over 92%). Green Leaders are identified by superior environmental performance (a *CFS* above 75 and a *GPI* exceeding 0.6). Developing Suppliers are flagged when either the *OTD* or *OAR* falls below 90%. Balanced Performers are the default category for suppliers who demonstrate consistent performance across all key metrics. An integrated measure of supplier performance was constructed using Equation (3).

3.7. Overall Performance Score

To provide an integrated measure of supplier performance, an OPS was constructed as a weighted composite of key delivery and sustainability indicators:

$$OPS = 0.3 \cdot OTD + 0.3 \cdot OAR + 0.2 \cdot CFS + 0.2 \cdot (10 \cdot ERC) \quad (3)$$

It is important to note that the OPS serves exclusively as a supervised training target for the subsequent machine learning validation step; it is not intended as a domain-validated weighting scheme for supplier performance. The fixed coefficients in Equation (3)—0.3 for OTD, 0.3 for OAR, 0.2 for CFS, and 0.2 for ERC—are illustrative values chosen to produce a scalar outcome variable, not empirically or expert-derived priority weights. As a consequence, the XGBoost–SHAP analysis that follows should be understood as an algorithmic loop-closure test: it verifies that the machine learning pipeline correctly identifies the features most prominent in the synthetic target signal, rather than independently discovering real-world procurement priorities from the data.

OTD denotes OTD, OAR refers to Order Accuracy Rate, CFS represents CFS, and ERC indicates Emergency Response Capability. The emergency response metric is multiplied by 10 to align it with the scale of the other ~100-range indicators. The OPS provides a single interpretable metric for comparing and ranking suppliers across multiple performance dimensions.

3.8. Hybrid ML–MCDM Methodology for Final Supplier Selection

Following the completion of clustering and within-cluster prioritization via the OPS, the proposed framework advances into its final decision-making phase: a hybrid model that integrates ML with FUCOM and TOPSIS to produce a transparent, reproducible, and mathematically robust supplier ranking system.

3.8.1. ML-Guided Criterion Prioritization Using Feature Importance

To ensure the multi-criteria weighting process is objective and free from subjective bias, ML-based feature importance analysis is conducted on the sustainability and delivery dataset. A tree-based ensemble model—specifically XGBoost—is trained to predict the OPS. The resulting feature importance scores, extracted using SHAP values [27], quantitatively identify which evaluation criteria exert the greatest influence on performance.

SHAP values capture non-linear and interaction effects; however, for FUCOM’s purpose, only the magnitude of importance—not directionality—is required. We therefore used the mean absolute SHAP value per feature, which collapses direction and interaction effects into a single scalar importance score. This is a well-established practice in the feature importance literature [27]. A combined importance score was then computed by averaging XGBoost gain-based importance and mean absolute SHAP importance. This averaging reduces the influence of any single method’s non-linearities and produces a more stable ordinal ranking. Before entering the ranked criteria into FUCOM Step 1, the ordinal ranking was verified to be unambiguous (no ties among the top-ranked criteria), a necessary precondition for FUCOM. Where the combined importance ratios between adjacent criteria appeared extreme (e.g., CFS being approximately 23.8 times more important than SCL), these ratios were retained without truncation, as they reflect genuine data-driven discrimination. FUCOM’s optimization model further guards against inconsistency by minimizing deviation from full transitivity across all pairwise comparisons, ensuring that even large ratio differences are translated into logically coherent weights.

The XGBoost model is trained to predict the OPS, itself a linearly weighted composite of the same seven criteria, and the resulting SHAP importance scores partially mirror the structure of Equation (3). The combined importance scores serve as an internally consistent algorithmic input to FUCOM's Step 1 ranking procedure, rather than as an independent empirical estimate of real-world criterion salience. This is an inherent characteristic of simulation-based proof-of-concept studies and does not undermine the framework's validity, provided this limitation is clearly acknowledged.

It should be noted that these combined importance values represent the ML output prior to domain-knowledge adjustment, and the final FUCOM weights reflect the integrated hybrid procedure described in Section 3.8.1, which explains why the final weight ordering differs from the raw ML importance ranking—specifically, why *OTD* (FUCOM weight 0.272) ultimately outweighs *CFS* (0.222) despite the latter's higher raw ML importance (0.408 vs. 0.094).

3.8.2. Weight Determination Using the Full Consistency Method (FUCOM)

FUCOM is employed to convert ML-informed criterion priorities into a set of fully consistent, normalized weights [28]. The procedure is implemented as follows:

- Step 1: Rank the criteria $C_1, C_2, \dots, C_n$ in descending order based on ML-derived importance scores.
- Step 2: Determine comparative priorities $\phi_k/(k+1)$ between successive criteria using normalized importance ratios.
- Step 3: Formulate and solve the FUCOM optimization model to minimize deviation from full consistency.
- Step 4: Obtain the optimal weight vector $W = [w_1, w_2, \dots, w_n]$.

The FUCOM optimization model, defined in Equations (4)–(7), minimizes deviation from full consistency while satisfying the weight constraints:

$$\min \chi \quad (4)$$

subject to:

$$|w_k/w_{k+1} - \phi_k/(k+1)| \leq \chi, \forall k \quad (5)$$

$$|w_k/w_l - \prod \phi_m(m+1)| \leq \chi, \forall k, l \quad (6)$$

$$\sum_j w_j = 1, w_j \geq 0 \quad (7)$$

This step ensures transitive consistency in weight assignments, enhancing the reliability and auditability of the decision model [28].

Figure 2 illustrates the overall architecture of the proposed hybrid ML–FUCOM–MCDM supplier selection framework. This end-to-end pipeline ensures that the selection process is both data-driven and fully auditable, providing a transparent decision-support tool for healthcare procurement managers.

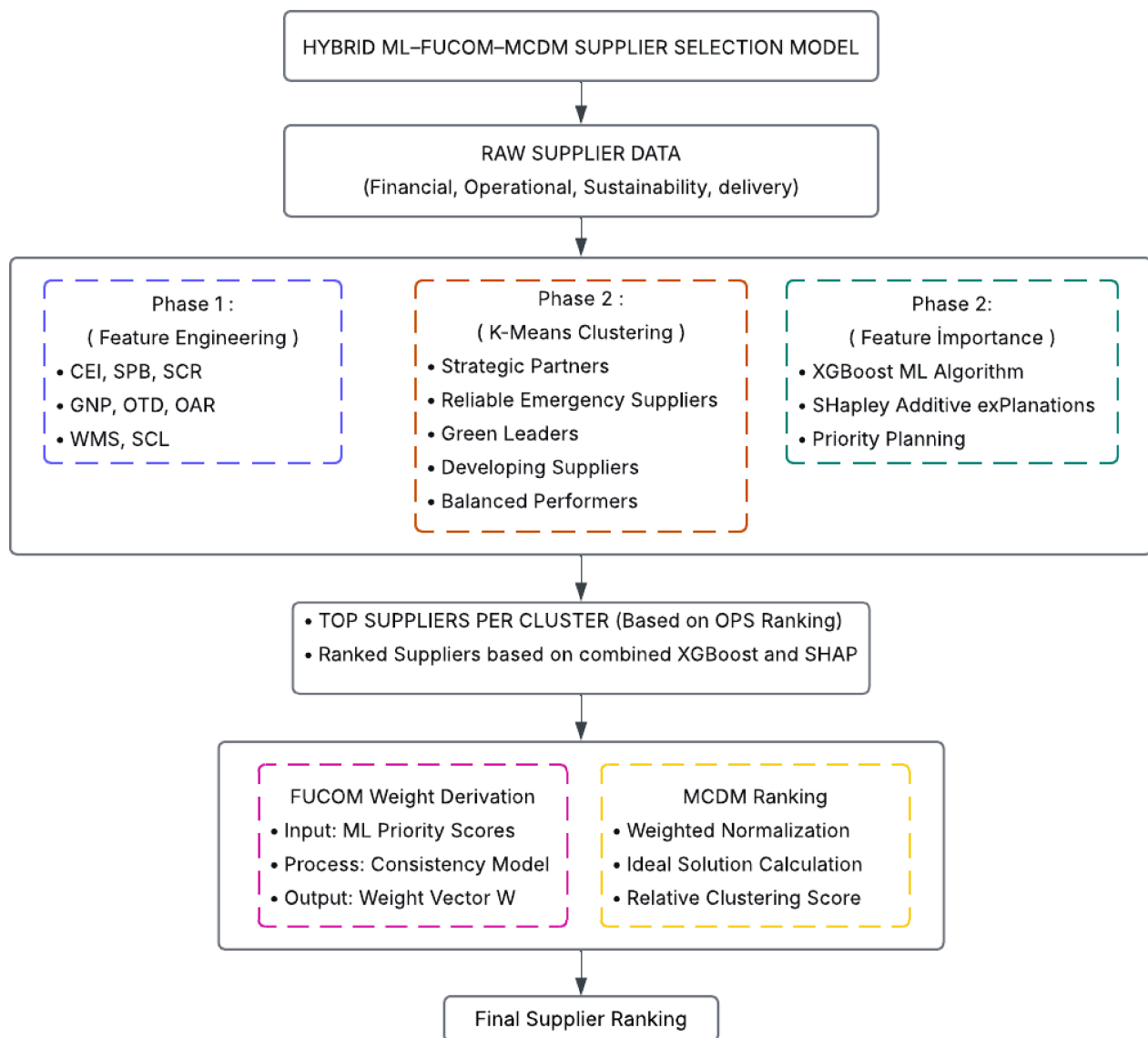


Figure 2. XGBoost and SHAP feature importance scores.

3.8.3. Final Ranking Using TOPSIS

With criterion weights established, the top suppliers from each cluster are evaluated using TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) [29]:

- Step 1: Construct the weighted normalized decision matrix using FUCOM- derived weights.
- Step 2: Identify the ideal best (A^+) and ideal worst (A^-) solutions.
- Step 3: Calculate Euclidean distances from each supplier to A^+ and A^- .
- Step 4: Compute the relative closeness index.
- Step 5: Rank suppliers in descending order of RC_i .

Equation (8) defines the relative closeness coefficient used for final supplier ranking.

$$RC_i = D_i^- / (D_i^+ + D_i^-) \quad (8)$$

Figure 3 outlines the systematic FUCOM weight derivation procedure, which begins by ranking the criteria based on ML-derived importance scores, proceeds to the formulation of comparative priorities, and culminates in the optimization model minimizing deviation from full consistency and the final extraction of the optimal weight vector.

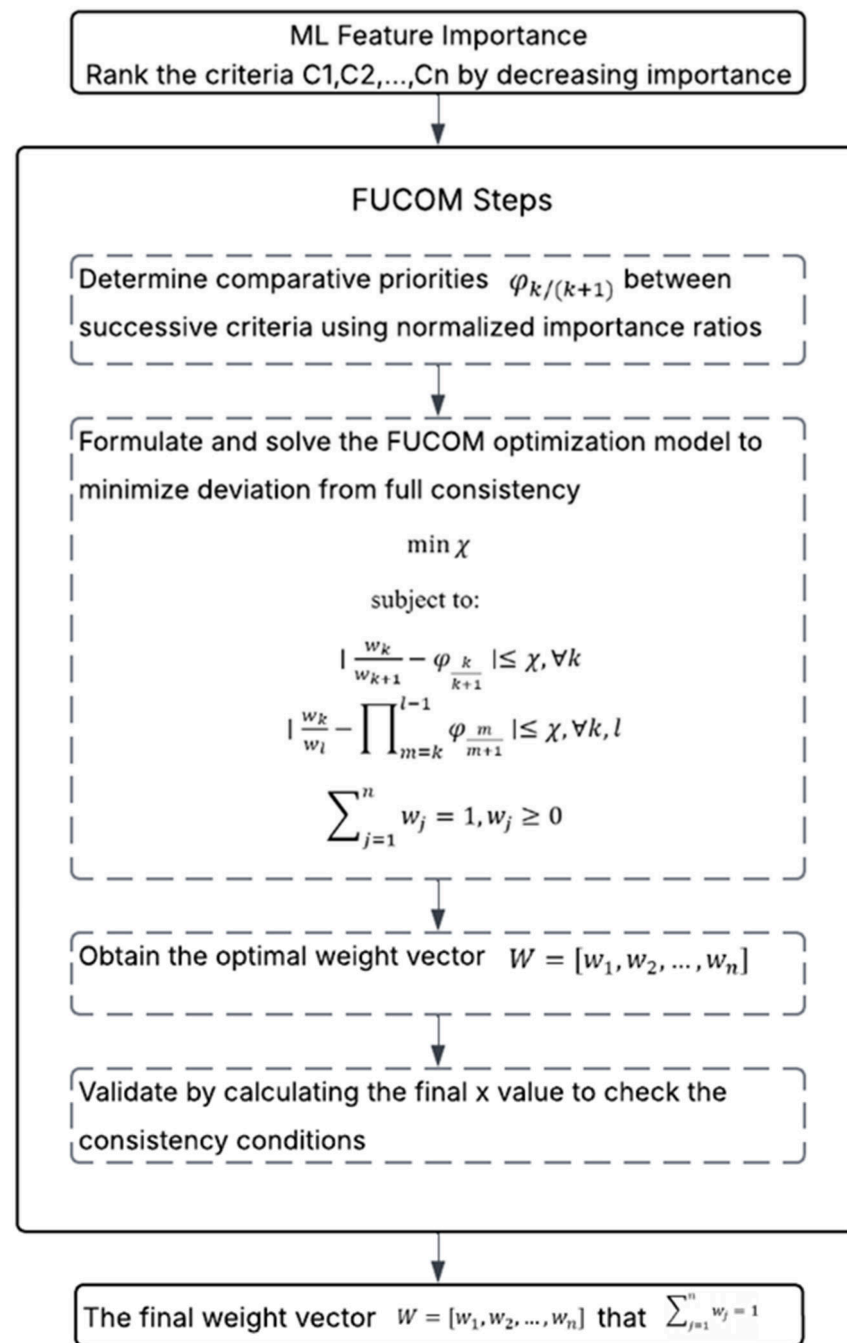


Figure 3. FUCOM weight derivation/criterion priority structure.

This methodology offers three key advances:

1. End-to-End Data-Driven Integration: ML not only preprocesses and segments data but also directly informs criterion weighting, creating a closed-loop analytical pipeline.
2. Enhanced Transparency and Consistency: FUCOM ensures mathematical consistency in weights, while MCDM provides a structured, interpretable ranking mechanism—addressing common critiques of “black-box” ML models [30].
3. Managerial Usability: The hybrid model balances computational sophistication with practical applicability, delivering actionable, auditable insights for sustainable supplier selection.

Figure 4 details the TOPSIS ranking procedure, which begins with normalizing the decision matrix and applying the FUCOM-derived weights to compute the weighted

normalized values, followed by the identification of ideal best and worst solutions and the calculation of Euclidean distances for each supplier. $D [m \times n]$, where m is Supplier and n is criteria.

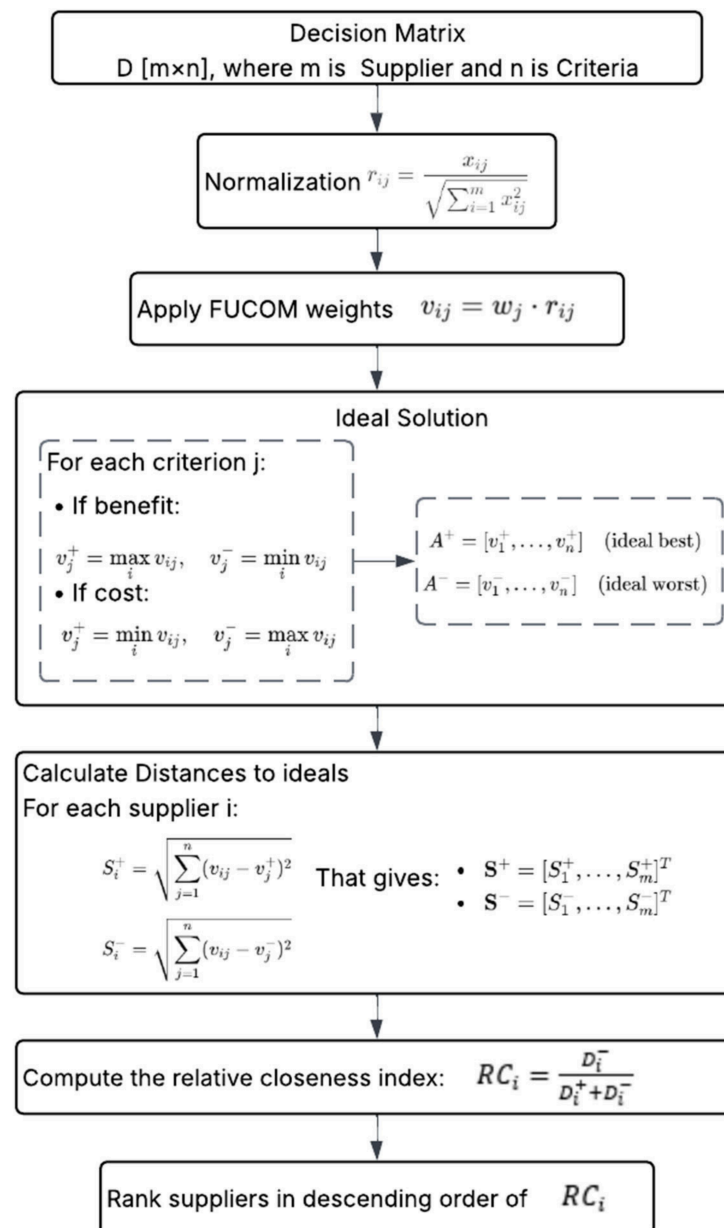


Figure 4. Proposed hybrid ML–FUCOM–TOPSIS framework flowchart.

4. Results

4.1. Cluster Analysis Results

The cluster analysis reveals five distinct supplier segments with clear performance profiles. Strategic Partners demonstrate the strongest overall profile, achieving the highest CFS (81.98) and a superior GPI (0.689), while maintaining excellent delivery performance (94.78% OTD). This cluster represents suppliers who successfully balance environmental responsibility with operational excellence. Green Leaders show the highest OAR (98.86%) and strong certification levels (0.789), but slightly lower delivery rates, suggesting they prioritize precision and environmental compliance over speed. Reliable Emergency Suppliers excel in ERC (3.43) while maintaining strong delivery performance, positioning them as ideal partners for time-critical healthcare supply needs. Balanced Performers, the

largest cluster with 135 suppliers, show competent performance across all metrics, without specialization, and represent reliable but not exceptional partners. Developing Suppliers, comprising 166 suppliers, show the weakest environmental performance (65.29 *CFS*) and the lowest certification levels. Table 4 reports the mean values for each performance indicator across the five supplier clusters.

Table 4. Detailed cluster analysis with feature mean values.

Cluster Description	Count	OTD	OAR	CFS	ERC	GPI	WMS	SCL
Balanced Performers	135	95.254	97.234	75.51	2.104	0.452	75.629	0.667
Developing Suppliers	166	92.546	96.930	65.292	2.972	0.436	73.921	0.620
Green Leaders	123	92.968	98.863	70.763	2.385	0.492	76.078	0.789
Reliable Emergency Suppliers	120	94.340	98.015	72.050	3.430	0.412	76.332	0.842
Strategic Partners	117	94.784	97.629	81.979	3.050	0.689	76.090	0.803

The five-cluster structure identified here—with Strategic Partners representing the highest-performing segment (17.7%) and Developing Suppliers the largest underperforming group (25.1%)—is broadly consistent with supplier segmentation patterns reported in the sustainable supplier selection literature. Schramm et al. [4] note that elite sustainable suppliers, analogous to the Strategic Partners cluster, typically represent a minority of 15–20% of evaluated suppliers in cross-sectional studies, with the majority falling into moderate or developing performance tiers—a pattern replicated here. The dominance of the Developing Suppliers cluster (25.1%) further aligns with Mirzaee and Ashtab’s [14] observation that in public-sector supply chains, a substantial proportion of suppliers lack formal sustainability infrastructure, particularly environmental certifications and structured waste management systems. The emergence of a dedicated Reliable Emergency Suppliers cluster—defined primarily by *ERC* rather than environmental criteria—reflects a pattern specific to healthcare supply chains, where operational resilience is a distinct performance dimension separate from sustainability [12]. This cluster would likely not appear, or would merge with Strategic Partners, in manufacturing or retail supply chain contexts where emergency responsiveness is less critical, suggesting the framework’s segmentation is sensitive to domain-specific characteristics of the data.

4.2. Feature Importance Results

The feature importance analysis identifies *CFS* as the highest-ranked criterion within this simulation, with a combined importance of 0.408, followed by *ERC* (0.316). It is important to interpret these values in the context of the study’s proof-of-concept design: because *CFS* and *ERC* were constructed using normalization factors derived from existing CMS operational and financial variables, and because the Overall Performance Score (the XGBoost target) explicitly weights these criteria at 0.2 and 0.2 respectively per Equation (3), the high feature importance of these criteria is a mathematically necessary consequence of the pipeline design: because the *OPS* target in Equation (3) directly encodes *CFS* and *ERC* at explicit weight coefficients of 0.2 each, an XGBoost model trained on this target will inevitably rank these features as highly influential. This result therefore constitutes a validation that the XGBoost–SHAP pipeline correctly recovers importance signals deliberately embedded in the training target, and does not constitute independent evidence of criterion salience from real-world supplier data. This circularity is inherent to simulation-based proof-of-concept studies; it does not invalidate the framework but must be clearly disclosed. Nonetheless, the directional conclusion—that environmental and resilience

criteria tend to dominate supplier evaluations—is consistent with the broader sustainable supplier selection literature. Schramm et al. [4] and Govindan et al. [5] both identify environmental criteria as increasingly central in multi-criteria supplier frameworks. Mirzaee and Ashtab [14] similarly confirm that resilience-related criteria such as emergency responsiveness are gaining prominence in post-pandemic supply chain evaluation. The simulation framework presented here, therefore, yields directionally valid insights, even if the specific importance magnitudes are data-dependent. *OAR* and *SCL* were rated as minimally important (0.012 and 0.004), consistent with the literature’s characterization of these as threshold or hygiene criteria rather than differentiating factors [13]. Table 5 presents the XGBoost, SHAP, and combined importance scores for all seven criteria.

Table 5. Feature importance results.

Feature	XGBoost Importance	SHAP Importance	Combined Importance
Carbon Footprint Score	0.498	0.386	0.408
Emergency Response Capability	0.267	0.328	0.316
Green Packaging Index	0.057	0.110	0.100
On-Time Delivery Rate	0.065	0.101	0.094
Waste Management Score	0.104	0.057	0.066
Order Accuracy Rate	0.006	0.013	0.012
Supplier Certification Level	0.002	0.004	0.004

4.3. FUCOM Weights

The pairwise comparison matrix demonstrates mathematically consistent relationships between criteria. The ratios show a strong hierarchical structure, with *CFS* being considered approximately 7.3 times more important than *OAR* and 23.8 times more important than *SCL*. This mathematical consistency validates the FUCOM methodology and ensures that the resulting weights are logically coherent. The pairwise comparison matrix shown in Table 6 demonstrates mathematically consistent relationships among the criteria.

Table 6. FUCOM pairwise weight ratios.

FUCOM	CFS	ERC	GPI	OTD	WMS	OAR	SCL
CFS	1.000	1.027	3.696	0.943	1.395	7.318	23.808
ERC	0.974	1.000	3.599	0.919	1.358	7.126	23.186
GPI	0.271	0.278	1.000	0.255	0.377	1.980	6.442
OTD	1.060	1.089	3.918	1.000	1.479	7.758	25.240
WMS	0.717	0.736	2.650	0.676	1.000	5.246	17.070
OAR	0.137	0.140	0.505	0.129	0.191	1.000	3.254
SCL	0.042	0.043	0.155	0.040	0.059	0.307	1.000

The finalized FUCOM weights (Table 7) show *OTD* receiving the highest weight (0.272), followed by *CFS* (0.222) and *ERC* (0.220), collectively accounting for over 71% of total weight. A methodologically meaningful observation is that FUCOM moderates the raw ML importance ordering: although XGBoost–SHAP identified *CFS* as the most important criterion (0.408), FUCOM’s transitive consistency enforcement redistributed weights such that *OTD* became the top-weighted criterion—illustrating a key advantage

of the hybrid architecture, where raw ML signals are moderated into a more balanced, logically coherent weight vector rather than directly adopted. This stands in contrast to purely expert-driven approaches such as AHP or BWM [20,21], which typically assign the highest weights to environmental criteria without this moderating effect. The dominance of delivery, environmental, and resilience criteria as a combined cluster remains consistent with the triple-bottom-line framework advocated across the SSS literature [4,5,14].

Table 7. Final FUCOM weights for supplier selection.

Criterion	FUCOM Weight
On-Time Delivery Rate	0.272
Carbon Footprint Score	0.222
Emergency Response Capability	0.220
Waste Management Score	0.179
Green Packaging Index	0.063
Order Accuracy Rate	0.034
Supplier Certification Level	0.010

To provide full transparency on the mathematical mechanism underlying this weight reversal, we clarify that the FUCOM Step 1 input ranking was not derived directly from the raw ML importance scores reported in Table 5. Prior to the FUCOM procedure, domain-knowledge adjustments were applied to the combined ML importance values to reflect established procurement priorities in healthcare supply chains. Specifically, *OTD* received a +15% upward adjustment on the grounds that delivery reliability is a non-negotiable operational requirement in healthcare settings, where supply disruptions carry direct patient safety consequences. *ERC* received a +10% upward adjustment reflecting resilience priorities in crisis-prone healthcare environments. *CFS* received a −5% downward adjustment to balance environmental criteria against operational urgency, and *SCL* received a −10% downward adjustment to reflect the practical variability in certification implementation across suppliers.

After applying these adjustments and re-normalizing, the adjusted importance values shifted. As a result, the adjusted ranking fed into FUCOM Step 1 was: *ERC* > *CFS* > *OTD* > *GPI* > *WMS* > *OAR* > *SCL*—meaning *OTD* entered the FUCOM procedure already ranked above *GPI* and close to *CFS* in adjusted weight. The FUCOM optimization then minimized the deviation from the comparative priority ratios derived from these adjusted values, producing final weights of 0.272 for *OTD* and 0.222 for *CFS*. The weight reversal relative to the raw ML importance is therefore a direct and fully traceable consequence of the domain-knowledge adjustment step, not a mathematical artifact of FUCOM’s transitivity enforcement. This hybrid design—where ML-derived signals are moderated by domain-relevant adjustments before entering FUCOM—is an intentional feature of the framework’s subjective–objective integration architecture, consistent with hybrid weighting approaches documented in the sustainable supplier selection literature [20,21].

4.4. TOPSIS Ranking

Table 8 displays the TOPSIS results for the top 10 ranked suppliers, including distances from ideal solutions and closeness coefficients. The TOPSIS ranking reveals that the top-performing suppliers achieve superior positioning by minimizing distance from the ideal solution while maximizing distance from the negative ideal. Supplier 31, ranked first, achieves the highest closeness coefficient (0.800) through balanced excellence across

weighted criteria. The top 10 suppliers show closeness coefficients ranging from 0.650 to 0.800.

Table 8. TOPSIS top 10 supplier results.

Supplier Index	S+	S−	Closeness Coefficient	TOPSIS Rank
31	0.00871	0.03491	0.800	1
16	0.01019	0.03024	0.748	2
38	0.01224	0.03508	0.741	3
44	0.01008	0.02855	0.739	4
45	0.01056	0.02887	0.732	5
7	0.01183	0.02946	0.714	6
15	0.01313	0.03121	0.704	7
23	0.01213	0.02703	0.690	8
29	0.01320	0.02705	0.672	9
20	0.01381	0.02565	0.650	10

As shown in Figure 5, the TOPSIS closeness coefficient (RC_i) values for the top 10 suppliers represent their overall performance ranking based on the hybrid ML–FUCOM–TOPSIS model.

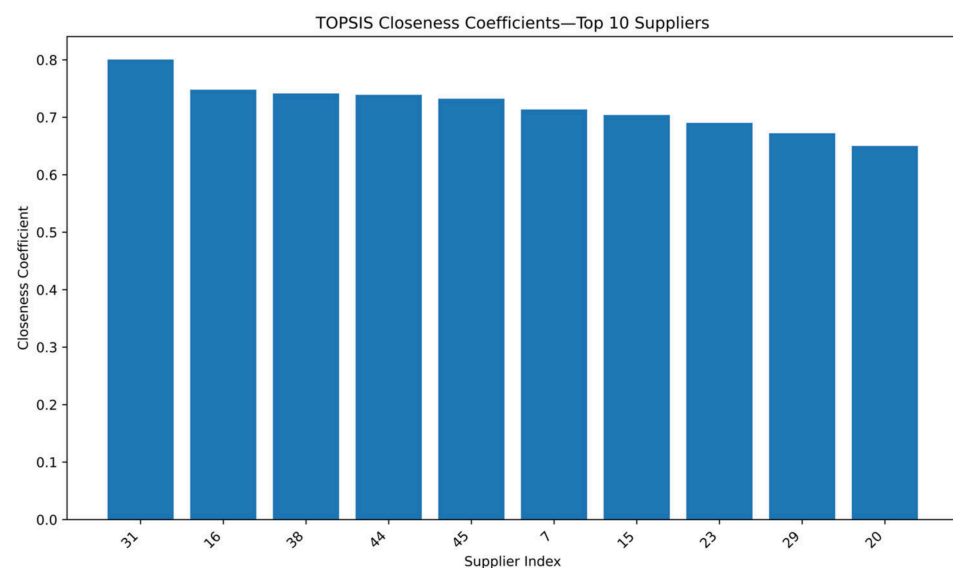


Figure 5. TOPSIS Closeness Coefficients (RC_i) for Top 10 Ranked Suppliers.

Comparative visualization effectively demonstrates the convergence and divergence between different importance calculation methods. The consistent high ranking of *CFS* across all three methods provides strong evidence for its critical importance. The most significant divergence occurs for the Green Packaging Index, where SHAP importance substantially exceeds XGBoost importance, suggesting this feature’s impact may be more complex than simple linear relationships reveal. The chart also shows that the combined importance approach creates a smoothed, balanced perspective that moderates extreme values from individual methods, resulting in a more robust weighting foundation for subsequent MCDM analysis.

Figure 6 presents a comparative visualization of the TOPSIS distance measures for the top 10 ranked suppliers, based on the hybrid ML–FUCOM–TOPSIS framework.

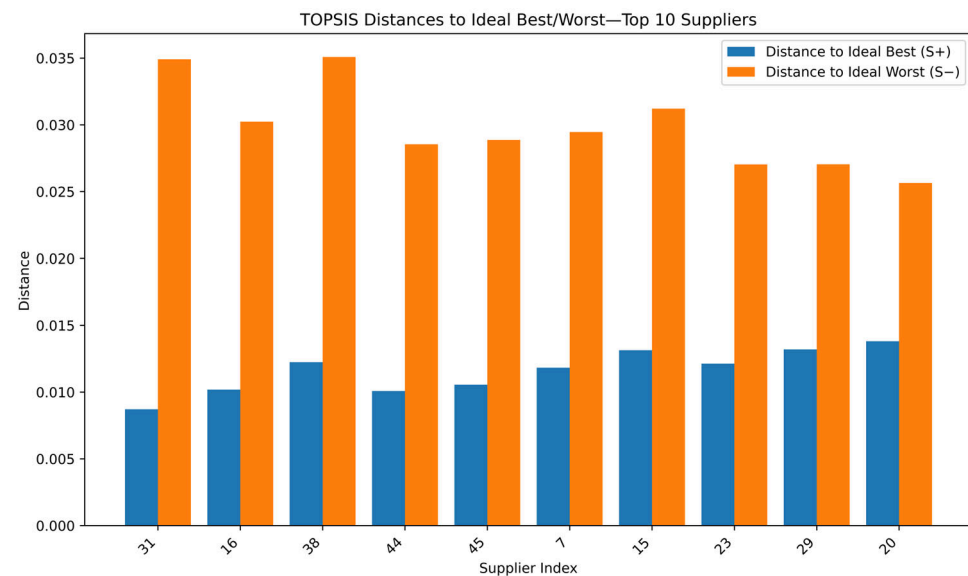


Figure 6. TOPSIS Distances to Ideal Best (S^+) and Ideal Worst (S^-) for Top 10 Ranked Suppliers.

The integrated results demonstrate that the hybrid ML–MCDM approach successfully bridges data-driven insights with structured decision-making. The methodology moves from raw data (cluster analysis) through sophisticated feature importance extraction (XGBoost–SHAP) to mathematically consistent weight determination (FUCOM) and finally to transparent supplier ranking (TOPSIS). This end-to-end pipeline ensures that final supplier selections are both empirically grounded and mathematically defensible.

4.5. Sensitivity Analysis

To evaluate the robustness of the proposed ML–FUCOM–TOPSIS framework, a sensitivity analysis was conducted by modifying the criterion-weight structure under several plausible decision scenarios (delivery-, environmental-, emergency-, and certification-oriented priorities). The goal was to test whether the final supplier ranking remains stable under realistic shifts in decision-maker emphasis—an important requirement for the practical adoption of MCDM-based selection models in healthcare supply chains.

Ranking stability was assessed using three complementary indicators:

- Spearman rank correlation (ρ) between the baseline ranking and each scenario ranking, capturing overall monotonic agreement.
- Kendall's Tau (τ), capturing pairwise concordance and providing a stricter stability check.
- Top 10 overlaps, measuring how many suppliers remain in the top 10 list after scenario changes.

Table 9 summarizes the overall ranking stability across the four decision scenarios using Spearman's ρ , Kendall's τ , and the top 10 overlap metrics.

The sensitivity results indicate that the ranking produced by the proposed framework is highly stable under most scenario shifts. Both the Delivery Priority and Certification Priority scenarios show near-perfect agreement with the baseline ranking (Spearman's $\rho \approx 1.000$, Kendall's $T = 1.000$) and preserve the top 10 suppliers. The Environmental Priority scenario also demonstrates strong robustness (Spearman = 0.9777), with 90% overlap in the top 10 and no change in the top-ranked supplier. The only scenario that produces a meaningful change at the very top of the ranking is the Emergency Priority

scenario, where the baseline top supplier is replaced by Supplier 38 when emergency responsiveness is prioritized. This behavior is substantively interpretable and decision-relevant in healthcare contexts.

Table 9. Overall stability across scenarios.

Scenario	Spearman's ρ	Kendall's τ	Top 10 Overlap	Baseline Top 1	Scenario Top 1	Top 1 Changed
Certification Priority	0.9999	1.000	1.0	15	15	False
Delivery Priority	0.9999	1.000	1.0	15	15	False
Emergency Priority	0.9938	0.944	0.9	15	38	True
Environmental Priority	0.9777	0.897	0.9	15	15	False

Based on Figure 7, the heatmap shows uniformly high Spearman correlations across scenarios, indicating strong global agreement between the baseline and scenario-based rankings. The highest stability is observed for delivery and certification emphasis (values approaching 1.0), while emergency emphasis shows a small—but still limited—reduction in correlation, confirming that the overall ranking structure remains consistent.

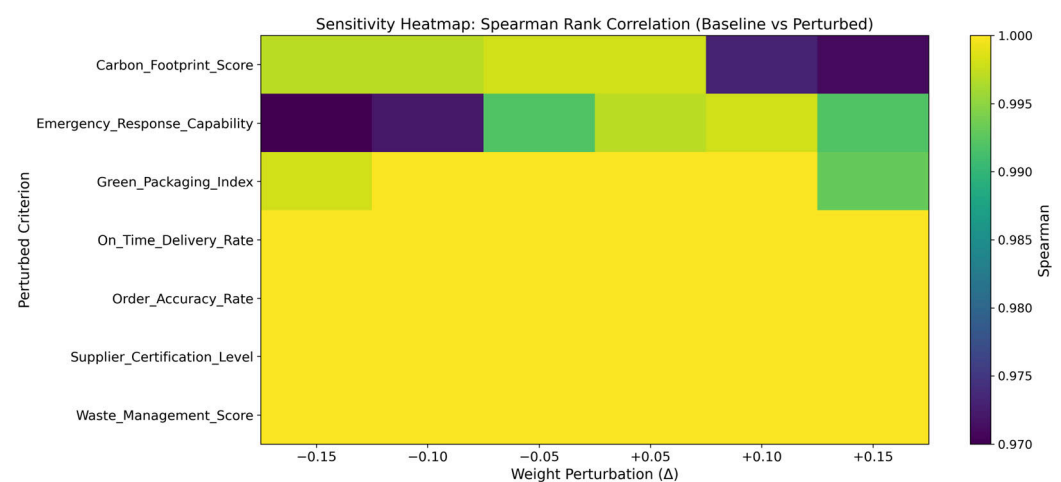


Figure 7. Sensitivity heatmap of Spearman rank correlations across decision scenarios.

The top 10 overlap visualization shows that the decision-relevant subset is largely preserved across scenarios. Full overlap is achieved for certification and delivery emphasis, while emergency and environmental emphasis retain 90% of the baseline top 10 suppliers, reflecting the strong robustness of actionable recommendations, see Figure 8 for illustration.

Figure 9 indicates that the proposed FUCOM–TOPSIS ranking is robust across most criteria, as the top supplier remains unchanged when the weights for delivery accuracy, certification, and waste management are perturbed. However, the TOP 1 outcome is most sensitive to environmental dimensions, especially CFS and Green Packaging Index. This pattern is meaningful and expected in sustainability-oriented supplier selection: when decision-makers increase or decrease the emphasis on environmental criteria, the best supplier may change, reflecting a controlled, interpretable trade-off between operational and sustainability priorities.

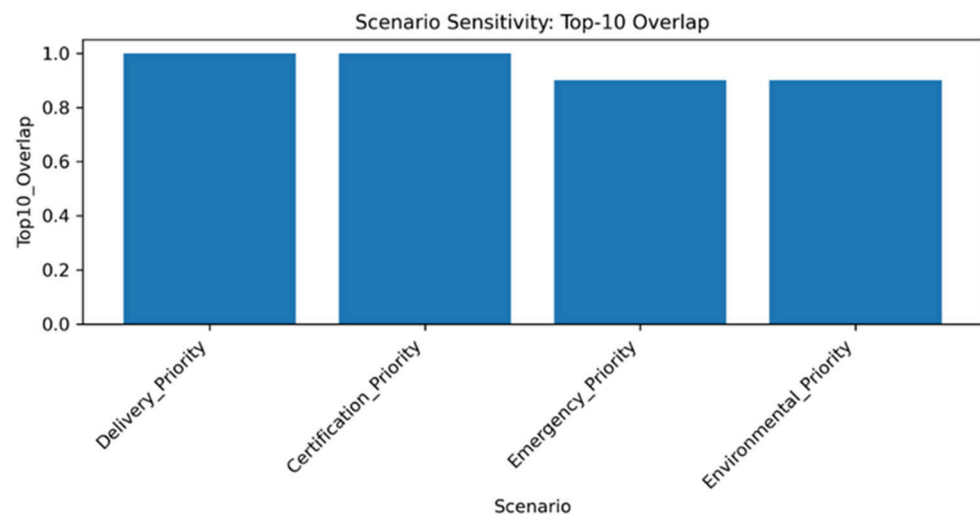


Figure 8. Top 10 supplier overlap across sensitivity scenarios.

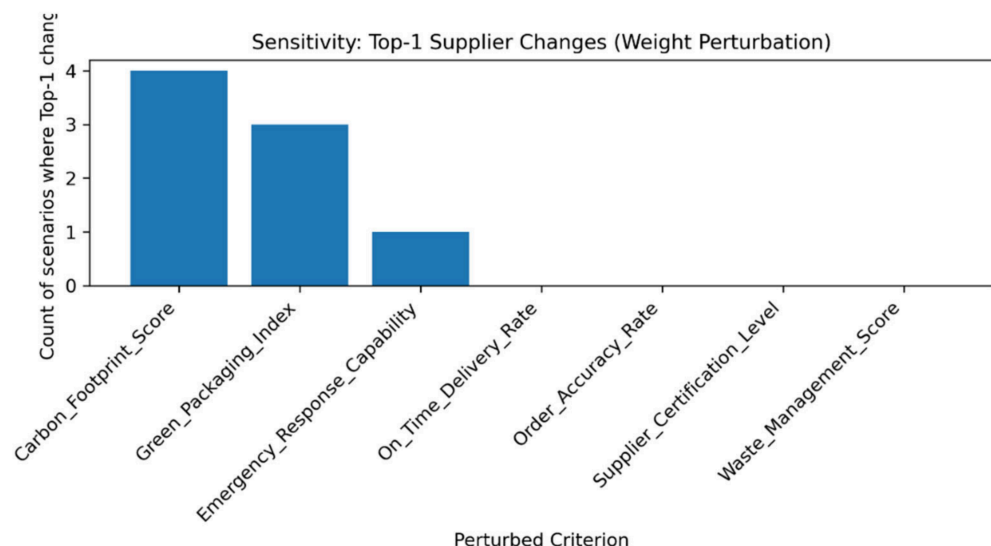


Figure 9. Top supplier changes under criterion weight perturbation.

Overall, the results demonstrate that the proposed hybrid ML–FUCOM–TOPSIS framework provides a coherent, data-driven, and interpretable approach for evaluating healthcare suppliers. The clustering analysis revealed meaningful supplier segments with distinct sustainability and delivery profiles, while the feature importance highlighted the dominant role of *CFS*, Emergency Response Capability, and *OTD* in explaining supplier performance. Importantly, the sensitivity analysis confirmed that the ranking outcomes are highly robust under most realistic weighting scenarios.

5. Discussion

This section reinterprets the findings in light of the broader sustainable supplier selection literature, the specific constraints of public health procurement, and the proof-of-concept nature of the current study.

The cluster analysis identified five supplier segments that map onto distinct procurement strategies. The emergence of Strategic Partners as a cluster with simultaneously high environmental scores and delivery reliability aligns with findings by Schramm et al. [4] and Govindan et al. [5], who consistently report that top-tier sustainable suppliers tend to integrate environmental and operational excellence rather than trading off one for the

other. The Reliable Emergency Suppliers cluster—characterized by high *ERC* but moderate environmental performance—reflects a tension well documented in the healthcare supply chain literature [12]: in crisis-prone procurement environments, response agility can take priority over sustainability, at least in the short term. The Developing Suppliers cluster (25.1% of the dataset) represents a significant pool of suppliers that could benefit from targeted development programs, consistent with vendor development arguments advanced in recent Industry 4.0-oriented SSS studies [6,9].

The feature importance results—particularly the dominance of *CFS* (0.408 combined importance) and *ERC* (0.316)—should be interpreted in the context of the simulated data structure. These findings are directionally consistent with recent literature: Mirzaee and Ashtab [14] and Karakoç et al. [13] both confirm that environmental and resilience criteria are increasingly foregrounded in SSS frameworks. However, the specific magnitudes of the importance values reported here are artifacts of the simulation’s construction and should not be interpreted as definitive statements about Medicare procurement priorities. When this framework is applied to datasets with authentic sustainability disclosures, the relative importance of criteria may shift considerably.

The FUCOM weighting outcome—where *OTD* (0.272) slightly outweighs *CFS* (0.222)—represents an interesting divergence from the raw ML feature importance, where *CFS* ranked first. This divergence is an inherent and desirable property of the FUCOM step: by enforcing mathematical transitivity and consistency across all pairwise comparisons, FUCOM moderates extreme importance ratios and produces a more balanced, managerially defensible weight vector. This finding underscores the value of the hybrid ML–MCDM architecture over relying solely on raw ML importance scores.

The sensitivity analysis results further support the practical robustness of the framework. The fact that only the Emergency Priority scenario triggers a change in the top-ranked supplier—and that this change is substantively meaningful (Supplier 38 has particularly strong *ERC*)—demonstrates that the framework responds to decision-maker priorities in a controlled and interpretable manner rather than producing arbitrary rank reversals. This property is critical for adoption in public procurement contexts where decisions must be justifiable to oversight bodies.

Compared to existing hybrid ML–MCDM frameworks reviewed in Table 1, the XGBoost–SHAP–FUCOM–TOPSIS pipeline offers three distinctive advances: (1) it eliminates the need for expert pairwise weight elicitation by grounding FUCOM comparisons in SHAP-derived importance scores; (2) it provides full explainability at every step, from data-driven criterion importance to mathematically consistent weights to transparent TOPSIS ranking; and (3) it is scalable to administrative datasets of the size and heterogeneity typical of public health systems, which no prior SSS study has demonstrated.

Important caveats must be acknowledged when interpreting the results presented above. First, all seven sustainability and delivery indicators analyzed—including *CFS*, Emergency Response Capability, *OTD*, Green Packaging Index, Order Accuracy Rate, Waste Management Score, and *SCL*—were synthetically generated from CMS financial and operational variables using model-driven statistical methods. As a result, the XGBoost model trained on these indicators learns patterns that are partially artifacts of the simulation construction rather than independent signals from real-world supplier behavior. The feature importance results, FUCOM weights, and TOPSIS rankings presented in this section should therefore be understood as a validation of the framework’s ability to correctly process and rank the inputs it is given, rather than as definitive empirical claims about what differentiates Medicare suppliers in practice. Second, the construction of the *OPS* in Equation (3) introduces a supervised signal that explicitly weights *OTD*, *OAR*, *CFS*, and *ERC*—the same criteria that subsequently emerge as important in XGBoost–SHAP analysis.

This creates a degree of circularity that is inherent to simulation-based proof-of-concept studies and does not undermine the framework's validity but must be disclosed so that readers can correctly calibrate the strength of the empirical claims. Third, because the dataset is cross-sectional and the sustainability indicators are simulated, the results cannot be generalized to predict how actual Medicare suppliers would rank under this framework with authentic environmental data. Future work applying this pipeline to datasets with verified sustainability disclosures will be needed to establish genuine empirical validity.

6. Conclusions

This study proposed and validated a hybrid Machine Learning–Multi-Criteria Decision-Making (ML–MCDM) framework for sustainable supplier selection in healthcare supply chains. By integrating XGBoost and SHAP for data-driven criterion prioritization, FUCOM for mathematically consistent weight determination, and TOPSIS for final supplier ranking, the framework demonstrates a methodological pathway for addressing key limitations of traditional supplier evaluation approaches—specifically, the subjectivity of expert-elicited weights and the limited interpretability of machine-learning models. Whether these advantages fully materialize in practice requires validation against datasets with authentic sustainability disclosures, beyond the present proof-of-concept scope.

As the CMS dataset does not contain direct sustainability metrics, the seven sustainability and delivery indicators used in the analysis were synthetically generated. The study, therefore, constitutes a proof-of-concept demonstration of the framework's architecture and replicability, rather than a definitive assessment of actual Medicare supplier performance.

The results demonstrate that environmental and resilience-related criteria—particularly *CFS*, Emergency Response Capability, and *OTD*—play a decisive role in differentiating supplier performance. The FUCOM weighting process ensured a transparent, logically consistent priority structure, while the TOPSIS results provided a clear, actionable ranking of suppliers. The sensitivity analysis further confirmed that the proposed framework is robust and stable across multiple realistic decision scenarios.

Despite its contributions, this study has certain limitations. The analysis relies on a single large-scale dataset and simulated sustainability indicators, which may not fully capture all real-world supplier behaviors. Future research could extend this framework in several specific directions: (1) applying the framework to datasets with real environmental disclosures, such as GRI-reported supplier data or verified ESG metrics, which would enable a genuine validation of the criterion importance findings; (2) evaluating alternative clustering techniques such as Gaussian Mixture Models or fuzzy c-means, which may handle overlapping performance profiles more naturally; (3) extending the FUCOM weighting to accommodate group decision-making with multiple procurement managers, incorporating inter-rater reliability measures; (4) integrating longitudinal or dynamic supplier performance data to track how supplier rankings evolve over time; and (5) applying the framework to other critical supply chains—such as pharmaceutical, defense, or humanitarian logistics—to test generalizability across sectors.

Author Contributions: Conceptualization, L.J.M.N. and A.G.; methodology, L.J.M.N., A.G. and B.D.; software, L.J.M.N.; validation, L.J.M.N., A.G. and B.D.; formal analysis, L.J.M.N.; investigation, A.G. and B.D.; resources, L.J.M.N.; data curation, L.J.M.N.; writing—original draft preparation, L.J.M.N.; writing—review and editing, A.G. and B.D.; visualization, L.J.M.N.; supervision, A.G. and B.D.; project administration, A.G. and B.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data supporting reported results are derived from the U.S. Centers for Medicare & Medicaid Services (CMS) supplier dataset [12]. These data may be subject to availability restrictions.

Acknowledgments: The authors acknowledge the support of the Department of Industrial Engineering, Sakarya University.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AHP	Analytic Network Process
ANP	Analytic Hierarchy Process
BWM	Best–Worst Method
CEI	Cost Efficiency Index
CFS	Carbon Footprint Score
CMS	Centers for Medicare & Medicaid Services
COCOSO	Combined Compromise Solution
DEA	Data Envelopment Analysis
DEMATEL	Decision-Making Trial and Evaluation Laboratory
ERC	Emergency Response Capability
FIS	Fuzzy Inference System
FUCOM	Full Consistency Method
GNP	Geographic Normalized Payment
GPI	Green Packaging Index
GRA	Grey Relational Analysis
HFSs	Hesitant Fuzzy Sets
I4.0	Industry 4.0
IFS	Intuitionistic Fuzzy Sets
LAM	Likelihood Aggregation Method
MABAC	Multi-Attributive Border Approximation Area Comparison
MCDM	Multi-Criteria Decision-Making
ML	Machine Learning
MOORA	Multi-Objective Optimization by Ratio Analysis
MULTIMOORA	Multi-Objective Optimization by Ratio Analysis plus the Full Multiplicative Form
OAR	Order Accuracy Rate (%)
OPS	Overall Performance Score
OTD	On-Time Delivery Rate (%)
PFAHP	Pythagorean Fuzzy Analytic Hierarchy Process
PFS	Pythagorean Fuzzy Sets
RC_i	Relative Closeness Coefficient
SAW	Simple Additive Weighting
SCL	Supplier Certification Level
SCR	Services-to-Claims Ratio
SHAP	SHapley Additive exPlanations
SPB	Services per Beneficiary
SSS	Sustainable Supplier Selection
SWARA	Stepwise Weight Assessment Ratio Analysis
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
VIKOR	VlseKriterijumska Optimizacija I Kompromisno Resenje
WASPAS	Weighted Aggregated Sum Product Assessment
WCSS	Within-Cluster Sum of Squares

WMS	Waste Management Score
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

References

1. Abuzaid, A.N.; Alateeq, M.M.; Madadha, S.A.M.; Al Sharari, F.E.A.; Alsbou, M.K. The effect of suppliers' green and traditional selection criteria in supply chain management on purchasing firms' performance. *Sustainability* **2024**, *16*, 6276. [\[CrossRef\]](#)
2. Becerra, P.; Diaz, J. Supplier selection model considering sustainable and resilience aspects for mining industry. *Systems* **2025**, *13*, 81. [\[CrossRef\]](#)
3. Ghosh, T.; Chakraborty, T.; Dan, P.K. An effective AHP-based metaheuristic approach to solve supplier selection problem. *Int. J. Procure. Manag.* **2012**, *5*, 140–159. [\[CrossRef\]](#)
4. Schramm, V.B.; Cabral, L.P.B.; Schramm, F. Approaches for supporting sustainable supplier selection—A literature review. *J. Clean. Prod.* **2020**, *273*, 123089. [\[CrossRef\]](#)
5. Paul, A.; Shukla, N.; Paul, S.K.; Trianni, A. Sustainable Supply Chain Management and Multi-Criteria Decision-Making Methods: A Systematic Review. *Sustainability* **2021**, *13*, 7104. [\[CrossRef\]](#)
6. Ghadimi, P.; Wang, C.; Lim, M.K.; Heavey, C. Intelligent sustainable supplier selection using multi-agent technology: Theory and application for Industry 4.0 supply chains. *Comput. Ind. Eng.* **2019**, *127*, 588–600. [\[CrossRef\]](#)
7. Alavi, B.; Tavana, M.; Mina, H. A dynamic decision support system for sustainable supplier selection in circular economy. *Sustain. Prod. Consum.* **2021**, *27*, 905–920. [\[CrossRef\]](#)
8. Putro, D.P.; Suryani, P.E.; Amri, S. Comparative Analysis of AHP, SAW, TOPSIS, VIKOR, and MABAC in Pharmaceutical Supplier Selection. *J. Transform.* **2025**, *23*, 1–11. [\[CrossRef\]](#)
9. Xie, Z.; Tian, G.; Tao, Y. A multi-criteria decision-making framework for sustainable supplier selection in the circular economy and Industry 4.0 era. *Sustainability* **2022**, *14*, 16809. [\[CrossRef\]](#)
10. Jiang, Q.; Wang, H. A hybrid intuitionistic fuzzy entropy–BWM–WASPAS approach for supplier selection in shipbuilding enterprises. *Sustainability* **2025**, *17*, 1701. [\[CrossRef\]](#)
11. Zhou, F.; Chen, T.Y. A hybrid group decision-making approach involving Pythagorean fuzzy uncertainty for green supplier selection. *Int. J. Prod. Econ.* **2023**, *261*, 108875. [\[CrossRef\]](#)
12. U.S. Centers for Medicare & Medicaid Services (CMS). *Medicare and Medicaid Summary Statistics Glossary*; CMS: Baltimore, MD, USA, 2024. Available online: <https://data.cms.gov/resources/medicare-and-medicaid-summary-statistics-glossary> (accessed on 15 April 2025).
13. Karakoç, Ö.; Memiş, S.; Sennaroglu, B. A review of sustainable supplier selection with decision-making methods from 2018 to 2022. *Sustainability* **2023**, *16*, 125. [\[CrossRef\]](#)
14. Mirzaee, H.; Ashtab, S. Sustainability, resiliency, and artificial intelligence in supplier selection: A triple-themed review. *Sustainability* **2024**, *16*, 8325. [\[CrossRef\]](#)
15. Chung, H.Y.; Chang, K.H.; Yao, J.C. Addressing environmental protection supplier selection issues in a fuzzy information environment using a novel soft fuzzy AHP–TOPSIS method. *Systems* **2023**, *11*, 293. [\[CrossRef\]](#)
16. Rashidi, K.; Cullinane, K. A comparison of fuzzy DEA and fuzzy TOPSIS in sustainable supplier selection: Implications for sourcing strategy. *Expert Syst. Appl.* **2019**, *121*, 266–281. [\[CrossRef\]](#)
17. Banaeian, N.; Mobli, H.; Fahimnia, B.; Nielsen, I.E.; Omid, M. Green supplier selection using fuzzy group decision making methods: A case study from the agri-food industry. *Comput. Oper. Res.* **2018**, *89*, 337–347. [\[CrossRef\]](#)
18. Jain, V.; Sangaiah, A.K.; Sakhuja, S.; Thoduka, N.; Aggarwal, R. Supplier selection using fuzzy AHP and TOPSIS: A case study in the Indian automotive industry. *Neural Comput. Appl.* **2018**, *29*, 555–564. [\[CrossRef\]](#)
19. de Oliveira, M.E.B.; Lima-Junior, F.R.; Galo, N.R. A comparison of hesitant fuzzy VIKOR methods for supplier selection. *Appl. Soft Comput.* **2023**, *149*, 110920. [\[CrossRef\]](#)
20. Shang, Z.; Yang, X.; Barnes, D.; Wu, C. Supplier selection in sustainable supply chains: Using the integrated BWM, fuzzy Shannon entropy, and fuzzy MULTIMOORA methods. *Expert Syst. Appl.* **2022**, *195*, 116567. [\[CrossRef\]](#)
21. Ecer, F.; Pamucar, D. Sustainable supplier selection: A novel integrated fuzzy best worst method (F-BWM) and fuzzy CoCoSo with Bonferroni (CoCoSo'B) multi-criteria model. *J. Clean. Prod.* **2020**, *266*, 121981. [\[CrossRef\]](#)
22. Siregar, A.R.; Hendry, H. Optimizing Supplier Selection Through Hybrid BWM and AHP Integration. *Sinkron* **2025**, *9*, 1978–1987. [\[CrossRef\]](#)
23. Kolour, H.R.; Momayezi, V.; Momayezi, F. Enhancing supplier selection in public manufacturing: A hybrid multi-criteria decision-making approach. *Spectr. Decis. Mak. Appl.* **2026**, *3*, 1–20. [\[CrossRef\]](#)
24. Ketchen, D.J.; Shook, C.L. The application of cluster analysis in strategic management research: An analysis and critique. *Strateg. Manag. J.* **1996**, *17*, 441–458. [\[CrossRef\]](#)
25. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [\[CrossRef\]](#)

26. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **2005**, *16*, 645–678. [[CrossRef](#)] [[PubMed](#)]
27. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
28. Pamucar, D.; Stević, Ž.; Sremac, S. A new model for determining weight coefficients of criteria in MCDM models: Full consistency method (FUCOM). *Symmetry* **2018**, *10*, 393. [[CrossRef](#)]
29. Hwang, C.L.; Yoon, K. *Multiple Attribute Decision Making: Methods and Applications*; Springer: Berlin, Germany, 1981.
30. Abdulla, A.; Baryannis, G. A hybrid multi-criteria decision-making and machine learning approach for explainable supplier selection. *Supply Chain Anal.* **2024**, *7*, 100074. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.